Advances in Spatial and Network Economics

Springer
Berlin
Heidelberg
New York
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Tokyo

Titles in the Series

Martin J. Beckmann and *Tönu Puu*
Spatial Structures

Aisling J. Reynolds-Feighan
The Effects of Deregulation on U.S. Air Networks

Åke E. Andersson, David F. Batten
Kiyoshi Kobayashi and *Kazuhiro Yoshikawa (Eds.)*
The Cosmo-Creative Society

Börje Johansson, Charlie Karlsson,
and *Lars Westin (Eds.)*
Patterns of a Network Economy

Bertrand M. Roehner
Theory of Markets

Ashish Sen · Tony E. Smith

Gravity Models of Spatial Interaction Behavior

With 13 Figures

Springer

Professor Dr. Ashish Sen
The University of Illinois at Chicago
School of Urban Planning and Policy (M/C 348)
College of Architecture, Art, and Urban Planning
1007 West Harrison Street
Chicago, IL 60607-7137
USA

Professor Dr. Tony E. Smith
University of Pennsylvania
Department of Regional Science
Philadelphia, PA 19104
USA

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Sen, Ashish:
Gravity models of spatial interaction behavior / Ashish Sen ;
Tony E. Smith. - Berlin ; Heidelberg ; New York ; Barcelona ;
Budapest ; Hong Kong ; London ; Milan ; Paris ; Tokyo :
Springer, 1995
(Advances in spatial and network economics)
ISBN-13:978-3-642-79882-5
NE: Smith, Tony E.:

ISBN-13: 978-3-642-79882-5 e-ISBN-13: 978-3-642-79880-1
DOI: 10.1007/ 978-3-642-79880-1

Softcover reprint of the hardcover 1st edition 1995

SPIN 10489086 42/2202 - 5 4 3 2 1 0 - Printed on acid-free paper

To the memory of our grandmothers:

Promodini Sen

and

Ethel D. Watts

Foreword

Ever since John Stewart proposed the concept of social physics based on the Newtonian gravitational framework, research on the gravity model and its spatial interaction implications has been an exciting activity for regional scientists, including geographers, city and regional planners, and others close to the regional science field. At the same time there has always been a reluctance by traditional social scientists, especially economists, to accept this approach to understanding spatial behavior. My reply to criticisms of the gravity model has always been that I have found no better general method for use in applied research dealing with spatial interaction — whether we consider journey-to-work phenomena, estimating demand for new marina facilities, or determining the best location of medical complexes, libraries, and other service-type functions. More recently in the area of trade among nations (a central and basic concern of economics) I find that there is significant improvement in understanding and projecting world trade patterns when a gravity-type component is added to existing models. Yet, it is still to be admitted that an adequate general theoretical basis for gravity models has not till now been set forth.

It is against this background that the present book by Sen and Smith is most welcomed. In Part I of the book, a *probabilistic* foundation for gravity-type behavior is developed which should appeal to theoreticians and practitioners alike. For theoreticians, this foundation provides axiomatic formulations of gravity models which characterize precisely the types of probabilistic spatial interaction behavior embodied in these models. Such axioms can also serve as inputs to more comprehensive theories of spatial interaction behavior exhibiting gravity-type components. From the practitioners' viewpoint, they provide useful guidelines for identifying the types of behavioral applications for which the gravity model is most appropriate. Moreover, the probabilistic foundation developed provides a unified framework for statistical estimation of gravity models in specific applications.

This theme is developed more fully in Part II of the book, where explicit estimation procedures are constructed for a general class of exponential specifications which includes most gravity models used in practice. In particular, a new maximum-likelihood estimation procedure is shown to be efficient for even very large systems of spatial interactions, and should provide a welcomed addition to the practitioners' tool kit. Moreover, the complete set of inferential tools for maximum likelihood estimation and the very simple and efficient least squares procedure presented in the book should also find their place among tools of the serious practitioner.

Walter Isard

Preface

Gravity models describe, and hence help predict, spatial flows of commuters, air-travelers, migrants, commodities and even messages. They are one of the oldest and most widely used of all social science models. This book presents an up-to-date, consistent and unified approach to the theory, methods and application of the gravity model — which spans from the axiomatic foundations of such models all the way to practical hints for their use. The work is largely that of the authors, most of it appearing in book form for the first time and much of it previously unpublished.

The book is intended primarily for scholars who are working on gravity models or who plan to use such models. Among the elements which should be of interest to the theoreticians and to those interested in applications are the various sets of conditions established for invariance of specific model parameters under changes in spatial structure. Additional tools for practitioners include efficient maximum likelihood procedures, for both the estimation of model parameters and the estimation of covariances, and highly efficient least squares procedures.

The book can also serve as a text in advanced courses. Earlier drafts have, in fact, been used in this way by both the authors. By skipping the more complex proofs, most of the book can be covered in a one semester graduate course.

This book would have been much more difficult, if not impossible, to write without the help of our colleagues and students. We are grateful to Cheri Heramb who assisted one of the authors on his first gravity model project and drew his attention to the need for a book on the gravity model. At that time there were none available on the subject. Along the way, several students have contributed to the work on which this book is based. Among them are Bob Gray, Wayne Miczek, Zbigniew Matuszewsky, Caesar Singh, Lih-Ching Sue, Jay Weber and Steve Wojtkiewicz. The work of Seong-Sun Yun and Hyung-Jin Kim has been referred to in the text; moreover, as the manuscript went through its many revisions, they, along with Linda Chambers, input hundreds of pages of text. We are also grateful to David Bernstein for his careful reading and comments on earlier drafts of Part I of the book. A colleague, Siim Sööt, collaborated with one of the authors on several papers and a long unpublished manuscript on the gravity model. That experience was invaluable in the writing of this book. A particular debt is owed to Piyushimita Thakuriah who input text, formatted text, took care of various LaTeX errors and line-breaks, checked and edited the manuscript and assisted in various other ways during the final stages of production of this book. We are also grateful to Walter Isard, David Boyce, Sven Erlander and Masahita Fujita for their advice and encouragement.

We gratefully acknowledge the several agencies which supported much of the work reported on in the book. This includes the National Science Foundation through Grant Numbers SES 85–21502, SES 85 06397, SES 86 06309, SES 87 09312 and, most recently, DMS 93 13013. Several other projects supported by the U.S Department of transportation and Illinois Department of Transportation provided experience and insights from which the book benefited. We are also indebted to the Computer Center at the University of Illinois at Chicago for making computer time freely available and to George Yanos for being always available to help with the more complex computational problems.

The manuscript was typeset using LaTeX. We are indebted to Eberhard Mattes for making it available for OS/2, where the dreaded message, 'TeX capacity exceeded,' never appears for error-free manuscripts.

Contents

Introduction

Since the early 1940's, efforts to model the spatial interaction behavior of human populations have been largely dominated by *gravity models.* The appeal of these models can be attributed both in the simplicity of their mathematical form and the intuitive nature of their underlying assumptions. For, as observed by Isard and Bramhall (1960, p. 515), these models amount to the simplest possible representation of the basic *gravity hypothesis* that, all else being equal, 'the interaction between any two populations can be expected to be directly related to their size; and ... inversely related to distance'. Thus, to the extent that interaction behavior is consistent with this hypothesis, one may expect gravity models to perform reasonably well empirically.

But when one attempts to go beyond these simple observations and develop sharper theories of spatial interaction behavior, a host of problems arise. First and foremost, the precise linkages between individual decision behavior and overall population behavior are notoriously complex and elusive. Moreover, even when collective individual behavior does lead to certain overall regularities in population behavior, such regularities are often difficult to quantify, let alone explain. In the present case, while the above gravity hypothesis seems intuitively appealing, its underlying concepts of 'interaction', 'size', and 'distance' are quite vague, and indeed, are capable of any number of alternative interpretations. In addition, even when explicit operational definitions of these concepts can be given, there exists a vast range of explicit model forms which are each consistent with this hypothesis. Hence from a theoretical viewpoint, it may safely be said that the basic gravity hypothesis raises many more questions than it answers.

But in spite of these theoretical difficulties, the simplicity and generality of the gravity concept has attracted the interest of researchers over a wide range of fields. For indeed, almost all human relationships can be said to involve interactions which are impeded by distances of one form or another — be they cultural, political, or socio-economic. Thus it is not surprising to find that gravity models have been employed in behavioral studies ranging all the way from marriage relations to international relations. In view of this growing interest, our primary objective in writing the present book has been to develop a rigorous analytical framework within which the behavioral properties of gravity models can be characterized explicitly, and hence applied in a more meaningful way to a wide range of social phenomena.

To motivate the basic philosophy of our approach, it is appropriate to begin by reviewing the classical applications of gravity models in the social sciences. Within this context, we then briefly consider the many theoretical approaches to gravity models which have appeared in the literature, and which set the stage for the present approach. We conclude this introductory discussion with a brief but critical discussion of methods of application of the gravity model.

Classical Gravity Models

Gravity models have a long history in the social sciences, and it is not possible to do justice to all this work here. Hence the following overview is necessarily selective, and is designed primarily to motivate the specific theoretical issues discussed below. For additional background information, the reader is referred to the wealth of historical material in Carrothers (1956b), Isard and Bramhall (1960), Olsson (1965), Philbrick (1973), Isard (1975a), Hua and Porell (1979), Erlander (1980), Batten and Boyce (1987), and Erlander and Stewart (1989), among others.

In the social sciences, the gravity concept dates back at least to the work of Carey (1858) who, in his three-volume essay on the social sciences, repeatedly drew physical analogies to human behavior. In particular, he emphasized the tendency of mankind to "gravitate to his fellow man" (Carey, 1858, Vol. 1, p. 42), and identified the direct effects of mass (size of attraction) and the inverse effects of distance. Similar ideas were developed by Ravenstein (1885, 1889) in his studies of migration. While there appear to have been certain early efforts to formalize these relationships [such as the work of Lill (1891) on railway traffic, cited by Erlander and Stewart (1989)], the formal analogy with Newtonian physics is believed to have first been proposed by Young (1924) in his study of the movement of farm populations. In particular, Young observed that while human migration "does not lend itself to exact mathematical formulation", it approaches the formula

$$M = kF/D^2, \tag{1}$$

where k is a constant, M is the movement of population, F is the intensity of the attraction of any other community, and D is the distance to the other community' (Young, 1924, p. 88). This formulation was subsequently applied by Reilly (1929, 1931) to study retail market areas. However, it was Stewart (1941) who carried through the Newtonian analogy to its limit and proposed a theory of 'demographic gravitation'. In particular, Stewart postulated that the *interaction*, T_{ij}, between population centers i and j was proportional to the *demographic force* between these centers, as defined by the product of their population masses, P_i and P_j, divided by the square of the distance, d_{ij}, between them, i.e., that

$$T_{ij} = GP_iP_j(d_{ij})^{-2}, \tag{2}$$

where G is the *demographic gravitational constant.* This basic relation was later refined by Stewart (1948, 1950) and Dodd (1950) to include weights,

$$T_{ij} = G(w_iP_i)(w_jP_j)(d_{ij})^{-2}, \tag{3}$$

reflecting the heterogeneity of population masses (in a manner consistent with Newtonian gravitation).

But in spite of the suggestive nature of this physical analogy, its theoretical underpinnings remained very weak. Moreover, unlike Newton's 'law of universal gravitation' which can safely be accepted on the basis of its empirical accuracy alone, the same has never been true of 'demographic gravitation'. Indeed from the very outset, the empirical accuracy of (2) [or (3)] was called into question. Initial criticism focused on the 'correct exponent' for distance. Stewart himself proposed an alternative formulation involving a unit exponent, reflecting *demographic energy* rather than demographic force. This same formulation was also explored by Zipf (1946, 1949), who applied it to a wide variety of socio-economic flows. But it was strenuously argued by many others [including Iklé (1954), Carroll (1955), and Carroll and Bevis (1957)] that this exponent should be treated as a statistical parameter which may vary from one context to another. Similar arguments were also made with respect to the implicit unit exponents on the population masses, P_i and P_j [as for example in Anderson (1955, 1956) and Carrothers (1956a)], leading to the following extended class of gravity models [as for example in Isard and Bramhall (1960, p. 543)]:

$$T_{ij} = G(w_iP_i^{\alpha})(w_jP_j^{\beta})(d_{ij})^{-\theta}, \tag{4}$$

where the population weights, w_i, w_j, and exponents, α, β, together with the distance exponent, θ, are all treated as statistical parameters to be estimated.

But while this parametric extension of course increased the empirical flexibility of the gravity model form, it also served to underscore the lack of any theoretical underpinnings. Indeed, as observed in the introductory remarks above, the intuition underlying all these models remains the same, namely that population interactions should vary directly with respect to population size and inversely with respect to distance. Following this intuition, it has been proposed by many authors [as for example in Hua and Porell (1979) and Sen and Sööt (1981)] that expression (4) is best viewed as simply one possible specification of the following general class of *gravity models,*

$$T_{ij} = A(i)B(j)F(d_{ij}), \tag{5}$$

where $A(\cdot)$ and $B(\cdot)$ are unspecified origin and destination *weight functions* [which may involve locational attributes other than population size, and which contain any relevant dimensional constants such as G in (4) above], and where $F(\cdot)$ is an unspecified *distance deterrence function* [possibly involving generalized measures of distance]. While even more general classes of models can be constructed [as in Chapter 2 below], the form in (5) will suffice for purposes of this introductory discussion.

Within this broader modeling context, interest has largely focused on the specification of the distance decay function, F, which is hypothesized to summarize all effects of *space* on interactions. In particular, the above specification

$$F(d_{ij}) = (d_{ij})^{-\theta} \tag{6}$$

in (4) is often designated as the *power deterrence function* with *distance sensitivity parameter,* $\theta > 0$. A large number of alternative specifications of F have been proposed [as detailed in Chapter 2 below]. But for our present purposes it suffices to list only one additional specification which has generated a great deal of interest, namely the *exponential deterrence function,*

$$F(d_{ij}) = \exp[-\theta d_{ij}], \tag{7}$$

in which θ may again be interpreted as a positive *distance sensitivity parameter.* This specification of F appears to have first been studied by Kulldorf (1955), who found it to perform better than the power deterrence function with respect to migration data [as later supported by the work of Morrill and Pitts (1967) and others]. However, primary interest in the exponential deterrence function has centered on its theoretical significance (as discussed below). Hence with this brief introduction to classical gravity models, we now turn to a consideration of the many theoretical arguments which have been put forth in support of these models.

Theoretical Approaches to Gravity Models

In view of the enormous interest generated by gravity models since the 1940's, it is not surprising that a host of different theoretical approaches to these models have been proposed. Hence in the present discussion we cannot provide an exhaustive listing of all such arguments. Rather, the intent here is to illustrate the main types of arguments which have been put forth, and to emphasize their wide diversity. [For additional discussion and critiques of various theoretical approaches to gravity models, see Beckmann and Golob (1972), Hua and Porell (1979), Erlander (1980), and Erlander and Stewart (1989), and Fotheringham and O'Kelly (1989), among others]. We begin with a brief consideration of certain deterministic approaches, and then focus on the probabilistic approaches which are of more interest for our purposes.

Deterministic Theories

In view of the Newtonian origins of gravity models, it is appropriate to begin by considering possible theoretical links between Newton's law of universal gravitation and Stewart's notion of demographic gravitation. Interestingly enough there appear to have been remarkably few efforts in this direction. Part of the reason for this, as observed by Schneider (1959) and Catton (1965), is that Newton's theory involves implicit equations of motion for physical masses which have no direct interpretation for cities [i.e., cities are rarely observed to 'move' in physical space]. However, if one considers more general spaces of attributes for cities (including for example, per capita income levels and employment opportunities) then it is quite meaningful to consider equations of motion for cities in such spaces. At least one serious attempt has been made in this direction by Griesinger (1979), who postulated the existence of a continuous *utility field* governing the movement of behaving units within an abstract space of attributes (which plays essentially the same role as gravitational potential in Newton's theory). In particular, the gradient of this utility field is interpreted as a *psychological force* (analogous to Stewart's demographic force) which induces movement by behaving units in the direction of highest utility increments. Hence, if this utility field is of the same mathematical form as Newton's gravitational potential, then one obtains precisely the Stewart model. But, while Griesinger does point out that the Newtonian potential function exhibits an interesting 'maximal smoothness' property with respect to its gradient field, the underlying behavioral rationale for this specification remains unclear. Moreover, the fundamental notions of 'utility fields' and 'psychological forces' remain undefined in an operational sense. Hence while these ideas are no doubt very suggestive, their practical relevance for spatial interaction modeling remains to be seen. [See also the comments by Genosko and Hasl (1981) and reply by Griesinger (1981)].

A more direct utility-theoretic approach to gravity models was first proposed by Niedercorn and Bechdolt (1969), and has since generated a great deal of interest [see for example, Mathur (1970), Niedercorn and Bechdolt (1970, 1972), Allen (1972), Niedercorn and Moorehead (1974), Choukroun (1975), and Colwell (1982)]. The essential idea of this approach is to model spatial interaction behavior within the classical microeconomic paradigm of utility-maximizing choice behavior subject to an appropriate budge constraint. In particular, it can be shown that under various forms of utility functions, the resulting 'demand function' for interaction with spatial opportunities takes the form of a gravity model. An alternative approach first proposed by Beckmann and Wallace (1969) [see also Golob and Beckmann (1971)] is to model interaction costs (travel time, etc.) as a form of 'disutility' which can be incorporated directly into the utility function. Finally, a closely related approach first proposed by Isard (1975b) [and subsequently axiomatized by Smith (1976a, 1976b)] is to model such disutilities as a form of subjective 'spatial discounting' which diminishes the value of interaction (in a manner analogous to subjective time discounting of future consumption). The difficulty with these approaches from a theoretical viewpoint is that each requires a host of strong assumptions about *individual* choice behavior. Moreover, from a practical viewpoint, these assumptions are generally difficult to test in any systematic way. [In particular, the classical microeconomic methods of revealed-preference analysis typically require observations on price and/or income changes which are difficult to observe in most spatial interaction contexts.]

Probabilistic Theories

With this in mind, a number of more flexible *probabilistic* approaches to individual spatial behavior have been proposed. While the elements of a probabilistic approach are evident in even the earliest studies of the empirical 'goodness of fit' of gravity models [beginning with work of Reilly (1929, 1931), Stewart (1941, 1948) and Zipf (1946, 1949)], the first formal statement of a statistical hypothesis was given by Dodd (1950), who postulated a gravity-type functional form for the 'expected interactance' between spatially separated populations. But the first probabilistic derivation of such an hypothesis was given by Carroll and Bevis (1957), who postulated that in the absence of any spatial separation effects on behavior, the probable destination of any trip should not depend on the origin of the trip [as developed in detail in Example 1 of Chapter 2 below]. This probabilistic hypothesis (together with simple proportionality hypotheses relating expected trip activity to population sizes) yields an explicit test of the null hypothesis that interaction frequencies are not influenced by distance. However, if this hypothesis can be rejected, then further assumptions are required in order to specify the form of this dependence. [In particular, Carroll and Bevis (1957) assumed a log-linear form for this dependence and obtained the classical power deterrence function in (6) above.]

Hence the most relevant probabilistic theories from our viewpoint are those in which the form of this distance dependence is derived entirely from explicit *behavioral* hypotheses. The earliest (and still one of the most interesting) efforts in this direction was inspired by the work of Stouffer (1940), who proposed a modification of the classical gravity model in which the relevant notion of behavioral distance is measured in terms of 'intervening opportunities' between locations. Subsequently, it was shown by Schneider (1959) that a continuous version of the Stouffer model could be derived from a simple 'search theory' of interaction behavior. A discrete version of this model was also developed by Schneider for use in the Chicago Area Transportation Study (1960). This simple search model yields distance deterrence functions of the exponential type in (7) above [as developed for a simple class of discrete search models in Example 4 of Chapter 2 below]. More general types of deterrence functions are obtainable from this basic model by introducing heterogeneous populations of actors exhibiting different types of search behavior [as shown for the continuous version of the Schneider model by Harris (1964) and Choukroun (1975)].

An alternative probabilistic approach to modeling spatial interaction behavior, based on Luce's (1959) theory of individual choice behavior, was first proposed by Ginsberg (1972), and given an axiomatic formulation by Smith (1975). In particular, it was shown by Smith (1975) that this approach allows behavioral notion of perceived distance to be treated as endogenous, and thus provides a statistical framework for testing specific perceived distance hypotheses, such as the 'intervening opportunity' hypothesis of Stouffer (1940) mentioned above. Subsequently, sharper theories of this type have been developed by applying random utility theory to spatial choice behavior. In particular, the multinomial logit model developed by McFadden (1973) and others is directly applicable to spatial choice behavior, and (under suitable utility hypotheses) yields gravity models with exponential deterrence functions [as discussed by Anas (1981), Batten and Boyce (1987), Fotheringham and O'Kelly (1989), and many others].

A second type of random utility approach, based on the simpler 'satisficing' concept of Simon (1957), was first proposed by Tellier and Sankoff (1975). In this approach, individual decisions to interact are modeled by a probabilistic 'threshold' variable depending on relevant measures of spatial separation. This type of threshold interaction model was subsequently extended by Smith (1988), and is developed in detail in Example 2 of Chapter 2 below.

A final class of probabilistic theories, based on statistical equilibrium concepts from statistical mechanics, was first proposed by Wilson (1967) and later extended by many others [including Wilson (1970), Charnes *et al.* (1972), Fisk and Brown (1975), Snickers and Weibull (1977), Roy and Lesse (1981), Fisk (1985) and Smith (1988, 1989)]. The essence of this 'most probable state' approach is to show that very simple probabilistic hypotheses about micro interactions can lead to very regular patterns of macro interaction behavior. In particular, if probabilistic variations in micro interactions are assumed to depend only on average interaction costs and activity levels, then the overwhelmingly most probable patterns of macro interaction frequencies are asymptotically approximated by gravity models of exponential type. Moreover, it turns out that under additional statistical independence hypotheses about individual interaction events, these macro interaction frequencies are *exactly* representable by gravity models of exponential type [as shown in Section 2.4.3 of Chapter 2 below]. These results thus suggest that at the macro level, gravity models with exponential deterrence functions may in fact arise from a wide range of micro interaction behavior. With these observations in mind and because the exponential model may be general enough for most practical purposes (see Section

5.1), much of the present book is devoted to the analysis of exponential gravity models.

Minimal Theories

It should be clear from the above overview of deterministic and probabilistic theories that gravity models can be derived from a variety of different behavioral assumptions. Hence, as mentioned at the beginning of this introduction, these many alternative sets of assumptions appear to raise more questions about gravity models than they answer. With this in mind, the approach adopted here proceeds in the opposite direction. In particular, we shall start with the model forms themselves and seek conditions on interaction probabilities which are both necessary and sufficient for the existence of such model representations. The necessity of such conditions ensures that they must be satisfied by every interaction theory leading to these model forms. Moreover, the sufficiency of such conditions ensures that the given model forms are derivable from these conditions alone. Hence each set of necessary and sufficient conditions can in this sense be regarded as a *minimal theory* of the given model form.

This axiomatic approach to the study of model representations is by no means new to the literature. With respect to deterministic theories of choice behavior, axiomatic formulations of utility-maximizing behavior have a long history in the behavioral sciences [as summarized for example in Fishburn (1970) and Hurwitz *et al.* (1971)]. Specific applications of these results to utility-maximizing spatial interaction behavior can be found in Choukroun (1975) and Smith (1976a, 1976b). With respect to probabilistic theories of choice behavior, the axiomatic approach goes back at least to the seminal work of Luce (1959) [and is summarized in a number of works, including Luce and Suppes (1965) and Luce (1977)]. As mentioned above, applications of Luce (1959) to spatial choice behavior include the work of Ginsberg (1972) and Smith (1975) [see also the discussion of the independence-of-irrelevant-alternatives axiom (axiom IIA) toward the end of Section 2.2.3 below]. Additional examples of axiomatic formulations of probabilistic spatial interaction behavior are summarized in Hua and Porell (1979) and Smith (1987). Hence the central objective of Chapters 2 and 4 below is to develop these results within a unified probabilistic framework.

Methods of Application

While, as the discussion above demonstrates, considerable attention has been paid to the theoretical development of the gravity model and to its algebraic form, relatively less attention has been directed towards the estimation of parameters of the model. Indeed, for some of the earlier versions of the model, little need seems to exist to develop procedures specific to the gravity model. For example, the only parameter G in (2) requires straightforward arithmetic (as mentioned above, there was some question as to whether statistical estimation of certain other parameters was even desirable). Some other models appeared to have been formulated so that linear least squares could be used for parameter estimation.

The first series of efforts to develop estimation methods specific to the gravity model appears to have come from transportation planners seeking to develop forecasting tools in the wake of the commencement of the building of the Interstate Highway System in the United States. This early work (see, for example, BPR, 1965, FHWA, 1974), has, perhaps not been adequately recognized for its innovativeness and for the fact that it clearly laid the groundwork for the contemporary gravity model. The model postulated was equivalent to the general form (5) in a straightforward way. Perhaps prompted by the form of the model and also by the need to accommodate very large data sets, an estimation method was put forward which (but for some intermediate smoothing steps) was surprisingly similar to a maximum likelihood approach.

Subsequently, a number of other gravity model specific estimation procedures were developed. In order to discuss them, it is necessary to consider forecasting in general, as we do next.

Assumptions Underlying Forecasting

Typically, two time periods are involved in forecasting: *a base period* which is a time period in the past for which observations are available, and *a forecast period* which is a period for which the forecasts need to be constructed. The forecast period is usually sometime in the future, although this need not necessarily be true. Sometimes, the forecast period may chronologically precede the base period — in which case we get back forecasts. In order to make a forecast, it is assumed that some aspect of the base period remains unchanged into the forecast period (or changes in some exogenously well defined way). Assumptions regarding which aspects remain unchanged determine different approaches to forecasting as well as (very frequently) to different forecasts.

An important case in this context is a forecasting procedure given by Wilson (1967). His theoretical contribution has already been mentioned in an earlier section, and his method of forecasting is consistent with his theoretical derivation. Wilson suggests computing average generalized cost for each trip from base period data. Using this information and also exogenously computed $T_{i\oplus} = \sum_j T_{ij}$ and $T_{\oplus j} = \sum_i T_{ij}$, he estimates the parameters $A(i)$, $B(j)$ (for all origins i and all destinations j) and θ of the exponential gravity model $T_{ij} = A(i)B(j)\exp[-\theta c_{ij}]$ for the *forecast period.* The method essentially consists of maximizing the limit of a probability subject to constraints: one constraint for the average cost and one each for the different $T_{i\oplus}$'s and different $T_{\oplus j}$'s. The estimates of $A(i)$'s, $B(j)$'s and θ's then emerge as Lagrange multipliers and are thus estimated. It is interesting to note that the model derivation and parameter estimation occur within a single step. No attempt is made to estimate base period parameter values, although the method given (by Wilson) could have been used for this purpose.

Similarly, Erlander (1980) uses a base period estimate of $T_{ij} \log[T_{ij}]$ along with exogenous $T_{i\oplus}$'s and $T_{\oplus j}$'s to estimate the same parameters — again for the forecast period. In fact, even the BPR approach is somewhat ambivalent on whether the estimation of parameters should be for the base or the forecast period. It estimates parameters for the base period, but then suggests checking if the empirical distribution (cumulative distribution function) of trip travel times for the base and forecast period are the same. (Notice that if the empirical distribution function of travel times are the same for the base and forecast periods, average trip travel times will be the same also).

The position taken in this book is different. Axioms are provided (Chapter 4) under which some or all parameters of the gravity model remain unchanged from the base to the forecast period. Forecasting would then consist of estimation of parameters for the base period followed by the use of these estimated values to construct forecasts for the forecast period. This approach, which is a more traditional statistical approach, is implicit in the works of a number of authors including Hyman (1969), Cesario (1975), Sen and Sööt (1981) and Fotheringham and O'Kelley (1989), among others.

Because of this difference, methods of application as well as nomenclature used by Wilson need to be revisited. Since in Wilson's work, theoretical development of the algebraic form of the model, parameter estimation and estimation of forecasts occurs in a single step for the forecast period, Wilson identified *four kinds of gravity models*. If $T_{i\oplus}$'s and $T_{\oplus j}$'s are not available, the corresponding constraints and the corresponding parameters which are the Lagrange multipliers are dispensed with, and we get *unconstrained gravity models*. Similarly if only one set, $T_{i\oplus}$'s or $T_{\oplus j}$'s is available, the other set of constraints is removed along with the corresponding parameters, and we get either *origin constrained* or *destination constrained models*. If both sets are available, we get a *doubly constrained model*.

Since in our approach, the data used is for the base period and observations on each T_{ij} are assumed available, we estimate a single gravity model. However, the application of model varies depending on which parameters are assumed to remain invariant into the forecast period. If $A(i)$, $B(j)$ and the function F in (5) remain unchanged then prediction is straightforward. If only $A(i)$ and F or $B(j)$ and F remain unchanged, then we would need additional information to make forecasts. For example if $A(i)$, F and forecast period $T_{\oplus j}$'s and d_{ij}'s are available, forecasts T_{ij} can be generated by using

$$B(j) = T_{\oplus j} / \sum_i A(i)F(d_{ij}), \tag{8}$$

which is a direct consequence of (5). Similarly with $B(j)$'s, F and $T_{i\oplus}$'s known

$$A(i) = T_{i\oplus} / \sum_j B(j)F(d_{ij}) \tag{9}$$

may be invoked. If neither $A(i)$'s nor $B(j)$'s are known, but, instead, $T_{i\oplus}$'s and $T_{\oplus j}$'s are known, (8) and (9) may be used alternately and iteratively to generate the $A(i)$'s and $B(j)$'s (Section 5.3.1). Thus, unlike Wilson's treatment, the model does not change; only the way it is used. We call the model (5), with $A(i)$ replaced by the right side of (8), an origin gravity model and if (9) is incorporated into (5), a destination gravity model.

Because of its importance, a description of our overall forecasting method bears restatement. Based on the theoretical development of the first four chapters, we take the following approach to forecasting: Using base period data, estimates are constructed. Some of these estimates are assumed to remain the same into the forecast period. These, perhaps augmented by exogenous forecasts, are then used to construct the forecasts themselves.

Use of Random Variables

It also needs to be emphasized that the flow N_{ij} between every origin i and every destination j is considered to be a random variable in this book. Thus, these random variables are converted into a set of random variables which are the estimators or estimates. These estimates, augmented by additional variables yield other random variables which are the forecasts. It turns out that many of the estimates are at least approximately normally distributed. Therefore, their distribution is largely characterized by their expectation and covariance matrix. The estimation of these means and covariance matrix constitute a large part of the work of the last two chapters of this book.

This overall approach is very standard in statistical theory, and not mired in controversies. Criteria for quality of estimates are well established and may be invoked to decide between procedures.

As mentioned earlier, Wilson's derivation and estimation method were both asymptotic. Thus, a point estimate of parameter values was appropriate, since all probability would condenses on this estimate asymptotically. This is also true of most other early derivations of the model.

Finally, multiple costs and, consequently, multiple parameters are accommodated within our development, so that the model used for estimation purposes is of the form

$$T_{ij} = A(i)B(j)\ \exp[\boldsymbol{\theta}^t \mathbf{c}_{ij}], \tag{10}$$

where observed flows N_{ij} are independent Poisson variates with $T_{ij} = \mathrm{E}[N_{ij}]$. The estimation of such a model was first presented in Sen and Sööt (1981), and indeed many of the estimation methods given in this book constitute generalizations, clarifications and elaborations of the procedures given in that paper. Maximum likelihood estimation is covered in Chapter 5 and least squares, both linear and non-linear, in Chapter 6.

Computation of Estimates

It has been said that the actual power of a test is its theoretical power times the number of times it is used (Lehmann, 1959). The computational demands of some gravity model estimation procedures have been a problem in the past (see Batty, 1976). However, better computational algorithms (e.g., see Yun and Sen, 1994, Evans and Kirby, 1974, Gray and Sen, 1983) together with faster computers have largely alleviated these problems. Today, better knowledge of existing routines and greater user-friendliness of software are perhaps all that are required to render spatial interaction computations largely routine.

However, since this was not always the case, some estimation programs were written with the aim of reducing computation time. Many of these programs are described only in users' manuals and appear to be of dubious ancestry. Often the only way available to assess them is to use them on a broad class of artificially generated data sets for which the true parameter values are known. Many of these programs do not fare too well. In fact, some of them perform quite poorly. Thus it is important that in applications only procedures be used whose performance has been assessed and found to be acceptable.

Several such procedures have been given in this book but we draw the readers' attention particularly to two of them. In Chapter 5, a procedure called the modified scoring procedure is presented. This is a method for computing maximum likelihood estimates. It is very fast, provides very accurate results and can handle very large data sets. The other procedure is a least squares procedure, which we have called Procedure 1. Unlike several other least squares procedures, it is faster than the modified scoring procedure and is simple enough for even hand computations for moderate-sized data sets. It has excellent properties when either the T_{ij}'s are large or when the T_{ij}'s are not too small and the number of origins and destinations are large. While the properties of estimates it provides are still inferior to the excellent properties possessed by maximum likelihood estimates, it remains a more efficient alternative to the modified scoring procedure.

This book provides a consistent and comprehensive treatment of the gravity model — from its theoretical foundations, through estimation procedures to computational algorithms — and the computational procedures appear to perform very well.

Part I

Theoretical Development

CHAPTER 1

Spatial Interaction Processes: An Overview

1.1 Introduction

In this chapter we introduce the basic theoretical framework to be employed throughout the rest of the book. The formal details of the framework will be developed in a more abstract theoretical setting in Chapter 3 below. Hence the objectives here are to introduce the main concepts in an informal manner, and to illustrate their meaning in terms of simple examples. To do so, we begin in Section 1.2 below with a consideration of the basic theoretical perspectives embodied in the present approach to spatial interaction behavior.

1.2 Theoretical Perspectives

Given the complexity of human beings, theories of human behavior tend to focus only on selected aspects of that behavior. Depending on the questions of interest, some theories focus on micro aspects of individual behavior, while others focus on macro aspects of overall population behavior. Some are concerned with temporal aspects of behavior, while others are more concerned with static or steady-state behavior. Some postulate deterministic regularities, while others adopt a probabilistic viewpoint. Hence, to illuminate the possible advantages and limitations of the present theoretical framework, it is useful to consider each of these perspectives in turn.

1.2.1 MACRO VERSUS MICRO THEORIES

The present framework is designed to encompass the types of behavioral theories of spatial interaction consistent with the class of *gravity models* outlined above. Hence the theoretical perspectives implicit in this work are derived mainly from the nature of gravity models themselves. In particular, as mentioned in the discussion of probabilistic theories in the Introduction, these models have proved to be most successful in describing *macro patterns* of spatial interaction behavior, involving large populations rather than single individuals [see also the discussions in Isard and Bramhall (1960, Section 11.E) and Alcaly (1967)]. For example, in the context of intra-urban shopping behavior, it is often observed that trip frequencies between origin and destination zones exhibit regularities which are well described by gravity models. But if one focuses on the shopping behavior of any given individual household, then such models are known to be far less successful. This is not because gravity models are without meaning at the micro level. Indeed, a variety of gravity-type models of individual choice behavior have been proposed in the literature [including the many deterministic utility maximizing models starting with Neidercorn and Bechdolt (1969), and the probabilistic models based on Luce's (1959) theory of individual choice behavior, as summarized in the Introduction]. But such attempts have for the most part met with little empirical success [as discussed for example in Luce (1977) and Hua and Porell (1979)]. Part of the reason for this failure is that most individual decision situations (such as shopping decisions) are influenced by a host of factors which are specific to the given individual. Hence meaningful results can only be obtained by employing more detailed models incorporating a range of individual-specific information (such as individual shopping needs, shopping habits, knowledge of shopping alternatives, access to transportation modes, etc.).

On the other hand, the success of gravity models in describing overall interaction frequencies does *not* imply that individual-specific information is without meaning at the macro level. Indeed, if one had available a reliable micro model of each individual shopper, there is little doubt that one could obtain a very accurate picture of macro shopping patterns. Rather, the relevant question for practical purposes is *how much information* about individual behavior is required in order to obtain reasonable estimates of overall shopping frequencies. From this viewpoint, the real power of gravity models has been their ability to yield such estimates on the basis of only very aggregate information about levels of travel activity (such as trip totals for each origin and destination zone) and very rough measures of spatial separation (such as average travel distances and travel times between zones). Hence the viewpoint adopted here amounts to an application of *Occam's razor,* namely, that the best model to use is the simplest model which works.

It should be emphasized, however, that this macro approach places severe limitations on the types of behavioral questions which can be addressed. For even though gravity models can in principle be disaggregated to any degree, such disaggregate analyses could require very large population samples to obtain statistically meaningful results. Hence the types of questions which can most easily be addressed within the present modeling framework are those which relate to overall changes in spatial interaction patterns. In the context of travel behavior, for example, gravity models can usefully be employed to study the overall impacts of changes in the relevant structure of spatial separation (such as road network improvements or the imposition of new tolls). Similarly, one can study the impact of factors affecting the relative distribution of population and interaction opportunities (such as population migration or the construction of new shopping facilities). More generally, one can study the overall impacts of those system changes which can be directly reflected by shifts in the aggregate explanatory variables of such models. [The axiomatic approach to gravity models developed in Chapter 2 is designed to facilitate the analysis of such questions, as discussed in Section 2.2.3 below].

1.2.2 Static versus Dynamic Theories

A second important theoretical perspective implicit in the present framework relates to the treatment of time. Gravity models have traditionally been most successful in describing stationary patterns of interaction behavior in time periods not involving significant structural changes. Typical examples include the stationary patterns of travel flows, mail flows, phone calls, or commodity shipments occurring during such periods. With this in mind, the temporal perspective adopted here is essentially *static* in nature. In particular, the behavioral impacts of structural changes are addressed with a *comparative static* framework, involving a shift from one stationary (or equilibrium) state to another.

It should be pointed out however, that gravity models have been employed in dynamical contexts. In particular, such models have been used to describe state transitions in dynamical interactive Markov processes [as for example in DePalma and Lefevre (1983), Papageorgiou and Smith (1983), Nijkamp and Poot (1987), Boots and Kanaroglou (1988), and Haag (1989)]. But even in such dynamical applications, it is generally assumed that the process is only observed at discrete time points, and that state transitions involve interactive decision processes which are sufficiently 'fast' to allow the resulting interaction patterns to be treated as realizations of a stationary (or equilibrium) state. Hence the present framework can be applied to analyze all such behavior, and in particular, to model the corresponding structure of state transitions.

1.2.3 Probabilistic versus Deterministic Theories

Finally, from a methodological viewpoint, perhaps the most important theoretical perspective of the present approach is the *probabilistic* nature of the framework developed. As with the analysis of many types of complex phenomena, our present analysis focuses on the identification of overall trends in human spatial interaction behavior. Such trends may in certain cases be sufficiently strong to permit deterministic model formulations. For example, deterministic network equilibrium models of travel behavior can be quite effective in describing typical flow patterns of commuter traffic, where time minimization is the dominant mode of behavior [as developed for example in Sheffi (1985)]. But such strong regularities in human spatial interaction behavior are clearly the exception. Even in the case of traffic patterns, there is usually a complex mixture of motives governing travel behavior which lead to probabilistic model formulations [as developed for example in Daganzo and Sheffi (1977) and Smith (1983, 1988)]. More generally, most spatial interaction phenomena tend to involve a multitude of individual decisions which defy any attempt at deterministic description. Hence to identify meaningful trends in such phenomena, it is generally appropriate to adopt a probabilistic viewpoint in which such trends can be modeled explicitly in terms of statistical averages.

With this in mind, we begin in the next section with an overview of the general analytical framework to be employed throughout the book. A formal development of this framework is given in Chapter 3 below.

1.3 Analytical Framework

The behavioral phenomena of interest in the present work are describable in their most general terms as interactions between populations of *actors* and *opportunities* distributed over some relevant space. More specifically, we are interested in those *patterns* (ensembles) of spatial interactions which may occur during some relevant time period. Such interactions may involve movements of people from place to place, such as daily traffic flows in which the relevant actors are individual travelers (commuters, shoppers, etc.) and the relevant opportunities are their destinations (jobs, stores, etc.). Similarly, one may consider annual migration flows, in which the relevant actors are migrators (individuals, family units, economic firms, etc.) and in which the relevant opportunities are their possible new locations. Interactions may also involve flows of information, such as telephone calls or mail deliveries. Here the callers or letter writers may be the relevant actors, and the possible receivers of calls or letters may be viewed as the relevant opportunities.

While all these examples of interactions are unilateral (directed) in nature, one may also consider a range of bilateral (symmetric) interactions. Examples here include the formation of friendships or marriages between individuals, economic contracts between firms, or even treaties between nations. In such bilateral interactions, the distinction between 'actors' and 'opportunities' is generally quite arbitrary. Hence these terms serve only as convenient labels for distinguishing the relevant parties. Such bilateral examples also illustrate the relevance of nonphysical spatial separation. In marriage, for example, differences in cultural or religious backgrounds may affect the likelihood that any given individuals will marry. Hence such attributes of individuals can in principle constitute relevant dimensions of spatial separation.

With this range of examples in mind, the purpose of the present section is to develop an analytical framework in which all such spatial interaction behavior can be studied in a unified way. To do so, we begin in Section 1.3.1 with a development of those measures of *spatial separation* which constitute the fundamental spatial data for the analysis. In Section 1.3.2 we then introduce certain *spatial aggregation* assumptions which will allow us to operationalize the macro-statistical approach outlined in Section 1.2 above. Finally, to facilitate the analysis of these spatial aggregates, we introduce a number of simplifying *structural independence* assumptions in Section 1.3.3.

1.3.1 Measures of Spatial Separation

The fundamental question of interest in the present analysis is how interaction behavior is influenced by various types of spatial separation between actors and opportunities. In particular, we are interested in identifying those specific types of spatial separation which tend to impede or enhance the likelihood of interactions in a given behavioral context. The most obvious types of separation of course involve physical space. For example, the physical separation between the locations of shoppers and of stores will surely influence the relative likelihoods among various possible shopping trips. However, such separation relationships need not involve physical space. For example, in the marriage illustration above, one may speak of the separation between potential marriage partners in an abstract space of 'cultural locations' or 'religious locations'.

To model this variety of separation possibilities in a simple manner, it is convenient to focus directly on numerical measures of separation, and to treat the underlying location space for actors and opportunities as implicit. To be more precise, we now consider a given set of *actors,* $\alpha \in A$, and *opportunities,* $\beta \in B$, and postulate the existence of a finite set of relevant *separation attributes,* $k \in K$, between actors in A and opportunities in B. Each separation attribute, k, is taken to be representable by a numerical function, c^k, of actor-opportunity pairs, where $c^k(\alpha, \beta)$ denotes the degree of *k-type separation* between α and β, and where higher values denote greater separation. In shopping behavior, for example, relevant separation attributes of physical space typically include the *travel time,* $c^1(\alpha, \beta)$, the (out of pocket) *travel cost,* $c^2(\alpha, \beta)$ between the locations of α and β, and even the number of *intervening opportunities,* $c^3(\alpha, \beta)$, between shopper α and store β [as developed in more detail in Example 4 of Chapter 2 below]. An additional behavioral measure of separation which is often highly correlated with physical separation is actor α's *information level,* $c^4(\alpha, \beta)$, about opportunity β [as developed in Section 2.5.3 of Chapter 2 below]. However, even in interaction situations involving physical movement through space, there may exist many relevant measures of nonphysical separation. Consider for example a shopping situation in which the relevant opportunities consist of a set, B, of ice cream stores, each offering possibly different types of ice cream. In this context, one may expect that (other things being equal) a shopper, $\alpha \in A$, will tend to frequent the ice cream store, $\beta \in B$, which offers his or her favorite types of ice cream. Hence we may imagine an implicit attribute space of ice cream qualities (say in terms of butter fat content levels), and may consider candidate measures of *qual-*

ity separation, $c^5(\alpha,\beta)$, such as the absolute difference between the butter fat content level most preferred by shopper α and the butter fat content level offered by store β. More generally there may exist a multidimensional space of product attributes which differentiate (separate) various shopping opportunities. [See for example, Shocker and Srinivasan (1979) and Choi (1988)]. Such nonspatial dimensions of separation are also present in interactions between individuals and groups [as exemplified, for example, by the measures of perceived *social distance* discussed in Laumann (1966)]. Even interactions between nations can be influenced by separation measures reflecting various types of *political distance* and *cultural distance* [as studied for example by Rummel (1972)]. Additional illustrations of social and cultural distance concepts can be found, for example, in Isard (1960, pp. 542–544) and Gatrell (1983, Chapter 3).

Finally, one may also consider measures of separation with respect to *nominal* (or *categorical*) attributes. In the ice cream illustration above, one may for example treat ice cream flavors as a relevant nominal variable, and consider a zero-one valued measure of *flavor separation,* $c^6(\alpha,\beta)$, which takes the value zero (representing no separation) if β carries the favorite flavor of shopper α, and takes the value one otherwise. Similarly, in the marriage illustration, one may represent *religious separation,* $c^8(\alpha,\beta)$ [or *racial separation,* $c^9(\alpha,\beta)$] by a zero-one variable which assumes the value zero if potential marriage partners, α and β, are of the same religious faith [or the same race], and assumes the value one otherwise. [Other examples of nominal attributes can be found in Laumann (1966) and Rummel (1972)]. More generally, one may often extend the notion of nominal attributes to *tree structures* of attributes, and thereby define a much richer class of associated separation measures in terms of the many concepts of *similarity* which have been developed in the literature for analyzing tree structures [as summarized for example in Shepard (1974) and Tversky (1977)].

With these many examples in mind, it is important to distinguish between those notions of separation which are quantifiable in terms of meaningful units of measurement (such as travel time and travel distance), and those which are only ordinal in nature (such as ordinal measures of social or political distance). While both types of separation attributes are relevant from a behavioral viewpoint, their consequences for *modeling* interaction behavior are quite different. For example, in the classical *gravity* model of expression (2) in the Introduction, it is implicitly assumed that distance, d_{ij}, is measurable in physical units. For if d_{ij} were to represent an ordinal measure of, say, 'social distance' between population centers i and j, then since the functional relation in (2) cannot possibly hold for all ordinal transformations of such distances, this model would fail to have any empirical content whatsoever. However, since physical distance quantities are unique up to a choice of units, it follows that if the functional form in (2) is valid for any choice of units, then (for appropriate choices of the dimensional constant, G) it must be valid for *all* choices of units. Hence, model (2) is said to be a *meaningfully parameterized relation* with respect to physical distance measures [as for example in Pfanzagl (1968, Section 2.6)]. For modeling purposes, it is thus important to distinguish those types of separation attributes which are quantifiable in terms of *extensive measurement,* i.e., which are uniquely measurable up to a choice of units [as developed, for example, in Krantz *et al.* (1971, Chapter 3)]. In so far as interactions can be said to involve expenditures of such quantities (such as miles traveled, hours spent, and money spent), it is convenient to designate such separation attributes as interaction *costs.* [Further discussion of functional specifications is given in Section 2.3 below.]

With these general observations, we now postulate that in any given spatial interaction context, all relevant aspects of spatial separation can be represented by some finite collection of *separation measures,* $\{c^k : k \in K\}$, where the subset of extensive measures are designated as *cost measures.* For each actor, $\alpha \in A$, and opportunity, $\beta \in B$, the vector of values, $c(\alpha, \beta) = (c^k(\alpha, \beta) : k \in K)$, is designated as the relevant *separation profile* for α and β. If all separation measures are cost measures, then $c(\alpha, \beta)$ is designated as a *cost profile.*

1.3.2 Spatial Aggregation Assumptions

In attempting to study the influence of spatial separation on interaction behavior, a number of practical and theoretical problems arise. To appreciate some of these difficulties, it is instructive to consider an idealized hypothetical experiment in which one has available a large number of identical actors, $\alpha_1, \ldots, \alpha_N$, identical opportunities, $\beta_1, \ldots, \beta_N$, and a spatial setting in which it is possible to choose the values of each separation profile, $c(\alpha_n, \beta_n)$, $n = 1, \ldots, N$, at will. In this context, if one were to set all profiles equal to a common profile value, c, and to observe the relative frequency of interactions (versus noninteractions) among the pairs, (α_n, β_n), $n = 1, \ldots, N$, then this statistic might be employed as a point estimate of the probability, $p(c)$, of a potential interaction occurring in the presence of separation profile c. By systematically changing the values of c, one might then study the influence of c on $p(c)$ for this type of actor and opportunity.

But even if we ignore all of the obvious practical difficulties in carrying out such an experiment, the results would generally have little meaning for our purposes. For in most cases of interest, the relevant spatial interaction events do not occur in isolation. In shopping behavior, for example, the important question for shoppers is often not whether to shop, but rather, *where* to shop. More generally, interaction decisions almost always involve a choice among *many* alternatives. Hence the influence of spatial separation on interaction behavior cannot usually be separated from this choice question (as was implicitly assumed in the above experiment). Moreover, in bilateral interaction situations such as marriage decisions or barter transactions, it is clear that both participants make choices, and hence that there exist choice sets for each. Thus the questions of most interest behaviorally tend to involve spatial separation relationships among (potentially large) *sets* of both actors and opportunities.

These effects can in principle be studied by expanding our hypothetical experiment above in the following way. Suppose that we are able to replicate given sets of actor types, $\{\alpha_i : i \in I\}$, and opportunity types, $\{\beta_j : j \in J\}$, as many times as is desired. Then for given spatial arrangement of actor locations and opportunity locations, as summarized by the associated configuration of separation profiles, $c = \{c(\alpha_i, \beta_j) : i \in I,\ j \in J\}$, one could in principle observe the interaction behavior of independent replications of actor and opportunity populations, $(\{\alpha_{in} : I \in I\},\ \{\beta_{jn} : j \in J\})$, $n = 1, \ldots, N$, situated in configuration c. Hence by counting the number of observed interactions, $N(\alpha_i, \beta_j)$, between each actor type, α_i, and opportunity type, β_j, over the N replications, one could employ the corresponding relative interaction frequency, $N(\alpha_i, \beta_j)/N$, to estimate the probability, $p_c(\alpha_i, \beta_j)$, that a randomly observed interaction in this configuration is between α_i and β_j. Finally, by systematically varying the configuration, c, and repeating these replicated experiments under each configuration, one could in principle study the influence of the underlying configuration, c, on each of these interaction probabilities, $p_c(\alpha_i, \beta_j)$.

While this idealized experiment would no doubt yield a rich body of statistical information about spatial interaction behavior, it should be abundantly clear that such experiments are simply out of the question in the social sciences. With this in mind, the practical task at hand is to construct a feasible approximation of this idealized setting which still permits some statistically meaningful inferences to be drawn. To do so, let us again consider the example of urban shopping behavior, and suppose that we are interested in studying the influence of travel-time configurations on behavioral patterns of shopping in a given urban area. In particular, let us postulate that there currently exists a large population of households, $\alpha \in A$, and shopping opportunities, $\beta \in B$, distributed throughout this urban area in some given travel-time configuration, $c = \big(c(\alpha, \beta) : \alpha \in A, \beta \in B\big)$. While it is not possible to study the behavior of all these actors in detail, we can still obtain a reasonable overall picture of interaction behavior in terms of appropriately stratified spatial aggregations as follows. First observe that while individual households have different locations, all households on the same block are locationally very similar in terms of their accessibility to shopping opportunities. More generally, there are many partitions of the household population, A, into finite collections, $\{A_i : i \in I\}$, of spatial subpopulations (residential zones), in which each subpopulation is reasonably homogeneous in terms of accessibility to shopping. Similarly, with respect to shopping opportunities, one may suppose that travel times to all stores on the same block (or in the same shopping center) are essentially identical. Hence there are also many partitions of the population, B, into finite collections, $\{B_j : j \in J\}$, of spatial subpopulations (shopping zones), in which each subpopulation is reasonably homogeneous in terms of accessibility to shoppers. Hence one may attempt to approximate the idealized replication scheme above by partitioning both households and stores into subpopulations which are sufficiently small to ensure spatial homogeneity, and yet which are sufficiently large to allow meaningful statistical sampling.

The choice of appropriate spatial aggregation units of course involves an implicit tradeoff between these two conflicting objectives, since the expansion of any spatial subpopulation to allow larger samples necessarily introduces greater spatial heterogeneity into that subpopulation. It should be equally clear that this fundemental tradeoff is inherent in all statistical aggregation procedures, and hence that a systematic analysis of this issue would take us far beyond the scope of the present work. Thus for our purposes, it suffices to say that the theoretical emphasis in the present work is heavily on the side of *spatial homogeneity*. In particular, it is postulated that all spatial subpopulations of both actors, A_i, and opportunities, B_j, are sufficiently small to allow the spatial separation between them to be approximated by a single representative separation profile, c_{ij}. On the statistical sampling side, our only explicit assumption is that all subpopulations are sufficiently large to ensure a positive probability of at least one interaction between each subpopulation pair, A_i and B_j.

These preliminary observations motivate the following formal scheme for spatial aggregation. Given an actor population, A, opportunity population, B, and set of separation measures, $\{c^k : k \in K\}$, we now define a *spatial aggregation scheme* to consist of a pair of *finite partitions,* $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$, of A and B respectively, together with a set of *aggregation functions,* $\{\Phi^k : k \in K\}$, which assign to each pair, A_i and B_j, a unique *representative separation profile,* $c_{ij} = (c_{ij}^k : i \in I, j \in J)$, of the form:

$$c_{ij}^k = \Phi^k\left[c^k(\alpha,\beta) : \alpha \in A_i,\ \beta \in B_j\right], \quad k \in K. \tag{1.1}$$

In some cases, the choice of spatial aggregation scheme is completely obvious. For example, in the case of travel between major cities, it is reasonable to assume that the separation between all actors in one city and all opportunities in another are identical, and hence that the natural partition of A and B is given by a division into city populations, where A_i and B_i denote the actors and opportunities in city $i \in I$ (and where $J = I$). In this case the relevant spatial separation values, c_{ij}^k, between distinct cities i and j are definable within sharp bounds (and the separation values, c_{ii}^k, are sufficiently small to be treated as zero).

In most cases, however, the appropriate choice of schemes is not obvious, and indeed, can be highly nonunique. Turning first to the choice of aggregation functions, it is again important to distinguish between situations involving extensively measurable interaction costs, and those involving more general measures of separation. For the case of cost variables, c^k, a wide range of aggregation functions are available. The most common, of course, is simply to take the average value of all possible costs which may be incurred in interactions between A_i and B_j. More precisely, if N_i and N_j denote the population sizes of A_i and B_j, respectively, then the *average cost*

$$c_{ij}^k = \frac{1}{N_i N_j} \sum_{\alpha \in A_i} \sum_{\beta \in B_j} c^k(\alpha, \beta), \; i \in I, j \in J, k \in K \tag{1.2}$$

yields a natural representative value [as for example in Beardwood and Kirby (1975)]. However, in the more general case of ordinal measures of separation, c^k, the choice of meaningful aggregation functions is much more limited. A common procedure is to adopt the *median value* with respect to the population of potential interactions between A_i and B_j, which in the present case is simply the median value of $c^k(\alpha, \beta)$ within the population of potential interaction pairs $(\alpha, \beta) \in A_i \times B_j$. Such values constitute invariant representatives of the overall population in the sense that if the median, c_{ij}^k, corresponds to the separation value, $c^k(\alpha, \beta)$, for a given actor-opportunity pair, (α, β), under one numerical representation, then this same correspondence must hold under all numerical representations. An alternative aggregation function which is applicable to all separation measures is simply to focus on individuals rather than separation values, and to set

$$c_{ij}^k = c^k(\alpha_i, \beta_j), \quad i \in I, j \in J, k \in K, \tag{1.3}$$

where $\alpha_i \in A_i$ and $\beta_j \in B_j$ denote choices of 'representative' actors and opportunities with respect to some appropriate criterion.

Turning next to the choice of appropriate partitions, $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$, of the actor and opportunity populations, this choice in most cases involves some implicit measure of within-group homogeneity. Again, in the case of extensively measurable costs, there is a wide variety of such measures. For example, with respect to the average cost values in (1.2) above, a standard 'clustering' procedure involves the identification of partitions which minimize the *total within-group variance,* as measured by the sum of squared deviations of individual interaction costs, $c^k(\alpha, \beta)$, about the mean value, c_{ij}^k, within each group, $A_i \times B_j$. In the case of more general ordinal measures, notions of clustering are much more difficult. One approach is to measure 'homogeneity' of actor populations in terms of the rank correlations (or other appropriate ordinal distance measures) between their individual separation profiles with all opportunities. Alternatively, the more general measures of similarity discussed in Shepard (1974) and Tversky (1977) may in principle also be employed here to define homogeneity of subgroups.

A more detailed analysis of spatial aggregation schemes is beyond the scope of this book. [For further discussion of this topic, see for example Beardwood and Kirby (1975), Masser (1977, 1979, 1981, the last contains a review of the literature), Masser and Brown (1975, 1977, 1978), Masser and Scheurwater Masser, Batey and Brown (1975), Hirst (1977) Openshaw (1977, 1978), Hensher and Johnson (1981, Section 7.2), Findlay and Slater (1981) and Batty and Sikdar (1982)]. Hence we simply postulate the existence of some appropriate aggregation scheme, and focus on the development of statistical models which are in principle applicable to all such schemes. In doing so, however, it should be born in mind that the choice of an underlying spatial aggregation scheme can profoundly influence the nature of the results obtained, and hence should always be regarded as an implicit part of each model developed.

With these general observations in mind, we now introduce certain simplifying terminology which will be employed throughout the book to treat all spatial aggregation schemes. First, since all interactions between members of subpopulations, A_i and B_j, are regarded as indistinguishable events within any given spatial aggregation scheme, we shall usually not refer to individual actors, $\alpha \in A_i$, or opportunities, $\beta \in B_j$. Rather we shall speak simply of an (ij)-*interaction* between subpopulation i and subpopulation j. [Exceptions are illustrated by the threshold interaction processes and simple search processes of Examples 2 and 4 in Chapter 2 below.] When the ordering of i and j is important, as in the case of unilateral interactions, we shall refer to subpopulation i as the *origin* of the interaction, and subpopulation j as the *destination* of the interaction. Hence the finite sets of actor populations, i, and opportunity populations, j, are now designated respectively as the *origin set,* I, and *destination set,* J, for the underlying spatial aggregation scheme. This transportation-based terminology provides a convenient set of labels for spatially separated population aggregates, and is not meant to have behavioral significance other than to indicate the direction of unilateral interactions (when appropriate). Moreover, as with actors and opportunities, the distinction between origin and destination populations is completely arbitrary in the case of bilateral interactions (such as marriages, employment contracts or trade agreements).

With these conventions, the vector of relevant separation values for each origin-destination pair, $ij \in I \times J$, is designated as the (ij)-*separation profile*,

$$c_{ij} = (c_{ij}^k : k \in K), \tag{1.4}$$

and the collection of all (ij)-separation profiles,

$$c = (c_{ij} : ij \in I \times J), \tag{1.5}$$

is designated as the *separation configuration* between origin set I and destination set J. Similarly, if all separation measures are cost measures, then (as in Section 1.3.1 above) we shall speak of *cost profiles* and *cost configurations.* [A more general treatment of separation measures is given in terms of measurable *locational attributes* of interactions in Section 3.3.2 below].

1.3.3 Structural Independence Assumptions

Given this formal setting, our ultimate objective is to develop an operational version of the idealized experiment outlined above, in which one can define meaningful interaction probabilities, $p_c(ij)$, between each origin-destination pair, $ij \in I \times J$, under separation configuration c. By analyzing the dependence of these probabilities, $p_c(ij)$, on the underlying configuration, c, one can in principle draw inferences about the dependence of interaction behavior on spatial separation. With this end in mind, it is appropriate to begin by specifying the relevant set of possible separation configurations in more detail. In doing so, our basic strategy is to develop a formal analytical framework which is broad enough to cover most interaction contexts of interest, and yet which is simple enough to allow a unified analysis of these different contexts. Hence we now make a number of simplifying assumptions which will greatly facilitate the analysis to follow.

First of all, while possible interactions among origin-destination pairs may in general be restricted (as for example when political conditions preclude economic trade between certain pairs of nations), we shall assume for simplicity that interactions between all origin-destination pairs, $ij \in I \times J$, are possible. [While such restrictions could in principle be treated by allowing infinite separation values, we choose to ignore such complications in the present analysis.] Moreover, while the feasible values of separation measures are generally restricted (for example, airline travel distances are surely limited by the size of the earth), we shall assume for simplicity that such values are unrestricted. Similarly, while certain types of separation measures can only assume discrete values [such as the separation measure based on 'intervening opportunities' in Example 4 of Chapter 2 below] we shall generally treat separation measures as continuous variables.

In addition, we shall assume that there exist no *structural dependencies* between different measures of separation. This assumption is particularly critical for separation measures such as travel time and travel cost, which tend to be highly correlated, and yet which are each behaviorally relevant in their own right. Whenever such dependencies exist, we shall implicitly assume that they have been eliminated by a prior reduction to a smaller set of measures (or statistically aggregated into orthogonal sets of composite measures by techniques such as principal components analysis). Hence we postulate the existence of a given set of *structurally independent* separation measures. In formal terms, this implies that the set of possible values for separation profiles is taken to be the cartesian product of the sets of values for each component measure. The only distinction we shall make between individual component measures is in terms of their signs. In particular, most separation measures are fundamentally nonnegative in nature (such as travel distance, travel time, and shipment expenditures). However, other relevant separation measures may assume negative values (such as age differences or income differences between individuals). In summary, if R denotes the real numbers, and R_+ denotes the nonnegative real numbers, and if the subset of nonnegative separation measures in K is denoted by K_+, then the set, V, of *possible separation profile* values is postulated to be a cartesian product set of the form

$$V = (R_+)^{K_+} \times (R)^{K-K_+}, \tag{1.6}$$

where $K - K_+$ denotes the set of separation components which are unrestricted in sign. [With a slight abuse of notation, we often denote the cardinality, $|K|$, of a given set K by the same symbol, K, and speak of the K-dimensional euclidean space, R^K.]

Finally, our single most important structural independence assumption relates to the nature of space itself. In particular, even when the separation between each origin-destination pair can assume any value, certain configurations of such values may not be possible. For example, if a given separation measure, c_k, corresponds to a metric distance in some space of possible locations for both origins and destinations (where $I = J$), then each triple of separation values $(c_{12}^k, c_{23}^k, c_{13}^k)$ must of course satisfy the triangle inequality condition that $c_{12}^k + c_{23}^k \geq c_{13}^k$. Hence, the set of feasible values of separation configurations in (1.5) can in general be exceedingly complex in any given context. [Even the set of realizable distance configurations among finitely many points in euclidean space is notoriously difficult to analyze.] Hence we shall again make the simplifying assumption that all configurations of separation values are meaningful, i.e., that the possible values of separation measures are given by the cartesian product of individual profile values. More formally, we now assume that the relevant *configuration class*, C, of all separation configuration values in (1.5) is given by

$$C = V^{I \times J} = \{c : c_{ij} \in V, ij \in I \times J\}. \tag{1.7}$$

To illustrate the practical significance of this assumption, it is of interest to consider the impact of certain road improvements on the separation configuration of travel times between cities. While it is clear that improvements on a road between cities i and j can influence travel times between other cities (as shortest-time routes between cities change), it is reasonable to postulate that such indirect effects are relatively small in comparison with direct effects. In addition, the aggregate nature of the separation values in (1.1) tends to lessen the effects of local changes in one separation profile, c_{ij}, on other profiles, c_{gh}. Hence, the product set $V^{I \times J}$ in (1.7) can often be regarded a reasonable approximation for studying the impact of (at least small) changes in separation values.

Finally, it should be emphasized that all of the above assumptions can be relaxed to varying degrees in each specific modeling context. In particular, the independence assumptions on V in (1.6) can be significantly relaxed in most of the representation theorems in Chapter 4 below. Moreover, the results developed for simple search processes in Example 4 of Chapter 2 below show that in many cases the independence assumptions on C in (1.7) can also be relaxed. [More general results of this type will be developed in subsequent works.] Hence for the present it suffices to say that the structural independence assumptions in (1.6) and (1.7) allow a unified development in subsequent chapters which avoid the need for lengthy case-by-case analyses. In essence, these simplifications allow us to focus on *behavioral* questions rather than structural questions.

1.4 Spatial Interaction Processes

Given the general analytical framework developed above, recall that our ultimate goal is to study the influence of various types of spatial separation on interaction behavior. In particular, recall from the introductory discussion in Section 1.3.3 that we seek to construct an operational version of the idealized experiment in Section 1.3.2 in which observed relative frequencies of interactions still yield meaningful estimates of interaction probabilities, $p_c(ij)$, between origin-destination pairs, $ij \in I \times J$, under any possible separation configuration, $c \in C$. To do so, we begin in Section 1.4.1 below by defining the relevant outcome space of possible interaction patterns between origins and destinations which might occur during some relevant time period. Each probability distribution on this outcome space will then constitute a possible probability model of interaction behavior in the given context. The specific probability models of interest for our purposes are designated as *spatial interaction processes*, and are developed in Section 1.4.2 below. Finally, certain statistical independence conditions are postulated for such processes in Section 1.4.3 below which in effect permit observed interaction patterns to be treated statistically as random samples. In particular, the realized interaction frequencies in such *independent spatial interaction processes* are shown to be Poisson distributed, which in turn implies that for each realized number of total interactions, the conditional relative frequencies provide classical (multinomially distributed) sample estimates of interaction probabilities, $p_c(ij)$, which can be interpreted as aggregate approximations of the idealized interaction probabilities, $p_c(\alpha_i, \beta_j)$, in Section 1.3.2 above.

1.4.1 Interaction Patterns

To allow flexibility in analyzing a wide variety of interaction phenomena, it is convenient to start with an abstract notion of *interaction events*, and to model all relevant attributes of interaction phenomena in a given context as random variables over these events. [A rigorous treatment of these concepts is given in Chapter 3 below.] Hence we begin by letting Ω_1 denote the relevant universe of possible *individual interaction events* (where the subscript '1' denotes single or individual events). For example, Ω_1 may consist of all possible shopping trips (phone calls, marriage decisions, etc.) which may occur within a given region. To model collections (or ensembles) of such events, observe that if the n-fold product of Ω_1 is denoted by $\Omega_n = (\Omega_1)^n$, then each possible occurrence of n individual interaction events in Ω_1 is representable by an element, $\omega = (\omega_r : r = 1, \ldots, n) \in \Omega_n$, and is designated as an interaction pattern of size n. In particular, if Ω_1 denotes possible individual shopping trips, then each interaction pattern, $\omega = (\omega_r : r = 1, \ldots, n) \in \Omega_n$, represents a list of n shopping trips which might occur during a given time period. Alternatively, each $\omega \in \Omega_n$ might represent a list of n telephone calls (marriages, migration decisions, etc.) occurring during some period, as in Section 1.3 above. To allow for the possibility that no interactions occur during the relevant period, it is convenient to let Ω_0 consist of the single *null interaction event*, o. In this context, we then take the relevant *outcome space*, Ω, for a given interaction context to consist of all finite interaction patterns which may occur, i.e., $\Omega = \cup_{n \geq 0} \Omega_n$.

In each specific situation, the attributes of interaction patterns which are most relevant for analysis may differ. For example, if in every non-null interaction pattern, $\omega = (\omega_r : r = 1, \ldots, n) \in \Omega$, we let $\alpha_r = \alpha_r(\omega)$ and $\beta_r = \beta_r(\omega)$ denote the specific actor and opportunity involved in each individual interaction, ω_r, (such as the name of the shopper and store involved in each individual shopping trip), then in certain cases, this type of detailed attribute information may be relevant for analysis [as illustrated in Examples 2 and 4 of Chapter 2 below]. Typically, however, we shall only be interested in the origin zone and destination zone of each interaction. More precisely, if for every non-null interaction pattern, $\omega = (\omega_r : r = 1, \ldots, n)$ we now let $i_r = i_r(\omega)$ and $j_r = j_r(\omega)$ denote the respective origin and destination of each individual interaction, ω_r, then these *origin attributes* and *destination attributes* will play a major role in our subsequent analyses. In so far as these attributes convey all the locational information typically required for analysis, it is convenient to designate the joint realization of these attributes as the *spatial interaction pattern*, $s = s(\omega) = [(i_r j_r) : r = 1, \ldots, n]$, associated with each interaction pattern, $\omega = (\omega_r : r = 1, \ldots, n)$. To model the relevant outcome space for spatial interaction patterns, it is convenient to let the set S_0 consist of the *null spatial interaction pattern*, $s_o = s(o)$, corresponding to the null interaction event, o, and similarly, to denote the set of all spatial interaction patterns of size $n > 0$ by $S_n = (I \times J)^n$. Then the *outcome space*, S, for possible spatial interaction patterns arising from interaction patterns in Ω is given by $S = \cup_{n \geq 0} S_n$.

1.4.2 General Interaction Processes

Given this interaction framework, any population behavior representable by outcomes in Ω can be described probabilistically in terms of a *probability measure* on Ω. Each such measure, P, assigns probabilities, $P(A)$, to certain subsets A of Ω which denote the probability that a realized outcome in Ω will belong to A. The subsets, A, for which $P(A)$ is defined are designated as *measurable events* in Ω. [A more precise definition of these concepts is given in Section 3.2 of Chapter 3 below.] For purposes of the present discussion, it suffices to require that the probable occurrence of each *spatial interaction pattern* be defined, i.e., that the subset of interaction patterns, $\Omega(s) = \{\omega \in \Omega : s(\omega) = s\}$, corresponding to each spatial interaction pattern, $s \in S$, be a measurable event in Ω. Hence if we now write $P(s) = P[\Omega(s)]$ for all $s \in S$, then each probability measure P on is seen to generate a unique probability function on S [i.e., a nonnegative function P on S satisfying the condition that $\sum_{s \in S} P(s) = 1$]. For most of the analysis to follow, this simple probability function on S will contain all the relevant probability information required from the underlying probability measure on Ω. Hence our present treatment will focus entirely on these probability functions. [Cases requiring a more explicit analysis of the underlying probability measure on Ω will be illustrated in Examples 2 and 4 of Chapter 2 below].

Within this context, observe next that since all relevant spatial relations between origins in I and destinations in J are postulated to be represented by a given separation configuration, $c = (c_{ij} : ij \in I \times J) \in C$, it is appropriate to treat c as a vector of parameters for each probability function on S, and hence to write P_c. Thus, for our present purposes, a *probability model* of spatial interaction behavior is taken to consist of a family of probability functions, $\{P_c : c \in C\}$, on S which describe interaction behavior between I and J under each possible separation configuration, $c \in C$. Each such model asserts an explicit dependency relation between spatial separation and interaction behavior, as formally implied by variations in *pattern probabilities*, $P_c(s)$, with respect to changes in the underlying separation configuration, c. For example, if in a given model it is true that $P_{c'}(s) > P_c(s)$, then this model asserts that interaction behavior corresponding to pattern s is more likely to occur under separation configuration c' than under c.

(A) Regularity Conditions

Within this general probabilistic framework, it is convenient to introduce three regularity conditions on probabilities which impose only mild restrictions on the types of behavior which can be considered, and which will greatly facilitate the analysis to follow. Observe first that the outcome space, S, by definition contains *all* finite spatial interaction patterns, and hence includes patterns which are much too large to be meaningful in any given practical context. In almost every conceivable case, there is some upper bound on the size of possible interaction patterns beyond which no pattern can have positive probability (such as 'total grid lock' in city traffic). But since patterns approaching these bounds are also very unlikely to occur, they are generally of little behavioral relevance. Hence it is often more convenient to treat each outcome as having a positive probability, and to model impossible occurrences simply as 'extremely unlikely' events. With this in mind, we now impose the following *positivity condition* on probabilities:

R1. (Positivity) *For all separation configurations, $c \in C$, and all spatial interaction patterns, $s \in S$, the pattern probability, $P_c(s)$, is positive.*

Note in particular that condition R1 requires that the *null* interaction event always have positive probability. So even though it is hard to imagine situations in which no shopping trips would occur during, say, a typical working day in a major city, we shall implicitly treat such events as having very small positive probabilities. [A much weaker version of condition R1 is employed in the formal analysis of Chapter 3 below (see condition R1 in Section 3.4.2). However, for the important case of independent interaction processes developed in Section 1.4.3 below, these two conditions are equivalent (as can be seen from Definition 3.5 together with the remark following Definition 3.1 in Chapter 3)].

To motivate our second condition, observe that in a given spatial interaction pattern, $s = (i_r j_r : r = 1, \ldots, n)$, the labels, $r = 1, \ldots, n$, are not assumed to have any behavioral significance, but serve only as a means of enumerating the list of interactions. Hence for the types of behavior we shall consider, the pattern probabilities, $P_c(s)$, are assumed to be completely independent of the order in which the individual interactions in s are labeled. With this in mind, we next impose the following *symmetry condition* on pattern probabilities:

R2. (Symmetry) *For any given separation configuration, $c \in C$, and pair of interaction patterns, $s, s' \in S$, if s and s' differ only by the ordering of their individual interactions, then $P_c(s) = P_c(s')$.*

The practical significance of this condition is to restrict analysis to those interaction contexts in which the only relevant behavioral information contained in each realized interaction pattern, s, consists of the frequencies of individual interactions, $ij \in I \times J$. To clarify the nature of this restriction, consider an interaction context such as the spread of a rumor or communicable disease over space during some given time period. If one records each successive contact between individuals, then the resulting list of, say, all contacts during the first week will yield an interaction pattern, $s = (i_r j_r : r = 1, \ldots, n)$, in which the order of interactions contains a great deal of important information. In particular, one can identify exactly where the rumor (or disease) started, and how it initially spread over space. More generally, when the dynamics of such a diffusion process are reflected in the ordering of individual interactions, then the symmetry assumption is obviously violated. However, suppose that in the same context, the given interaction pattern represents, say, all contacts during the tenth week of the process. By this time, the population of potential transmitters is likely to be so diffuse that the particular ordering of additional contacts conveys little information. However, the contact frequencies between specific individual types can still reveal a great deal of information regarding general interaction patterns within the population (i.e., the types of individuals who tend to come into contact with one another). It is this latter type of information which is of primary importance here. In addition, it is important to note that in most studies of population interactions, the types of data available are typically unordered collections which may already be in the form of frequencies. Hence in such cases, the symmetry assumption amounts simply to a formal recognition that no particular ordering of the data can meaningfully be distinguished from any other.

Our final assumption about general pattern probabilities, $P_c(s)$, relates directly to the dependency of these probabilities on the underlying separation configuration, c. In particular, we assume that small changes in separation values result only in small changes in pattern probabilities. More formally, if the 'closeness' of two separation configurations, $c, c' \in C$, is measured in terms of euclidean distance, $||c - c'||$, then we postulate that pattern probabilities satisfy the following *continuity condition*:

R3. (Continuity) *For any given interaction pattern, $s \in S$, separation configuration, $c \in C$, and scalar $\epsilon > 0$, there exists a $\delta > 0$ sufficiently small to ensure that $|P_c(s) - P_{c'}(s)| < \epsilon$ holds for all configurations, $c' \in C$ with $||c - c'|| < \delta$.*

From a modeling viewpoint, this condition simply asserts that each pattern probability, $P_c(s)$, is taken to be a continuous function of the parameter vector, c. At the level of individuals, it is possible that this type of assumption may be inappropriate in cases where threshold or 'last straw' phenomena are observed (such as when a small rise in bridge tolls or bus fares produces a radical change in an individual's travel behavior). However, at the aggregate level of our present analysis, such individual effects tend to be insignificant. In particular, even when all individuals exhibit threshold behavior, their threshold levels generally differ, and hence yield a smoothed response pattern in the aggregate. [A formal model of such smoothing in terms of *threshold distributions* is developed in Example 2 of Chapter 2 below.]

By way of summary, we may now formalize the desired class of probability models in terms of these three conditions as follows:

Definition 1.1 A family of probability functions, $\boldsymbol{P} = \{P_c : c \in C\}$, on S is designated as a *spatial interaction process* iff $\boldsymbol{P}$ satisfies regularity conditions R1, R2, and R3.

Given this notion of spatial interaction processes, it is also of interest for our later purposes to specify the class of probability measures on Ω which implicitly generate these processes on S. Hence, in terms of the concepts introduced in Section 1.4.2 above, we now say that:

Definition 1.2 A family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω with common measurable events is designated as an *interaction process* on Ω if and only if it generates a corresponding family of probability functions on S satisfying conditions R1, R2, and R3.

In other words, *interaction processes* on Ω consist precisely of those families of probability measures on Ω which generate *spatial interaction processes* on S. [Such interaction processes are illustrated in Examples 2 and 4 of Chapter 2, and are formalized in Section 3.4 of Chapter 3 below].

(B) INTERACTION FREQUENCIES

To analyze measurable properties of a given interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω, we begin by designating a numerical function, X, on Ω as a *random variable* for process $\boldsymbol{P}$ if and only if for each possible value x the set of interaction patterns ω with $X(\omega) \leq x$ is a common measurable event for all probability measures in $\boldsymbol{P}$. In other words, X is a random variable for P if and only if for each x and P_c in $\boldsymbol{P}$, the probability, $P_c(X \leq x)$, that the realized value of X does not exceed x is well defined. [A more precise definition is given in Section 1 of Chapter 3 below.] Each function X which depends solely on spatial interaction patterns, i.e., which is of the form $X(\omega) = X[s(\omega)]$ for all $\omega \in \Omega$ is automatically a random variable. In particular if we let $S(x, \leq) = \{s \in S : X(s) \leq x\}$ then the probability $P_c(X \leq x)$ is given by

$$P_c(X \leq x) = \sum_{s \in S(x,\leq)} P_c(s). \tag{1.8}$$

Similarly, if $S(x) = \{s \in S : X(s) = x\}$ then the probability that X equals x under configuration c is given by

$$P_c(x) = P_c(X = x) = \sum_{s \in S(x)} P_c(s). \tag{1.9}$$

The associated *mean value* (or *expected value*) of X under c is given by

$$E_c(X) = \sum_{s \in S} X(s) P_c(s), \tag{1.10}$$

whenever it exists (i.e., whenever $\sum_{s \in S} |X(s)| P_c(s) < \infty$). Of special interest for our purposes are the random variables representing frequencies of spatial interactions between given origin-destination pairs, i.e., the frequencies of *(ij)-interactions* for each pair $ij \in I \times J$. In particular, if the number of (ij)-interactions in each spatial interaction pattern, s, is denoted by $N_{ij}(s)$, then these *(ij)-interaction frequencies* define an integer-valued random variable for $\boldsymbol{P}$ with distribution given for configuration, c, and nonnegative integer, n_{ij}, by

$$P_c(n_{ij}) = P_c(N_{ij} = n_{ij}) = \sum_{s \in S(n_{ij})} P_c(s). \tag{1.11}$$

The associated *mean interaction frequency* is given by

$$E_c(N_{ij}) = \sum_{s \in S} N_{ij}(s) P_c(s) = \sum_{n_{ij}} n_{ij} P_c(n_{ij}). \tag{1.12}$$

Similarly, if for each pattern, $s \in S$, we denote the *total interaction frequency* for s by,

$$N(s) = \sum_{ij \in I \times J} N_{ij}(s). \tag{1.13}$$

then these total frequencies also define a random variable, N, for each separation configuration, c, with associated *mean total frequency* given by

$$E_c(N) = \sum_{ij \in I \times J} E_c(N_{ij}). \tag{1.14}$$

[Note in addition that $N(s)$ is formally equivalent to the size of s, since $N(s) = n$ if and only if $s \in S_n$]. The importance of these frequency variables in the present context can be seen from the symmetry property (R2) of spatial interaction processes. In particular, observe that if two spatial interaction patterns, s and s', differ only by the ordering of their individual interactions then they must have precisely the same (ij)-interaction frequencies for all $ij \in I \times J$. Indeed, these two conditions are seen to be equivalent. Hence it follows at once from condition R2 that all patterns with identical interaction frequencies must be *equiprobable*, i.e., that for all $s, s' \in S$,

$$[N_{ij}(s) = N_{ij}(s') : ij \in I \times J] \Rightarrow P_c(s) = P_c(s'). \tag{1.15}$$

This fundamental property of symmetric processes is easily seen to imply that all probabilistic information about interaction patterns is expressible in terms of their associated (ij)-interaction frequency probabilities [as is evident for example in expressions (1.36) and (1.37) below]. More generally, each spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is uniquely expressible in terms of an associated *frequency process*, $\boldsymbol{N} = (N^c : c \in C)$, as formalized in Chapter 3 below.

(C) Interaction Probabilities

While the general definition of spatial interaction processes above provides a useful conceptual framework for analysis, there are a number of practical problems which remain. First of all, in view of the complexity of the outcome space, S, it is difficult to develop meangingful behavioral models directly in terms of general pattern probabilities. Even more important from an empirical viewpoint is the difficulty of estimating such pattern probabilities directly. Indeed, for the types of large scale interaction systems of interest in the present analysis, it is generally not possible to observe complete interaction patterns at all. For example, in studying urban traffic patterns during a typical day, it is virtually impossible to record every trip which is taken. Hence (in a manner analogous to the study of temporal stochastic processes) it is desirable to be able to draw meaningful behavioral inferences from only partial observations of interaction patterns.

As mentioned in the previous sections, our approach to resolving these difficulties is to derive a reduced set of *interaction probabilities*, $p_c(ij)$, from general pattern probabilities, $P_c(s)$, which yield useful summary information about interaction propensities between origin-destination pairs, $ij \in I \times J$, under given separation configurations, $c \in C$. As one simple approach along these lines, observe that each individual interaction, ij, formally constitutes an interaction pattern of *size one*. Hence if we let $P_c(ij)$ denote the pattern probability of $ij \in I \times J (= S_1)$, and denote the probability of a singleton interaction pattern by $P_c(S_1) = \sum_{ij \in I \times J} P_c(ij)$, then one could simply take $p_c(ij)$ to be the conditional probability of an (ij)-interaction given that a singleton interaction pattern occurs, i.e.,

$$p_c(ij) = P_c(ij)/P_c(S_1) \tag{1.16}$$

[recall from regularity condition that all singleton interaction patterns are possible, and hence that $P_c(S_1) > 0$]. But while this definition does yield a simple measure which is directly derivable in terms of pattern probabilities, it is exceedingly restrictive both empirically and behaviorally. First, (in a manner similar to the discussion of null interaction patterns following condition R3 above), the occurrence of a singleton interaction pattern is an extremely rare event in most cases of interest (such as a single urban shopping trip during a typical day). Moreover, even if such an event were to occur, it could not generally be expected to represent 'typical' interaction behavior in the system of interest.

Thus we seek a meaningful summary measure of (ij)-interaction propensities which involves the full range of possible system behavior. In view of the practical difficulty of observing full interaction patterns, s, the most natural approach here is to consider a single randomly sampled observation from pattern s and to ask what is the probability that an (ij)-interaction will be observed. If each realized interaction is equally likely to be drawn in such a sample, then the desired probability is of course proportional to the frequency of (ij)-interactions in pattern s. More formally, the conditional probability of observing an (ij)-interaction given a single random sample from a non-null interaction pattern, $s \in S$, can be defined in terms of the frequency attributes in (1.13) by

$$p_c(ij|s) = N_{ij}(s)/N(s), \quad s \neq s_0. \tag{1.17}$$

To avoid the degenerate case in which no interactions occur (and hence in which the above sampling procedure is undefined), we focus on the conditional event that at least one interaction occurs, i.e., that the null interaction pattern, s_0, does not occur. Hence, observing that the probability of this event is given by $1 - P_c(s_0)$, it follows that the conditional probability of sampling an (ij)-interaction given the occurrence of at least one interaction under configuration c is given by

$$p_c(ij) = \begin{cases} [1 - P_c(s_0)]^{-1} \sum_{s \neq s_0} p_c(ij|s) P_c(s) \\ [1 - P_c(s_0)]^{-1} \sum_{s \neq s_0} [N_{ij}(s)/N(s)] P_c(s). \end{cases} \tag{1.18}$$

It is this sampling probability which constitutes our operational definition of the *interaction probabilities*, $p_c(ij)$, derived from spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$. From a statistical viewpoint, $p_c(ij)$ represents the likelihood that any randomly selected interaction realized under configuration c will be an (ij)-interaction. In this sense, $p_c(ij)$ can be interpreted behaviorally as the propensity of (ij)-interactions to occur under configuration c within the given spatial interaction process, $\boldsymbol{P}$.

Notice also that these interaction probabilities can be written more succinctly in terms of conditional expectations as follows. Recall that the occurrence of a non-null interaction pattern, $s \neq s_0$, is equivalent to the occurrence of a *positive* value for the total frequency variable, N. Hence in terms of the general expectation of a numerical attribute, X, in (1.12), it follows that the *conditional expectation* of X given the occurrence of a non-null pattern under configuration c can be written as

$$E_c(X|N > 0) = [1 - P_c(s_0)]^{-1} \sum_{s \neq s_0} X(s) P_c(s). \tag{1.19}$$

In terms of this notation it follows at once from (1.18) and (1.19) that for all $ij \in I \times J$ and $c \in C$,

$$p_c(ij) = E_c[N_{ij}/N | N > 0]. \tag{1.20}$$

In other words, each (ij)-interaction probability can be equivalently interpreted as the *expected relative frequency* of (ij)-interactions given the occurrence of at least one interaction.

1.4.3 Independent Interaction Processes

While the above interaction probabilities yield useful summary measures of interaction propensities, there still remains the practical question of estimating these probabilities. In particular, since interaction probabilities are functions of pattern probabilities, it would appear that one must first estimate pattern probabilities themselves. However, if the interdependencies among individual interactions are sufficiently weak to permit certain statistical independence approximations, then these interaction probabilities can be estimated on the basis of very limited observations on overall interaction patterns. Hence the central purpose of this final section is to motivate these independence assumptions, and to summarize their consequences for the statistical estimation of interaction probabilities.

(A) Independence Hypotheses

Recall from the discussion of theoretical perspectives in Section 1.2 that the gravity models of interest for our purposes are designed to capture the average spatial interaction behavior exhibited by large populations of individuals. In particular, these models focus on the mean frequencies of interactions between various origin-destination pairs. Hence it is implicitly assumed that the spatial interaction behavior of interest can be successfully described in terms of mean interaction frequencies. As mentioned in the discussion of the symmetry condition (R2) above, this requires that the probability of any realized population state be independent of the particular ordering of its individual states. In addition, this type of model implicitly requires that the statistical dependencies between individual interactions be sufficiently weak to allow meaningful behavioral description in terms of averages. With this in mind, we now formulate two statistical independence hypotheses which will form the underpinning for most of the statistical analyses to follow.

To motivate our first hypothesis, consider again the case of shopping behavior and suppose that $s = (i_r j_r : r = 1, \ldots, n)$ denotes a given realized trip pattern. In this context, there may in principle be many types of behavioral interdependencies between the individual realized trips in s. As an extreme example, observe that (for any reasonable number of trips) it is virtually impossible for all trips to go to the same store. More generally, the presence of various *congestion effects* may render it highly unlikely that shopping destinations would be extremely concentrated in space. On the other hand, if friends like to shop together, then the occurrence of a shopping trip from i to j may imply the occurrence of several trips from i to j. More generally, there may exist a variety of *contagion effects* or *band wagon effects* which can lead to identical interaction choices by many individuals. While each possible type of strong behavioral interdependency over space is clearly of interest in its own right, our present objective is to model the overall effect of *spatial separation* on interaction behavior. Hence, we hypothesize that the influence of particular types of interdependencies among spatial actors is minimal, and thus that individual interaction decisions can be treated as statistically independent events. More specifically, we hypothesize that the specific origin-destination locations of individual interactions are statistically independent. To state this hypothesis precisely, we first denote the *conditional probability* of spatial interaction pattern, $s \in S_n$, given the occurrence of some pattern of size n under configuration c by

$$P_c^n(s) = P_c(s)/P_c(S_n). \tag{1.21}$$

Next, if for any $r = 1, \ldots, n$ and $i_r j_r \in I \times J$ we let $S_n(i_r j_r)$ denote the set of all interaction patterns of size n in which the r-th interaction is an $(i_r j_r)$-interaction, then the *marginal probability* of such an interaction is given by

$$P_c^n(i_r j_r) = \sum_{s \in S_n(i_r j_r)} P_c^n(s). \tag{1.22}$$

With this notation, the desired *locational independence hypothesis* for spatial interaction processes, $\boldsymbol{P}$, can be stated as follows:

A1. (Locational Independence) *For all non-null interaction patterns,* $s = (i_r j_r : r = 1, \ldots, n) \in S$, *and separation configurations,* $c \in C$,

$$P_c^n(s) = \prod_{r=1}^{n} P_c^n(i_r j_r). \tag{1.23}$$

[A weaker version of condition A1 not requiring positivity of $P_c(S_n)$ for all n is given in Definition 3.7 of Chapter 3 below.] From a behavioral viewpoint it is also useful to interpret locational independence in terms of conditional probabilities. In particular, if for any interaction pattern, $s = (i_r j_r : r = 1, \ldots, n)$, of size n and any $r = 1, \ldots, n$ we denote the pattern obtained from s by omitting the r-th interaction by $\tilde{s}_r = (i_\sigma j_\sigma : \sigma \neq r)$, then $P_c^n(i_r j_r | \tilde{s}_r)$ denotes the *conditional probability* that the r-th interaction in a realized interaction pattern, s, of size n will be an $(i_r j_r)$-interaction, given the realization, $\tilde{s}_r$, of all other interactions. In terms of this notation, it follows at once that locational independence is equivalent to the condition that for all $c \in C$, $s \in S_n$, and $r = 1, \ldots, n$,

$$P_c^n(i_r j_r | \tilde{s}_r) = P_c^n(i_r j_r). \tag{1.24}$$

In other words, for any given interaction, $(i_r j_r)$, in an interaction pattern of size n it is hypothesized that no properties of the other realized interactions, $(i_\sigma j_\sigma : \sigma \neq r)$, influence the likelihood that $(i_r j_r)$ will occur. Note also that at the micro level of individual actors, $\alpha \in A$, and opportunities, $\beta \in B$, this type of independence can be very problematic indeed. For example while shopping-trip decisions by different individuals may be quite independent, such decisions by the same individual are generally not. Hence the above independence hypothesis implicitly assumes that both the actor populations, A_i, $i \in I$, and opportunity populations, B_j, $j \in J$, are sufficiently large to minimize the influence of any individual actors or opportunities.

In addition, these implicit aggregation assumptions serve to minimize the many types of frequency dependencies among interaction types which may exist at the micro level. For example, if an individual, α, regularly frequents a given barber shop, β, then it may generally be inferred that α does not frequent other barber shops. However, if shopping trips from zone i to zone j are high, this need not imply that shopping trips from i to other zones are necessarily low. More generally, at the present aggregate level of analysis, it may also be hypothesized that (ij)-frequencies between different origin-destination pairs, $ij \in I \times J$, are statistically independent. To be more precise, if we now let $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$ denote a typical *frequency profile* of realized (ij)-frequencies for each $ij \in I \times J$, and if for each separation configuration, $c \in C$, we let $P_c(\boldsymbol{n})$ denote the probability of frequency profile, $\boldsymbol{n}$, under configuration c, then the *desired frequency independence* hypothesis for spatial interaction processes, $\boldsymbol{P} = \{P_c : c \in C\}$, can be stated in terms of (1.11) as follows:

A2. (Frequency Independence) *For all interaction frequency profiles, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, and separation configurations, $c \in C$,*

$$P_c(\boldsymbol{n}) = \prod_{ij \in I \times J} P_c(n_{ij}). \tag{1.25}$$

As in the case of locational independence, this hypothesis can also be given a conditional probability interpretation. In particular, if for any frequency profile, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, and interaction pair, $ij \in I \times J$, we denote by $\tilde{\boldsymbol{n}}_{ij} = (n_{gh} : gh \neq ij)$ the profile obtained from $\boldsymbol{n}$ by deleting the (ij)-interaction frequency component, then frequency independence is easily seen to be equivalent to the condition that for all $c \in C$, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, and $ij \in I \times J$,

$$P_c(n_{ij} | \tilde{\boldsymbol{n}}_{ij}) = P_c(n_{ij}). \tag{1.26}$$

In other words, the frequency independence hypothesis asserts that the realized value of each (ij)-interaction frequency is not influenced by the realized values of any other interaction frequencies. Given these two independence hypotheses, we now say that:

Definition 1.3 Each spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying independence hypotheses A1 and A2 is designated as an *independent spatial interaction process.*

Independent spatial interaction processes will play a central role in the analysis to follow. The practical advantage of these processes (as discussed more fully in the next section) is that the realized interaction frequency profiles in such processes can be treated as independent random samples from the *interaction probability distribution* in (1.18) [or equivalently (1.20)] above. Hence this distribution can be estimated without any additional knowledge of the pattern probabilities in (1.18). Before developing these results, it should be emphasized that neither of these independence hypotheses can be expected to hold exactly in any given spatial interaction context. However, they can be shown to provide reasonable approximations to a wide variety of interaction phenomena (as discussed in Section 1.5.3 below).

More importantly, they provide an initial set of working *null hypotheses* for the investigation of almost any type of spatial interaction behavior. In cases where these hypotheses cannot be rejected on the basis of subsequent observations, such interaction behavior may be well described by the types of mean-interaction models developed in Chapter 2 below. On the other hand, in cases where either or both of these hypotheses can be rejected, one must introduce more explicit models of the types of behavioral interdependencies observed. A brief discussion of the many possibilities here is given in Section 1.5 below.

(B) Poisson Characterization Theorem

From a theoretical viewpoint, the significance of the locational and frequency independence axioms is that together they yield an exact distribution theory for interaction frequencies. In particular, if we recall from expression (1.12) the definition of mean (ij)-interaction frequencies, $\mathrm{E}_c(N_{ij})$, under each separation configuration $c \in C$, then in terms of this notation we have the following characterization of independent spatial interaction processes (established in Theorem 3.2 of Chapter 3 below):

Poisson Characterization Theorem. *A spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is independent if and only if for all frequency profiles, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, and separation configurations, $c \in C$,*

$$P_c(\boldsymbol{n}) = \prod_{ij \in I \times J} [\frac{\mathrm{E}_c(N_{ij})^{n_{ij}}}{n_{ij}!} \exp\{-\mathrm{E}_c(N_{ij})\}]. \qquad (1.27)$$

Thus the (ij)-interaction frequencies in independent spatial interaction processes are seen to be independently *Poisson distributed*. Conversely, no other spatial interaction processes satisfy this condition. Hence the location and frequency independence axioms (A1,A2) together may be said to characterize such Poisson frequencies (which are formalized as *Poisson frequency processes* in Chapter 3 below). Such Poisson processes have long been employed in frequency analyses of categorical data [as treated for example in Bishop, Fienberg, and Holland (1975) and Haberman (1978, 1979)]. With respect to gravity models in particular, this distributional assumption has been studied by Sen and Sööt (1981), Flowerdew and Aitkin (1982), Fotheringham and Williams (1983), Baxter (1984, 1985) and Davies and Guy (1987), among others. Hence the Poisson Characterization Theorem may be said to provide a behavioral foundation for this distributional assumption. [A comparison of this result with alternative measure-theoretic characterizations of Poisson frequency processes for spatial and temporal continua is given in Smith (1986).]

From a behavioral viewpoint, one of the most important consequences of the Poisson Characterization Theorem is to show that all probabilistic behavior in independent spatial interaction processes can be represented entirely in terms of *average* interaction levels. More precisely, expression (1.27) shows that the only statistical parameters governing such behavior are the *mean interaction frequencies*, $\mathrm{E}_c(N_{ij}), ij \in I \times J$. This fundamental property of independent spatial interaction processes provides the formal basis for essentially all models of spatial interaction behavior studied in this book. In particular, it shows that such behavior can be adequately described in terms of *mean interaction models*, i.e., explicit parametric models relating mean interaction frequencies to appropriate attributes of both actor and opportunity populations. Such models will be developed in Chapter 2 below.

From a statistical sampling viewpoint, the central consequence of this theorem is to show that for each given size of realized spatial interaction patterns, the conditional interaction frequencies are *multinomially distributed* in terms of the interaction probabilities in expression (1.18). In addition, this theorem implies that the interaction probabilities in (1.18) can be written in a number of useful equivalent forms. To state these results, observe first from the symmetry property (R2) of spatial interaction processes that for any given spatial interaction pattern of size n, the marginal interaction probabilities, $P_c^n(i_r j_r)$, in A1 must be the *same* for every component $r = 1, \ldots, n$. Hence for each interaction pair, $ij \in I \times J$, we now let

$$P_c^n(ij) = \sum_{s \in S_{n-1}} P_c^n(ij, s). \tag{1.28}$$

denote the common marginal probability for each component of an n-interaction pattern. With this notation, we now have the following corollary of the Poisson Characterization Theorem (established in Corollary 3.3 of Chapter 3 below):

Multinomial Sampling Theorem. *If* $\boldsymbol{P} = \{P_c : c \in C\}$ *is an independent spatial interaction process, then for each separation configuration,* $c \in C$, *and frequency profile,* $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, *with* $\sum_{ij} n_{ij} = n$,

$$P_c^n(\boldsymbol{n}) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}, \tag{1.29}$$

where for each $ij \in I \times J$, $c \in C$, *and positive integer* n,

$$p_c(ij) = \mathrm{E}_c[N_{ij}/N | N > 0] = \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N)} = P_c^n(ij). \tag{1.30}$$

It follows at once from (1.29) and the first equality in (1.30) that each realized frequency profile, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, with $n = \sum_{ij} n_{ij}$ is formally identical to a *random sample of size* n from the interaction probability distribution in (1.18) [as reinterpreted in (1.20)]. Thus, under the independence hypotheses A1 and A2, the interaction probabilities, $p_c(ij)$ [or any finite-parameter specification of $p_c(ij)$], can be estimated by standard maximum likelihood techniques. In particular, it follows [in a manner paralleling (1.17)] that the *sample relative frequencies*, $\hat{p}_{ij}(s)$, defined for all $s \in S$ and $ij \in I \times J$ by

$$\hat{p}_{ij}(s) = N_{ij}(s)/N(s) \tag{1.31}$$

yield *maximum likelihood estimates* of the interaction probabilities, $p_c(ij)$, under each possible separation configuration, $c \in C$. Hence such estimates exhibit all the optimality properties of general maximum likelihood estimates (as discussed in Chapter 5). In addition, it also follows from the first equality in (1.30) that whenever such estimates are defined (i.e., whenever $N > 0$), they are always *unbiased.* A full development of maximum likelihood estimation procedures for parametric specifications of $p_c(ij)$ is given in Chapter 5.

Next observe from the second equality in (1.30) that interaction probabilities can also be written in terms of *unconditional mean frequencies.* In particular, it follows from (1.14) that for all $ij \in I \times J$and $c \in C$,

$$p_c(ij) = \frac{\mathrm{E}_c(N_{ij})}{\sum_{gh \in I \times J} \mathrm{E}_c(N_{gh})}. \tag{1.32}$$

This fundamental identity provides a key link between spatial interaction probabilities and mean interaction frequencies. In particular, it yields a direct relation between gravity models of interaction probabilities and gravity models of mean interaction frequencies, which permits a simple comparative analysis of these model types (as discussed in Section 2.2.1 of Chapter 2 below).

Finally, the last equality in (1.30) shows that the probability of an (ij)-interaction occurring in any position of a realized interaction pattern is independent of the size of the pattern. Notice in particular that since $P_c^1(ij)$ is precisely the quantity $P_c(ij)/P_c(S_1)$ in (1.16), it follows that for independent spatial interaction processes the alternative definitions of interaction probabilities in (1.16) and (1.18) are *identical.* More generally, this *scale independence property* implies that individual interaction decisions are not influenced by the overall level of interaction activity in the system of interest. [We shall return to this property in Section 1.5.1 below.]

1.5 Relaxations of Independence

The above results show that in situations where the locational and frequency independence axioms are appropriate, they provide a powerful probabilistic framework for modeling population interaction behavior in terms of mean interaction frequencies. However, as stressed in the discussion, these hypotheses represent at best a very simplified approximation of population interaction behavior. Hence it is of interest to consider possible relaxations of these assumptions which introduce certain elements of interdependence between interactions, but which also preserve many of the central features of the present analysis. To do so, we first consider separate relaxations of each of the basic independence axioms in Sections 1.5.1 and 1.5.2, respectively. In each case, a number of additional parameters must be introduced which reflect various types of dependencies among interaction decisions. Finally in Section 1.5.3 we briefly consider certain processes involving relaxations of both independence axioms.

1.5.1 Relaxations of Frequency Independence

For most of the analysis employed in subsequent chapters, the only properties of independent interaction processes which are required are those contained in the Multinomial Sampling Theorem above. Moreover, since these conditional sampling properties rely primarily on the locational independence axiom (A1), it is possible to relax the frequency independence axiom (A2) without affecting these properties. The only consequence of frequency independence which we require is the *scale independence property* mentioned in the discussion of the Multinomial Sampling Theorem above. To be more precise, if we now designate a spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, as *scale independent* if and only if

$$P_c^n(ij) = P_c^1(ij), \tag{1.33}$$

for all $c \in C$, $ij \in I \times J$, and $n > 0$, then it is shown in Chapter 3 below (Proposition 3.8) that the Multinomial Sampling Theorem continues to hold for all scale independent spatial interaction processes satisfying A1. In particular, this is seen to imply [from (1.21), (1.23) and (1.30)] that for each spatial interaction pattern, $s \in S_n$, with $N_{ij}(s) = n_{ij}$, $ij \in I \times J$, the associated pattern probabilities, $P_c(s)$, $c \in C$, can be written in terms of the interaction probabilities $p_c(ij)$ in (1.18) as

$$P_c(s) = P_c(n)[\prod_{ij} p_c(ij)^{n_{ij}}]. \tag{1.34}$$

Hence for such processes, interest focuses on the nature of the *total frequency probabilities*, $P_c(n)$. If $\boldsymbol{P}$ satisfies frequency independence (A2), then by the Poisson Characterization Theorem, these probabilities must be of the form

$$P_c(n) = \frac{\mathrm{E}_c(N)^n}{n!} \exp[-\mathrm{E}_c(N)], \tag{1.35}$$

for all $c \in C$and $n \geq 0$. Thus we see that frequency independence must fail to hold for any other specification of total frequency probabilities.

The most common types of frequency dependencies which tend to occur in practice are those related to sizes of potential actor and opportunity populations. [For example, if one considers the number of marriages occurring within a given population during a given year, then the numbers of males and females within this population obviously governs the possible number of marriages.] Moreover, variations in the sizes of such populations may be influenced by a number of structural factors which are clearly not consistent with Poisson randomness. Hence in such cases, probabilistic models of total interaction frequencies must incorporate explicit models of these population sizes. With respect to *actor populations* in particular, such structural factors may include birth and death rates, as well immigration to and emigration from the population. Thus explicit demographic birth-death models and migration models may constitute relevant components of the appropriate actor population model. In addition, other demographic factors (such as age levels) and economic factors (such as income levels) can influence the overall interaction activity of actor populations. [A particular class of such activity models is developed in terms of 'threshold interaction behavior' in Section 3.9.1 below.] Similarly, the relevant *opportunity population* (such as shopping opportunities or job opportunities) can also be influenced by structural factors such as overall economic growth and technological change. In general then, each possible type of overall population effect which simultaneously influences all interaction activity can create statistical dependencies which violate the frequency independence assumption. A number of more specific types of non-Poisson frequencies are discussed in the next section.

1.5.2 Relaxations of Locational Independence

As a second type of relaxation, it is also shown in Chapter 3 (Proposition 3.10) that for any spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying frequency independence (A2), the probability, $P_c(s)$, of each spatial interaction pattern, $s \in S_n$, with $N_{ij}(s) = n_{ij}$, $ij \in I \times J$, must be of the form:

$$P_c(s) = \frac{1}{n!} \prod_{ij} n_{ij}! P_c(n_{ij}). \tag{1.36}$$

Hence each such process, $\boldsymbol{P}$, is completely specified by the family of (ij)-*interaction frequency probabilities*, $P_c(n_{ij})$. If $\boldsymbol{P}$ also satisfies locational independence (A1), then again by the Poisson Characterization Theorem, each of these probabilities must be Poisson, i.e.,

$$P_c(n_{ij}) = \frac{\mathrm{E}_c(N_{ij})^{n_{ij}}}{n_{ij}!} \exp[-\mathrm{E}_c(N_{ij})]. \tag{1.37}$$

Hence, any specification of (ij)-interaction frequency probabilities other than (1.37) must necessarily yield a spatial interaction process, $\boldsymbol{P}$, which fails to satisfy A1.

As with the specification of total interaction frequencies above, our main interest here is in those specifications of (ij)-interaction frequency probabilities which arise from potentially relevant types of behavioral dependencies leading to non-Poisson frequencies. To illustrate such dependencies it is appropriate to begin by observing that the most characteristic (and indeed most restrictive) feature of the Poisson distribution is the equality between its mean and variance. Indeed, types of non-Poisson frequency behavior are often classified into those with variances less than their means and those with variances greater than their means. The former are typically designated as processes with *restricted variation*, and the latter as processes with *extra variation.*

(A) Processes with Restricted Variation

Many types of spatial interaction phenomena are much too regular in nature to exhibit Poisson randomness. An extreme example is daily commuting behavior, in which essentially the same work trips occur each working day. Here the variation in total trip frequencies is so insignificant in comparison with their means that models of such behavior which have been applied in practice treat such flows as completely deterministic in nature [as exemplified by the many models surveyed in Sheffi (1985)].

From a probabilistic viewpoint, such frequency processes can be modeled as sums of individual interaction decisions, each of which occurs with probability close to one. [A simple binomial model of such a process is illustrated in Section 3.9.2(A) below]. More general types of models with restricted variation include probabilistic *superpositions* of both discretionary and nondiscretionary types of decision behavior, such as superpositions of discretionary shopping trips and nondiscretionary commuting trips. [A simple Poisson-binomial model of this type is illustrated in Section 3.9.2(A) below.]

Many other types of processes with restricted variation are also possible. Suppose for example that the relevant population of actors is taken to be the set of residents within a randomly sampled city block, or other appropriate spatial unit. Then this population can be regarded as a 'random cell count' from a larger population dispersed over space. Here the appropriate class of probabilistic models of population size includes the many cell-count models which have been developed in the stochastic theory of spatial point processes [as summarized for example in Cox and Isham (1980) and Diggle (1983)]. In particular, if the randomly sampled spatial unit is taken to be a residential block, then the typically regular dispersion of households over residential areas will tend to result in variation which is much too *small* to qualify as Poisson variation. Many specific models of such regular variation have been developed in the literature. Of particular relevance here is the class of *spatial inhibition processes* [Diggle (1983, Section 4.8)], in which some minimum spacing between population members is required. In addition, if regular dispersion of households is a result of typical grid patterns of urban streets, then the class of *lattice-based processes* [Cox and Isham (1980, Section 6.3.3) and Diggle (1983, Section 4.11)] may also be relevant.

(B) Processes with Extra Variation

Turning next to processes with extra variation, we may illustrate such processes within the spatial sampling context above by supposing that the relevant spatial population is taken to be the set of residents within a randomly sampled grid square (say a square mile) from a map of a large metropolitan region. At this scale of analysis, one will tend to find clusterings of residential areas within the overall pattern of land uses (commercial areas, industrial areas, parks, etc.), and hence may observe variations in population sizes which are much to *large* to be Poisson. Hence the many types of *spatial clustering processes* which have been developed in the literature on spatial point processes [as in Cox and Isham (1980, Section 6.2.3) and Diggle (1983, Section 4.6)] might be more relevant.

Alternatively, extra variation can result from various types of heterogeneity within the population itself. Such heterogeneities are well illustrated by spatial migration frequencies in which movement tends to involve family units rather than individuals [as analyzed for example by Flowerdew and Aitken (1982) and Aufhauser and Fischer (1985, Section 4.2), among others]. Here, even if the decisions to migrate satisfy both the locational and frequency independence hypotheses, the actual numbers of migrants will depend on the distribution of family sizes within the relevant population. Hence an appropriate model here might correspond to a *compound Poisson process* [as in Cox and Isham (1980, Section 3.1.3) and Diggle (1983, Section 4.5)], in which the number of family units is Poisson distributed, but where the number of family members in each unit may have an arbitrary distribution. [A number of specific illustrations of such processes with extra variation are given in Section 3.9.2(B) below.]

1.5.3 More Complex Types of Interdependencies

Having discussed a number of partial relaxations of each independence assumption, it should again be emphasized that very few types of interaction behavior can be expected to satisfy either of these assumptions *exactly*. Hence our main interest in these assumptions stems from their *robustness* in approximating a wide range of population interaction behavior. This robustness is born out by the many types of asymptotic results which establish the Poisson distribution as the unique limiting form (or 'domain of attraction') for a wide range of interaction processes when the population size increases and the overall influence of each individual interaction decreases. Hence it is appropriate in this final section to mention a few of these approximation results for more complex types of spatial interaction processes. In particular, we shall briefly consider the general classes of *superposition processes* and *spatial Markov processes* which have played major roles in modeling various types of population behavior.

(A) Superimposed Processes

Recall from the Poisson-binomial example of travel behavior (in the discussion of processes with restricted variation) that many actual interaction processes may be composed of many different types of interaction behavior. More generally, it is often possible to decompose interacting populations into a number of subpopulations in such a way that most interdependencies among individual interactions occur within each subpopulation rather than *between* subpopulations. In such cases, overall interaction behavior can be modeled as a *superposition* of (possibly very many) mutually independent subpopulations. Hence to model such phenomena, one is led to consider the limiting behavior of superpositions of large numbers of subpopulations, as the contribution of each individual subpopulation becomes small.

The simplest conceivable model of such a process is the classic model of independent *coin flips*, in which each coin is regarded as an independent 'subpopulation'. If one takes the relevant frequency in this case to be the observed number of 'heads', and regards the probability of 'heads' for each coin as a measure of the probable contribution of that subpopulation to the total frequency, then one is led to consider the limiting frequency of 'heads' as the number of coins becomes large and the probability of heads for each coin becomes small. Within this probabilistic setting, it is well known that under very general conditions the limiting frequencies of 'heads' is always *Poisson distributed* [as for example in Feller (1957, Section XI.6, Example b)]. In particular, if all probabilities are the same (and the mean number of 'heads' is held constant), then this approximation result is precisely the classic Poisson approximation of the binomial distribution [as in Feller (1957, Section VI.5)]. This simple example has far reaching generalizations. Indeed, it can be shown [as developed more explicitly in Section 3.9.3(A) below] that essentially all spatial interaction processes involving superpositions of large numbers of 'small' independent subpopulations are well approximated by Poisson processes.

(B) Spatial Markov Processes

While the above class of processes is very broad, the underlying types of behavioral interdependence in such processes is left completely unspecified. Hence it is of interest to consider a class of spatial interaction processes in which a more explicit model of spatial interdependence is developed. In particular, we now consider the well known class of *spatial Markov processes* in which interdependencies are postulated to involve 'neighborhood' effects among behaving units [see for example Kinderman and Snell (1980, Section 6), Cox and Isham (1980, Section 6.3.2), and Diggle (1983, Section 4.9)]. To motivate the basic idea here, it is instructive to consider an example involving travel behavior, in which certain trips may interact with one another. For example two trips between the same origin, i, and destination, j, in a given trip pattern may well impose congestion costs on one another. Even if these trips share only origin, i, or destination, j, such congestion effects may still be present. In addition, 'head-on' traffic between j and i may also impose various types of congestion costs (head-light effects, potential-collision effects, etc.). Hence if we now designate the set of all trips which impose congestion effects on ij as the relevant *congestion neighborhood*, $\langle ij \rangle$, of ij in $I \times J$, then it may be hypothesized that the cost of trip ij is influenced by those trips in $\langle ij \rangle$. The simplest example is the 'congestion-deterrence hypothesis' in which the congestion costs created by trips in $\langle ij \rangle$ are hypothesized to discourage ij trips. Under this hypothesis, it is natural to expect that large numbers of interdependent trips would be relatively less likely than under the locational and frequency independence hypotheses. A class of *congestive spatial interaction processes* is developed in Section 3.9.3(B) which is shown to exhibit this property.

In addition, it is natural to expect that if congestion interdependencies are very weak then the resulting process should be well approximated by an independent spatial interaction process, so that all interaction frequencies are approximately Poisson. For the congestive processes in Section 3.9.3(B) it is shown that this is the case if spatial actors are sufficiently 'insensitive' to such interdependencies. In addition it is shown that even when actors are very sensitive to congestion effects, Poisson approximations still hold for overall interaction frequencies when the degree of congestion in the system is sufficiently 'small'. Hence this class of spatial Markov processes illustrates a range of more complex types of spatial interaction behavior in which interaction frequencies are well approximated by the Poisson distribution when the degree of interdependency among behaving units is sufficiently weak in some appropriate sense.

CHAPTER 2

Gravity Models: An Overview

2.1 Introduction

Given the general class of spatial interaction processes outlined in Chapter 1, we are now ready to develop the specific class of behavioral models which form the central focus of this book — namely *gravity models* of spatial interaction behavior. To do so, we begin by recalling from the discussion following the Poisson Characterization Theorem in Chapter 1 that each independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is completely characterized by its associated *mean interaction frequencies*, $\mathrm{E}_c(N_{ij})$, $ij \in I \times J$, for each separation configuration, $c \in C$. Hence each explicit model of mean interaction frequencies yields a complete specification of probabilistic interaction behavior in this context. With this observation in mind, recall from the Introduction that gravity models are precisely of this type. In particular, if the 'interaction levels', T_{ij}, in expressions (2) through (4) in the Introduction are now interpreted as mean interaction frequencies for the separation configuration defined by distances, d_{ij}, then each of these expressions is seen to constitute an explicit (finite parameter) model of mean interaction frequencies. More generally, even for spatial interaction processes in which the axioms of frequency independence and/or locational independence are not appropriate, gravity models may still be interpreted as representations of average interaction behavior within such processes.

With these general observations, our specific objectives in this chapter are to formalize gravity models as representations of mean interaction behavior, and to characterize these representations in terms of explicit behavioral axioms on spatial interaction processes. The overall organization of the chapter is divided into two parts. We begin in Section 2.2 with a consideration of the most general types of gravity models, in which the exact form of the basic deterrence function is left unspecified. Within this general framework, specific types of gravity models are then classified according to their degree of dependency on specific separation configurations. These various model types are then illustrated in terms of two broad classes of spatial interaction processes which have appeared in the literature. Finally, each model type is characterized in terms of explicit behavioral axioms on spatial interaction processes. In Section 2.4 we then concentrate on gravity models with exponential deterrence functions, which form the major focus of the present work. These exponential gravity models are classified, illustrated, and characterized in a manner paralleling the development of general gravity models above. Finally, a number of possible generalizations of gravity model forms are considered briefly in the concluding Section of the chapter.

2.2 General Gravity Models

Recall from the Introduction that the basic gravity hypothesis essentially asserts a multiplicative relationship between mean interaction frequencies and the effects of origin, destination, and separation attributes, respectively. This idea is formalized in Section 2.2.1 below, where a classification of such multiplicative *gravity models* is proposed. In Section 2.2.2 these concepts are illustrated in terms of two broad classes of spatial interaction processes, each of which gives rise to mean interaction frequencies representable by gravity models. With these illustrations as a background, we next develop in Section 2.2.3 general characterizations of all spatial interaction processes consistent with such model forms. First we consider a range of *aggregate* behavioral characterizations which relate gravity model forms to certain aggregate properties of spatial interaction patterns. In addition, we consider certain *local* behavioral characterizations which focus on pairwise origin-destination dependencies within overall separation configurations. Finally in Section 2.3 we consider the many specifications of origin, destination, and deterrence functions which have appeared in the literature. With respect to deterrence functions in particular, the class of exponential specifications are shown to provide a flexible unifying framework for a broad range of specifications. This development serves to motivate the second part of the chapter, which is devoted to exponential gravity models.

2.2.1 Model Specifications

To motivate the most general class of gravity models to be considered in this book, we begin with expression (5) in the Introduction. In terms of the notation in Chapter 1, this model specifies that all mean interaction frequencies, $\mathrm{E}_c(N_{ij})$, for a given spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, are of the form

$$\mathrm{E}_c(N_{ij}) = A(i)B(j)F(c_{ij}) \tag{2.1}$$

for some appropriate choice of an origin function, A, destination function, B, and separation function, F. Note also that in the context of Chapter 1, expression (2.1) asserts much more than a simple multiplicative relationship. In particular, it implicitly asserts that the functions A and B are completely independent of the separation configuration, $c \in C$, and that the function F is independent of all separation profiles in c other than c_{ij}. However, there are many behavioral contexts in which these independence assumptions can fail, even though the basic multiplicative gravity hypothesis holds (as illustrated in the examples of Section 2.2.2 and 2.4.2 below).

Hence in the most general version of the multiplicative gravity hypothesis, we shall allow each of these functions to depend on the full configuration, $c \in C$, and hence write A_c, B_c, and F_c. In particular, if for each separation configuration $c \in C$ we now designate the set of separation profiles in c by $V_c = \{c_{ij} \in V : ij \in I \times J\}$, then the relevant deterrence function, F_c, may in general only be defined for those separation profiles in V_c. With these observations, we may now formalize the most general class of *gravity models* for our purposes as follows:

Definition 2.1 (Model G1) A spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be mean-representable by a *gravity model* if and only if for each separation configuration, $c \in C$, there exists a positive origin function, A_c, destination function, B_c, and deterrence function, F_c, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)F_c(c_{ij}). \tag{2.2}$$

We shall often refer to this general gravity model form as Model G1. Several additional remarks about this general model are appropriate. First of all, while Model G1 is defined for all spatial interaction processes, our primary interest is in *independent interaction processes.* For this class of processes it was observed in the introduction to this chapter that each specification of the functions A_c, B_c, and F_c completely determines the entire process. Hence if (2.2) holds for an independent interaction process, $\boldsymbol{P}$, we shall simply say that $\boldsymbol{P}$ is *representable* by a gravity model.

Next observe from the positivity of the functions, A_c, B_c, and F_c that each gravity model can be expressed equivalently as a *log-additive model* of the form

$$\log \mathrm{E}_c(N_{ij}) = a_c(i) + b_c(j) + f_c(c_{ij}), \tag{2.3}$$

where $a_c(i) = \log A_c(i)$, $b_c(j) = \log B_c(j)$, and $f_c(c_{ij}) = \log F_c(c_{ij})$. [The more specialized term *log-linear model* is reserved for the log form of exponential gravity models in expression (2.128) below. For further discussion of additive versus linear models, see, for example, Hastie and Tibshirani (1990).] Hence even though separation profiles generally involve continuous variables, it should be clear that much of the statistical theory of log-additive (or log-linear) models of categorical data can be applied to the study of gravity models [as developed in detail in Chapters 5 and 6 below].

From a behavioral viewpoint, it is of particular importance to observe that the *interaction probabilities* defined in (1.18) also have a simple gravity form. In particular, it follows from expression (1.32) of Chapter 1 that each gravity representation of mean frequencies in (2.2) yields the following associated gravity representation of interaction probabilities:

$$p_c(ij) = \frac{\mathrm{E}_c(N_{ij})}{\sum_{gh} \mathrm{E}_c(N_{gh})} = \frac{A_c(i)B_c(j)F_c(c_{ij})}{\sum_{gh} A_c(g)B_c(h)F_c(c_{gh})}. \tag{2.4}$$

Such gravity models in terms of interaction probabilities have a long history in the literature [as discussed for example in Hua and Porell (1979) and Batten and Boyce (1987)]. However the relation between these two types of gravity models is seldom made explicit. Hence an important analytical advantage of our present framework is that within the context of *independent interaction processes*, one can simultaneously model both mean interaction frequencies and interaction probabilities in terms of common gravity model forms. More generally, recall from Section 1.5.1 that (in the presence of scale independence) these results continue to hold for all interaction processes satisfying locational independence [as shown in Proposition 3.9 of Chapter 3 below].

Gravity models of interaction probabilities have appeared most often in the context of *probabilistic choice models* which focus on the conditional interaction probabilities associated with each given origin, $i \in I$ [as discussed for example in Batten and Boyce (1987) and Fotheringham and O'Kelly (1989, Chapter 4)]. More precisely, if for the interaction probabilities in (1.18) we now define for each $i \in I$ the corresponding *conditional destination probabilities* by

$$p_c(j \mid i) = \frac{p_c(ij)}{\sum_{h \in J} p_c(ih)}, \qquad j \in J, \tag{2.5}$$

then (2.4) yields the following *destination gravity model:*

$$p_c(j \mid i) = \frac{B_c(j)F_c(c_{ij})}{\sum_{h \in J} B_c(h)F_c(c_{ih})}, \qquad j \in J. \tag{2.6}$$

In a parallel manner, if one defines for each $j \in J$ the corresponding *conditional origin probabilities* by

$$p_c(i \mid j) = \frac{p_c(ij)}{\sum_{g \in I} p_c(gj)}, \qquad i \in I, \tag{2.7}$$

then (2.4) also yields the following *origin gravity model:*

$$p_c(i \mid j) = \frac{A_c(i)F_c(c_{ij})}{\sum_{g \in I} A_c(g)F_c(c_{gj})}, \qquad i \in I. \tag{2.8}$$

Models of this type have been most frequently applied to model the probable origins of work trips to given job locations [as discussed for example in Wilson (1971, 1974), White (1988), and Fotheringham and O'Kelly (1989, Section 2.3.3)]. Such models are also implicit in 'choice-based sampling' approaches to the statistical estimation of probabilistic choice models [as discussed for example in Manski and Lerman (1977), Manski and McFadden (1981) and Ben-Akiva and Lerman (1985)]. Certain additional formal properties of both origin and destination gravity models are developed in Smith (1984).

Given this general class of gravity models, observe that aside from positivity, the functions A_c, B_c, and F_c are completely arbitrary. In particular, the nature of the dependency between mean interaction frequencies and separation profiles, as expressed by the deterrence function, F_c, is left unspecified. But it should be clear from the opening paragraph of the Introduction that the original gravity hypothesis asserts an *inverse* relationship between interaction levels and spatial separation [as indicated by the term 'deterrence function']. Indeed the vast majority of deterrence functions employed in practice do exhibit such inverse relationships [an exception is illustrated in expression (2.125) below]. Hence it is of interest to distinguish those gravity models exhibiting this classic property. In particular, if for any two separation profiles, $v = (v^k : k \in K)$ and $w = (w^k : k \in K)$ we write $v \geq w$ whenever $v^k \geq w^k$ for all $k \in K$, and designate a deterrence function F_c as *nonincreasing* on V_c if and only if for all $v, w \in V_c$ it is true that $F_c(v) \leq F_c(w)$ whenever $v \geq w$, then we now have the following *monotone* version of Model G1:

Definition 2.2 (Model G1*) A spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be mean-representable by a *monotone gravity model* if and only if for each $c \in C$, $\boldsymbol{P}$ satisfies (2.2) for a deterrence function F_c which is nonincreasing on V_c.

As in the case of general gravity models, we shall often refer to such monotone gravity models as Model G1*. Within this general modeling framework, our interest focuses on those specifications in which the relevant component functions, A_c, B_c, and F_c, exhibit some degree of *invariance* with respect to the underlying separation configuration, $c \in C$.

(A) Deterrence-Invariant Gravity Models

Of most interest for our purposes are models in which (at least) the deterrence function, F_c, is invariant with respect to the underlying configuration, $c \in C$. Indeed, the very concept of a 'deterrence' function is usually taken to reflect those behavioral influences of *spatial separation* which can be distinguished from overall configuration effects. With this in mind, we now formalize those *deterrence-invariant* gravity models in which the relevant deterrence function, F, is independent of the specific configuration, $c \in C$, and hence which is taken to be defined on the full set of separation profiles, V:

Definition 2.3

(i) (Model G2) A spatial interaction process, $\boldsymbol{P}$, is said to be mean-representable by a *deterrence-invariant gravity model* if and only if there exists a positive deterrence function, F, together with positive origin functions, A_c, and destination functions, B_c, for each separation configuration,

$c \in C$, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)F(c_{ij}). \tag{2.9}$$

(ii) (Model G2*) If in addition the deterrence function F in (2.9) is nonincreasing on V, then $\boldsymbol{P}$ is said to be representable by a *monotone deterrence-invariant gravity model.*

From a practical viewpoint, it is important to note the following additional invariance property exhibited by deterrence-invariant gravity models. In particular, observe from (2.9) that for all origins, $i, g \in I$, destinations, $j, h \in J$, and configurations, $c \in C$, we must have

$$\frac{\mathrm{E}_c(N_{ij})\mathrm{E}_c(N_{gh})}{\mathrm{E}_c(N_{ih})\mathrm{E}_c(N_{gj})} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gj})}. \tag{2.10}$$

Hence the 'cross-product ratios' of mean interaction frequencies on the left hand side of (2.10) are seen to depend *only* on the separation profiles involving (i, g, j, h), and thus are invariant with respect to the global properties of configuration, c. From an estimation viewpoint, this invariance property provides a natural set of estimation equations for both specifying and calibrating the deterrence function, F, as discussed further in Chapter 6 below. Moreover, in the context of independent spatial interaction processes, an extension of this invariance property leads to a general behavioral characterization of deterrence-invariant gravity models [Theorem 4.9 of Chapter 4].

(B) Destination-Deterrence-Invariant Gravity Models

Within this class of deterrence-invariant gravity models, it is of interest to distinguish those models which also exhibit invariance properties with respect to origin or destination functions. Turning first to destination functions, we have the following class of gravity models which are invariant both with respect to destination and deterrence functions:

Definition 2.4
(i) (Model G3) A spatial interaction process, $\boldsymbol{P}$, is said to be mean-representable by a *destination-deterrence-invariant gravity model* if and only if there exists a positive destination function, B, and if positive deterrence function, F, together with positive origin functions, A_c, for each separation configuration, $c \in C$, such that for all $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B(j)F(c_{ij}). \tag{2.11}$$

(ii) (Model G3*) If in addition the deterrence function F in (2.11) is nonincreasing on V, then $\boldsymbol{P}$ is said to be representable by a *monotone destination-deterrence-invariant gravity model.*

As a parallel to (2.10), it follows at once from (2.11) that for each origin, $i \in I$, such gravity models exhibit the following configurational invariance property with respect to relative mean interaction frequencies for all destinations $j, h \in J$:

$$\frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{ih})} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})}. \tag{2.12}$$

For independent spatial interaction processes, this invariance property can be given a more direct behavioral interpretation in terms of interaction probabilities. In particular, it follows at once from (1.30) and (2.12) that

$$\frac{p_c(ij)}{p_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} \tag{2.13}$$

and hence that regardless of the overall properties of a configuration, c, the *relative likelihood*, $p_c(ij)/p_c(ih)$, of (ij)-interactions versus (ih)-interactions depends only on the corresponding separation profiles, c_{ij} and c_{ih} [together with $B(j)$ and $B(h)$]. From a modeling viewpoint, this implies that the *conditional destination probabilities* in (2.6) now have the simple form

$$p_c(j \mid i) = \frac{B(j)F(c_{ij})}{\sum_{h\in J} B(h)F(c_{ih})}, \qquad j \in J. \tag{2.14}$$

Hence, for destination-deterrence-invariant gravity models, the associated destination gravity model in (2.14) is seen to be *invariant* with respect to the underlying separation configuration, $c \in C$.

(C) Origin-Deterrence-Invariant Gravity Models

In a completely parallel manner, we may also consider the following class of models which are invariant with respect to both origin and deterrence functions:

Definition 2.5
(i) (Model G4) A spatial interaction process, $\boldsymbol{P}$, is said to be mean-representable by an *origin-deterrence-invariant gravity model* if and only if there exists a positive origin function, A, and positive deterrence function, F, together with positive destination functions, B_c, for each separation configuration, $c \in C$, such that for all $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A(i)B_c(j)F(c_{ij}). \tag{2.15}$$

(ii) (Model G4*) If in addition the deterrence function F in (2.15) is nonincreasing on V, then $\boldsymbol{P}$ is said to be representable by a *monotone origin-deterrence-invariant gravity model.*

As a parallel to (2.12), it follows that for each destination, $j \in J$, these gravity models exhibit the following configurational invariance property with respect to relative mean interaction frequencies for all origins, $i, g \in I$:

$$\frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{gh})} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})}. \tag{2.16}$$

In a similar way, for independent spatial interaction processes, this invariance property implies that the *relative likelihoods*,

$$\frac{p_c(ij)}{p_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})}, \tag{2.17}$$

of (ij)-interactions versus (gj)-interaction depend only on the corresponding separation profiles, c_{ij} and c_{gj}, and hence that the *conditional origin probabilities* in (2.8) now have the simple form

$$p_c(i \mid j) = \frac{A(i)F(c_{ij})}{\sum_{g\in I} A(g)F(c_{gj})}, \qquad i \in I. \tag{2.18}$$

Hence, for origin-deterrence-invariant gravity models, the associated origin gravity model in (2.18) is *invariant* with respect to the underlying separation configuration, $c \in C$.

(D) Relatively Invariant Gravity Models

Given these types of partially invariant gravity models, we now consider models in which origin, destination, and deterrence functions are all free of configuration effects, as in (2.1) above. As a first step, we consider the somewhat more general class of gravity model in which *relative* mean interaction frequencies are representable by such invariant functions:

Definition 2.6
(i) (Model G5) A spatial interaction process, $\boldsymbol{P}$, is said to be mean-representable by a *relatively invariant gravity model* if and only if there exists a positive origin function, A, destination function, B, deterrence function, F, and configuration function, λ, such that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = \lambda(c)A(i)B(j)F(c_{ij}). \tag{2.19}$$

(ii) (Model G5*) If in addition the deterrence function F in (2.19) is nonincreasing on V, then $\boldsymbol{P}$ is said to be representable by a *monotone relatively invariant gravity model.*

To appreciate the analytical advantage of these models, observe that each mean frequency ratio, $\mathrm{E}_c(N_{ij})/\mathrm{E}_c(N_{gh})$, within any given configuration, c, depends only on the *local* configuration properties defined by the separation profiles c_{ij} and c_{gh}. Hence if the deterrence function F is known (or can be estimated), then the local impact of changes in these separation profiles can be analyzed. In particular, observe from (2.4) and (2.19) that for the case of independent interaction processes, the corresponding gravity representation of *interaction probabilities* depends only on the functions A, B, and F:

$$p_c(ij) = \frac{\mathrm{E}_c(N_{ij})}{\sum_{gh} \mathrm{E}_c(N_{gh})} = \frac{A(i)B(j)F(c_{ij})}{\sum_{gh} A(g)B(h)F(c_{gh})}. \tag{2.20}$$

Hence if mean interaction frequencies in such processes are representable by relatively invariant gravity models, then the corresponding interaction probabilities in (2.20) are now completely free of configuration effects.

(E) Invariant Gravity Models

While relatively invariant gravity models allow one to study the local impacts of configurational changes on relative mean frequencies, the impacts on actual mean frequency levels cannot be determined without knowledge of the configuration function λ in (2.19). In particular, it follows from (2.19) together with (1.14) that

$$\begin{aligned} \mathrm{E}_c(N) = \sum_{ij} \mathrm{E}_c(N_{ij}) &= \lambda(c) \sum_{ij} A(i)B(j)F(c_{ij}) \\ \Rightarrow \lambda(c) &= \frac{\mathrm{E}_c(N)}{\sum_{ij} A(i)B(j)F(c_{ij})} \end{aligned} \tag{2.21}$$

and hence that, aside from the functions, A, B, and F, the configuration function, λ, also depends on the functional relationship between total mean frequencies, $\mathrm{E}_c(N)$, and the underlying separation configurations, c. With this in mind, it is of interest to consider models in which the configuration function, λ, is constant across all configurations (and can thus be absorbed into A, B, or F). More formally, we now designate the class of models given by expression (2.1) as follows:

Definition 2.7
(i) (Model G6) A spatial interaction process, $\boldsymbol{P}$, is said to be mean-representable by an *invariant gravity model* if and only there exists a positive origin function, A, destination function, B, and deterrence function, F, such that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A(i)B(j)F(c_{ij}). \tag{2.22}$$

(ii) (Model G6*) If in addition the deterrence function F in (2.22) is nonincreasing on V, then $\boldsymbol{P}$ is said to be representable by a *monotone invariant gravity model.*

In such models, knowledge of the origin, destination, and deterrence functions completely determines the mean interaction frequency levels in every possible configuration. In particular, the impact of configurational changes on mean frequency levels depends only on changes in the separation profile, c_{ij}, and thus is seen to be completely localized in such models.

2.2.2 Illustrative Examples

Given these general model forms, we now develop two specific classes of spatial interaction processes in which mean interaction frequencies are representable by such models. To do so, we begin with the earliest probabilistic derivation of gravity models, due to Carroll and Bevis (1957), and formalize this derivation in terms of a class of spatial interaction processes designated as *Carroll-Bevis processes.* Next we develop a more explicit example of spatial interaction behavior giving rise to gravity models. In particular, a type of threshold-decision behavior is postulated for spatial actors, which is rooted in the general concept of satisficing behavior first proposed by Simon (1957). These ideas are formalized in terms of a class of interaction processes designated as *threshold interaction processes.* Finally, it should be noted that while instances of each of the above model types can be derived within the framework of either Carroll-Bevis or threshold interaction processes, we have chosen to develop only a selection of model types. In particular, we shall focus on the more common types of gravity models (Models G5 and G6) in which all functional forms are *configuration free.* With respect to *configuration-specific forms*, we shall develop only the most general gravity model (Model G1) in which all functions may depend on the underlying configuration. [Selected illustrations of intermediate forms are given in Sections 2.2.2 and 2.4.2 below.]

(A) Example 1: Carroll-Bevis Processes

As mentioned in the discussion of probabilistic theories in the Introduction, the first explicit probabilistic derivation of a gravity model was given by Carroll and Bevis (1957) [and extended by Isard and Bramhall (1960)]. Their essential idea was to begin with a null hypothesis about interaction behavior in the absence of any spatial influences, and then to postulate that (mean) deviations of actual behavior from this null hypothesis could be explained solely in terms of distance effects. This idea can be given a more precise formulation in terms of spatial interaction processes as follows. Suppose that in a given spatial interaction context the null hypothesis of spatially insensitive interaction behavior is taken to be representable by a particular spatial interaction process, $\boldsymbol{P}^o = \{P_c^o : c \in C\}$. While there are in principle many ways to model the process $\boldsymbol{P}^o$, the approach of Carroll and Bevis (1957) focuses on the relative interaction levels of a given origin with all possible destinations. In particular, if spatial actors are assumed to be insensitive to spatial locations (and to differ only in terms of their origin locations), then one might expect these relative interaction levels to be the same at every origin. To be more precise, if the mean interaction frequencies corresponding to each distribution P_c^o in $\boldsymbol{P}^o$ are denoted by $\mathrm{E}_c^o(N_{ij})$, and if the ratios of mean interaction frequencies, $\mathrm{E}_c^o(N_{ij})/\mathrm{E}_c^o(N_{ih})$, are taken to reflect the relative interaction levels between any given origin, $i \in I$, and pair of destinations, $j, h \in J$, then it might be hypothesized that these relative interaction levels are invariant across all origins, i.e., that for each configuration, $c \in C$, and destination pair, $j, h \in J$, process $\boldsymbol{P}^o$ satisfies the following condition for all origins, $i, g \in I$:

$$\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N_{ih})} = \frac{\mathrm{E}_c^o(N_{gj})}{\mathrm{E}_c^o(N_{gh})}. \tag{2.23}$$

Hence, if we now designate each spatial interaction process, $\boldsymbol{P}^o$, satisfying (2.23) as a *location-insensitive process*, then one may ask how actual interaction behavior differs from such processes. In particular, if actual interaction behavior is influenced by spatial separation effects then, again following Carroll and Bevis (1957), it might be postulated that variations in actual mean frequencies from those predicted by $\boldsymbol{P}^o$ can be explained in terms of spatial separation. To formalize this idea, suppose that actual interaction behavior is taken to be representable by a spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with mean interaction frequencies, $\mathrm{E}_c(N_{ij})$, and that variations in actual mean frequencies relative to those in $\boldsymbol{P}^o$ are taken to be representable by the ratios, $\mathrm{E}_c(N_{ij})/\mathrm{E}_c^o(N_{ij})$, for each $ij \in I \times J$ and $c \in C$. Then it might be postulated that these relative variations depend only on the corresponding spatial separation between i and j, i.e., that the following condition holds for all origin-destination pairs, $ij, gh \in I \times J$, and separation configurations, $c \in C$,

$$c_{ij} = c_{gh} \Rightarrow \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c^o(N_{ij})} = \frac{\mathrm{E}_c(N_{gh})}{\mathrm{E}_c^o(N_{gh})}. \tag{2.24}$$

Spatial interaction behavior consistent with these hypotheses can then be formalized in terms of the following class of interaction processes:

Definition 2.8 Each spatial interaction process, $\boldsymbol{P}$, satisfying (2.24) with respect to some location-insensitive process, $\boldsymbol{P}^o$, is designated as a *Carroll-Bevis process (CB-process).*

In terms of CB-processes, we now have the following illustration of Model G1 above [established in part (i) of Proposition 4.1 of Section 4.3.1 below]:

G1-Representability. *Every CB-process,* $\boldsymbol{P}$*, is mean-representable by a gravity model.*

In other words, conditions (2.23) and (2.24) imply the existence of origin functions, A_c, destination functions, B_c, and deterrence functions, F_c, for each separation configuration, $c \in C$, such that mean interaction frequencies for each CB-process, $\boldsymbol{P}$, are representable as in (2.2).

Uniform CB-Processes. Given this representational property of general CB-processes, we next observe that the location-insensitivity hypothesis in (2.23) is implicitly *configuration specific*, i.e., that relative interaction levels,

$E^o_c(N_{ij})/E^o_c(N_{ih})$, may in principle still vary between different configurations, c. But if behavior is truly insensitive to spatial separation, then one might reasonably expect such relative mean frequencies to be invariant not only with respect to the location of spatial actors, but also with respect to the separation configuration itself. More formally, one may postulate that $\boldsymbol{P}^o$ satisfies the following stronger version of (2.23) for all pairs of separation configurations, $c, c' \in C$, as well as origins, $i, g \in I$, and destinations, $j, h \in J$,

$$\frac{E^o_c(N_{ij})}{E^o_c(N_{ih})} = \frac{E^o_{c'}(N_{gj})}{E^o_{c'}(N_{gh})}. \tag{2.25}$$

Moreover, given this type of insensitivity to configurational effects, one might also expect the relative interaction activity levels among actors at different origins to remain invariant under configurational changes. More formally, if at each origin, $i \in I$, the *total origin-activity frequency* is denoted by $N_i = \sum_j N_{ij}$, with corresponding *mean origin-activity frequency*, $E^o_c(N_i)$, under each distribution, P^o_c, then it might be postulated that $\boldsymbol{P}^o$ satisfies the following additional invariance condition for all origins, $i, g \in I$, and configurations, $c, c' \in C$,

$$\frac{E^o_c(N_i)}{E^o_c(N_g)} = \frac{E^o_{c'}(N_i)}{E^o_{c'}(N_g)}. \tag{2.26}$$

In a manner paralleling location-sensitive processes, we now designate any spatial interaction process, $\boldsymbol{P}^o$, satisfying the stronger conditions (2.25) and (2.26) as a *configuration-insensitive process.* Given this stronger type of spatially insensitive behavior, observe next that if variations in actual mean interaction frequencies relative to those in $\boldsymbol{P}^o$ are again hypothesized to be explainable in terms of spatial separation effects, then as a parallel to condition (2.24), it may be postulated that the spatial interaction process, $\boldsymbol{P}$, representing actual behavior satisfies the following condition for all origins, $i, g \in I$, destinations, $j, h \in J$, and configurations, $c, c' \in C$,

$$c_{ij} = c'_{gh} \Rightarrow \frac{E_c(N_{ij})}{E^o_c(N_{ij})} = \frac{E_{c'}(N_{gh})}{E^o_{c'}(N_{gh})}. \tag{2.27}$$

With these hypotheses, we now have the following stronger version of Definition 2.8:

Definition 2.9 Each spatial interaction process, $\boldsymbol{P}$, satisfying (2.27) with respect to some configuration-insensitive process, $\boldsymbol{P}^o$, is designated as a *uniform CB-process.*

As an illustration of Model G5 above, we also have the following stronger representation property of uniform CB-processes [established in part (ii) of Proposition 4.1 of Section 4.3.1]:

G5-Representability. *Every uniform CB-process, $\boldsymbol{P}$, is mean-representable by a relatively invariant gravity model.*

Simple CB-Processes. Finally, to illustrate Model G6 within the context of CB-processes, suppose that in addition to the relative insensitivity hypotheses in (2.25) and (2.26) it is postulated that the *total* mean interaction levels in (1.13) are also insensitive to configurational changes. More formally, if each configuration-insensitive process, $\boldsymbol{P}^o$, satisfying the additional condition that

$$\mathrm{E}_c^o(N) = \mathrm{E}_{c'}^o(N), \tag{2.28}$$

for all configurations, $c, c' \in C$, is designated as a *strongly configuration-insensitive process*, then we may hypothesize that spatially insensitive interaction behavior is representable by such a process. Hence, as a further strengthening of Definition 2.8 we now say that:

Definition 2.10 Each spatial interaction process, $\boldsymbol{P}$, satisfying (2.27) with respect to a strongly configuration-insensitive process is designated as a *simple CB-process.*

For simple CB-processes, we then have the following stronger representation property of mean interaction frequencies [established in part (iii) of Proposition 4.1 of Section 4.3.1]:

G6-Representability. *Each simple CB-process, $\boldsymbol{P}$, is mean-representable by an invariant gravity model.*

As an explicit illustration of a simple CB-process, suppose it is postulated that spatially insensitive behavior is characterized by a spatial interaction process, $\boldsymbol{P}^o$, in which mean origin-activity frequencies, $\mathrm{E}_c^o(N_i)$, are proportional to the *actor population level*, a_i, at each origin, $i \in I$, so that for some proportionality factor, $\alpha > 0$,

$$\mathrm{E}_c^o(N_i) = \alpha a_i, \tag{2.29}$$

for all $i \in I$. Similarly, suppose that for each given origin, $i \in I$, the relative mean interaction frequencies with all destinations are given by the relative *opportunity population levels*, b_j, at each destination, $j \in J$, so that

$$\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N_{ih})} = \frac{b_j}{b_h}, \tag{2.30}$$

for all $i \in I$ and $j, h \in J$. Then the resulting process, $\boldsymbol{P}^o$, is strongly configuration-insensitive. To see this, observe simply that (2.25) and (2.26) follow immediately from (2.30) and (2.29), respectively, and that (2.28) also follows from (2.29) since $\mathrm{E}_c^o(N) = \sum_i \mathrm{E}_c^o(N_i) = \alpha \sum_i a_i$ for all $c \in C$. In addition, if we denote the *opportunity population total* by $b = \sum_j b_j$, then it also follows from (2.29) and (2.30) that for all $ij \in I \times J$ and $c \in C$,

$$\begin{aligned} \alpha a_i = \mathrm{E}_c^o(N_i) &= \sum_h \mathrm{E}_c^o(N_{ih}) = \mathrm{E}_c^o(N_{ij}) \sum_h [\mathrm{E}_c^o(N_{ih})/\mathrm{E}_c^o(N_{ij})] \\ &= \mathrm{E}_c^o(N_{ij}) \sum_h (b_h/b_j) = \mathrm{E}_c^o(N_{ij})(\sum_h b_h)/b_j \\ &\Rightarrow \mathrm{E}_c^o(N_{ij}) = \alpha a_i b_j/b, \end{aligned} \tag{2.31}$$

Given this null hypothesis about spatially insensitive interaction behavior, suppose next that relative variations of actual mean frequencies, $\mathrm{E}_c(N_{ij})$, from those in $\boldsymbol{P}^o$ are postulated to be a decreasing log-linear function of separation profiles, $c_{ij} = (c_{ij}^k : k \in K)$, i.e., that there exist positive parameters $(\theta_k : k \in K)$ such that

$$\log[\mathrm{E}_c(N_{ij})/\mathrm{E}_c^o(N_{ij})] = -\sum_{k \in K} \theta_k c_{ij}^k, \tag{2.32}$$

for all $ij \in I \times J$ and $c \in C$. Then it follows at once from (2.32) that the spatial interaction process, $\boldsymbol{P}$, representing actual behavior must be a simple CB-process with

$$\frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c^o(N_{ij})} = \exp[-\sum_k \theta_k c_{ij}^k]. \tag{2.33}$$

Moreover, it follows by combining (2.31) and (2.33) that

$$\mathrm{E}_c(N_{ij}) = (\alpha a_i)(b_j/b)\exp[-\sum_k \theta_k c_{ij}^k], \tag{2.34}$$

for all $ij \in I \times J$ and $c \in C$. Thus, as a special case of the general representation result for simple CB-processes above, we see from (2.34) that in the present case $\boldsymbol{P}$ is mean-representable by an (monotone) invariant gravity model with $A(i) = \alpha a_i$, $B(j) = b_j/b$, and $F(c_{ij}) = \exp[-\sum_k \theta_k c_{ij}^k]$ for all $ij \in I \times J$ and $c \in C$. As a specific instance of (2.34), observe that if we set $I = J$, $a_i = b_i$, and let $c_{ij} = \log(d_{ij})$ for some positive scalar measure of distance, then (2.34) reduces to

$$\mathrm{E}_c(N_{ij}) = \alpha(b_i b_j/b) d_{ij}^{-\theta}, \tag{2.35}$$

which is the original model derived by Carroll and Bevis (1957).

(B) Example 2: Threshold Interaction Processes

While the Carroll-Bevis processes above serve to illustrate the full range of gravity model representations, they involve no explicit hypotheses about the nature of actual spatial interaction behavior. Rather, hypotheses are made about *spatially insensitive behavior*, and actual behavior is treated only indirectly in terms of deviations from such hypothesized behavior. Hence, though this approach has a certain appeal from a testing viewpoint, it leaves open the fundamental question of how spatial separation actually influences interaction decisions. With this in mind, it is of interest to consider a class of processes in which such decisions are modeled more explicitly.

In particular, we shall focus on *discretionary* types of interaction behavior (such as shopping trips, recreational trips, visits to friends) in which there is a genuine possibility that any given interaction considered by an individual will not be taken. Basically, individual behavior is here postulated to involve an implicit two-stage process in which a variety of potential interaction situations arise (such as the desire to buy a certain commodity, see a certain movie, visit a certain friend), and are either acted upon or not, depending on the individual's current attitudes toward spatial interaction. More specifically, it is hypothesized that a given interaction will occur if and only if the anticipated travel costs (time, effort, stress) do not exceed the individual's current tolerance levels, here designated as his interaction *threshold levels.*

To model such interaction processes explicitly, we begin (as in Section 2 of Chapter 1) with a set of *actors*, $\alpha \in A$, distributed over a spatial configuration of *origin zones*, $i \in I$, and similarly, a set of opportunities, $\beta \in B$, distributed over a configuration of *destination zones*, $j \in J$, where A_i and B_j denote the sets of actors at i and opportunities at j, respectively. Each possible *interaction* from i to j then involves the decision of some actor, $\alpha \in A_i$, to interact with some opportunity, $\beta \in B_j$. For any opportunity currently being considered by an actor, his decision of whether or not to interact with that opportunity depends on his current willingness to incur all associated costs of overcoming the spatial separation between i and j (such as monetary costs, time costs, psychological costs). Assuming that the relevant types of spatial separation are representable by a K-dimensional *separation profile* with values in $V \subseteq R^K$ [as in expression (1.6) of Chapter 1], these attitudes toward spatial separation for individual, α, are here modeled in terms of a *threshold vector*, $t = (t_k : k \in K) \in R^K$, in which each component, t_k, is taken to represent α's current maximum tolerable level for separation attribute, k (where for example, t_1 = maximum tolerable travel time, t_2 = maximum tolerable travel cost, and so on). Each situation in which an actor, α, with tolerance levels, t, considers a possible interaction with opportunity, β, is designated as a *potential interaction situation*, (α, β, t). Given any potential interaction situation, (α, β, t) with $\alpha \in A_i$ and $\beta \in B_j$, if the prevailing levels of separation between i and j are given by separation profile, $c_{ij} = (c_{ij}^k : k \in K)$, then this situation will lead to a realized interaction if and only if the prevailing separation levels between i and j do not exceed α's current tolerance levels, i.e., if and only if $c_{ij} \leq t$ [equivalently, if and only if $c_{ij}^k \leq t^k$ for each component $k \in K$].

To analyze probabilistic behavior of this type within the framework of

Chapter 1, we begin with a probabilistic model of potential interaction situations. Since each such situation is described by a triple, $(\alpha, \beta, t) \in A \times B \times R^K$, the relevant *individual interaction space* (Section 1.4.1) for this case is given by $\Omega_1 = A \times B \times R^K$. The corresponding *outcome space*, Ω, then consists of all finite *potential interaction patterns*, $\omega = (\omega_r : r = 1, \ldots, n)$, with $\omega_r = (\alpha_r, \beta_r, t_r) \in \Omega_1$ for each $r = 1, \ldots, n$. Hence probabilistic behavior in this setting is taken to be representable by an interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω (Definition 1.2 of Chapter 1) which is designated as *a threshold interaction process (TI-process)* [a more precise definition of such processes is given in Definition 3.6 of Chapter 3 below].

Within this general setting, we again focus on *independent* interaction processes. However in the present case we shall introduce certain stronger independence hypotheses. To develop these hypotheses, we employ the following concepts. For each potential interaction pattern, $\omega = [(\alpha_r, \beta_r, t_r) : r = 1, \ldots, n]$, of size n, let $T_r(\omega) = t_r$ denote the threshold vector for the r-th potential interaction in ω. For each pattern of size n, these threshold attributes then define a matrix of random variables, $T^n = (T_r : r = 1, \ldots, n)$, on Ω_n [defined more precisely in Section 3.7.1 below]. The joint distribution of these random variables is then given for all possible *threshold patterns*, $t^n = (t_r : r = 1, \ldots, n)$, by

$$\begin{aligned} P_c^n(T^n \geq t^n) &= P_c^n(T_r \geq t_r : r = 1, \ldots, n) \\ &= P_c^n[\{\omega \in \Omega_n : T_r(\omega) \geq t_r\},\ r = 1, \ldots, n]. \end{aligned} \tag{2.36}$$

Next, if for each potential interaction pattern, $\omega = [(\alpha_r, \beta_r, t_r) : r = 1, \ldots, n]$, we let i_r and j_r denote the origin and destination zones of actor α_r and opportunity β_r, respectively, then we may denote the corresponding *spatial pattern* of potential interactions for ω by $s(\omega) = (i_r j_r : r = 1, \ldots, n) \in S$. With this notation, it follows that for each spatial pattern, $s \in S_n$, and threshold pattern, $t^n = (t_r : r = 1, \ldots, n)$, the probable occurrence of a threshold pattern at least as large as t^n, given that spatial pattern s occurs, must be of the form:

$$\begin{aligned} P_c^n(T^n \geq t^n \mid s) &= P_c^n[s, (T^n \geq t^n)]/P_c^n(s) \\ &= \frac{P_c^n[\{\omega \in \Omega_n : s(\omega) = s, T_r(\omega) \geq t_r, r = 1, \ldots, n\}]}{P_c^n[\{\omega \in \Omega_n : s(\omega) = s\}]}. \end{aligned} \tag{2.37}$$

Finally, if for each potential interaction pattern, $\omega \in \Omega$, we denote the corresponding *frequency profile* of potential interactions by $\boldsymbol{n} = \boldsymbol{n}(\omega) = (n_{ij} : ij \in I \times J)$, then the desired class of independent processes can be specified as follows:

Definition 2.11 A threshold interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is designated as a *fully independent TI-process* if and only if $\boldsymbol{P}$ satisfies the following three conditions:

T1. (Origin-Destination Independence) *For all non-null spatial potential-interaction patterns, $s = (i_r j_r : r = 1, \ldots, n) \in S$, and separation configurations, $c \in C$,*

$$P_c^n(s) = \prod_{r=1}^{n} P_c^n(i_r) P_c^n(j_r). \tag{2.38}$$

T2. (Frequency Independence) *For all potential-interaction frequency profiles, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, and separation configurations, $c \in C$,*

$$P_c(\boldsymbol{n}) = \prod_{ij \in I \times J} P_c(n_{ij}). \tag{2.39}$$

T3. (Threshold Independence) *For all threshold patterns, $t^n = (t_r : r = 1, \ldots, n)$, spatial interaction patterns, $s \in S_n$, and separation configurations, $c \in C$,*

$$P_c^n(T^n \geq t^n \mid s) = \prod_{r=1}^{n} P_c^1(T_1 \geq t_r). \tag{2.40}$$

[The term *fully independent* TI-process is employed here to distinguish these processes from the more general classes of *independent* TI-processes and *weakly independent* TI-processes developed below in Sections 3.7.2 and 3.9.1, respectively.] Given this definition, observe first that the origin-destination independence condition (T1) implies locational independence (A1), and that the frequency independence conditions (T2) and (A2) are identical. Hence each fully independent TI-process is an independent interaction process in the sense of Definition 1.3. The stronger condition of origin-destination independence essentially implies that potential interactions between actors and opportunities arise independently of any locational considerations. The threshold independence condition (T3) implies that attitudes toward spatial separation are intrinsic to individual actors, and are influenced neither by locational factors nor by the presence or absence of other individuals. [Note also that this condition combines conditions P2 and C3 in Definition 3.6 and 3.7 of Chapter 3.] Hence within any given separation configuration, $c \in C$, *all* effects of space on interaction behavior are assumed to be captured by these attitudinal threshold variables alone.

Realized-Interaction Frequencies. Given this class of processes, we now consider the probability that a given potential interaction situation, (α, β, t), will result in a *realized interaction*, i.e. that the separation profile between α and β does not exceed α's threshold profile, t, in any component. For each origin-destination pair, $ij \in I \times J$, separation configuration, $c \in C$, and each r-th component, $\omega_r = (\alpha_r, \beta_r, t_r)$, of potential interaction patterns, $\omega \in \Omega$, this *realized-interaction property* can be represented by a zero-one variable of the form:

$$\delta_{ij}^{cr}(\omega) = \begin{cases} 1, & \alpha_r \in A_i,\ \beta_r \in B_j,\ \text{and } T_r \geq c_{ij} \\ 0, & \text{otherwise.} \end{cases} \tag{2.41}$$

In terms of these zero-one variables, it follows that for each origin-destination pair, $ij \in I \times J$, the corresponding number, $N_{ij}(\omega)$, of realized interactions between i and j in pattern ω is given by:

$$N_{ij}(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^{n} \delta_{ij}^{cr}(\omega), & \omega \in \Omega_n, n > 0. \end{cases} \tag{2.42}$$

Hence, in a manner paralleling the general definition of interaction frequencies [in Section 1.4.2 above], the random variables defined by (2.42) are now designated as the *realized-interaction frequencies* between each origin-destination pair, $ij \in I \times J$. The probability distribution of N_{ij} under each probability measure, P_c, in process $\boldsymbol{P}$ is then definable in a manner paralleling expression (1.11) of Chapter 1.

With these definitions, our interest focuses on the representation properties of *mean realized-interaction frequencies*, $\mathrm{E}_c(N_{ij})$, for threshold interaction processes. In particular, if we now say that a given TI-process, $\boldsymbol{P}$, is *mean representable* by a given gravity model if and only if the mean realized-interaction frequencies in $\boldsymbol{P}$ are representable by that gravity model, then we have the following illustration of Model G1* in terms of such processes [established in Proposition 4.2 of Section 4.3.2 below]:

G1*-Representability. *Each fully independent TI-process, $\boldsymbol{P}$, is mean representable by a monotone gravity model.*

In particular, it is shown [in Proposition 4.2] that the (monotone) deterrence function, F_c, in (2.2) is in the present case given for all separation profiles, $v \in V_c$, by the *threshold probabilities:*

$$F_c(v) = P_c^1(T_1 \geq v). \tag{2.43}$$

Hence for such processes, the notion of *deterrence* has a clear behavioral interpretation. In particular, for any potential interaction between origin i and destination j, larger component values of the separation profile, c_{ij}, will reduce the probability that actors' threshold levels will exceed these values, and hence reduce the probability of a realized interaction.

Conditionally Configuration-Free Processes. Given the general monotone representability property, we next consider stronger conditions on TI-processes which lead to the correspondingly stronger types of monotone representability defined in Section 2.2.1 above. First, in a manner paralleling the notion of *configuration-insensitive* processes in Example 1 above, we consider TI-processes in which separation configurations may influence the overall frequency of potential interaction events, but not the individual nature of these events. To do so, recall from the development of general interaction processes in Section 1.4.1 above that for any given number of potential interaction situations, n, all probabilistic properties of individual interactions are describable under each configuration, $c \in C$, by the *conditional probability distribution*, P_c^n, on Ω_n. Hence we now consider TI-processes in which these conditional distributions are independent of the underlying configuration, c:

Definition 2.12 A TI-process, $\boldsymbol{P} = (P_c : c \in C)$, is said to be *conditionally configuration-free* if and only if for all configurations, $c, c' \in C$, pattern sizes, $n > 0$, and measurable events, $A \subseteq \Omega_n$,

$$P_c^n(A) = P_{c'}^n(A). \tag{2.44}$$

This stronger type of threshold interaction behavior yields the following illustration of Model G5* for mean realized-interaction frequencies [established in part (i) of Proposition 4.3 in Section 4.3.2 below]:

G5*-Representability. *If a fully independent TI-process, $\boldsymbol{P}$, is conditionally configuration-free, then $\boldsymbol{P}$ is mean-representable by a monotone relatively invariant gravity model.*

As a parallel to (2.43) above, the unique (monotone) deterrence function, F, in (2.19) is now given for all $v \in V$ by the configuration-free threshold probabilities:

$$F(v) = P^1(T_1 \geq v). \tag{2.45}$$

Configuration-Free Processes. Finally, to illustrate Model G6*, we consider threshold interaction behavior which is completely free of any overall configuration effects. To do so, we may strengthen Definition 2.12 as follows:

Definition 2.13 A TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, is called *configuration-free* if and only if for all configurations, $c, c' \in C$, and measurable events, $A \subseteq \Omega$,

$$P_c(A) = P_{c'}(A). \tag{2.46}$$

Hence, in configuration-free threshold interaction behavior, all spatial influences are assumed to be captured by the intrinsic attitudes of actors toward the individual separation profiles which are relevant for their particular interaction situations. Given this strong type of local configuration sensitivity, we have the following sharper representation property of such behavior [established in part (ii) of Proposition 4.3 in Section 4.3.2 below]:

G6*-Representability. *If a fully independent TI-process, $\boldsymbol{P}$, is configuration-free, then $\boldsymbol{P}$ is representable by a monotone invariant gravity model.*

2.2.3 Behavioral Characterizations

Given the specific illustrations above, we turn now to a general consideration of the types of population interaction behavior representable by such gravity models. To do so, we shall concentrate on population behavior representable by *independent spatial interaction processes*, in which the axioms of *locational independence* (A1) and *frequency independence* (A2) are implicitly adopted as basic behavioral hypotheses. Within this context, the central objective of the present section is to identify those additional properties of population behavior which specifically characterize gravity models of mean-interaction frequencies. To do so, we begin with an examination of each model form, and derive additional properties of interaction behavior that are implied by these representations. Each such property must necessarily be exhibited by all interaction behavior consistent with these models, i.e., must constitute a *necessary* condition for such behavior. But while many different necessary conditions are generally derivable from a given model form, we shall only be concerned with those conditions which are also *sufficient* for the existence of such model forms. Hence, in the context of independence axioms A1 and A2, these additional behavioral axioms may be said to constitute the *minimal theories* of behavior implied by these models (as discussed in the Introduction).

With these general observations in mind, our first objective is to derive certain *aggregate* behavioral properties of overall pattern probabilities which characterize each type of general gravity model form in Section 2.2.1 above. We next derive certain *local* behavioral properties of the relative likelihoods of given pairs of interactions which also characterize gravity-type behavior. Finally, we conclude this section with a number of observations on the relationships between these two types of axioms. Throughout the development to follow, a spatial interaction structure consisting of origin-destination pairs, $I \times J$, and configuration class, C, is assumed to be given.

(A) Aggregate Behavioral Characterizations

To identify the most general properties of population interaction behavior implied by gravity models, we begin by considering population behavior representable by an independent spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, and observe from (1.21), (1.23) and (1.30) in Chapter 1 that for separation configurations, $c \in C$, population sizes, $n > 0$, and spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n)$, $s' = (i'_r j'_r : r = 1, \ldots, n) \in S_n$,

$$\frac{P_c(s)}{P_c(s')} = \frac{P_c^n(s)}{P_c^n(s')} = \frac{\prod_{r=1}^n P_c^n(i_r j_r)}{\prod_{r=1}^n P_c^n(i'_r j'_r)} \frac{\prod_{r=1}^n \mathrm{E}_c(N_{i_r j_r})}{\prod_{r=1}^n \mathrm{E}_c(N_{i'_r j'_r})}. \tag{2.47}$$

Hence if we now designate a pair of interaction patterns, s and s', as *comparable patterns* whenever they are of the same size [i.e., whenever $N(s) = N(s')$], then it follows that the relative likelihoods of comparable patterns are determined entirely by the mean frequencies of their individual interactions.

Separation Dependence. With this general observation in mind, observe next that if mean interaction frequencies, $\mathrm{E}_c(N_{ij})$, are representable by any gravity model, i.e., are of the form

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)F_c(c_{ij}), \tag{2.48}$$

for all $ij \in I \times J$ and $c \in C$ (as in Model G1 of Definition 2.1), and if the origin and destination activity levels for each interaction pattern, $s \in S$, are denoted respectively by

$$N_i(s) = \sum_{j \in J} N_{ij}(s), \qquad i \in I, \tag{2.49}$$

$$N_j(s) = \sum_{i \in I} N_{ij}(s), \qquad j \in J, \tag{2.50}$$

then for any comparable patterns, $s, s' \in S$, it follows from (2.47) and (2.48) that

$$\begin{aligned} \frac{P_c(s)}{P_c(s')} &= \frac{\prod_{r=1}^n [A_c(i_r)B_c(j_r)F_c(c_{i_r j_r})]}{\prod_{r=1}^n [A_c(i'_r)B_c(j'_r)F_c(c_{i'_r j'_r})]} \\ &= \frac{\prod_{i\in I} A_c(i)^{N_i(s)} \prod_{j \in J} B_c(j)^{N_j(s)} \prod_{r=1}^n F_c(c_{i_r j_r})}{\prod_{i\in I} A_c(i)^{N_i(s')} \prod_{j \in J} B_c(j)^{N_j(s')} \prod_{r=1}^n F_c(c_{i'_r j'_r})}. \end{aligned} \tag{2.51}$$

Given this ratio of pattern probabilities, observe finally that if for each spatial interaction pattern, $s \in S$, we designate the vector

$$A(s) = [(N_i(s) : i \in I), (N_j(s) : j \in J)] \tag{2.52}$$

as the *(origin-destination)* activity profile for s, then for those pairs of patterns, $s, s' \in S$, which are *activity-equivalent* in the sense that $A(s) = A(s')$, it follows from (2.51) that

$$\frac{P_c(s)}{P_c(s')} = \frac{\prod_{r=1}^n F_c(c_{i_r j_r})}{\prod_{r=1}^n F_c(c_{i'_r j'_r})} \tag{2.53}$$

[where by definition $A(s) = A(s')$ implies that s and s' are comparable]. Thus the relative likelihood of activity-equivalent patterns is seen to depend entirely on their *separation* properties. To be more precise, if for each interaction pattern, $s = (i_r j_r : r = 1, \ldots, n) \in S$, we now designate the corresponding array of separation profiles

$$c_s = (c_{i_r j_r} : r = 1, \ldots, n) \tag{2.54}$$

as the *separation array* for s, then for any two activity-equivalent interaction patterns, $s, s' \in S_n$, it follows from (2.53) and (2.54) that

$$c_s = c_{s'} \Rightarrow F_c(c_{i_r j_r}) = F_c(c_{i'_r j'_r}),\ r = 1, \ldots, n$$
$$\Rightarrow \prod_{r=1}^{n} F_c(c_{i_r j_r}) = \prod_{r=1}^{n} F_c(c_{i'_r j'_r}) \Rightarrow P_c(s) = P_c(s'). \tag{2.55}$$

Hence we may conclude that in every independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, which is representable by a gravity model, pattern probabilities always satisfy the following *separation dependence* condition:

A3. (Separation Dependence) *For all separation configurations, $c \in C$, and activity-equivalent interaction patterns, $s, s' \in S$,*

$$c_s = c_{s'} \Rightarrow P_c(s) = P_c(s'). \tag{2.56}$$

Thus whenever the activity profiles and separation arrays of any two spatial interaction patterns, $s, s' \in S$, are the same for any given separation configuration, $c \in C$ [i.e., whenever $A(s) = A(s')$ and $c_s = c_{s'}$], their corresponding pattern probabilities, $P_c(s)$ and $P_c(s')$, must be equal. It is also important to note that (2.55) must hold for any spatial interaction patterns, $s, s' \in S$, in which c_s and $c'_{s'}$ differ only by the ordering of the individual interactions in s and s' [since products of deterrence values are independent of their ordering]. Hence while the order-specific definition of separation arrays, c_s, in (2.54) is convenient for analysis, such arrays can in fact be regarded as an *aggregate* property of interaction patterns, s, reflecting only the frequencies with which each separation profile appears in the individual interactions of s [as is made more explicit in Lemma 4.10 of Chapter 4 below]. But since activity profiles are also aggregate frequency properties of interaction patterns by definition, it follows from the separation dependence condition (A3) that all independent spatial interaction behavior representable by gravity models is entirely determined by the *aggregate* activity and separation properties of that behavior.

With these observations, it can also be shown [as in Theorem 4.1 of Chapter 4 below] that every independent interaction process satisfying Separation Dependence must be representable by a gravity model. Thus we now have the following characterization of such processes:

G1-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a gravity model if and only if $\boldsymbol{P}$ satisfies separation dependence* (A3).

Hence, we may conclude that independent interaction behavior is representable by a gravity model *if and only if* that behavior is entirely determined by its aggregate activity and separation properties as in axiom A3.

Relative Separation Dependence. Given this aggregate characterization of all gravity models, we next consider those models of mean interaction frequencies which exhibit the *deterrence-invariant condition* that

$$E_c(N_{ij}) = A_c(i)B_c(j)F(c_{ij}), \tag{2.57}$$

for all $ij \in I \times J$ and $c \in C$ (as in Model G2 of Definition 2.3). To derive an appropriate characterization of this behavior, observe first that if for each spatial interaction pattern, $s = (i_r j_r : r = 1, \ldots, n)$, with corresponding separation array, c_s, we denote the product of deterrence function values in (2.57) by

$$\Pi F(c_s) = \prod_{r=1}^{n} F(c_{i_r j_r}), \tag{2.58}$$

then for an interaction patterns, $s, t \in S$, satisfying the activity-equivalence condition that $A(s) = A(t)$, it follows by substituting (2.58) into (2.51) (with $s' = t$) that

$$\frac{P_c(s)}{P_c(t)} = \frac{\Pi F(c_s)}{\Pi F(c_t)}. \tag{2.59}$$

Moreover, since F is independent of c, then it follows that for all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, s' \in S$,

$$c_s = c'_{s'} \Rightarrow \Pi F(c_s) = \Pi F(c'_{s'}). \tag{2.60}$$

Hence, by combining (2.59) and (2.60), we see that for any pair of separation configurations, $c, c' \in C$, and quadruple of comparable interaction patterns, $s, t, s', t' \in S$, if $A(s) = A(t)$, $A(s') = A(t')$, $c_s = c'_{s'}$ and $c_t = c'_{t'}$, then we must have

$$(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{\Pi F(c_s)}{\Pi F(c_t)} \frac{\Pi F(c'_{s'})}{\Pi F(c'_{t'})} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.61}$$

Thus, every independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, representable by a deterrence-invariant gravity model must satisfy the following stronger version of separation independence in A3 above:

A3$'$. (Relative Separation Dependence) *For all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, if $A(s) = A(t)$ and $A(s') = A(t')$ then*

$$(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.62}$$

Hence, independent interaction behavior representable by deterrence-invariant gravity models satisfies the stronger condition that for any pair of activity-equivalent interaction patterns, $s, t \in S$, the *relative likelihood* of s and t depends only on their separation properties, c_s and c_t, within any given separation configuration, $c \in C$. Observe in particular, by setting $s' = t' = t$ and $c = c'$, that condition A3$'$ implies A3. But unlike A3, condition A3$'$ focuses on *relative* likelihoods of spatial interaction patterns, and asserts an *invariance* property of these relative likelihoods across configurations. To illustrate the behavioral significance of this invariance property, let us consider travel behavior between two possible residential zones, $i, k \in I$, and shopping zones $j, h \in J$, and in particular, consider the pair of activity-equivalent trip patterns, $s = (ij, kh)$ and $t = (ih, kj)$. In this context, suppose that the prevailing separation configuration, c, is altered by improving the roads from i and k to a third shopping zone, $d \in J$, while leaving the travel costs between $\{i, k\}$ and $\{j, h\}$ the same, so that the new configuration, c', satisfies $(c_{ij}, c_{ih}, c_{kj}, c_{kh}) = (c'_{ij}, c'_{ih}, c'_{kj}, c'_{kh})$. If the reduced costs to d induce some shoppers to switch from destinations j and h to d, then one might expect the likelihoods of both patterns s and t to decrease, i.e., $P_{c'}(s) < P_c(s)$ and $P_{c'}(t) < P_c(t)$. However, given that these trip patterns both involve origin and destination pairs for which costs do not change, i.e., $c_s = c'_s$ and $c_t = c'_t$, one might also expect that individual choices *between* these two destinations do not change, and hence that the *relative* likelihood of these patterns remains the same, i.e., $P_c(s)/P_c(t) = P_{c'}(s)/P_{c'}(t)$. This is seen to be precisely the type of behavioral invariance property asserted by A3$'$ (where in this illustration we have chosen $s = s'$ and $t = t'$). More generally, it should be noted in the development to follow that whenever an invariance property across configurations is asserted, a similar type of behavioral interpretation can be given.

As with separation dependence (A3), the importance of this relative separation dependence property (A3$'$) is that it is not only exhibited by all behavior consistent with deterrence-invariant gravity models, it essentially *characterizes* this behavior. To be more precise, in all spatial interaction contexts involving at least three distinct origins and/or three distinct destinations (i.e., with either $|I| \geq 3$ or $|J| \geq 3$), this behavior yields the following aggregate behavioral characterization of Model G2 [as shown in part (ii) of Theorem 4.3 in Chapter 4]:

G2-Characterization Theorem. *For any spatial interaction structure in which either $|I| \geq 3$ or $|J| \geq 3$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a deterrence-invariant gravity model if and only if $\boldsymbol{P}$ satisfies relative separation dependence* (A3$'$).

Recall from our general assumption on partitions (in Section 1.3.3 above) that $|I| \geq 2$ and $|J| \geq 2$ always hold. Hence the only case not covered by this result is the case in which $|I| = 2 = |J|$. In Example 4.1 of Chapter 4 below, it is shown that relative separation dependence is not generally sufficient to ensure the existence of deterrence-invariant gravity representations when $|I| = 2 = |J|$. Hence a stronger version of relative separation dependence (designated as *strong separation dependence*) is developed in Chapter 4 which is shown to yield a complete characterization of deterrence-invariant gravity models in all cases [see part (i) of Theorem 4.3 below]. However, in view of its limited applicability, we choose to defer this more technical development until Chapter 4.

The special classes of destination-deterrence-invariant and origin-deterrence-invariant gravity models will be considered in Section (B) below. Hence we turn now to models exhibiting full invariance of functional forms:

Uniform Separation Dependence. Let us begin by considering those models of mean-interaction frequencies which exhibit the *relative invariance* condition that

$$\mathrm{E}_c(N_{ij}) = \lambda(c)A(i)B(j)F(c_{ij}), \tag{2.63}$$

for all $ij \in I \times J$ and $c \in C$ (as in Model G5 of Definition 2.6). To derive an appropriate characterization of this behavior, observe that for any pair of separation configurations, $c, c' \in C$, and quadruple of comparable interaction patterns, $s, t, s', t' \in S_n$, if $A(s) = A(s')$, $A(t) = A(t')$, $c_s = c'_{s'}$ and $c_t = c'_{t'}$, then it now follows by substituting (2.63) and (2.58) into (2.51) that

$$\begin{aligned} \frac{P_c(s)}{P_c(t)} &= \frac{\lambda(c)^n \prod_i A(i)^{N_i(s)} \prod_j B(j)^{N_j(s)} \Pi F(c_s)}{\lambda(c)^n \prod_i A(i)^{N_i(t)} \prod_j B(j)^{N_j(t)} \Pi F(c_t)} \\ &= \frac{\lambda(c')^n \prod_i A(i)^{N_i(s')} \prod_j B(j)^{N_j(s')} \Pi F(c'_{s'})}{\lambda(c')^n \prod_i A(i)^{N_i(t')} \prod_j B(j)^{N_j(t')} \Pi F(c'_{t'})} = \frac{P_{c'}(s')}{P_{c'}(t')}. \end{aligned} \tag{2.64}$$

Thus, every independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, representable by a relatively invariant gravity model must satisfy the following *uniform* version of separation independence:

A3″. (Uniform Separation Dependence) *For all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, if $A(s) = A(s')$ and $A(t) = A(t')$ then*

$$(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.65}$$

Hence independent interaction behavior representable by relatively-invariant gravity models satisfies the stronger invariance condition that relative likelihoods of *all* pairs of comparable interaction patterns, $s, t \in S$, in any configuration, $c \in C$, depend only on their aggregate activity and separation properties, $[A(s),\ A(t),\ c_s,\ c_t]$, and are independent of any other properties of s, t, or c. [Observe again by setting $s' = t' = t$ and $c = c'$, that condition A3″ implies A3, and in the presence of A1 and A2 also implies A3′, as discussed in Remark 4.2 of Chapter 4]. Conversely, this uniform separation dependency condition (A3″) guarantees the existence of such relatively invariant gravity model representations in all cases [as shown in Theorem 4.5 of Chapter 4]. Hence we obtain the following aggregate behavioral characterization of Model G5:

G5-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a relatively invariant gravity model if and only if $\boldsymbol{P}$ satisfies uniform separation dependence* (A3″).

Separation Efficiency. Given these characterizations of general gravity models in terms of separation dependency conditions, recall next that the nature of the separation deterrence function in these models is left unspecified. Hence it is of interest to ask what additional conditions will ensure that spatial separation does act as a *deterrent* to interaction behavior. To derive such conditions, let us begin by considering independent interaction behavior, $\boldsymbol{P} = \{P_c : c \in C\}$, which is representable by a monotone gravity model (Model G1*), i.e., one in which the deterrence function, F_c, in (2.48) is nonincreasing for each $c \in C$. Then by the same arguments as in (2.51) through (2.55) above, it now follows that for any two activity-equivalent interaction patterns, $s = (i_r j_r : r = 1, \ldots, n)$, $s' = (i'_r j'_r : r = 1, \ldots, n) \in S$, with separation profiles satisfying, $c_s \leq c_{s'}$, we must have:

$$\begin{aligned} c_s \leq c_{s'} &\Rightarrow c_{i_r j_r} \leq c_{i'_r j'_r},\ r = 1, \ldots, n \\ &\Rightarrow F_c(c_{i_r j_r}) \geq F_c(c_{i'_r j'_r}) \quad , r = 1, \ldots, n \\ &\Rightarrow \Pi_{r=1}^n F\big(c_{i_r j_r}\big) \geq \Pi_{r=1}^n F_c\big(c_{i'_r j'_r}\big) \Rightarrow P_c(s) \geq P_c(s'). \end{aligned} \tag{2.66}$$

Hence, for any given levels of overall interaction activity, those patterns involving lower levels of separation in each interaction are at least as likely to occur as those involving higher levels. In this sense, spatial separation can indeed be said to have a deterrent effect on interaction behavior.

As an alternative behavioral interpretation of (2.66), observe that from among all interaction patterns giving rise to the same levels of overall interaction activity, the most likely patterns must always involve individual interactions with minimal levels of separation. In this sense, such populations may be said to exhibit *efficient* behavior in overcoming spatial separation. More formally, such behavior is seen to be consistent with the following *separation efficiency* property:

A4. (Separation Efficiency) *For all separation configurations, $c \in C$, and activity-equivalent spatial interaction patterns, $s, s' \in S$,*

$$c_s \leq c_{s'} \Rightarrow P_c(s) \geq P_c(s'). \tag{2.67}$$

This separation efficiency property can also be shown to imply that the gravity representation in (2.48) must be monotone [as shown in Theorem 4.2 of Chapter 4]. Hence we obtain the following behavioral characterization of Model G1* for independent spatial interaction processes:

G1*-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a monotone gravity model if and only if $\boldsymbol{P}$ satisfies separation efficiency* (A4).

Relative Separation Efficiency. To characterize independent spatial interaction behavior representable by gravity models which are both deterrence-invariant and monotone, we begin by observing from the argument in (2.66) that if the deterrence function, F, in (2.57) is nonincreasing then for all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$,

$$\begin{aligned}(c_s \leq c'_{s'}, c_t \geq c'_{t'}) &\Rightarrow [\Pi F(c_s) \geq \Pi F(c'_{s'}), \Pi F(c_t) \leq \Pi F(c'_{t'})] \\ &= \frac{\Pi F(c_s)}{\Pi F(c_t)} \geq \frac{\Pi F(c'_{s'})}{\Pi F(c'_{t'})}.\end{aligned} \tag{2.68}$$

Hence, if it is also true that $A(s) = A(t)$ and $A(s') = A(t')$, then by combining (2.59) and (2.68) we now have the following parallel to (2.61)]:

$$(c_s \leq c'_{s'}, c_t \geq c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{\Pi F(c_s)}{\Pi F(c_t)} \geq \frac{\Pi F(c'_{s'})}{\Pi F(c'_{t'})} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.69}$$

Hence we see that each independent spatial interaction process, $\boldsymbol{P}$, which is representable by a monotone deterrence-invariant gravity model must satisfy the following stronger version of separation efficiency:

A4′. (Relative Separation Efficiency) *For all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(t)$ and $A(s') = A(t')$,*

$$(c_s \leq c'_{s'}, c_t \geq c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.70}$$

As a parallel to the characterization of Model G2 above, is can be shown [as in part (ii) of Theorem 4.4 in Chapter 4] that this relative separation efficiency condition in turn implies that $\boldsymbol{P}$ satisfies (2.57) with respect to some nonincreasing deterrence function whenever $|I| \geq 3$ or $|J| \geq 3$, and hence yields the following characterization of Model G2* for independent interaction processes:

G2*-Characterization Theorem. *For any spatial interaction structure in which either $|I| \geq 3$ or $|J| \geq 3$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a monotone deterrence-invariant gravity model if and only if $\boldsymbol{P}$ satisfies relative separation efficiency* (A4′).

Uniform Separation Efficiency. To characterize independent spatial interaction behavior representable by gravity models which are both relatively invariant and monotone, we begin by observing that the argument in (2.68) must continue to hold for all comparable patterns, $s, t, s', t' \in S$. Hence if it is also true that $A(s) = A(s')$ and $A(t) = A(t')$, then by combining (2.68) together with the argument in (2.64) we may conclude that when $(c_s \leq c'_{s'}, c_t \geq c'_{t'})$ and F is monotone

$$\begin{aligned} \frac{P_c(s)}{P_c(t)} &= \frac{\prod_i A(i)^{N_i(s)} \prod_j B(j)^{N_j(s)}}{\prod_i A(i)^{N_i(t)} \prod_j B(j)^{N_j(t)}} \cdot \frac{\Pi F(c_s)}{\Pi F(c_t)} \\ &= \frac{\prod_i A(i)^{N_i(s')} \prod_j B(j)^{N_j(s')}}{\prod_i A(i)^{N_i(t')} \prod_j B(j)^{N_j(t')}} \cdot \frac{\Pi F(c_s)}{\Pi F(c_t)} \\ &\geq \frac{\prod_i A(i)^{N_i(s')} \prod_j B(j)^{N_j(s')}}{\prod_i A(i)^{N_i(t')} \prod_j B(j)^{N_j(t')}} \cdot \frac{\Pi F(c'_{s'})}{\Pi F(c'_{t'})} = \frac{P_{c'}(s')}{P_{c'}(t')}. \end{aligned} \tag{2.71}$$

Hence each independent spatial interaction process, $\boldsymbol{P}$, which is representable by a monotone relatively invariant gravity model must satisfy the following uniform version of separation efficiency:

A4″. (Uniform Separation Efficiency) *For all separation configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(s')$ and $A(t) = A(t')$,*

$$(c_s \leq c'_{s'}, c_t \geq c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.72}$$

This uniform separation efficiency condition in turn implies that $\boldsymbol{P}$ satisfies (2.63) with respect to some nonincreasing deterrence function [as shown in Theorem 4.6 in Chapter 4], and hence yields the following characterization of Model G5* for independent interaction processes:

G5*-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a monotone relatively invariant gravity model if and only if $\boldsymbol{P}$ satisfies uniform separation efficiency* (A4″).

Sub-Configuration Dependence. Finally, to characterize independent interaction behavior representable by invariant gravity models, suppose that mean interaction frequencies, $\mathrm{E}_c(N_{ij})$ are representable as in (2.63) with $\lambda(c) = 1$ for all separation configurations, $c \in C$. Then if for each separation configuration, $c \in C$, we designate the array,

$$c_i = (c_{ij} : j \in J), \tag{2.73}$$

of all separation profiles involving origin, i, as the *sub-configuration* at i, it follows at once that for any two separation configurations, $c, c' \in C$, with the same sub-configurations at i,

$$\begin{aligned} c_i = c'_i &\Rightarrow c_{ij} = c'_{ij}, \quad j \in J \\ &\Rightarrow A(i)B(j)F(c_{ij}) = A(i)B(j)F(c'_{ij}), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_{ij}) = \mathrm{E}_{c'}(N_{ij}), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_i) = \sum_j \mathrm{E}_c(N_{ij}) = \sum_j \mathrm{E}_{c'}(N_{ij}) = \mathrm{E}_{c'}(N_i). \end{aligned} \tag{2.74}$$

Hence if $\boldsymbol{P}$ is representable by an invariant gravity model, then each mean interaction activity level, $\mathrm{E}_c(N_i)$, is seen to depend only on the separation between origin, i, and all destinations, $j \in J$, i.e., on the *sub-configuration* at i. More formally, such mean interaction activity levels exhibit the following *sub-configuration dependence* property

A5. (Sub-Configuration Dependence) *For all origins, $i \in I$, and separation configurations, $c, c' \in C$,*

$$c_i = c'_i \Rightarrow \mathrm{E}_c(N_i) = \mathrm{E}_{c'}(N_i). \tag{2.75}$$

In addition, since each invariant gravity model is of course relatively invariant, each independent interaction process representable by an invariant gravity model must also satisfy uniform separation dependence (A3$''$). Conversely, these two conditions together guarantee the existence of invariant gravity model representations [as shown in Theorem 4.7 of Chapter 4], and hence yield the following characterization of Model G6 for independent spatial interaction processes:

G6-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by an invariant gravity model if and only if $\boldsymbol{P}$ satisfies both uniform separation dependence* (A3$''$) *and sub-configuration dependence* (A5).

In addition, it should be clear from the argument above that each independent interaction process, $\boldsymbol{P}$, representable by the stronger class of monotone invariant gravity models must also satisfy sub-configuration dependence. Hence as a direct parallel of the above characterization, we also obtain the following characterization of Model G6* [as in Theorem 4.8 of Chapter 4]:

G6*-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a monotone invariant gravity model if and only if $\boldsymbol{P}$ satisfies both uniform separation efficiency* (A4$''$) *and sub-configuration dependence* (A5).

(B) Local Behavioral Characterizations

While the above aggregate characterizations of gravity models are important from a behavioral viewpoint [and indeed serve to motivate our macro behavioral approach to gravity models (as discussed in Section 1.2.1)], these characterizations necessarily depend on *global* properties of spatial interaction patterns which are difficult to observe in practice. Hence, from a testing viewpoint, it is of interest to consider characterizations of gravity models in terms of *local* pattern properties which are more readily observable. To motivate our approach, we begin by observing first that if for any independent spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, and any choice of single-interaction pattern, $s = (ij) \in S_1$, we write $P_c(s) = P_c(ij)$, then it follows from expression (1.21) together with the Multinomial Sampling Theorem in Chapter 1 that for all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c \in C$,

$$\frac{P_c(ij)}{P_c(gh)} = \frac{P_c(ij)/P_c(S_1)}{P_c(gh)/P_c(S_1)} = \frac{P_c^1(ij)}{P_c^1(gh)} = \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{gh})}. \tag{2.76}$$

In other words, the relative likelihoods among single-interaction patterns are expressible entirely in terms of their corresponding mean interaction frequencies. This local relationship between relative likelihoods and mean interaction frequencies in independent interaction processes forms the basis of the gravity model characterizations to follow. With these general observations in mind, we first consider *destination-deterrence-invariant* and *origin-deterrence-invariant gravity models*, respectively, and derive the fundamental *proportionality* and *separability* properties which characterize each of these models. Next, we observe that the proportionality properties of these two models, designated respectively as *destination proportionality* and *origin proportionality*, together yield a complete local characterization of all relatively invariant gravity models. Finally, we derive the appropriate monotone versions of these proportionality axioms, designated respectively as *destination monotonicity* and *origin monotonicity*, which characterize monotone destination-deterrence-invariant and origin-deterrence-invariant gravity models.

Proportionality Properties. Let us begin by considering an independent spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with mean-interaction frequencies exhibiting the following *destination-deterrence-invariant* form

$$\mathrm{E}_c(N_{ij}) = A_c(i)B(j)F(c_{ij}), \tag{2.77}$$

for all $ij \in I \times J$ and $c \in C$ (as in Model G3 of Definition 2.4). Then it follows at once from (2.76) and (2.77) that for any origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$, if $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$, then

$$\frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{2.78}$$

Thus we see that the relative likelihood of an interaction with destination j versus h from any origin i under any configuration c is independent of all properties of i and c other than the separation profiles c_{ij} and c_{ih}, and hence that all destination-deterrence-invariant gravity models exhibit the following *destination proportionality* property:

A6. (Destination Proportionality) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c'_{gj}, c_{ih} = c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{2.79}$$

In a similar manner, suppose next that mean-interaction frequencies for $\boldsymbol{P}$ are of the following *origin-deterrence-invariant* form

$$\mathrm{E}_c(N_{ij}) = A(i)B_c(j)F(c_{ij}), \tag{2.80}$$

for all $ij \in I \times J$ and $c \in C$ (as in Model G4 of Definition 2.5). Then it now follows from (2.80) and (2.76) that if for any origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$, it is true that $c_{ij} = c'_{ih}$ and $c_{gj} = c'_{gh}$, then we must have

$$\frac{P_c(ij)}{P_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} = \frac{A(i)F(c'_{ih})}{A(g)F(c'_{gh})} = \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{2.81}$$

Thus for any origins, $i, g \in I$, and destination, $j \in J$, the relative likelihood of (ij)-interactions versus (gj)-interactions under any configuration c is independent of all properties of j and c other than the separation profiles c_{ij} and c_{gj}, so that origin-deterrence-invariant gravity models are seen to exhibit the following *origin proportionality* property:

A7. (Origin Proportionality) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c'_{ih}, c_{gj} = c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{2.82}$$

These fundamental proportionality properties are a direct consequence of the multiplicative form of the gravity model and, not surprisingly, have a long history in the literature [see for example the discussions in Cesario (1973) and Smith (1984), and in Section 2.2.3(C) below]. In particular, since each *relatively invariant* model is both destination-deterrence-invariant and origin-deterrence-invariant, it follows at once that such models must satisfy *both* of these proportionality conditions. More importantly, these two local behavioral properties together actually yield a complete characterization of relatively invariant gravity models [as shown in Theorem 4.13 of Chapter 4]. Hence, if we designate the conjunction of these two conditions (A6,A7) as *biproportionality,* then we now have the following local characterization of relatively invariant gravity models:

G5-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a relatively invariant gravity model if and only if $\boldsymbol{P}$ satisfies biproportionality* (A6,A7).

Note finally that since the sub-configuration dependence condition (A5) is local with respect to origins, we also have the following local behavioral characterization of Model G6 [as discussed in Remark 4.5 of Chapter 4]:

G6-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by an invariant gravity model if and only if $\boldsymbol{P}$ satisfies sub-configuration dependence* (A5) *and biproportionality* (A6,A7).

Separability Properties. In view of this characterization, it is natural to conjecture that the individual properties of destination proportionality and origin proportionality should characterize the respective classes of destination-deterrence-invariant and origin-deterrence-invariant gravity models. To see that this is not the case, it is of interest to derive one additional property exhibited by each of these models. Turning first to *destination-deterrence-invariant* models, notice again from (2.76) and (2.77) that if for any origin, $i \in I$, and destinations, $j, h \in J$, it is true that the (ij)-separation profile and (ih)-separation profile are the same in each of two separation configuration, $c, c' \in C$, i.e., if both $c_{ij} = c_{ih}$ and $c'_{ij} = c'_{ih}$, then we must have

$$\frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)}{B(h)} = \frac{B(j)F(c'_{ij})}{B(h)F(c'_{ih})} = \frac{P_{c'}(ij)}{P_{c'}(ih)}. \tag{2.83}$$

Hence, if (ij)-separation and (ih)-separation are the same [e.g., the travel distance to stores at j and h], then the relative likelihood of (ij)-interactions and (ih)-interactions depends only on the nonspatial attributes of j and h [e.g., the measurable qualities of stores at j and h] as reflected in the destination factors $B(j)$ and $B(h)$. In other words, the influence of spatial versus nonspatial attributes on interactions involving a common origin and alternative destinations can be effectively *separated.* Hence, all interaction behavior representable by destination-deterrence-invariant gravity models is seen to satisfy the following *destination separability* property:

A8. (Destination Separability) *For any origin, $i \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c_{ih}, c'_{ij} = c'_{ih}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(ij)}{P_{c'}(ih)}. \tag{2.84}$$

Given this additional property of such models, suppose we consider an independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with mean-interaction frequencies of the form

$$\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(N) \cdot \frac{(1 + ||c_{ij}||)^{-j}}{\sum_{gh \in I \times J} (1 + ||c_{gh}||)^{-h}}, \tag{2.85}$$

where $|| \cdot ||$ denotes *Euclidean norm*, and where destinations, $j \in J = \{1, 2, \ldots, |J|\}$, are represented by their rank orders with respect to certain desirable nonspatial qualities (so that destination 1 is 'best'). Then by (2.76) it follows that for all $i, g \in I$, $j, h \in J$, and $c, c' \in C$, with $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$, we must have

$$\frac{P_c(ij)}{P_c(ih)} = \frac{(1 + ||c_{ij}||)^{-j}}{(1 + ||c_{ih}||)^{-h}} = \frac{(1 + ||c'_{gj}||)^{-j}}{(1 + ||c'_{gh}||)^{-h}} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{2.86}$$

and hence that $\boldsymbol{P}$ satisfies destination proportionality (A6). However, it should be clear that $\boldsymbol{P}$ cannot satisfy destination separability. For example if we let $h = j + 1$ in (2.86), and if for any given separation profiles, $v, w \in V$, with $||v|| \neq ||w||$ we choose separation configurations, $c, c' \in C$, with $c_{ij} = v = c_{ih}$ and $c'_{gj} = w = c'_{gh}$, then it follows from (2.86) that $P_c(ij)/P_c(ih) = 1 + ||v|| \neq 1 + ||w|| = P_{c'}(ij)/P_{c'}(ih)$, and hence that $\boldsymbol{P}$ fails to satisfy destination separability (A8). Thus mean-interaction frequencies for $\boldsymbol{P}$ are not representable by any destination-deterrence-invariant gravity model, even though $\boldsymbol{P}$ satisfies destination proportionality (and exhibits reasonable behavior both with respect to spatial separation and nonspatial attributes).

In view of this counterexample, it is of interest to ask whether the additional requirement of destination separability is sufficient to ensure the existence of such representations. Unfortunately, even in the presence of this additional condition, there exist counterexamples involving spatial interactions with only two distinct destinations [as shown in Example 4.2 in Section 4.5.2(B) of Chapter 4 below]. However, these special cases turn out to be the only possible counterexamples [as shown in part (ii) of Theorem 4.9 in Chapter 4]. Hence we now have the following characterization of destination-deterrence-invariant gravity models (Model G3) in terms of the local behavioral axioms of destination proportionality and destination separability:

G3-Characterization Theorem. *For any spatial interaction structure with $|J| \geq 3$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a destination-deterrence-invariant gravity model if and only if $\boldsymbol{P}$ satisfies both destination proportionality* (A6) *and destination separability* (A8).

Finally, we note that local characterizations of Model G3 are possible even without the restriction, $|J| \geq 3$. In particular, since Model G3 has already been observed to be characterized by the aggregate behavioral axiom of relative separation dependence (A3′), it is not surprising that a local version of this axiom can be constructed which, together with destination proportionality, yields a local behavioral characterization of Model G3 in all cases [as shown in part (i) of Theorem 4.9 in Chapter 4]. However, since the case $|J| = 2$ is of limited practical interest, this more technical development is again deferred until Chapter 4.

Turning next to *origin-deterrence-invariant* models, observe from (2.80) that for any origins, $i, g \in I$, and destination, $j \in J$, it is true that the (ij)-separation profile and (gj)-separation profile are the same in each of two separation configurations, $c, c' \in C$, i.e., if both $c_{ij} = c_{gj}$ and $c'_{ij} = c'_{gj}$, then as a parallel to (2.83) we now have

$$\frac{P_c(ij)}{P_c(ih)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} = \frac{A(i)}{A(g)} = \frac{A(i)F(c'_{ij})}{A(g)F(c'_{gj})} = \frac{P_{c'}(ij)}{P_{c'}(gj)}. \tag{2.87}$$

Thus in this case, the influence of spatial versus nonspatial attributes on interactions involving a common destination and alternative origins can be effectively *separated*, and we see that all interaction behavior representable by origin-deterrence-invariant gravity models must satisfy the following *origin separability* property:

A9. (Origin Separability) *For any destination, $j \in J$, origins, $i, g \in I$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c_{gj}, c'_{ij} = c'_{gj}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ij)}{P_{c'}(gj)}. \tag{2.88}$$

As a parallel to the characterization of Model G3 above, we now have the following characterization of origin-deterrence-invariant gravity models (Model G4) in terms of origin proportionality and separability [as in part (ii) of Theorem 4.11 in Chapter 4]:

G4-Characterization Theorem. *For any spatial interaction structure with $|I| \geq 3$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by an origin-deterrence-invariant gravity model if and only if $\boldsymbol{P}$ satisfies both origin proportionality* (A7) *and origin separability* (A9).

As with Model G3, a local behavioral characterization of Model G4 is also possible which covers the case $|I| = 2$ [as shown in part (i) of Theorem 4.11 in Chapter 4].

Monotonicity Properties. Turning next to the monotone versions of these models, it should be clear that such models can be characterized by appropriate monotone versions of the destination and origin proportionality axioms. To derive these monotone versions, observe simply that if mean interaction frequencies are representable as in (2.63) with respect to some nonincreasing deterrence function, F, and if for any origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$, it is true that

$c_{ij} \leq c'_{gj}$ and $c_{ih} \geq c'_{gh}$, then it follows from the monotonicity of F that $F(c_{ij}) \geq F(c'_{gj})$ and $F(c_{ih}) \leq F(c'_{gh})$, so that we now have

$$\frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} \geq \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{2.89}$$

Hence the destination proportionality property in (2.78) is now replaced by the stronger property that the relative likelihood of (ij)-interactions versus (ih)-interactions is monotone nonincreasing in the separation profile, c_{ij}, and monotone nondecreasing in the separation profile, c_{ih}. More formally we now have the following *destination monotonicity property:*

A10. (Destination Monotonicity) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} \leq c'_{gj}, c_{ih} \geq c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} \geq \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{2.90}$$

In other words, if the separation between i and j is decreased, and the separation between i and h is increased, then (all else being equal) the relative likelihood of (ij)-interactions versus (ih)-interactions never decreases. As a parallel to the characterization of Model G3 above, we now have the following characterization of Model G3* in terms of destination monotonicity [as shown in part (ii) of Theorem 4.10 in Chapter 4]:

G3*-Characterization Theorem. *For any spatial interaction structure with $|J| \geq 3$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a monotone destination-deterrence-invariant gravity model if and only if $\boldsymbol{P}$ satisfies both destination separability* (A8) *and destination monotonicity* (A10).

Turning next to the appropriate monotone version of origin proportionality, observe that if $\boldsymbol{P}$ is representable by a monotone origin-deterrence-invariant gravity model, and if $c_{ij} \leq c'_{ih}$ and $c_{gj} \geq c'_{gh}$, then it follows from the monotonicity of F in (2.15) that $F(c_{ij}) \geq F(c'_{ih})$ and $F(c_{gj}) \leq F(c'_{gh})$, and hence that we now have the following parallel to (2.89):

$$\frac{P_c(ij)}{P_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} \geq \frac{A(i)F(c'_{ih})}{A(g)F(c'_{gh})} = \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{2.91}$$

Therefore all independent interaction behavior representable by monotone origin-deterrence-invariant gravity models is seen to satisfy the following stronger *origin monotonicity* condition:

A11. (Origin Monotonicity) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} \leq c'_{ih}, c_{gj} \geq c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} \geq \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{2.92}$$

As a parallel to the characterization of Model G4 above, we have the following local behavioral characterization of Model G4* [as shown in part (ii) of Theorem 4.12 in Chapter 4]:

G4*-Characterization Theorem. *For any spatial interaction structure with* $|I| \geq 3$*, an independent spatial interaction process,* $\boldsymbol{P}$*, on* $I \times J$ *is representable by a monotone origin-deterrence-invariant gravity model if and only if* $\boldsymbol{P}$ *satisfies both origin separability* (A9) *and origin monotonicity* (A11).

Next observe that, as with the local characterization of relatively invariant gravity models above, it follows by definition that every monotone relatively invariant gravity model must satisfy both of these monotonicity conditions. Hence, if we now designate the conjunction of these two conditions (A10,A11) as *bimonotonicity*, then as a parallel to the characterization of Model G5 above, we have the following local characterization of Model G5* [as in Theorem 4.14 of Chapter 4]:

G5*-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$*, is representable by a monotone relatively invariant gravity model if and only if* $\boldsymbol{P}$ *satisfies bimonotonicity* (A10,A11).

Note finally that since the sub-configuration dependence condition (A5) is local with respect to origins, we also have the following local behavioral characterization of monotone invariant gravity models (Model G6*), [as noted in Remark 4.5 of Chapter 4]:

G6*-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$*, is representable by a monotone invariant gravity model if and only if* $\boldsymbol{P}$ *satisfies sub-configuration dependence* (A5) *and bimonotonicity* (A10,A11).

(C) Comparisons Among Axioms

Before proceeding, it is of interest to consider certain relations between the aggregate axioms and local axioms developed above. To begin with, we consider relationships between the local and aggregate characterizations of relatively invariant gravity models given above. We then turn to a consideration of relationships between local and aggregate properties of more general gravity models.

Properties of Relatively Invariant Gravity Models. Let us begin by observing that the separate aggregate and local characterizations of relatively invariant gravity models above must imply that, within the context of independent interaction behavior (as defined by the locational and frequency independence axioms, A1 and A2), uniform separation dependence (A3″) is formally equivalent to biproportionality (A6,A7) [and similarly that uniform separation efficiency (A4$''$) is equivalent to bimonotonicity (A8,A9)]. But since axiom (A3$''$) implicitly involves global properties of separation configurations (for sufficiently large spatial interaction patterns), and since A6 and A7 involve only local properties of such configurations, it should be clear that axiom A3$''$ can be significantly weakened. Indeed, since biproportionality depends only on properties of interaction pairs [such as the pairs (ij, ih) and (gj, gh) in (2.78)], it is natural to expect that A3$''$ can be restricted to interaction patterns of size two without affecting the proof of existence of relatively invariant gravity representations [as discussed in Remark 4.4 of Chapter 4 below]. Hence, the present version of this aggregate axiom is designed primarily to emphasize the behavioral parallel between separation dependence (A3) and uniform separation dependence (A3$''$) [and similarly between separation efficiency (A4) and uniform separation efficiency (A4$''$)].

In addition, this uniform separation dependence axiom (A3$''$) can be given the following alternative interpretation in terms of its local consequences. Observe that if for any origin-destination pairs, $ij, gh \in I \times J$, we consider the comparable spatial interaction patterns, $s = (ij) = s' \in S_1$ and $t = (gh) = t' \in S_1$, then since $A(s) = A(s')$ and $A(t) = A(t')$, it follows from (2.65) that for all configurations, $c, c' \in C$,

$$\begin{aligned}(c_{ij}, c_{gh}) = (c'_{ij}, c'_{gh}) &\Rightarrow (c_s = c'_{s'}, c_t = c'_{t'}) \\ &\Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} \Rightarrow \frac{P_c(ij)}{P_c(gh)} = \frac{P_{c'}(ij)}{P_{c'}(gh)}.\end{aligned} \tag{2.93}$$

Hence the relative likelihood of any pair of interactions, (ij) and (gh), is seen to be independent of any properties of other interactions. In particular, if we set $i = g$ in (2.93) then it follows that the relative likelihoods of interactions with distinct destinations exhibits the following *independence-of-irrelevant-alternatives* (IIA) property at each origin:

IIA. (Independence-of-Irrelevant-Alternatives) *For all origins, $i \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij}, c_{ih}) = (c'_{ij}, c'_{ih}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(ij)}{P_{c'}(ih)}. \tag{2.94}$$

In other words, the properties of interaction alternatives at destinations other than j and h have no influence on the relative likelihood of (ij)-interactions versus (ih)-interactions by actors at any origin i. This IIA property has a long history in probabilistic choice modeling, starting with the work of Luce (1959) [mentioned in the Introduction]. In the context of spatial interaction models, versions of this property have been employed to derive gravity-type representations by Ginsberg (1972) and Smith (1975). Additional discussion can be found in Hua and Porell (1979) and Fotheringham and O'Kelly (1989, Section 4.72). [See also the discussion in Section 2.5.3(B) below].

Notice in addition that this IIA property is obtainable directly from the destination proportionality axiom (A6) by simply setting $i = g$ in (2.79). Hence it should be clear that both uniform separation dependence (A3″) and biproportionality (A6,A7) are much stronger than the IIA property. In particular, by setting $j = h$ in (2.93) [or equivalently, by setting $j = h$ in expression (2.82) of the origin proportionality condition (A7)], we also obtain the following dual version of the IIA property

$$(c_{ij}, c_{gj}) = (c'_{ij}, c'_{gj}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ij)}{P_{c'}(gj)}. \tag{2.95}$$

This additional property, which might be designated *independence-of-irrelevant-actors*, implies that with respect to interactions at any destination, the relative likelihood of interactions involving actors at origin i versus actors at origin g is independent of all properties of actors at other origins.

Properties of General Gravity Models. Given these additional local properties of independent interaction behavior representable by relatively invariant gravity models, we next consider various local properties of behavior implied by general gravity model representations, i.e., by Model G1 (and Model G1*). In particular, we begin by noting the conspicuous absence of any *local characterization* of such models. Indeed, the existence of local axioms which characterize this full class of models remains an open question. To illuminate the nature of the problem, we first observe that the above characterization theorems appear to suggest a likely candidate for such an axiom. In particular, since the aggregate characterization of general gravity models is seen to be obtainable from the aggregate characterization of relatively invariant gravity models by relaxing the uniform separation dependence condition (A3″) to separation dependence (A3), it would appear that a similar weakening of biproportionality might also yield this general model. To be more precise, let us consider the following weaker (configuration-specific) version of the biproportionality condition (A6,A7):

WB. (Weak Biproportionality) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c \in C$,*

$$(c_{ij} = c_{gj}, c_{ih} = c_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_c(gj)}{P_c(gh)}, \tag{2.96}$$

$$(c_{ij} = c_{ih}, c_{gj} = c_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_c(ih)}{P_c(gh)}. \tag{2.97}$$

By substituting (A_c, B_c, F_c) for (A, B, F) in (2.78) and (2.81), we see that every independent interaction process representable by a gravity model must satisfy weak biproportionality (WB), and hence that this axiom yields a natural necessary condition for such representability. However, weak biproportionality fails to be sufficient for such representability, as can be seen from the following simple counterexample. Consider a spatial interaction context involving four distinct origins, $I = (i, g, i', g')$, and four distinct destinations, $J = (j, h, j', h')$, with corresponding separation configuration class, $C = V^{I\times J}$. If we choose a separation configuration, $c \in C$, for which the only equalities between separation profiles involving distinct origin-destination pairs are the following:

$$c_{ij} = c_{i'j'}, c_{ih} = c_{i'h'}, c_{gj} = c_{g'j'}, c_{gh} = c_{g'h'}, \tag{2.98}$$

then we see from (2.96) and (2.97) that for any spatial interaction process, $\boldsymbol{P}$ weak biproportionality holds *vacuously* for this separation configuration [since no equalities in (2.98) involve either common origins or destinations]. However, if $\boldsymbol{P}$ is representable by a gravity model, as in (2.48) above, then the equalities in (2.98) imply that

$$\begin{aligned} \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} &= \frac{F_c(c_{ij})F_c(c_{gh})}{F_c(c_{ih})F_c(c_{gj})} \\ &= \frac{F_c(c_{i'j'})F_c(c_{g'h'})}{F_c(c_{i'h'})F_c(c_{g'j'})} = \frac{P_c(i'j')P_c(g'h')}{P_c(i'h')P_c(g'j')}. \end{aligned} \tag{2.99}$$

Hence, if one chooses any independent spatial interaction process, $\boldsymbol{P}$, satisfying (2.48) for all configurations in C other than c, and with probabilities violating condition (2.98) and (2.99) for c [say with $P_c(ij) = 2P_c(ih) > 0$ and $P_c(gh) = P_c(gj) = P_c(i'j') = P_c(i'h') = P_c(g'j') = P_c(g'h') > 0$] then it is clear that $\boldsymbol{P}$ will satisfy weak biproportionality (WB) but will fail to be representable by any gravity model for configuration c. This counterexample in fact shows that the 'cross-product invariance' condition in (2.98) and (2.99) yields an additional necessary condition for representability [which is also referred to in the literature as the 'odds ratio' property of gravity models, as for example in Kirby (1970), Sen and Sööt (1983) and Gray and Sen (1983); see also Chapter 6, equations (6.6), (6.9)]. This condition [which is also related to the local version of the relative separation dependence axiom employed to characterize G3 Models in Theorem 4.9 of Chapter 4] may be formalized within our framework as follows:

CPI. (Cross-Product Invariance) *For all origins, $i, g, i', g' \in I$, destinations, $j, h, j', h' \in J$, and separation configurations, $c \in C$,*

$$\begin{bmatrix} c_{ij} & c_{ih} \\ c_{gj} & c_{gh} \end{bmatrix} = \begin{bmatrix} c_{i'j'} & c_{i'h'} \\ c_{g'j'} & c_{g'h'} \end{bmatrix} \Rightarrow \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_c(i'j')P_c(g'h')}{P_c(i'h')P_c(g'j')}. \tag{2.100}$$

Moreover, since (2.100) is seen to imply both (2.96) and (2.97) [by setting $(i', g', j', h') = (g, g, j, h)$ and $(i', g', j', h') = (i, g, h, h)$, respectively, in (2.100)] it follows that CPI implies WB. Hence this new local axiom, CPI, appears to be a possible candidate for characterizing Model G1. However, to see that this is not the case, observe that CPI is simply a special case of the separation dependence axiom (A3). In particular, observe that the pair of spatial interaction patterns, $s = (ij, gh, i'h', g'j')$ and $s' = (i'j', g'h', ih, gj)$, are activity equivalent [since the origin labels, (i, g, i', g'), and destination labels, (j, h, j', h'), appear exactly once in each pattern]. Hence, by clearing denominators in (2.100) [and recalling the definition of $P_c(ij)$ together with both the locational independence axiom (A1) and Multinomial Sampling Theorem in Chapter 1], it follows that this condition asserts precisely that

$$\begin{aligned} c_s = c_{s'} &\Rightarrow \prod_{r=1}^{4} P_c(i_r j_r) = \prod_{r=1}^{4} P_c(i'_r j'_r) \\ &\Rightarrow \prod_{r=1}^{4} P_c^4(i_r j_r) = \prod_{r=1}^{4} P_c^4(i'_r j'_r) \Rightarrow P_c^4(s) = P_c^4(s') \\ &\Rightarrow P_c(s)/P_c(S_4) = P_c(s')/P_c(S_4) \Rightarrow P_c(s) = P_c(s'), \end{aligned} \tag{2.101}$$

and thus is seen to be a special case of the separation dependence axiom (A3) for patterns of size four. This observation shows at once that CPI cannot be equivalent to A3, since separation dependence for patterns of size 4 says nothing about separation dependence for general patterns of size larger than four. Moreover, since the proof of the G1-Characterization Theorem (Theorem 4.1 in Chapter 4) requires the full set of separation dependence conditions in A3 (which is not only infinite in number, but is also dependent on the *global* properties of separation configurations for sufficiently large pattern sizes), this observation also suggests that no local characterization of general gravity models may be possible.

2.3 Functional Specifications

Given the above characterizations of general gravity models, it is of interest to consider some of the more common types of functional specifications which have been employed in the literature. We begin with a brief consideration of typical specifications of origin and destination functions. This is followed by a more detailed examination of the many types of deterrence functions which have been employed.

2.3.1 Origin and Destination Functions

We begin by observing that there is a basic formal distinction implicit in the definitions of origin and destination functions, on the one hand, and deterrence functions on the other. This difference is seen most clearly in the relatively invariant gravity models of Definition 2.6, where deterrence functions are postulated to be explicit functions of numerical separation variables, while origin and destination functions are formally only weights associated with *nominal* (or *categorical*) origin and destination variables.

Hence one may well ask what is meant by 'specifications' of such origin and destination functions. Indeed, each value of such a function can in principle be treated as an unknown parameter to be estimated. This is done for example in the many 'Lagrange multiplier' derivations of such origin and destination weights in the literature [illustrated by the derivation of gravity representations for Kulback-Leibler processes in Proposition 4.7 of Chapter 4 below]. Estimates of the values of $A_c(i)$ and $B_c(j)$ can also be obtained by maximum likelihood methods as shown in Chapter 5. Indeed this is the approach taken by the authors in several practical applications and the treatment of parameter estimation in Chapters 5 and 6 also regards $A_c(i)$ and $B_c(j)$ as parameters to be estimated.

But while such abstract origin and destination weights are allowed by our general definitions, it is important to emphasize that this in no way precludes the possibility of relevant origin and destination attributes. With this in mind, it is of interest to consider the types of functional specifications of origin and destination weights which have been employed in the literature, as we do next. However, note that, because of the relationship, $A(i) = \mathrm{E}[N_i]/\sum_{j\in J} B(j)F(c_{ij})$, which follows by summing (2.1) over all j, any specification of $A_c(i)$ on the basis of i alone is highly restrictive. A similar observation can be made for $B_c(j)$.

To begin with, recall from the historical discussion in the Introduction that in the earliest applications of gravity models to migration, both origin and destination weights were postulated to be explicit functions of population levels. These population levels may be taken (in the context of Chapter 1 above) to reflect the numbers of actors at each origin and interaction opportunities at each destination, respectively. More formally, if we employ the same notation as in the development of Carroll-Bevis processes (Example 1 of Section 2.2.2 above), and again let a_i denote the *actor population size* at each origin, $i \in I$, and let b_j denote the *opportunity population size* at each destination, $j \in J$, then the classical specifications for relatively invariant gravity models (2.19) are given [as in expression (4) of the Introduction] by the *population power functions*

$$A(i) = (a_i)^r, \tag{2.102}$$

$$B(j) = (b_j)^s, \tag{2.103}$$

where the population exponents, r and s, are parameters to be estimated [and where all dimensional constants are implicitly assumed to be absorbed in the configuration function, λ, in (2.19)]. A detailed discussion of the structural influence of these population exponents on the resulting gravity parameters in given in Haynes and Fotheringham (1984, pp.12-16). As for their behavioral interpretation, observe first that if $r = s = 1$ (as in the Carroll-Bevis example), the product of functions, $A(i)B(j) = a_i b_j$, in (2.19) can be interpreted as simply the number of distinct (ij)-interactions which are possible. Hence for origin-destination pairs with the same levels of separation, it follows from (2.19) that mean interaction levels are simply proportional to the number of possible interactions between such origin-destination pairs. The motivation for more general specifications of power functions is typically based on the homogeneity properties of such functions. In particular, if the numbers of opportunities at every destination were to double, then it may be argued that the *relative* interaction activity levels among destinations should remain the same. More generally, it follows from (2.19) and (2.103) that if all opportunity population levels, b_j, are scaled by a common positive factor, ν , then for each pair of destinations, $j, h \in J$, the equality

$$\frac{(\nu b_j)^s}{(\nu b_h)^s} = \frac{\nu^s (b_j)^s}{\nu^s (b_h)^s} = \frac{(b_j)^s}{(b_h)^s} \tag{2.104}$$

implies that the relative interaction activity levels, $\mathrm{E}_c(N_j)/\mathrm{E}_c(N_h)$, remain the same [and indeed that for each origin, $i \in I$, the relative mean interaction activity levels, $\mathrm{E}_c(N_{ij})/\mathrm{E}_c(N_{ih})$, remain the same]. Moreover, this invariance property with respect to uniform scale changes, often designated as invariance under *similarity transformations*, is only exhibited by power functions (as discussed more fully in the next section on deterrence functions), and hence completely characterizes this class of functions. [Finally, it should be noted this type of invariance property is sometimes justified in terms of a desired invariance with respect to the choice of measurement units (such as monetary units or temporal units). However, since the choice of measurement units generally has no behavioral implications (and can always be achieved by adding appropriate dimensional constants), we choose to stress the *behavioral invariance* properties implied by power function representations.]

The simple specifications (2.102) and (2.103), in terms of population levels, can of course be extended to include other relevant attributes of actors and opportunities (all implicitly assumed to be *extensively measurable*, as in the discussion of Section 1.3.1 above). In the case of shopping behavior, for example, it is clear that family size and income levels may be relevant attributes of actors. Similarly, store sizes, product variety, and price levels may be relevant attributes of opportunities. Hence if all relevant *actor attributes* at each origin, $i \in I$, are represented by the finite family of positive measures, $(x_{it} : t \in T)$, and similarly, if all relevant *opportunity attributes* at each destination, $j \in J$, are represented by the finite family of positive measures, $(z_{jw} : w \in W)$, then one may consider the following family of *composite power functions*

$$A(i) = \prod_{t \in T} (x_{it})^{r_t}, \tag{2.105}$$

$$B(j) = \prod_{w \in W} (z_{jw})^{s_w}, \tag{2.106}$$

where the exponents, $(r_t : t \in T)$ and $(s_w : w \in W)$, are parameters to be estimated. [For illustrations of a range of explicit attribute specifications including shopping, migration and other examples, see for example Fotheringham and O'Kelly (1989, Chapter 6) and Putman and Ducca (1978a)]. Instead of multiplicative specifications, linear specifications have also been used [as for example in Porell (1980)]. In exploratory analysis using such origin and destination functions, one of the authors has applied the following method: estimate $A(i)$'s and $B(j)$'s first [using maximum likelihood procedures, for example] and then use these parameters as the dependent variable values in a linear least squares, or other model fitting, procedure. Porell (1980) has also applied this approach.

The above specifications can, of course, be extended to the more general case of *configuration-dependent* origin and destination functions by simply adding the subscript 'c' to all parameters. However, it is also possible to consider a number of more interesting types of configuration dependencies in terms of explicit configuration-dependent attributes of actors and opportunities. With respect to actors, it may be argued for example that if in a given separation configuration, $c \in C$, some origin location, $i \in I$, is highly separated from *all* relevant destinations (other than location i itself), then this type of 'isolation effect' may inhibit the overall interaction activity of actors at i. In the case of a scalar separation measure (such as travel time) this type of isolation effect might be reflected by a *minimum-separation measure* of the form

$$s_{ic} = \min\{c_{ij} : j \in J/i\}, \tag{2.107}$$

where $J/i = J - \{i\}$. If the diminishing effect of such separation on interaction activity is reflected by a negative exponent, $-r$, then (2.105) might be extended to include *configuration-dependent origin functions* of the form

$$A_c(i) = (s_{ic})^{-r} \prod_{t \in T} (x_{it})^{r_t}. \tag{2.108}$$

A more comprehensive approach might be to include the overall accessibility of each origin to all opportunities. The simplest measure of this type, first proposed by Hansen (1959) [and subsequently axiomatized for more general types of deterrence functions by Weibull (1980) and Smith (1980)], is given by the following class of *origin-accessibility measures:*

$$a_{ic} = \sum_{j \in J/i} b_j (c_{ij})^{-\theta}, \tag{2.109}$$

where b_j again denotes the number of opportunities at j, and where accessibility to j is reflected by a simple power deterrence function (as discussed more fully in the next section). If the enhancing effect of such accessibility on interaction activity is reflected by a positive exponent, r, then in this case (2.105) might again be extended to include the following class of *configuration-dependent origin functions* [first proposed by Cesario (1975)]:

$$A_c(i) = (a_{ic})^{r} \prod_{t \in T} (x_{it})^{r_t}. \tag{2.110}$$

Similarly, one may consider configuration dependencies in destination functions as well. For example, it has often been observed that the presence of agglomerations of opportunities in space tends to enhance their individual attractiveness in terms of mutual accessibility to one another (as for example in comparison shopping). For the case of a scalar measure of separation, one may again employ a Hansen-type measure,

$$a_{jc} = \sum_{h \in J/j} b_h (c_{jh})^{-\theta}, \tag{2.111}$$

of the *destination accessibility* of j to all other destinations, where in this case it is implicitly assumed that each destination is itself a possible origin (i.e., that $J \subseteq I$). If the enhancing effect of accessibility is again reflected by a positive exponent, s, then (2.106) might be extended to include *configuration-dependent destination functions* [first proposed by Fotheringham (1983a,1986)]:

$$B_c(j) = (a_{jc})^s \prod_{w \in W} (z_{jw})^{s_w}. \tag{2.112}$$

Alternatively, it has also been suggested by Fotheringham (1983a,1983b) that 'competition effects' among opportunities in close proximity might in some cases diminish their individual attractiveness relative to more prominent isolated opportunities. Hence the sign of s might actually be negative in such cases. [Further discussion of 'prominence' models is given in Section 2.5.3(B) below.]

It should be noted that when such configuration-dependent attributes appear multiplicatively in origin or destination functions, it is in principle possible to treat these attributes as part of the relevant separation deterrence function. For example, if the minimum-separation attribute in (2.107) is combined with any deterrence function, F_c, then one obtains an *origin-dependent deterrence function* of the form

$$F_{ic}(c_{ij}) = (s_{ic})^{-\sigma} F_c(c_{ij}). \tag{2.113}$$

However, we choose to maintain the classical separable form of Model G1 in which the relevant deterrence functions, F_c, are taken to involve only those aspects of deterrence which are common to all spatial interactions (i.e., involve no origin-specific or destination-specific effects). [Possible relaxations of these separability assumptions are discussed in more detail in Section 2.5 below.]

2.3.2 Deterrence Functions

Turning next to deterrence functions, it should be clear from our emphasis on separation measures that the specification of deterrence functions is of central importance for our purposes. Indeed, in so far as these functions reflect the attitudes of actors toward spatial separation, they constitute (in our view) the very core of gravity models. However, it is important to recall that our general concept of 'separation' can involve *ordinal* measures (such as the 'cultural separation' or 'religious separation' between groups or individuals discussed in Section 1.3.1 of Chapter 1). Indeed, in our general formulation of gravity models above (Definitions 2.1 through 2.6), the specific nature of individual separation variables can be quite arbitrary. But if 'separation' is measurable only in ordinal terms, then for each numerical representation of this measure, one can obtain a *different* deterrence function. Hence, while the notions of multiplicative separability and (ordinal) deterrence are perfectly meaningful in such cases, the specific *measurement status* of deterrence functions is very weak.

With this in mind, we shall focus on those types of separation which are extensively measurable, i.e., on interaction *cost measures* for which all classical specifications of deterrence functions are meaningfully parameterized (as in Section 1.3.1 above; non-cost separation functions have been considered very briefly in this section and in Chapter 5, in the context of step function deterrence functions). Such interaction costs may, for example, include expenditures of time and money, as well as distance traversed between origins and destinations. In addition to these measurement restrictions, we choose (for sake of notational simplicity) to focus only on specifications of deterrence functions, F, for *deterrence invariant* gravity models in which all parameters are assumed to be independent of the underlying separation configuration. In this context, we begin below by considering the classical *power* specification of deterrence functions for scalar measures of separation. This is followed by a consideration of *exponential* specifications, which will occupy a central role in our subsequent analyses. Finally, a number of *multivariate* specifications are discussed which may involve more than one type of interaction cost.

Openshaw and Connolly (1977) have compiled a list of possible functions which includes several forms not mentioned below [see also Chojnicki, 1966]. Different forms have also been empirically compared in Openshaw and Connolly (1977, see also Baxter, 1978) and also in Black and Larson (1972). A number of empirically obtained plots of F as a function of travel time are given in FHWA (1974).

Power Deterrence Functions

Recall from the development in the Introduction that the classical deterrence functions were postulated to be negative powers of distance (typically with values -1 and -2). More generally, for any positive scalar cost measure, c_{ij}, and positive parameter, θ, we have the following power-function specification of F,

$$F(c_{ij}) = (c_{ij})^{-\theta}. \tag{2.114}$$

Each such function, F, is designated as a *power-deterrence function* with *cost sensitivity parameter,* θ. In view of the formal relation of (2.114) to the "Pareto distribution of the first kind" [Johnson and Kotz, 1970, Chapter 19, expression (2.2)], this specification of F may also be designated as a *Type-1 Pareto deterrence function* [see for example in Morrill and Pitts (1967)].

From a behavioral viewpoint, a possible justification for this specification of deterrence functions is again suggested by the homogeneity property of such functions [as discussed for origin and destination functions above]. In particular, if it is true that relative mean interaction levels, $\mathrm{E}_c(N_{ij})/\mathrm{E}_c(N_{ih})$, remain invariant under similarity transformations, where all separation values are scaled by a common positive factor, ν, then it follows from (2.9) that

$$\begin{aligned} &\frac{\mathrm{E}_{\nu c}(N_{ij})}{\mathrm{E}_{\nu c}(N_{ih})} = \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{ih})} \\ &\Rightarrow \frac{\lambda(\nu c)A_{\nu c}(i)B_{\nu c}(j)F(\nu c_{ij})}{\lambda(\nu c)A_{\nu c}(i)B_{\nu c}(h)F(\nu c_{ih})} = \frac{\lambda(c)A_{\nu c}(i)B_{\nu c}(j)F(c_{ij})}{\lambda(c)A_{\nu c}(i)B_{\nu c}(h)F(c_{ih})} \\ &\Rightarrow \frac{F(\nu c_{ij})}{F(\nu c_{ih})} = \frac{F(c_{ij})}{F(c_{ih})}. \end{aligned} \tag{2.115}$$

Hence if for any positive number, x, we choose a cost configuration with $c_{ij} = x$ and $c_{ih} = 1$, then it follows from (2.115) that for all $x, \nu > 0$,

$$\begin{aligned} F(\nu x) = \frac{F(\nu)F(x)}{F(1)} &\Rightarrow \frac{F(\nu x)}{F(1)} = \frac{F(\nu)}{F(1)} \cdot \frac{F(x)}{F(1)} \\ &\Rightarrow \phi(\nu x) = \phi(\nu)\phi(x), \end{aligned} \tag{2.116}$$

where $\phi(x) = F(x)/F(1)$. But if F is assumed to be eventually decreasing on some interval (for sufficiently large cost values) then it is well known [see for example Eichhorn (1978, Theorem 1.9.13 and Remark 1.9.23)] that the only functions, ϕ, satisfying (2.116) for all positive values of x and ν must be of the form, $\phi(x) = x^{-\theta}$ for some $\theta > 0$. Hence we may conclude from (2.116) that F must be of the form (2.114) [where it is implicitly assumed that the proportionality constant $1/F(1)$ in (2.116) is absorbed into the origin or destination weights in (2.9)].

From a behavioral viewpoint, however, this type of invariance under similarity transformations is very questionable for *small cost values.* Indeed, since $x \to 0$ implies that $x^{-\theta} \to \infty$ for all $\theta > 0$, it follows from (2.115) that mean interaction activity between origins and destinations with very small values of interaction costs must be *overwhelmingly* larger than all other mean interaction levels. But since this type of interaction behavior is generally not observed, it has been widely recognized that power deterrence functions are not appropriate for modeling interactions involving small cost values [as discussed for example in Fotheringham and O'Kelly (1989, pp.12-13)].

Hence a number of alternative specifications of F have been proposed to remedy this situation. Indeed, this problem was already noted by Stewart (1948, p.48), who proposed that within-city interaction costs be taken to be the average cost of interactions within the city itself. An alternative approach, first suggested by Anderson (1956), is to postulate the existence of some positive *start-up cost*, $\epsilon > 0$, implicit in every interaction. One version of this approach is to include ϵ as an additive constant in the definition of interaction costs, c_{ij}, and thus obtain the following two-parameter *shifted power-deterrence function*:

$$F(c_{ij}) = (\epsilon + c_{ij})^{-\theta}. \tag{2.117}$$

In view of its formal relation to the "Pareto distribution of the second kind" [Johnson and Kotz, 1970, Chapter 19, expression (2.3)] this deterrence function might also be designated as a *Type-2 Pareto deterrence function.* [Such deterrence functions have been obtained in a different context by Harris (1964), who designated them as *modified gravity models.*] A second version of the start-up cost approach is to include ϵ as an additive constant distinct from distance, and thus to obtain the following alternative specification

$$F(c_{ij}) = [\epsilon + (c_{ij})^{\theta}]^{-1}. \tag{2.118}$$

In view of its relation to the 'Champernowne distribution' [Johnson and Kotz, 1970, Chapter 19, expression (8.53)], this deterrence function may be designated as the *Champernowne deterrence function.*

Exponential Deterrence Functions

One further specification which resolves this 'small distance problem' without requiring additional parameters is the popular *exponential deterrence function*

$$F(c_{ij}) = \exp[-\theta c_{ij}], \tag{2.119}$$

in which θ may again be interpreted as a *cost sensitivity parameter*. However it is important to emphasize that, unlike the power specification above, the parameter θ in this case must be a *dimensional parameter* with specific value depending on the choice of units for interaction costs. In other words, (2.119) is a meaningfully parameterized function with respect to extensively measurable interaction costs if and only if all transformations of measurement units are taken to be absorbed in the definition of θ. This specification of F appears to have first been studied by Kulldorf (1955), who found it to perform better than the power deterrence function in fitting migration data. These empirical findings were later supported by the work of Morrill and Pitts (1967) [see also the studies cited in Fotheringham and O'Kelly (1989, p.13)]. However, our primary interest in exponential deterrence functions centers on their theoretical significance from a behavioral viewpoint. Hence we shall consider this specification in much more detail in Section 2.4 below.

MULTIVARIATE DETERRENCE FUNCTIONS

Recall that our definition of separation profiles allows for the possibility of multiple measures of separation. Hence it is important to extend the specifications above to multidimensional profiles of interaction costs. First, in a manner completely paralleling the discussion of composite power specifications of origin and destination functions above, it should be clear that for any set of positive cost profiles, $c_{ij} = (c_{ij}^k : k \in K) \in R_{++}^K$ [where in this case, $K = K_+$ in (1.6) of Chapter 1], one can extend the power specification in (2.114) above to the following class of *multivariate power deterrence functions*,

$$F(c_{ij}) = \prod_{k \in K} (c_{ij}^k)^{-\theta_k}, \tag{2.120}$$

with corresponding *cost sensitivity vector*, $\theta = (\theta_k : k \in K) \in R^K$ [where, of course, F is nonincreasing if and only if $\theta \in R_+^K$]. As in the scalar case, this functional form is characterized by invariance of relative mean interactions under arbitrary scale transformations of any (or all) separation components [see for example the more general formulation of this result in Krantz, *et al.* (1971, Theorem 10.4)].

Similarly, for *any* set of separation profiles, $c_{ij} = (c_{ij}^k : k \in K) \in R^K$, one may extend the exponential specification in (2.119) above to the multivariate case. In particular, following Sen and Sööt (1981), we may define the class of *multivariate exponential deterrence functions*,

$$F(c_{ij}) = \exp[-\sum_{k \in K} \theta_k c_{ij}^k], \tag{2.121}$$

with *cost sensitivity vector*, $\theta = (\theta_k : k \in K) \in R^K$ [where we again employ the sign convention that F is nonincreasing if and only if $\theta \in R_+^K$, and where we again note that θ is a vector of *dimensional parameters*]. As in the scalar case, we shall study the behavioral properties of multivariate exponential deterrence functions in detail in Sections 2.4.2 and 2.4.3 below. Hence for the present, it suffices to observe that from a practical viewpoint, this class of multivariate deterrence functions provides a very flexible representational framework for modeling purposes. In particular, observe that since the sign of separation values in (2.121) is *unrestricted*, one can consider linear combinations, $\sum_k \theta_k c_{ij}^k$, of arbitrary cost measures as 'composite measures' of separation cost. For example, the multivariate power deterrence function is itself seen to be *log linear*, and hence representable in this form by simply replacing each positive separation measure c_{ij}^k with its log.

More importantly, this general log-linear form yields a framework within which the above classes of *scalar* deterrence functions can be considerably expanded. For example, it should be clear that a more flexible class of scalar specifications can be obtained by simply combining the power specification (2.114) and the exponential specification (2.119) [for positive separation values, c_{ij}] into a single two-parameter family of *power-exponential deterrence functions,*

$$F(c_{ij}) = (c_{ij})^{-\sigma} \exp[-\theta c_{ij}]. \qquad (2.122)$$

This specification is particularly advantageous from a statistical viewpoint when comparing the relative appropriateness of power specifications and exponential specifications [as for example in Morrill and Pitts (1967)]. In addition, this combined form may in certain cases have behavioral meaning in its own right [see for example the derivation of (2.122) in terms of 'marginal perceived costs' in Zaryouni and Liebman (1976)]. To express this composite form within the framework of multidimensional exponential deterrence functions, observe simply that if one defines the two-dimensional separation profile, $c_{ij} = (c_{ij}^1, c_{ij}^2)$, by $c_{ij}^1 = \log(c_{ij}^2)$, then (2.122) is equivalent to the following two-dimensional instance of (2.121):

$$F(c_{ij}) = \exp[-(\theta_1 c_{ij}^1 + \theta_2 c_{ij}^2)] = (c_{ij}^2)^{-\theta_1} \exp[-\theta_2 c_{ij}^2]. \qquad (2.123)$$

As a second illustration of such composite scalar deterrence functions, Harris (1964) obtained the following class of deterrence functions which combine (2.117) and (2.119),

$$F(c_{ij}) = (\epsilon + c_{ij})^{-\sigma} \exp[-\theta c_{ij}], \qquad (2.124)$$

and hence might be designated as *shifted power-exponential deterrence functions.* Here, if one defines $c_{ij}^1 = \log(\epsilon + c_{ij}^2)$, then, for each prespecified ϵ, (2.124) is also seen to be a two-dimensional instance of (2.121).

A final illustration is given by the following two-parameter family of *gamma deterrence functions*

$$F(c_{ij}) = (c_{ij})^{\sigma} \exp[-\theta c_{ij}], \qquad (2.125)$$

which is seen to differ from (2.122) only with respect to the *sign* of the exponent σ, and hence is again seen to be a two-dimensional instance of (2.121). The importance of this particular case is that (2.125) yields an explicit example of a *non-monotone* deterrence function. The first derivation of this type of deterrence function was given by Schneider (1959) for a continuous version of the simple search model developed in Example 4 of Section 2.4.2 below. In this case, the relevant non-monotonicity arises simply from the nature of two-dimensional space, and can in fact be eliminated by redefining the relevant notion of separation in terms of *intervening opportunities* (as discussed in Example 4 of Section 2.4.2 together with its extensions in Section 2.5.1 below). However, such non-monotone deterrence functions can arise directly from *behavioral* considerations as well. For example, with respect to travel by specific modes (say by car), one is generally unlikely to see very short trips (say one or two city blocks). Moreover, even in cases where individuals exhibit monotone deterrence behavior, population *heterogeneities* can lead to significant non-monotonicities in aggregate travel behavior. For example, it has been shown by Harris (1964) and Choukroun (1975) that mixtures of subpopulations characterized by exponential deterrence functions with different sensitivity parameters can result in aggregate interaction behavior characterized by gamma-type deterrence functions. Hence, the general multivariate exponential specification in (2.121) is seen to provide a flexible unifying framework within which a variety of different specifications can be treated. Of course, (2.125) can also be shifted by the addition of a known shift-quantity ϵ to c_{ij}. Section 5.3.2 presents an example where the use of such a *shifted gamma deterrence function* considerably improves the fit of a gravity model over the use of a simple gamma deterrence function.

Multivariate Exponential Deterrence Functions

Clearly, a key advantage of the multivariate exponential deterrence function (2.121) is that it includes as special cases multivariate power functions, power-exponential functions and gamma functions and their shifted versions if the shift quantity is preassigned. The exponential deterrence function (2.121) can be further extended by allowing transformations. In particular it follows by definition that any function of a separation measure, ${c_{ij}}^{(k)}$, is also a separation measure. While data transformations are common in applied statistical work, such methods do not appear to have been used in the context of gravity models. In our own work, for example, we have found that the square root of travel time performs better in exponential gravity models than does travel time itself. [A behavioral rationale for this transformation as well as some numerical examples are given in Chapter 5; see especially the material near the end of Section 5.8.1].

An even greater advantage of this exponential form stems from its ability to approximate a wide variety of deterrence functions. In particular notice that (2.1) may be written as

$$E(N_{ij}) = A(i)B(j)\ \exp\{\log[F(c_{ij})]\}.$$

Hence, if $\log[F(c_{ij})]$ is any continuous function of c_{ij}, then [by the Stone-Weirstrass Theorem (Dieudonné, 1960, p. 131)] it can be uniformly approximated on finite intervals by a polynomial function [and a number of other families of functions which are also linear in their parameters]. Such a polynomial would be of the form $\tilde{\theta}^t\tilde{c}_{ij}$ where $\tilde{c}_{ij}$ is a vector containing powers of $c_{ij}^{(k)}$ for $c_{ij} = (c_{ij}^{(1)}, \ldots, c_{ij}^{(K)})^t$. For example, suppose

$$\log[F({c_{ij}}^{(1)}, {c_{ij}}^{(2)})] \approx -{c_{ij}}^{(1)} - {c_{ij}}^{(2)} + 3{c_{ij}}^{(1)}{c_{ij}}^{(2)} - 2({c_{ij}}^{(2)})^4.$$

Then

$$\tilde{c}_{ij} = ({c_{ij}}^{(1)}, {c_{ij}}^{(2)}, {c_{ij}}^{(1)}{c_{ij}}^{(2)}, ({c_{ij}}^{(2)})^4)^t$$

and

$$\tilde{\theta}^t = (\tilde{\theta}_1, \ldots, \tilde{\theta}_4) = (-1, -1, 3, -2).$$

Alternatively, $[F(c_{ij})]$ may be approximated by a step function, which also may be shown to be of the form $\tilde{\theta}^t \tilde{c}_{ij}$, where the components of $\tilde{c}_{ij}$ are suitably defined Kronecker δ-functions (see Section 5.3.2 below; see also the beginning of the proof of Theorem 4.1). Therefore, if we set

$$F(c_{ij}) = \exp[\theta^t c_{ij}] \tag{2.126}$$

and define the vectors θ and c_{ij} appropriately, a wide class of $F(c_{ij})$'s will at least be approximated. In fact when step functions are considered, separation measures do not need to be costs.

In the context of the above-mentioned step function and polynomial approximations, we need to draw the readers' attention to the fact that in making predictions using the gravity model, we frequently require implicitly that the function $F(c_{ij})$ be configuration-free. It is reasonable to expect that if it is indeed so, an approximation to it also should be approximately configuration-free. However, the method of parameterization clearly affects the property of parameters being invariant with respect to configurations. The issue frequently reduces to that of the range of configurations over which the approximation is valid, and needs to be addressed for every specific situation.

Before proceeding to the analysis of exponential gravity models, there is an additional important caveat which should be mentioned with respect to the types of composite cost functions illustrated above. In particular, recall from the general development of separation profiles in Chapter 1 that the individual components of such profiles were assumed to be *structurally independent*, so that the set of possible profiles defines a full product set [as in expression (1.6)]. However, if two or more component measures are functionally related, as in (2.123) above, then this structural independence assumption is clearly violated. Hence, with respect to the behavioral characterizations of exponential gravity models in Section 2.4.3 below, it should be borne in mind that while the behavioral conditions developed are *always necessary* for the existence of exponential gravity models, the proofs of sufficiency (in Theorems 4.15 through 4.23 of Chapter 4 below) generally require the assumption of structural independence.

2.4 Exponential Gravity Models

Given the general class of gravity models above, we now focus on those models with exponential deterrence functions, designated as *exponential gravity models*. As in the general discussion of deterrence functions above, we shall assume throughout this section that all spatial separation is representable by *cost configurations*, $c \in C$. Within this measurement context, our development of exponential gravity models will closely parallel that of Section 2.2 above. First we begin with a specification of exponential gravity models in Section 2.4.1, and then illustrate these models with two examples in Section 2.4.2. Finally these models are characterized in Section 2.4.3 in terms of aggregate behavioral axioms paralleling those for general gravity models above.

2.4.1 Model Specifications

For convenience in the following development, we shall often employ vector notation. In particular, if for any two vectors, $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ we denote the *inner product* of x and y by $x^t y$ $(= \sum_{i=1}^n x_i y_i)$, then we may now specialize Definition 2.1 above to the case of exponential deterrence functions as follows:

Definition 2.14 (Model E1) A spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ is said to be mean-representable by an *exponential gravity model* if and only if for each cost configuration, $c \in C$, there exists a positive origin function, A_c, positive destination function, B_c, and cost sensitivity vector, $\theta_c \in R^K$, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)\exp[-\theta_c^t c_{ij}]. \tag{2.127}$$

We shall employ the symbol 'E' for exponential gravity models, and hence shall refer to the model form in (2.127) as Model E1. [Notice also that the sign of θ_c in (2.127) is unspecified, and hence that the inclusion of a minus sign in (2.127) is only for convenience in defining the class of monotone exponential gravity models below. In Chapters 5 and 6, where we consider parameter estimation, it will be convenient to drop minus sign] As in the case of general gravity models, when $\boldsymbol{P}$ is an independent spatial interaction process, then it is appropriate to say that $\boldsymbol{P}$ is *representable* by an exponential gravity model whenever (2.127) holds for some A_c, B_c, and θ_c. Moreover, as a parallel to expression (2.3) above, it now follows that each exponential gravity model can be expressed equivalently as a *log-linear model* of the form:

$$\begin{aligned}\log \mathrm{E}_c(N_{ij}) &= a_c(i) + b_c(j) - \theta_c^t c_{ij} \\ &= a_c(i) + b_c(j) - \sum_{k \in K} \theta_{ck} c_{ij}^k,\end{aligned} \tag{2.128}$$

where $a_c(i) = \log A_c(i)$ and $b_c(j) = \log B_c(j)$. Here 'linearity' refers to the fact that (2.128) is a linear function of each cost component value, c_{ij}^k.

As in the case of general gravity models, we shall be primarily interested in those models for which larger values of each cost component, c_{ij}^k, diminish the likelihood of (ij)-interactions (all else being equal). In particular, observe that for exponential gravity models this is equivalent to the requirement that each component of the cost sensitivity vector, θ_c, in (2.128) be *nonnegative*. Hence as a parallel to Definition 2.2, we now have:

Definition 2.15 (Model E1*) A spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ is said to be mean-representable by a *monotone exponential gravity model* if and only if for each $c \in C$, $\boldsymbol{P}$ satisfies (2.128) for some nonnegative cost sensitivity vector, $\theta_c \in R_+^K$.

The following types of invariance conditions on exponential gravity models are seen to be exact parallels to Definitions 2.3 through 2.7 above, and hence require no further explanation.

Definition 2.16
(i) (Model E2) A spatial interaction process, $\boldsymbol{P}$ is said to be mean-representable by a *deterrence-invariant exponential gravity model* if and only if there exists a cost sensitivity vector, $\theta \in R^K$, together with positive origin functions, A_c, and positive destination functions, B_c, for each cost configuration, $c \in C$, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)\exp[-\theta^t c_{ij}]. \tag{2.129}$$

(ii) (Model E2*) If in addition the cost sensitivity vector in (2.129) is nonnegative, i.e., if $\theta \in R_+^K$, then $\boldsymbol{P}$ is said to be representable by a *monotone deterrence-invariant exponential gravity model.*

Definition 2.17
(i) (Model E3) A spatial interaction process, $\boldsymbol{P}$ is said to be mean-representable by a *destination-deterrence-invariant exponential gravity model* if and only if there exists a cost sensitivity vector, $\theta \in R^K$, and a positive destination function, B, together with positive origin functions, A_c, for each cost configuration, $c \in C$, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A_c(i)B(j)\exp[-\theta^t c_{ij}]. \tag{2.130}$$

(ii) (Model E3*) If in addition the cost sensitivity vector in (2.130) is nonnegative, i.e., if $\theta \in R_+^K$, then $\boldsymbol{P}$ is said to be representable by a *monotone destination-deterrence-invariant exponential gravity model.*

Definition 2.18
(i) (Model E4) A spatial interaction process, $\boldsymbol{P}$ is said to be mean-representable by an *origin-deterrence-invariant exponential gravity model* if and only if there exists a cost sensitivity vector, $\theta \in R^K$, and a positive origin function, A, together with positive destination functions, B_c, for each cost configuration, $c \in C$, such that for all origin-destination pairs, $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A(i)B_c(j)\exp[-\theta^t c_{ij}]. \tag{2.131}$$

(ii) (Model E4*) If in addition the cost sensitivity vector in (2.131) is nonnegative, i.e., if $\theta \in R_+^K$, then $\boldsymbol{P}$ is said to be representable by a *monotone origin-deterrence-invariant exponential gravity model.*

Definition 2.19
(i) (Model E5) A spatial interaction process, $\boldsymbol{P}$ is said to be mean-representable by a *relatively invariant exponential gravity model* if and only if there exists a cost sensitivity vector, $\theta \in R^K$, together with a positive origin function, A, a destination function, B, and a configuration function, λ, such that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = \lambda(c)A(i)B(j)\exp[-\theta^t c_{ij}]. \tag{2.132}$$

(ii) (Model E5*) If in addition the cost sensitivity vector in (2.132) is nonnegative, i.e., if $\theta \in R_+^K$, then $\boldsymbol{P}$ is said to be representable by a *monotone relatively invariant exponential gravity model.*

Definition 2.20
(i) (Model E6) A spatial interaction process, $\boldsymbol{P}$ is said to be mean-representable by an *invariant exponential gravity model* if and only if there exists a cost sensitivity vector, $\theta \in R^K$, together with a positive origin function, A, and a destination function, B, such that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}_c(N_{ij}) = A(i)B(j)\exp[-\theta^t c_{ij}]. \tag{2.133}$$

(ii) (Model E6*) If in addition the cost sensitivity vector in (2.133) is nonnegative, i.e., if $\theta \in R_+^K$, then $\boldsymbol{P}$ is said to be representable by a *monotone invariant exponential gravity model.*

2.4.2 ILLUSTRATIVE EXAMPLES

Given these exponential model forms, we next consider two specific examples of spatial interaction processes which give rise to such models. The first example specializes the class of Carroll-Bevis processes developed in Example 1 above. In particular, a class of *Kullback-Leibler processes* is here developed in terms of a probabilistic measure of deviations between actual and hypothesized frequencies which is based on the notion of 'minimum discrimination information' first introduced by Kullback and Leibler (1951). The second example develops a class of processes which involve an explicit type of spatial search behavior first proposed by Schneider (1959). These *simple search processes* yield exponential gravity models in which the underlying measure of separation is given by the notion of 'intervening-opportunity distance' first proposed by Stouffer (1940).

(A) EXAMPLE 3: KULLBACK-LEIBLER PROCESSES

Recall from the development of Carroll-Bevis processes in Example 1 above that while deviations of actual interaction behavior from hypothesized spatially insensitive behavior were postulated to depend only on spatial separation, the exact form of this dependency was left unspecified. With this in mind, the central purpose of the present example is to develop a more explicit model of probabilistic deviations which yields an exponential form for such dependencies. To motivate the main idea, we again start with the same null hypothesis about spatially insensitive behavior given in Example 1, namely, that such behavior is representable [as in (2.23)] by a *location-insensitive process*, $\boldsymbol{P}^o = \{P_c^o : c \in C\}$, satisfying the condition that

$$\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N_{ih})} = \frac{\mathrm{E}_c^o(N_{gj})}{\mathrm{E}_c^o(N_{gh})}, \tag{2.134}$$

for all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c \in C$. In the present case, however, we focus exclusively on independent spatial interaction processes, $\boldsymbol{P} = \{P_c : c \in C\}$ in which conditional interaction frequencies for each given population size, n, are multinomially distributed with respect to the *interaction probabilities,*

$$p_c(ij) = \mathrm{E}_c(N_{ij})/\mathrm{E}_c(N), \tag{2.135}$$

given in the Multinomial Sampling Theorem of Chapter 1 above. For such processes it can be shown [see Proposition 4.4 of Chapter 4 below] that (2.134) is equivalent to the condition that these interaction probabilities be equal to the product of their marginal probabilities, i.e., that the probable origins and destinations of interactions are independent. Hence, if we now designate an independent spatial interaction process, $\boldsymbol{P}^o = \{P_c^o : c \in C\}$, as an *origin-destination independent (OD-independent) process* iff the associated interaction probabilities, $p_c^o(ij)$, in (2.135) satisfy

$$p_c^o(ij) = p_c^o(i\,\cdot)p_c^o(\cdot\, j) \tag{2.136}$$

with respect to their associated marginal probabilities

$$p_c^o(i\,\cdot) = \sum_h p_c^o(ih) = \sum_h [\mathrm{E}_c^o(N_{ih})/\mathrm{E}_c^o(N)] = \mathrm{E}_c^o(N_{i\,\cdot})/\mathrm{E}_c^o(N), \tag{2.137}$$

$$p_c^o(\cdot\, j) = \sum_g p_c^o(gj) = \sum_g [\mathrm{E}_c^o(N_{gj})/\mathrm{E}_c^o(N)] = \mathrm{E}_c^o(N_{\cdot j})/\mathrm{E}_c^o(N), \tag{2.138}$$

for all $ij \in I \times J$ and $c \in C$. This class of OD-independent processes is then taken to be the appropriate model of locationally insensitive interaction behavior.

Given this notion of locationally insensitive behavior, we next develop a probabilistic measure of 'deviations' from such behavior. To begin with, observe that those deviations from spatially insensitive behavior which are of most interest for our purposes relate to the *distribution* of interaction activity among origin-destination pairs, and not to the overall levels of such activity. Hence we now focus on comparisons between interaction processes which exhibit the same overall levels of interaction activity. More formally, if for any two interaction processes, $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$, we let $[\mathrm{E}_c(N_i), \mathrm{E}_c(N_j)]$ and $[\tilde{\mathrm{E}}_c(N_i), \tilde{\mathrm{E}}_c(N_j)]$ denote the corresponding *mean interaction activity levels* at each origin, $i \in I$, and destination, $j \in J$, under configuration, $c \in C$, then we now focus only on comparisons between processes, $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$, which are *mean-activity-equivalent* in the sense that

$$[\mathrm{E}_c(N_i), \mathrm{E}_c(N_j)] = [\tilde{\mathrm{E}}_c(N_i), \tilde{\mathrm{E}}_c(N_j)], \tag{2.139}$$

for all $ij \in I \times J$ and $c \in C$. With this convention, it can be shown [Proposition 4.5 in Chapter 4] that for any given independent spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, there is exactly one OD-independent process, $\boldsymbol{P}^o$, which is mean-activity-equivalent to $\boldsymbol{P}$ in the sense of (2.139). Hence, if actual interaction behavior is postulated to be representable by a given independent spatial interaction process, $\boldsymbol{P}$, then the unique *OD-independent version* of $\boldsymbol{P}$ is now taken to constitute the appropriate model of locationally insensitive behavior in this context. To compare behavior under $\boldsymbol{P}$ and $\boldsymbol{P}^o$, we next recall from the Multinomial Sampling Theorem that for any given separation configuration, $c \in C$, and realized frequency profile, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, of interactions with $\sum_{ij} n_{ij} = n$, the relative likelihood of $\boldsymbol{n}$ under $\boldsymbol{P}$ versus $\boldsymbol{P}^o$ is given by

$$\frac{P_c(\boldsymbol{n} \mid N = n)}{P_c^o(\boldsymbol{n} \mid N = n)} = \frac{\prod_{ij} p_c(ij)^{n_{ij}}}{\prod_{ij} p_c^o(ij)^{n_{ij}}}. \tag{2.140}$$

With respect to these relative likelihoods, it can be shown [as in expression (4.62) of Chapter 4] that for large sample sizes, n, the overwhelmingly most probable frequency profiles under $\boldsymbol{P}$ are well approximated by the conditional mean frequencies of the multinomial distribution in (1.29), i.e., by the quantities,

$$\mathrm{E}_c(N_{ij} \mid N = n) = p_c(ij)n, \tag{2.141}$$

for each $c \in C$. Hence, if we now designate the corresponding profile of values,

$$\boldsymbol{m}_n^c = (m_{nij}^c : ij \in I \times J) = [p_c(ij)n : ij \in I \times J], \tag{2.142}$$

as the *conditional $\boldsymbol{P}$-modal profile* for n and c, then we take this unique profile, $\boldsymbol{m}_n^c$, to represent the overwhelmingly most probable frequency behavior under $\boldsymbol{P}$ for each cost configuration, c, and large population size, n. With these conventions, it then follows from (2.140) and (2.142) that the relative likelihood of this modal frequency behavior under $\boldsymbol{P}$ versus $\boldsymbol{P}^o$ is well approximated by the ratio

$$\frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^o(\boldsymbol{m}_n^c \mid n)} = \frac{\prod_{ij} p_c(ij)^{p_c(ij)n}}{\prod_{ij} p_c^o(ij)^{p_c(ij)n}}. \tag{2.143}$$

Since $\boldsymbol{P}$-modal behavior is by definition most likely under $\boldsymbol{P}$ itself, it should be clear that these relative likelihoods are always greater than one, unless $\boldsymbol{P}$

is itself OD-independent (as demonstrated in Proposition 4.6 of Chapter 4). More generally, the smaller these relative likelihoods, the closer $\boldsymbol{P}$ is to OD-independence. With these observations, we now adopt (2.143) as the desired measure of deviation from OD-independent behavior.

Given this measure of deviation, we again hypothesize (as in Example 1) that deviations from OD-independent behavior are entirely attributable to spatial separation factors. More specifically, we assume in the present case that the relevant separation factors are representable by a set of *interaction costs* (with well defined units of measure), and that the influence of these separation factors on actual behavior is reflected by the *mean cost levels* associated with interaction activity. In this context, we hypothesize that among all behavior consistent with any given set of mean cost levels, actual behavior deviates 'as little as possible' from OD-independence. To formalize this hypothesis, let $\{c^k : k \in K\}$ denote the relevant set of interaction costs in the given interaction context [where, for example, c^1 = 'travel costs' and c^2 = 'time costs']. Then, for each cost configuration, $c \in C$, and number of realized interactions, n, we may define the associated *conditional mean-cost profile*, $\mathrm{E}_c(c \mid n) = [\mathrm{E}_c(c^k \mid n) : k \in K]$, for each $k \in K$ under process $\boldsymbol{P}$ by

$$\mathrm{E}_c(c^k \mid n) = \mathrm{E}_c[\sum_{ij} c_{ij}^k N_{ij} \mid N = n] = \sum_{ij} c_{ij}^k \mathrm{E}_c(N_{ij} \mid N = n) \tag{2.144}$$

[so that, for example, $\mathrm{E}_c(c^1 \mid n)$ = 'mean travel cost' and $\mathrm{E}_c(c^2 \mid n)$ = 'mean travel time' for interaction patterns of size n under configuration c]. Note from (2.144) together with (2.141) and (2.142) that

$$\mathrm{E}_c(c \mid n) = n \sum_{ij} c_{ij} p_c(ij) = \sum_{ij} c_{ij} m_{nij}^c, \tag{2.145}$$

and hence that the most probable cost profiles associated with behavior under $\boldsymbol{P}$ are well approximated by these mean cost profiles (for large n). With this notation, we may now formalize the above hypothesis in terms of the following class of *Kullback-Leibler processes.* In particular, if for any other independent interaction process, $\tilde{\boldsymbol{P}} = (\tilde{P}_c : c \in C)$, we now let $\tilde{\boldsymbol{m}}_n^c$, $\tilde{\mathrm{E}}_c(c \mid n)$, and $\tilde{\boldsymbol{P}}^o$ denote the corresponding conditional $\tilde{\boldsymbol{P}}$ -modal profiles, conditional mean-cost profiles, and OD-independent version of $\tilde{\boldsymbol{P}}$ under c, respectively, then we now say that:

Definition 2.21 An independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ is designated as a *Kullback-Leibler process (KL-process)* if and only if for each cost configuration, $c \in C$, interaction pattern size, $n > 0$, and mean-activity-equivalent independent interaction process, $\tilde{\boldsymbol{P}} = (\tilde{P}_c : c \in C)$,

$$\tilde{\mathrm{E}}_c(c \mid n) = \mathrm{E}_c(c \mid n) \Rightarrow \frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^o(\boldsymbol{m}_n^c \mid n)} \leq \frac{\tilde{P}_c(\tilde{\boldsymbol{m}}_n^c \mid n)}{\tilde{P}_c^o(\tilde{\boldsymbol{m}}_n^c \mid n)}. \tag{2.146}$$

In other words, $\boldsymbol{P}$ is a KL-process if and only if from among all mean-activity-equivalent processes with the same conditional mean-cost levels, $\boldsymbol{P}$ exhibits minimal deviations from OD-independence. Equivalently, if one rewrites (2.146) as

$$\tilde{\mathrm{E}}_c(c \mid n) = \mathrm{E}_c(c \mid n) \Rightarrow \frac{P_c^o(\boldsymbol{m}_n^c|n)}{P_c(\boldsymbol{m}_n^c \mid n)} \geq \frac{\tilde{P}_c^o(\tilde{\boldsymbol{m}}_n^c|n)}{\tilde{P}_c(\tilde{\boldsymbol{m}}_n^c|n)}, \tag{2.147}$$

then it follows from (2.143) that for any process, $\tilde{\boldsymbol{P}}$, with the same mean activity levels and same mean cost levels as $\boldsymbol{P}$, the relative likelihood of $\boldsymbol{P}$-modal frequencies under OD-independence is always at least as great as that for $\tilde{\boldsymbol{P}}_c$-modal frequencies. Hence in this sense, KL-processes may be said to exhibit interaction behavior which is *maximally OD-independent*, given the mean activity and cost levels associated with those processes.

Alternative Interpretations of Kullback-Leibler Processes. Before proceeding to the representation properties of KL-processes, it is important to point out several alternative interpretations of such processes. First and foremost is of course the original information-theoretic interpretation given by Kullback and Leibler (1951). In particular, the measure of deviation from OD-independence in (2.143) above is seen to be a special case of the more general information-theoretic measure of 'discrimination information' first proposed by Kullback and Leibler (1951). To motivate their information-theoretic approach in the present context, suppose that one considers two possible probability hypotheses, p and q, about the likelihood of various interactions in $I \times J$. If a given outcome, ij, is more likely to occur under p than q [i.e., if $p(ij) > q(ij)$] then the occurrence of ij can be said to provide evidence in favor of hypothesis p over q. This observation suggests that the quantity,

$$I_{p|q}(ij) = \log[p(ij)/q(ij)], \tag{2.148}$$

might be interpreted as a measure of the *information in outcome ij for discrimination in favor of p against q* [or equivalently, the *weight of evidence* in favor of p against q (Good, 1950)]. In particular, $I_{p|q}(ij)$ is positive if and only if $p(ij) > q(ij)$. More generally, if one extends (2.148) to sequences of independent outcomes $(i_r j_r : r = 1, \ldots, n)$ by setting

$$I_{p|q}(i_r j_r : r = 1, \ldots, n) = \log\left[\prod_{r=1}^{n} p(i_r j_r) / \prod_{r=1}^{n} q(i_r j_r)\right], \tag{2.149}$$

then this measure is also seen to exhibit the reasonable (and indeed characteristic) property that information is *additive* over independent events, i.e., that

$$I_{p|q}(i_r j_r ; r = 1, \ldots, n) = \sum_{r=1}^{n} I_{p|q}(i_r j_r). \tag{2.150}$$

Next observe from the symmetry of (2.150) that the information in a given sequence of independent outcomes depends only on the associated outcome frequency profile, $\boldsymbol{n} = (n_{ij} : ij \in I \times j)$, so that (2.150) may be equivalently written in terms of frequency profiles as

$$I_{p|q}(\boldsymbol{n}) = \sum_{ij} n_{ij} I_{p\,|\,q}(ij). \tag{2.151}$$

Finally, observe that if p is in fact the true distribution, then the argument preceding (2.142) shows that for large sample sizes, n, the overwhelmingly most probable information value for p against q is again given by the corresponding modal frequency profile, $\boldsymbol{m}_n^p = (m_{nij}^p : ij \in I \times J)$, as

$$\begin{aligned} I_{p|q}(\boldsymbol{m}_n^p) &= \sum_{ij} m_{nij}^p I_{p\,|\,q}(ij) = \sum_{ij} [p(ij)n] I_{p\,|\,q}(ij) \\ &= n \sum_{ij} p(ij) \log[p(ij)/q(ij)]. \end{aligned} \tag{2.152}$$

Hence, these observations suggest that the associated mean information value,

$$I_{p|q} = \sum_{ij} p(ij) \log[p(ij)/q(ij)] \approx (1/n) I_{p\,|\,q}(\boldsymbol{m}_n^p), \tag{2.153}$$

under distribution hypothesis, p, provides a reasonable summary measure of the *discrimination information* in favor of p against q, as proposed by Kullback and Leibler (1951).

To relate this discrimination-information measure to the deviation measure proposed in (2.143), simply observe that by setting $p = p_c$, $q = p_c^o$, and $m_n^p = m_n^c$ in (2.153) we obtain

$$\begin{aligned} I_{p_c|p_c^o} &= \log\Big[\prod_{ij} [p_c(ij)/p_c^o(ij)]^{p_c(ij)}\Big] \\ &= \frac{1}{n} \log\left[\frac{\prod_{ij} p_c(ij)^{p_c(ij)n}}{\prod_{ij} p_c^o(ij)^{p_c(ij)n}}\right] = \frac{1}{n} \log\left[\frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^o(\boldsymbol{m}_n^c \mid n)}\right]. \end{aligned} \tag{2.154}$$

Hence we may conclude [from the positive monotonicity of the function $(1/n)\log(\cdot)$] that Kullback-Leibler processes can equivalently be interpreted from an information-theoretic viewpoint as those independent interaction processes which yield *minimum discrimination information* against the hypothesis of OD-independence, in the sense of (2.146) above. This minimum-discrimination-information (MDI) approach to deriving the 'least informative' distributions consistent with various types of mean cost constraints is often referred to in the literature as *MDI-analysis* [or more specifically, as *MDI-analysis with external constraints* (Gokhale and Kullback, 1978)].

In the context of spatial interaction modeling, MDI-analysis has been employed by a host of researchers, including Charnes, Raike, and Bettinger (1972), March and Batty (1975), and Snickars and Weibull (1977), among many others. Excellent overviews of this work, together with further discussions of the philosophy of the approach, can be found in Webber (1979) and Haynes and Phillips (1982).

A closely related interpretation of Kullback-Leibler processes can be given directly in terms of hypothesis testing. In particular, if for any given cost configuration, $c \in C$, and sample size, n, one employs the standard *likelihood-ratio test* for the hypothesis, p_c, against the (OD-independence) hypothesis, p_c^o, and observes that the maximum value of the corresponding *likelihood-ratio statistic* is well approximated by the quantity

$$\frac{L_n(m_n^c \mid p_c)}{L_n(m_n^c \mid p_c^o)} = \frac{\prod_{ij} p_c(ij)^{m_{nij}^c}}{\prod_{ij} p_c^o(ij)^{m_{nij}^c}} = \frac{P_c(m_n^c \mid n)}{P_c^o(m_n^c \mid n)}, \tag{2.155}$$

then it follows at once from (2.143) that from among all processes with the same mean activity and cost levels, KL-processes must (with overwhelming probability) yield the lowest values of these test statistics for large samples sizes [as observed by Kullback (1959, Section 5.4) and many others]. Hence from an hypothesis-testing viewpoint, KL-processes can also be interpreted as yielding the most conservative possible tests against the null hypothesis of OD-independence.

As an alternative hypothesis-testing interpretation, it has been argued by Good (1963) that such processes should themselves be regarded as null hypotheses, namely, as posterior forms of prior hypotheses based on new information (in our case, new cost information about prior OD-independence hypotheses). A more powerful formulation of this approach is given by Snickars and Weibull (1977), who show that (in our case) the modal frequencies of KL-processes are in fact the *conditional modal frequencies* arising from a prior OD-independence hypothesis constrained by new cost information. In other words, if the hypothesis of OD-independence were true, then the most probable frequency profiles consistent with this new information would be precisely the modal frequencies of the corresponding KL-process. Hence in this sense, the modal frequencies of KL-processes are indeed the most natural test statistics for the hypothesis of OD-independence. A fuller discussion of this 'most probable state' interpretation is given in Smith (1989).

In contrast to these information-theoretic and testing interpretations, our present approach is designed to give a more *behavioral* interpretation of KL-processes. In particular, as in (2.147) above, KL-processes are here interpreted as models of 'maximally independent' spatial interaction behavior consistent with given mean activity and cost constraints.

Representations of Mean Interaction Frequencies. In view of the equivalence in (2.154), the representational properties of KL-processes are seen to be directly obtainable from standard results. In particular, since all minimum-information distributions subject to mean-value constraints are well known to be members of the *exponential family* of distributions (Kullback, 1959, Theorem 3.2.1), it follows in particular that the interaction probabilities, $p_c(ij)$, in (2.154) must be exponential functions of the associated cost profiles, c_{ij} (as shown in Proposition 4.7 of Section 4.3.3 in Chapter 4). Hence, in view of expression (1.30) in the Multinomial Sampling Theorem for independent spatial interaction processes, this property is necessarily inherited by the corresponding mean interaction frequencies, $\mathrm{E}_c(N_{ij})$, and we obtain the following instance of Model E1 for KL-processes:

E1-Representability. *Every KL-process, $\boldsymbol{P} = \{P_c : c \in C\}$, is representable by an exponential gravity model.*

Given this exponential property of deterrence functions for KL-processes, we next show that such processes are indeed an instance of the more general class of CB-processes in Example 1 above. To see this, recall from Definition 2.8 that to establish that any interaction process, $\boldsymbol{P}$, is a CB-process, it suffices to find some location-insensitive process, $\boldsymbol{P}^o$, satisfying (2.24) with respect to $\boldsymbol{P}$. Hence, suppose that $\boldsymbol{P}$ is a KL-process, with exponential gravity representation given for all $c \in C$ and $ij \in I \times J$ by

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)\exp[-\theta_c c_{ij}]. \tag{2.156}$$

Then, recalling from the Poisson Characterization Theorem in Chapter 1 that each independent interaction process is completely defined by its associated mean interaction frequencies, it follows that there exists a unique independent interaction process, $\boldsymbol{P}^o = \{P_c : c \in C\}$, with mean interaction frequencies given in terms of (2.156) for all $c \in C$ and $ij \in I \times J$ by

$$\mathrm{E}_c^o(N_{ij}) = A_c(i)B_c(j). \tag{2.157}$$

Moreover, since (2.157) implies at once that

$$\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N_{ih})} = \frac{B_c(j)}{B_c(h)} = \frac{\mathrm{E}_c^o(N_{gj})}{\mathrm{E}_c^o(N_{gh})}, \tag{2.158}$$

for all $c \in C$, $i, g \in I$, and $j, h \in J$, it follows from (2.134) [or equivalently (2.23)] that $\boldsymbol{P}^o$ is a location-insensitive process. But for this choice of $\boldsymbol{P}_c$ we then see from (2.156) and (2.157) that for all $c \in C$ and $ij, gh \in I \times J$,

$$\begin{aligned} c_{ij} = c_{gh} &\Rightarrow \exp[-\theta_c c_{ij}] = \exp[-\theta_c c_{gh}] \\ &\Rightarrow \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c^o(N_{ij})} = \frac{\mathrm{E}_c(N_{gh})}{\mathrm{E}_c^o(N_{gh})}. \end{aligned} \tag{2.159}$$

Hence $\boldsymbol{P}$ is seen to be a CB-process with respect to this choice of $\boldsymbol{P}^o$. However, it should be noted that while $\boldsymbol{P}^o$ is necessarily an OD-independent process [as in (2.136)], it is generally *not the OD-independent version* of $\boldsymbol{P}$. Indeed, the OD-independent version of $\boldsymbol{P}$ will usually fail to satisfy (2.27) with respect to $\boldsymbol{P}$. Hence, if the requirement of activity equivalence between $\boldsymbol{P}$ and $\boldsymbol{P}^o$ were to be added to the general definition of CB-processes, then the type of dependency postulated in (2.27) is seen to be a very stringent condition.

(B) Example 4: Simple Search Processes

As in the case of Carroll-Bevis processes, the above class of Kullback-Leibler processes treat actual spatial interaction behavior only indirectly in terms of deviations from null hypotheses about spatially insensitive behavior. Hence it is of interest to consider more explicit behavioral processes in which mean interaction frequencies are representable by exponential gravity models. The present class of processes is inspired by the behavioral distance concept of 'intervening opportunities' between locations, first proposed by Stouffer (1940). While intervening-opportunity distances have been applied in a variety of spatial interaction contexts [as for example in Baxter (1984) and Wills (1986)], this behavioral concept arises most naturally in the context of *spatial search processes.* Such processes typically involve spatial actors who search for opportunities which satisfy their current needs [such as a search for stores satisfying certain shopping needs]. If the first satisfactory opportunity found is taken to define the relevant *spatial interaction* for the actor, then the number of intervening opportunities searched may well constitute the most appropriate notion of distance involved in the interaction. The first explicit model of search behavior incorporating these concepts was developed in a continuous spatial setting by Schneider (1959) [and later extended by Harris (1964)]. A discrete version of this model was also developed by Schneider, for use in the Chicago Area Transportation Study (1960). Subsequent extensions and clarifications of this discrete model have been made by Ruiter (1967), Okabe (1976), Weibull (1978), and others [as discussed further in Section 2.5.1 below].

To develop this model within our present framework, we begin (as in Example 2 above) with a set of *actors*, $\alpha \in A$, distributed over a set of *origin locations*, $i \in I$, and a set of *opportunities*, $\beta \in B$, distributed over a set of *destination locations*, $j \in J$, where A_i and B_j again denote the sets of actors at each origin i and opportunities at destination j, respectively. Within this setting, the relevant individual interaction space, Ω_1, is here interpreted as the set of *individual search events*, $\omega \in \Omega_1$, each involving some actor, $\alpha \in A$, who searches among opportunities, $\beta \in B$, until he either finds one which satisfies his current needs or exhausts all opportunities. For our purposes, the single most important property of each actor, α, is his *origin location*, i. In addition to location, the most important property of each opportunity, $\beta \in B$, for actor, α, is of course the ability of β to satisfy his current needs. This property can be represented by a zero-one *satisfaction variable*, $\delta_\beta \in \{0,1\}$, where $\delta_\beta = 1$ if β satisfies the current needs of the actor, and $\delta_\beta = 0$ otherwise. Hence if each vector of such attributes for all opportunities is denoted by $\delta = (\delta_\beta : \beta \in B) \in \{0,1\}^B$, then the relevant class of individual search events can be represented by $\Omega_1 = I \times \{0,1\}^B$, with individual events, $\omega = (i, \delta)$, involving an actor at location, $i \in I$, whose current needs are satisfiable precisely by those opportunities, $\beta \in B$, with $\delta_\beta = 1$ in δ. Given these individual search events, we designate each joint occurrence of n individual search events, $\omega = (\omega_r : r = 1, \ldots, n) \in \Omega_n = (\Omega_1)^n$, as a *population search event of size*, n, and let $\Omega = \cup_{n \geq 0} \Omega_n$ denote the class of all finite *population search events* [where the *null search event* is again represented by $\Omega_0 = \{o\}$].

The relevant notion of spatial separation in this search context is taken to be implicitly generated by the underlying *search scheme*, which defines the order in which spatial opportunities are searched by actors from each origin location. Typically the order in which opportunities are searched by a given actor will be based both on efficiency considerations (such as 'shortest paths') and on the actor's limited knowledge of those opportunities. For our present purposes, it is enough to assume that some well-defined search sequence is given. [For more detailed analyses of search behavior, see for example Karni and Schwartz (1977), Weibull (1978), Rogerson (1982), and Phipps and Laverty (1983), Smith *et al.* (1992).] Moreover, it is assumed for simplicity that all actors at a common location employ the same search scheme. [Alternatively, one may treat each actor as a distinct location.] With these assumptions, each *search scheme* is formally representable by an integer-valued vector, $c_i = (c_i(\beta) : i\beta \in I \times B)$, in which the value of $c_i(\beta)$ denotes the order in which opportunity β is searched by actors at i. For example, if opportunity β is the third opportunity searched by actors at i, then $c_i(\beta) = 3$. Our only additional assumption about such search schemes is that all opportunities at a given destination are searched before a new destination is explored. More formally, for any destinations, $j, h \in J$, and opportunities $\beta_j, \beta'_j \in B_j$ and $\beta_h, \beta'_h \in B_h$, each search scheme, c, is assumed to satisfy the following condition for all origins, $i \in I$,

$$c_i(\beta_j) < c_i(\beta_h) \Leftrightarrow c_i(\beta'_j) < c_i(\beta'_h). \tag{2.160}$$

The class, C, of all search schemes c satisfying (2.160) plays the role of the relevant class of *separation configurations* for this case (as will be seen below). Hence, in a manner paralleling Definition 1.2 above, we now take each probabilistic model of search behavior to be representable by a family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω, where each measure P_c describes search behavior under search scheme, $c \in C$. Within this general class of *search processes* (as formalized in Section 4.3.4 below), we now develop a class of *simple search processes* (based on the discrete model of Schneider). First we require that $\boldsymbol{P}$ satisfy the following version of the regularity conditions, R1 and R2, in Section 1.3.2 above. [The continuity condition, R3, always holds in the present setting, as shown in Proposition 3.7 of Chapter 3 below]:

S1. (Event Positivity) *For each search scheme, $c \in C$, and population search event, $\omega \in \Omega$, the event probability, $P_c(\omega)$, is positive.*

S2. (Population Symmetry) *For each search scheme, c, and pair of population search events, $\omega, \omega' \in \Omega$, if ω and ω' differ only by the ordering of their individual search events, then $P_c(\omega) = P_c(\omega')$.*

[A slightly weaker version of S1 is stated in Definition 3.8 of Chapter 3 below.] Next we state a number of independence conditions, paralleling conditions A1 and A2 in Section 1.4.2. To do so, we employ the following additional notation. First, observe from the event positivity condition (S1) above that the probability, $P_c(\Omega_n)$, of population search events of size n is positive for all n. Hence, for each measurable event, $A \subseteq \Omega_n$, the conditional probabilities, $P_c^n(A) = P_c(A)/P_c(\Omega_n)$, are well defined. In particular, the marginal probabilities, $P_c^n(i_r)$ and $P_c^n(\delta_{r\beta})$, that the r-th individual search event will involve an actor at origin, i_r, and a satisfaction value of $\delta_{r\beta}$ for opportunity β, respectively, are well defined. Hence, as a parallel to the locational independence condition (A1) in Section 1.4.2 we now assume that $\boldsymbol{P}$ satisfies:

S3. (Location-Satisfaction Independence) *For all search schemes, $c \in C$, population sizes, $n > 0$, and population search events, $\omega = (\omega_r : r = 1, \ldots, n) = [(i_r, (\delta_{r\beta} : \beta \in B)) : r = 1, \ldots, n] \in \Omega_n \subseteq \Omega$,*

$$P_c^n(\omega) = \prod_{r=1}^{n} \left[P_c^1(i_r) \prod_{\beta \in B} P_c^1(\delta_{r\beta}) \right]. \tag{2.161}$$

In other words, it is assumed that there exist no dependencies among the locations of individual actors and the need-satisfaction properties of opportunities in population search events of any given size. In addition, the superscript '1' on the right hand side of (2.161) implies that the properties of individual search events are assumed to be independent of population size, and in particular, that population search events of size n can be treated as n independent random samples of population search events of size 1. [A slightly weaker version of this condition is given in Definition 3.9 of Chapter 3.] Next observe that [in a manner paralleling the definition of (ij)-interaction frequencies preceding expression (1.11) of Chapter 1 above] if for each population search event, $\omega \in \Omega$, we let $N_i(\omega)$ denote the total number of individual search events in ω involving actors from origin i, then these values define an integer random variable, N_i, representing the *origin frequency* at i (as in Example 1 above). Hence, if each specific realization, $\boldsymbol{n} = (n_i : i \in I)$, of such frequencies is designated as an *origin-activity profile*, and if for each search scheme, $c \in C$, the probability of profile $\boldsymbol{n}$

is denoted by $P_c(\boldsymbol{n})$, with corresponding marginal probabilities, $P_c(n_i)$, for each $i \in I$, then, as a parallel to the frequency independence condition (A2) in Section 1.4.2, we now assume that $\boldsymbol{P}$ satisfies:

S4. (Origin-Frequency Independence) *For all search schemes, $c \in C$, and origin frequency profiles, $\boldsymbol{n} = (n_i : i \in I)$,*

$$P_c(\boldsymbol{n}) = \prod_{i \in I} P_c(n_i). \tag{2.162}$$

Hence it is assumed that there exist no dependencies among the search activity levels at each origin. Finally, in order to focus exclusively on the spatial aspects of search behavior, we assume for simplicity that all opportunities are *homogeneous* in the sense that each has the same chance of satisfying any actor's needs, regardless of the search scheme employed. More formally, if we let the binary random variable, $\Delta_{r\beta}$, denote the *acceptability* of opportunity β in the r-th individual component of any population search event (i.e., $\Delta_{r\beta}(\omega) = 1$ if and only if β meets the needs of the r-th actor in search event ω), then we now assume that $\boldsymbol{P}$ satisfies:

S5. (Opportunity Homogeneity) *For all search schemes, $c, c' \in C$, and interaction opportunities, $\beta, \beta' \in B$,*

$$P_c^1(\Delta_{1\beta} = 1) = P_{c'}^1(\Delta_{1\beta'} = 1). \tag{2.163}$$

With these assumptions, we may now define the desired class of simple search processes as follows [a more general definition is given in Definition 4.7 of Chapter 4 below]:

Definition 2.22 A family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$ on Ω is designated as a *simple search process* if and only if $\boldsymbol{P}$ satisfies conditions S1, S2, S3, S4, and S5.

Realized-Interaction Frequencies. In a manner similar to the class of threshold interaction processes developed in Example 2 above, our primary interest in the present class of simple search processes focuses on the probable interactions between actors and opportunities. In the present case, an *interaction* is assumed to occur whenever an actor finds an opportunity satisfying his needs. More formally, for any given search scheme, $c \in C$, we now say that the actor, α, involved in any search event, $\omega = (i, (\delta_\beta : \beta \in B)) \in \Omega_1$, *interacts* with opportunity β in that event if and only if β is the first opportunity found by α which satisfies his current needs, i.e., if and only if $\delta_\beta = 1$, and in addition, $\delta_{\beta'} = 0$ for all opportunities, $\beta' \in B$, with $c_i(\beta') < c_i(\beta)$. In a manner similar to expression (2.41) above, these interaction properties can be represented by zero-one variables of the following form:

$$\delta_{i\beta}^{cr}(\omega) = \begin{cases} 1, & i_r = 1, \delta_{r\beta} = 1, \text{ and } c_i(\beta') < c_i(\beta) \Rightarrow \delta_{r\beta'} = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.164}$$

In terms of these zero-one variables, it follows that for each search scheme, $c \in C$, and origin-destination pair, $ij \in I \times J$, the corresponding number, $N_{ij}^c(\omega)$, of interactions between i and β in each population search event, $\omega \in \Omega$, is given by

$$N_{ij}^c(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^n \sum_{\beta \in B_j} \delta_{i\beta}^{cr}(\omega), & \omega \in \Omega_n, n > 0. \end{cases} \tag{2.165}$$

Hence, as in Example 2 above, we now designate the random variables defined by (2.165) as the *realized-interaction frequencies* in each population search event, $\omega \in \Omega$, and focus our attention on representation properties of the associated *mean realized-interaction frequencies*, $\mathrm{E}_c(N_{ij}^c)$, for simple search processes. To do so, we introduce the following notion of *spatial separation* induced by search schemes. If for each search scheme, $c \in C$, and origin-destination pair, $ij \in I \times J$, the set of destinations reached from i prior to destination j is denoted by

$$J_c[j \mid i] = \{h \in J : (\beta' \in B_h, \beta \in B_j) \Rightarrow c_i(\beta') < c_i(\beta)\}, \tag{2.166}$$

then the order of the last opportunity searched before reaching destination j is given by

$$c_{ij} = \max\{c_i(\beta) : \beta \in B_h, h \in J_c[j \mid i]\}. \tag{2.167}$$

Equivalently, if b_j denotes the cardinality of each opportunity set B_j, then c_{ij} is seen to be the total number of opportunities searched by actors at i before reaching j, as given by

$$c_{ij} = \sum_{h \in J_c[j \mid i]} b_h. \tag{2.168}$$

Hence we now designate c_{ij} as the *intervening-opportunity distance* from i to j under search scheme c. In terms of this notion of spatial separation, we have the following instance of Model E3* for simple search processes [established in Corollary 4.2 of Section 4.3.4 in Chapter 4 below]:

E3*-Representability. *For each simple search process, $\boldsymbol{P} = \{P_c : c \in C\}$, there exists a positive scalar, θ, together with a positive destination function, B, and family of positive origin functions, $\{A_c : c \in C\}$, such that for all $c \in C$, and $ij \in I \times J$,*

$$\mathrm{E}_c(N_{ij}^c) = A_c(i)B(j)\exp[-\theta c_{ij}]. \tag{2.169}$$

Thus, if a simple search process, $\boldsymbol{P}$, is now said to be *mean-representable* by a given gravity model if and only if all mean realized-interaction frequencies for $\boldsymbol{P}$ are representable by that gravity model, then it follows from (2.169) that each such process is mean-representable by a *monotone destination-deterrence-invariant exponential gravity model.* In the present case, these functions have the following behavioral interpretations. First, if in each population search event, $\omega \in \Omega$, we let $L_r(\omega)$ denote the number of opportunities searched in the r-th individual search event before a satisfactory opportunity is found, then these values define random variables, L_r, which can be viewed as the *search lengths* in each individual event. In terms of these random variables, the deterrence function in (2.169) has the following probabilistic interpretation [as shown in expression (4.97) of Chapter 4]:

$$\exp[-\theta c_{ij}] = P^1(L_1 > c_{ij} \mid i_1 = i). \tag{2.170}$$

Hence for each origin-destination pair, ij, the value of the deterrence function in (2.170) is precisely the probability that destination j will be reached by a searcher starting at origin i (in a population search event of size one). Next, to interpret the destination function, B, let the zero-one variable, $\Delta_{rj}(\omega)$, be defined to have value one if and only if some opportunity at destination j satisfies the needs of the actor in the r-th individual search event of each population search event, $\omega \in \Omega$ [i.e., $\Delta_{rj}(\omega) = \max\{\Delta_{r\beta}(\omega) : \beta \in B_j\}$]. Then the destination function, B, in (2.169) has the following probabilistic interpretation in terms of these *destination-satisfaction variables*, Δ_{rj} [as shown in expressions (4.91) and (4.93) of Chapter 4]:

$$B(j) = P^1(\Delta_{1j} = 1). \tag{2.171}$$

Hence $B(j)$ is simply the probability that in any individual search event, the actor's need will be satisfied by some opportunity at destination j. Finally, each origin function A_c in (2.169) can be expressed directly in terms of mean origin-activity frequencies [as shown in expression (4.98) of Chapter 4],

$$A_c(i) = \mathrm{E}_c(N_i), \tag{2.172}$$

and hence is interpretable as simply the mean level of search activity by actors at each origin, $i \in I$.

Search-Neutral Processes. Given the destination-invariant representation in (2.169), it should be clear that if search activity levels are not influenced by the nature of the underlying search scheme, c, then the resulting gravity model representations will exhibit stronger invariance properties with respect to c. With this in mind, we now consider the following parallel to Definition 2.12 for TI-processes,

Definition 2.23 A simple search process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be *conditionally search-neutral* if and only if for all search schemes, $c, c' \in C$, population sizes, $n > 0$, and measurable events, $A \subseteq \Omega_n$,

$$P_c^n(A) = P_{c'}^n(A). \tag{2.173}$$

For this class of search processes, we then have the following illustration of Model E5* [established in part (i) of Proposition 4.9 in Chapter 4]:

E5*-Representability. *If a simple search process, $\boldsymbol{P} = \{P_c : c \in C\}$, is conditionally search-neutral, then $\boldsymbol{P}$ is mean-representable by a monotone relatively invariant exponential gravity model.*

Similarly, if the search process involves major decisions (such as housing search and job search) they it may also be reasonable to assume that the overall level of search activity is independent of the underlying search scheme. For such cases, we now have the following parallel to Definition 2.13 for TI-processes,

Definition 2.24 A simple search process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be *search-neutral* if and only if for all search schemes, $c, c' \in C$, and measurable events, $A \subseteq \Omega$,

$$P_c(A) = P_{c'}(A). \tag{2.174}$$

In terms of this definition, we then obtain the following instance of Model E6* [established in part (ii) of Proposition 4.9 in Chapter 4]:

E6*-Representability. *If a simple search process, $\boldsymbol{P} = \{P_c : c \in C\}$, is search-neutral, then $\boldsymbol{P}$ is mean-representable by a monotone invariant exponential gravity model.*

2.4.3 Behavioral Characterizations

Given the examples above, we now consider the most general conditions under which independent spatial interaction behavior is representable by exponential gravity models. As in Section 2.2.3 above, we shall begin by deriving the desired conditions from the model representations themselves, and then proceed to interpret these conditions behaviorally.

(A) Cost Dependence Axioms

To derive the most general property implied by exponential gravity models, observe first the if population interaction behavior is representable by an independent spatial interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with mean interaction frequencies of the form (2.127), then for any activity-equivalent spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n)$, $s' = (i'_r j'_r : r = 1, \ldots, n) \in S$, it follows from (2.51) through (2.53) that for each $c \in C$,

$$\frac{P_c(s)}{P_c(s')} = \frac{\prod_{r=1}^n \exp[-\theta_c^t c_{i_r j_r}]}{\prod_{r=1}^n \exp[-\theta_c^t c_{i'_r j'_r}]} = \frac{\exp[-\theta_c^t \sum_{r=1}^n c_{i_r j_r}]}{\exp[-\theta_c^t \sum_{r=1}^n c_{i'_r j'_r}]}. \tag{2.175}$$

This general observation leads to the following 'cost dependence' property.

Cost Dependence. If for any cost configuration, $c \in C$, and spatial interaction pattern, $s = (i_r j_r : r = 1, \ldots, n)$, we denote the total cost generated by pattern, s, under configuration, c, by

$$c(s) = \sum_{r=1}^{n} c_{i_r j_r} = \sum_{ij \in I \times J} c_{ij} N_{ij}(s), \tag{2.176}$$

then it follows at once from (2.175) that for any activity-equivalent spatial interaction patterns, $s, s' \in S$, and cost configuration, $c \in C$,

$$\begin{aligned} c(s) = c(s') &\Rightarrow \exp[-\theta_c^t c(s)] = \exp[-\theta_c^t c(s')] \\ &\Rightarrow P_c(s) = P_c(s'). \end{aligned} \tag{2.177}$$

Hence we may conclude that for all independent interaction behavior representable by exponential gravity models, pattern probabilities must satisfy the following *cost dependence* condition:

A12. (Cost Dependence) *For all cost configurations, $c \in C$, and activity-equivalent spatial interaction patterns, $s, s' \in S$,*

$$c(s) = c(s') \Rightarrow P_c(s) = P_c(s'). \tag{2.178}$$

In other words, independent interaction behavior representable by exponential gravity models is entirely determined by aggregate population levels of interaction activity and interaction costs. Note moreover that since the size of all activity-equivalent patterns is the same, it follows that such behavior is equivalently determined by its *average* levels of interaction activity and interaction costs. Conversely, it can be shown [as in Theorem 4.15 of Chapter 4] that all behavior which is so determined must be representable by exponential gravity models. Hence we obtain the following characterization of Model E1 in terms of cost dependent behavior:

E1-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by an exponential gravity model if and only if $\boldsymbol{P}$ satisfies cost dependence* (A12).

Hence, the fundamental characteristic of independent interaction behavior representable by exponential gravity models is that all such behavior is determined entirely by certain of its *system averages*, namely its average levels of interaction activity and interaction costs. Several additional observations can be made about this type of behavior. First it is of interest to compare *cost dependent* behavior (A12) with *separation dependent* behavior (A3) in Section 2.2.3 above. To do so, observe first that for any pair of activity-equivalent spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n)$, $s' = (i'_r j'_r : r = 1, \ldots, n) \in S$, it follows from A12 that if $c_s = c_{s'}$ for some cost configuration, $c \in C$, then

$$\begin{aligned} c_s = c_{s'} &\Rightarrow c_{i_r j_r} = c_{i'_r j'_r}, \quad r = 1, \ldots, n \\ &\Rightarrow \sum_{r=1}^{n} c_{i_r j_r} = \sum_{r=1}^{n} c_{i'_r j'_r}, \\ &\Rightarrow c(s) = c(s') \\ &\Rightarrow P_c(s) = P_c(s'). \end{aligned} \tag{2.179}$$

Thus we see that cost-dependent behavior is indeed a special case of separation-dependent behavior (A3) [i.e., that A12 $\Rightarrow$ A3]. In addition, the relationship in (2.179) also shows that the key difference between separation dependence and cost dependence is in terms of the *level of aggregation* involved in these conditions. To see this, recall from the discussion of separation-dependent behavior [following the formal statement of axiom (A3) in Section 2.2.3 above] that the aggregate property implicitly defined by equalities between separation arrays, $c_s = c_{s'}$, is precisely the common frequency distribution of separation profiles exhibited by patterns s and s'. In these terms, the separation dependence axiom (A3) asserts that all activity-equivalent interaction patterns consistent with the same aggregate *profile frequencies* are equally likely, i.e., are probabilistically indistinguishable. Similarly, the cost dependence axiom (A12) asserts that all activity-equivalent interaction patterns consistent with the same aggregate *profile totals* are probabilistically indistinguishable. Hence cost-dependent behavior only exhibits probabilistic variations at a higher level of aggregation than separation-dependent behavior, and in this sense, constitutes a somewhat more uniform type of spatial interaction behavior.

It is of interest to note that similar axioms involving average system behavior have a long history in the physical sciences, and evidently first appeared in the classical equilibrium theory of simple particle systems (ideal gases). In this physical context, our cost-dependency axiom corresponds precisely to the *Boltzmann hypothesis* that, in equilibrium, all gas particle ensembles (population states) which exhibit the same total energy level are equally likely to occur [see for example Huang (1963, Section 9.6) and Erlander and Smith (1990)]. The first widely recognized application of this hypothesis (also known as the 'maximum entropy' principle) to spatial interaction modeling was made by Wilson (1967) and has since been extended by a host of researchers [including Wilson (1970,1974), Fisk and Brown (1975), Snickars and Weibull (1977), Roy and Lesse (1981), Fisk (1985), and Smith (1989), among many others]. The most important consequence of this Boltzmann-type hypothesis for spatial interaction processes is that the overwhelmingly most probable realized interaction frequencies are *always* approximated asymptotically by exponential gravity models as population sizes become large. More precisely, for large spatial interaction patterns, s, consistent with any given overall activity levels (2.52) and total cost profile

(2.176), the overwhelmingly most probable values, $(N_{ij}^* : ij \in I \times J)$, of realized interaction frequencies, $(N_{ij}(s) : ij \in I \times J)$, under this hypothesis are asymptotically of the form (2.127) with $\mathrm{E}_c(N_{ij})$ replaced by N_{ij}^*. [See for example Wilson (1967) and Smith (1989)].

By comparison, the characterization theorem above is seen to yield a somewhat stronger result. In particular, this result shows that for *independent* spatial interaction processes, mean interaction frequencies are *exactly* consistent with an exponential gravity model. Hence under independence axioms, A1 and A2, one obtains exponential gravity representations of *all* behavior, and not just large population behavior. [Note also that for independent interaction behavior, the correspondence between these two types of results is given essentially by the classical Law of Large Numbers].

Relative Cost Dependence. Given this characterization of exponential gravity models, it is natural to expect that a strengthening of cost dependence paralleling that for separation dependence in axiom A3′ should yield a corresponding characterization of deterrence-invariant exponential gravity models. To see that this is the case, one need only observe that if the representation in (2.129) holds for $\boldsymbol{P}$, then for all spatial interaction patterns, $s, t \in S$, with $A(s) = A(t)$, and all cost configurations, $c \in C$, it follows from (2.175) and (2.176) that

$$\frac{P_c(s)}{P_c(t)} = \frac{\exp[-\theta^t c(s)]}{\exp[-\theta^t c(s)]} = \exp\{-\theta^t [c(s) - c(t)]\}, \tag{2.180}$$

and hence that for any comparable interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(t)$ and $A(s') = A(t')$ we now have the following parallel to (2.61):

$$\begin{aligned}
[c(s) = c(s'), c(t) = c(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} &= \exp\{-\theta^t [c(s) - c(t)]\} \\
&= \exp\{-\theta^t [c(s') - c(t')]\} \\
&= \frac{P_{c'}(s')}{P_{c'}(t')}.
\end{aligned} \tag{2.181}$$

Thus, every independent interaction process, $\boldsymbol{P}$, representable by a destination-deterrence-invariant exponential gravity model satisfies the following cost version of the relative separation property (A3′):

A12′. (Relative Cost Dependence) *For all cost configurations,* $c, c' \in C$, *and comparable spatial interaction patterns,* $s, t, s', t' \in S$, *with* $A(s) = A(t)$ *and* $A(s') = A(t')$,

$$[c(s) = c(s'), c(t) = c(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.182}$$

With this definition it can be shown [as in part (ii) of Theorem 4.17 of Chapter 4] that by replacing relative separation dependence with relative cost dependence, we obtain a characterization of Model E2 completely paralleling that of Model G2. However, the *log linear* structure of exponential models yields one added bonus. In particular, for the case of *scalar* separation costs (i.e., with $|K| = 1$) this relative cost dependence axiom is sufficient to ensure the existence of deterrence-invariant exponential representations of mean-interaction frequencies [as shown in Theorem 4.18 of Chapter 4]. Hence for this exponential case, we obtain the somewhat stronger characterization result:

E2-Characterization Theorem. *For any spatial interaction structure where either $|I| \geq 3$, $|J| \geq 3$, or $|K| = 1$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a deterrence-invariant exponential gravity model if and only if $\boldsymbol{P}$ satisfies relative cost dependence* (A12′).

As in the case of Model G2 above, it is possible to establish stronger behavioral conditions which yield a complete characterization of deterrence-invariant exponential gravity models [as in part (i) of Theorem 4.17 in Chapter 4]. The development of this more technical result is again deferred until Chapter 4.

Next we observe that by combining these cost dependence axioms (A12, A12′) with the proportionality axioms (A6,A7) and separability axioms (A8,A9) we may also characterize the exponential versions of both destination-deterrence-invariant and origin-deterrence-invariant gravity models. Turning first destination-deterrence-invariant exponential gravity models, observe that by definition all such models must satisfy A12 and A12′, as well as A6 and A8. In addition, since the primary role of the separability axiom (A8) was to guarantee the invariance of deterrence functions for general destination-deterrence-invariant models [as illustrated by the example in expression (2.85) above], it is reasonable to expect that this class of exponential models can be completely characterized by simply adding destination proportionality (A6) to the above characterization of exponential deterrence-invariant models. In the present case, however, no special cardinality restrictions are needed [as shown in part (i) of Theorem 4.20 in Chapter 4]. Finally, since A6 and A8 are sufficient to guarantee the existence of destination-deterrence-invariant representations for all spatial interaction structures with $|J| \geq 3$, one may also expect that the exponential form of such models can be characterized by adding *cost dependence* A12 to this set of axioms [as shown in part (ii) of Theorem 4.20 in Chapter 4]. Hence, by employing the cost dependence axiom A12 and relative cost dependence axiom (A12′), we obtain the following two characterizations of destination-deterrence-invariant exponential gravity models:

E3-Characterization Theorem.
(i) *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a destination-deterrence-invariant exponential gravity model if and only if $\boldsymbol{P}$ satisfies destination proportionality* (A6) *together with relative cost depen-*

dence (A12′).
(ii) *If in addition it is true that* $|J| \geq 3$, *then* $\boldsymbol{P}$ *is representable by a destination-deterrence-invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies destination proportionality* (A6) *and destination separability* (A8) *together with cost dependence* (A12).

Given these characterizations, it should be clear that completely parallel results can be established for origin-deterrence-invariant exponential gravity models [as shown in parts (i) and (ii) of Theorem 4.21 of Chapter 4]. In particular, we now have the following two characterizations of these models:

E4-Characterization Theorem.
(i) *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by an origin-deterrence-invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies origin proportionality* (A7) *together with relative cost dependence* (A12′).
(ii) *If in addition it is true that* $|I| \geq 3$, *then* $\boldsymbol{P}$ *is representable by an origin-deterrence-invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies origin proportionality* (A7) *and origin separability* (A9) *together with cost dependence* (A12).

Uniform Cost Dependence. Turning now to exponential models which exhibit the *relative invariance* condition in (2.132), observe again from (1.21), (1.23) and (1.30) that for all spatial interaction patterns, $s \in S$ and $c \in C$,

$$\begin{aligned} P_c(s) &= \mu(c)^{N(s)} \prod_i A(i)^{N_i(s)} \prod_j B(j)^{N_j(s)} \exp[-\theta^t(\sum_{ij} c_{ij} N_{ij}(s))] \\ &= \mu(c)^{N(s)} \prod_i A(i)^{N_i(s)} \prod_j B(j)^{N_j(s)} \exp[-\theta^t c(s)], \end{aligned} \tag{2.183}$$

where $\mu(c) = \lambda(c)/\mathrm{E}_c(N)$, and hence that for any comparable interaction patterns, $s, t, s', t' \in S_n$, with $A(s) = A(s')$ and $A(t) = A(t')$, and cost configurations, $c, c' \in C$, with $c(s) = c'(s')$ and $c(t) = c'(t')$, we now have the following parallel to (2.64) with $c(s)$ replacing $\Pi F(c_s)$:

$$\begin{aligned} \frac{P_c(s)}{P_c(t)} &= \frac{\mu(c)^n \prod_i A(i)^{N_i(s)} \prod_j B(j)^{N_j(s)} \exp[-\theta^t c(s)]}{\mu(c)^n \prod_i A(i)^{N_i(t)} \prod_j B(j)^{N_j(t)} \exp[-\theta^t c(t)]} \\ &= \frac{\mu(c')^n \prod_i A(i)^{N_i(s')} \prod_j B(j)^{N_j(s')} \exp[-\theta^t c'(s')]}{\mu(c')^n \prod_i A(i)^{N_i(t')} \prod_j B(j)^{N_j(t')} \exp[-\theta^t c'(t')]} \\ &= \frac{P_{c'}(s')}{P_{c'}(t')}. \end{aligned} \tag{2.184}$$

Thus we obtain the following *uniform cost dependence condition* which must be satisfied by all independent interaction behavior representable by relatively invariant exponential gravity models:

A12″. (Uniform Cost Dependence) *For all cost configurations,* $c, c' \in C$, *and comparable spatial interaction patterns,* $s, t, s', t' \in S$, *with* $A(s) = A(s')$ *and* $A(t) = A(t')$,

$$[c(s) = c'(s'), c(t) = c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.185}$$

As in the more general case of relatively invariant gravity models, this uniform cost dependence condition (A12″) can also be shown to be sufficient for the existence of relatively invariant exponential gravity models [as in Theorem 4.22 of Chapter 4]. Hence we obtain the following characterization of Model E5 in terms of this condition:

E5-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by a relatively invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies uniform cost dependence* (A12″).

In a similar manner, it again follows that by adding the sub-configuration independence axiom (A5), we obtain the following characterization of invariant exponential gravity models [as shown in Theorem 4.23 of Chapter 4]:

E6-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by an invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies sub-configuration independence* (A5) *together with uniform cost dependence* (A12″).

(B) Cost Efficiency Axioms

As in the case of general gravity models, the exponential gravity models above make no assertion about the diminishing effect of interaction costs on the likelihood of interactions. Hence it is of interest to consider additional behavioral conditions under which this is the case. Again, it is natural to expect that the appropriate strengthening of cost dependence parallels that for separation dependence. To see that this is the case, observe that if behavior is representable by a monotone exponential gravity model as in (2.127) with $\theta_c \in R_+^K$ for all cost configurations, $c \in C$, then for any activity-equivalent trip patterns, $s, s' \in S$, the argument of (2.176) now shows that for each $c \in C$,

$$\begin{aligned} c(s) \le c(s') &\Rightarrow \theta_c^t c(s) \le \theta_c^t c(s') \\ &\Rightarrow \exp[-\theta_c^t c(s)] \ge \exp[-\theta_c^t c(s')] \\ &\Rightarrow P_c(s) \ge P_c(s'). \end{aligned} \tag{2.186}$$

Cost Efficiency. Thus, as a parallel to the separation efficiency condition (A4), we now see that all independent interaction behavior representable by monotone exponential gravity models must satisfy the following *cost efficiency* condition:

A13. (Cost Efficiency) *For all cost configurations, $c \in C$, and activity-equivalent spatial interaction patterns, $s, s' \in S$,*

$$c(s) \leq c(s') \Rightarrow P_c(s) \geq P_c(s'). \tag{2.187}$$

In other words, among all spatial interaction patterns consistent with the same origin and destination activity levels, those involving lower total interaction costs (in each cost component) are at least as likely to occur as those involving higher total costs. This cost efficiency condition can also be shown to guarantee the existence of monotone exponential gravity models for independent interaction behavior [as in Theorem 4.16 of Chapter 4], and thus yields the following characterization of Model E1*:

E1*-Characterization Theorem. *An independent spatial interaction process, $\boldsymbol{P}$, is representable by a monotone exponential gravity model if and only if $\boldsymbol{P}$ satisfies cost efficiency* (A13).

Hence all independent interaction behavior representable by monotone exponential gravity models is completely characterized by this aggregate cost efficiency property of behavior. Several additional observations about this behavioral property should be made. First observe that if cost efficiency holds, then it follows from (2.185) that for any cost configuration, $c \in C$, and activity-equivalent interaction patterns, $s, s' \in S$,

$$\begin{aligned} c(s) = c(s') &\Rightarrow [c(s) \leq c(s'), c(s') \leq c(s)] \\ &\Rightarrow [P_c(s) \geq P_c(s'), P_c(s') \geq P_c(s)] \\ &\Rightarrow P_c(s) = P_c(s'), \end{aligned} \tag{2.188}$$

and hence that cost efficiency implies cost dependency [i.e., A13 $\Rightarrow$ A12]. In particular, this implies that the asymptotic version of the results for cost dependency also hold for cost efficiency. More precisely, it can be shown [as for example in Smith (1989)] that for large spatial interaction patterns, s, consistent with any given overall activity levels (2.52) and total cost profile (2.176), the overwhelmingly most probable values, $(N_{ij}^* : ij \in I \times J)$, of realized interaction frequencies, $(N_{ij}(s) : ij \in I \times J)$, are asymptotically approximated by a *monotone* exponential gravity model.

Finally, it is important to emphasize that this cost efficiency property involves *no normative assertions* about overall population behavior. For example, it might seem tempting to argue that such populations exhibit 'welfare maximizing' tendencies with respect to their total 'social interaction costs'. However, this view is quite misleading, as can be seen by a comparison of this behavior with that exhibited in comparable deterministic models of 'network equilibrium'. In particular, for steady-state equilibria in which travel costs are taken to be *constant* (and thus which are structurally compatible with our present framework, as discussed in Section 1.2.2 above), the resulting user equilibrium patterns are well known to be precisely those which minimize total travel costs [as discussed for example in

Smith (1988, Section 5.2)]. But this does *not* mean that the population of trip makers is acting to maximize its 'welfare'. Rather, it means simply that in the presence of no externalities, populations of individual cost minimizers tend in the aggregate to minimize total costs. An even more extreme example is provided by the classical Boltzmann hypothesis discussed above. In particular, while many systems of interacting gas particles tend to 'seek minimum total energy levels' in equilibrium, this type of aggregate behavior clearly has no 'welfare' implications for such systems. [For a more detailed discussion of the relation between cost efficiency and the Boltzmann hypothesis, see Erlander and Smith (1990)].

Relative Cost Efficiency. Next observe that by simply replacing the product function, $\Pi F(\cdot)$, of separation arrays in both (2.68) and (2.69) with the exponential function, $\exp[-\theta^t(\cdot)]$, of total cost profiles, the same argument now shows that every monotone deterrence-invariant exponential gravity model must satisfy the following *relative cost efficiency* condition:

A13′. (Relative Cost Efficiency) *For all cost configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(t)$ and $A(s') = A(t')$,*

$$[c(s) \leq c'(s'), c(t) \geq c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{2.189}$$

Hence, as a strengthening of the above characterization of deterrence-invariant exponential gravity models, we now have the following monotone version of this result [established in part (ii) of Theorem 4.19 in Chapter 4]:

E2*-Characterization Theorem. *For any spatial interaction structure in which either $|I| \geq 3$, $|J| \geq 3$, or $|K| = 1$, an independent spatial interaction process, $\boldsymbol{P}$, on $I \times J$ is representable by a monotone deterrence-invariant exponential gravity model if and only if $\boldsymbol{P}$ satisfies relative cost efficiency* (A13′).

In a completely parallel manner, these cost efficiency axioms yield the following monotone characterization of destination-deterrence-invariant gravity models [established in Theorems 4.20]:

E3*-Characterization Theorem.
(i) *An independent spatial interaction process, $\boldsymbol{P}$ is representable by a monotone destination-deterrence-invariant exponential gravity model if and only if $\boldsymbol{P}$ satisfies destination proportionality* (A6) *together with relative cost efficiency* (A13′).
(ii) *If in addition it is true that $|J| \geq 3$, then $\boldsymbol{P}$ is representable by a monotone destination-deterrence-invariant exponential gravity model if and only if $\boldsymbol{P}$ satisfies destination proportionality* (A6) *and destination separability* (A8) *together with cost efficiency* (A13).

Similarly, we have the following monotone characterization of origin-deterrence-invariant gravity models [established in Theorem 4.21 of Chapter 4]:

E4*-Characterization Theorem.
(i) *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by a monotone origin-deterrence-invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies origin proportionality* (A7) *together with relative cost efficiency* (A13′).
(ii) *If in addition it is true that* $|I| \geq 3$, *then* $\boldsymbol{P}$ *is representable by a monotone origin-deterrence-invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies origin proportionality* (A7) *and origin separability* (A9) *together with cost efficiency* (A13).

Uniform Cost Efficiency. Turning next to monotone relatively invariant exponential gravity models, observe that by replacing the product function, $\Pi F(\cdot)$, of separation arrays in (2.71) with the exponential function, $\exp[-\theta^t(\cdot)]$, of total cost profiles, the same argument now shows that all such models must satisfy the following *uniform cost efficiency* condition:

A13″. (Uniform Cost Efficiency) *For all cost configurations,* $c, c' \in C$, *and comparable spatial interaction patterns,* $s, t, s', t' \in S$, *with* $A(s) = A(s')$ *and* $A(t) = A(t')$,

$$[c(s) \leq c'(s'), c(t) \geq c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \qquad (2.190)$$

Hence the characterization results for Model E1* and Model E5 are seen to imply that this uniform cost efficiency property must ensure the existence of monotone relatively invariant exponential gravity models for independent interaction behavior [as shown in Theorem 4.22 in Chapter 4], so that we now obtain the following characterization of Model E5* in terms of uniform cost efficiency:

E5*-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by a monotone relatively invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies uniform cost efficiency* (A13″).

Finally, the appropriate characterization of monotone invariant exponential gravity models is obtained in precisely the same manner as in the case of general gravity models by simply introducing the condition of sub-configuration dependence (A5) [as shown in of Theorem 4.23 in Chapter 4]:

E6*-Characterization Theorem. *An independent spatial interaction process,* $\boldsymbol{P}$, *is representable by a monotone invariant exponential gravity model if and only if* $\boldsymbol{P}$ *satisfies sub-configuration dependence* (A5) *together with uniform cost efficiency* (A13″).

2.5 Generalizations of the Gravity Models

In this final section we consider a number of possible generalizations of the basic gravity model form in Definition 2.1. To do so, we begin by considering possible formal relaxations of the basic separability assumptions implicit in this model. This is followed by a more explicit consideration of types of spatial interaction behavior leading to more complex model formulations. A major objective of this development is to show (especially Sections 2.5.2 and 2.5.3 below) how the basic behavioral axioms characterizing gravity models can be employed to obtain representations of more complex models as well.

Turning first to the formal structure of gravity models themselves, recall that our fundamental hypothesis about the effect of spatial separation on interaction behavior is that within any given configuration, $c \in C$, all effects of separation on mean interaction frequencies can be captured by a *deterrence function*, F_c, which is common to all origins and destinations. Hence, a wide range of generalizations can be achieved by simply allowing additional spatial variation in these functions. As a first possibility, suppose it is postulated that the significance of separation effects differs across origins, perhaps due to heterogeneities among actor populations. In this case, one might wish to consider various types of *origin-dependent deterrence functions*, F_{ic}, at each origin, $i \in I$. For example, one could define a class of *origin-dependent exponential gravity models* by simply replacing the cost sensitivity vectors, θ_c, with origin-dependent vectors, θ_{ic}, in (2.127) of Definition 2.14 above. As a second illustration, recall from the discussion of origin function specifications in Section 2.3 that the *minimum-separation measure* (2.107) was there employed to construct an explicit origin-dependent deterrence function in (2.113). A similar example could be constructed in terms of the Hansen-type *origin accessibility measure* in (2.109).

As a second possible relaxation of deterrence functions, observe that if there are significant heterogeneities in opportunity populations with respect to separation effects, then one may wish to consider *destination-dependent deterrence functions*, F_{jc}, at each destination, $j \in J$. For example, a class of *destination-dependent exponential gravity models* could be defined by replacing θ_c with θ_{jc} in (2.127). A more explicit example of this type could be constructed [in a manner paralleling (2.113)] by employing the Hansen-type measure of *destination accessibility* in (2.111) to construct the following destination-dependent version of a given deterrence function, F_c,

$$F_{jc}(c_{ij}) = [\sum_{h \in J/j} b_h(c_{jh})^{-\theta}] \cdot F_c(c_{ij}), \quad j \in J \tag{2.191}$$

[where it is assumed in addition that $J \subseteq I$ and that c_{ij} is a positive scalar measure of interaction cost].

Finally, one may postulate that variations in the perceptions of separation effects are significant across *both* origins and destinations, and thus consider more general specifications of deterrence functions of the form, F_{ijc}, for each origin-destination pair, $ij \in I \times J$. One explicit instance of this type of deterrence function is obtained by postulating that perceptions of destination accessibility vary across origins. In particular, if such heterogeneity effects are captured by replacing the exponent, θ, in (2.191) above with an origin-dependent exponent, θ_i, [as for example in Fotheringham (1983a, Section 4)] then one obtains the following extension of (2.191) which depends on both origins and destinations:

$$F_{ijc}(c_{ij}) = [\sum_{h \in J/j} b_h(c_{jh})^{-\theta_i}] \cdot F_c(c_{ij}), \quad ij \in I \times J. \tag{2.192}$$

One may also consider a broad class of model specifications within the general framework of log-additive (or log-linear) models [as applied to spatial interaction modeling by Fingleton (1981,1983) and Aufhauser and Fischer (1985), among others]. In particular, if each possible subset of the origin, destination, and configuration indices, $\{i, j, c\}$, is treated as a possible *interaction effect* influencing the mean frequencies, $\mathrm{E}_c(N_{ij})$ then each specification of such interaction effects may be said to yield a distinct class of spatial interaction models. In this context, our general gravity model (Model G1), for example, can be viewed as a type of log-additive model with interaction effects, $(\{i, c\}, \{j, c\}, \{c\})$, where the pure 'configuration effect', $\{c\}$, represents the deterrence function, F_c.

Numerous other specifications of spatial interaction models have also been proposed which are not captured by the above framework. Most notable among these is the specification of mean interaction frequencies proposed by Tobler (1983). In particular, if P_i and P_j denote the relevant 'population sizes' of locations i and j, respectively, then (in terms of our present framework) the Tobler model can be expressed as

$$\mathrm{E}_c(N_{ij}) = [G_c(i) + H_c(j)]P_i P_j F_c(c_{ij}) \tag{2.193}$$

where it is postulated that the relevant origin 'push' effects, $G_c(i)$, and destination 'pull' effects, $H_c(j)$, are *additive* rather than multiplicative [see also Ledent (1985) and Fotheringham and O'Kelly (1989, Section 2.6)]. But while such specifications are often quite suggestive [Tobler (1983), for example, derives (2.193) by minimizing a certain quadratic 'per-capita work function'], the types of spatial interaction behavior giving rise to these specifications remains unclear.

With this in mind, our primary interest in this final section is to illustrate explicit types of spatial interaction behavior which lead to model forms not encompassed by the class of gravity models in Definition 2.1. In particular we shall illustrate three types of behavior which exhibit this property. To do so, we begin in Section 2.5.1 below by considering certain extensions of the simple search behavior studied in Example 4 of Section 2.4.2

above. This is followed in Section 2.5.2 by a consideration of certain hierarchical decision structures in spatial interaction, resulting from perceived destination-clustering effects. Finally, we conclude in Section 2.5.3 with a consideration of certain more general types of contextual effects which may influence the relevant sets of destinations for spatial actors.

2.5.1 Generalized Search Processes

Recall that in the class of simple search processes constructed in Example 4 of Section 2.4.2, one of the strongest assumptions was the *opportunity homogeneity assumption* (S5) that all opportunities have the same likelihood of satisfying the needs of any actor. Hence we begin by showing that if the notion of 'intervening-opportunity distance' is broadened in an appropriate way, then gravity model representations of mean realized-interaction frequencies are possible under the following weaker version of the assumption S5 in which acceptance probabilities, $P_c^1(\Delta_{1\beta} = 1)$, are invariant only across search schemes, $c \in C$, as follows:

S5′. *For all search schemes, $c, c' \in C$, and opportunities, $\beta \in B$,*

$$P_c^1(\Delta_{1\beta} = 1) = P_{c'}^1(\Delta_{1\beta} = 1). \tag{2.194}$$

To do so, observe first that if we again let $\Delta_{1j} = \max\{\Delta_{1\beta} : \beta \in B_j\}$, then it follows from S5′ that the probability that some opportunity at destination j satisfies the needs of actor, α_1, is still given by $P^1(\Delta_{1j} = 1)$, in expression (2.171). However, since opportunities are no longer homogeneous with respect to their acceptance probabilities, the probability that searchers from origin i find no satisfactory opportunity prior to destination j in search scheme c now depends on the specific destinations in the set, $J_c[j \mid i]$, defined in (2.166). Hence the probability that a searcher from i reaches j in search scheme c is no longer a simple function of the number of intervening opportunities, c_{ij}, as in (2.170). Rather, this probability now depends explicitly on the opportunity set, $J_c[j \mid i]$, and in particular, is given by the joint probability that $\Delta_{1h} = 0$ for all destinations, $h \in J_c[j \mid i]$, which may be denoted by

$$P_c[j \mid i] = \prod_{h \in J_c[j \mid i]} P^1(\Delta_{1h} = 0) = \prod_{h \in J_c[j \mid i]} \prod_{\beta \in B_h} P^1(\Delta_{1\beta} = 0), \tag{2.195}$$

where the equality follows from the identity $P^1(\Delta_{1h} = 0) = P^1(\Delta_{1\beta} = 0 : \beta \in B_h)$, together with the independence assumption in (2.161). Hence, if we now replace the definition c_{ij} in (2.168) by

$$c_{ij} = -\log P_c[j \mid i] = \sum_{h \in J_c[j \mid i]} \sum_{\beta \in B_h} [-\log P^1(\Delta_{1\beta} = 0)], \tag{2.196}$$

then it follows (from the same argument as in Proposition 4.8 of Chapter 4) that

$$\begin{aligned} \mathrm{E}_c(N_{ij}^c) &= \mathrm{E}_c(N_i)P^1(\Delta_{1j} = 1)P_c[j \mid i] \\ &= \mathrm{E}_c(N_i)P^1(\Delta_{1j} = 1)\exp\{\log P_c[j \mid i]\} \qquad (2.197) \\ &= \mathrm{E}_c(N_i)P^1(\Delta_{1j} = 1)\exp[-c_{ij}], \end{aligned}$$

and hence that $\mathrm{E}_c(N_{ij})$ is again representable as in (2.169) [with the same origin and destination weights and with $\theta = 1$]. Here, we see from (2.196) that the notion of 'intervening-opportunity distance' in (2.168) is now extended to reflect not only the number of intervening opportunities, but also the relative 'power' of these opportunities to stop the given search process. To be more precise, observe from (2.196) that the contribution to c_{ij} of each intervening opportunity, β, is given by the nonnegative quantity, $c_{ij}(\beta) \equiv -\log P^1(\Delta_{1\beta} = 0)$, which may be interpreted as the *stopping power* of β in the search process. In particular, observe that $c_{ij}(\beta) = 0$ if and only if $P^1(\Delta_{1\beta} = 0) = 1$, i.e., if and only if β has 'no stopping power' with respect to the searcher's needs. Similarly, $c_{ij}(\beta) = \infty$ if and only if $P^1(\Delta_{1\beta} = 1) = 1$, i.e., if and only if β has 'full stopping power' with respect to the searcher's needs. Hence, from a behavioral viewpoint, the cumulative values c_{ij} in (2.196) might be designated as *effective intervening-opportunity distances*, where $c_{ij} = 0$ indicates that no intervening opportunities between i and j have stopping power (so that j will *always* be reached from i), and $c_{ij} = \infty$ indicates that at least one intervening opportunity between i and j has full stopping power (so that j will *never* be reached from i).

Observe, however, that this broader notion of effective intervening-opportunity distance stretches the basic concept of 'spatial separation'. In particular, the value c_{ij} depends not only on the spatial properties of intervening opportunities between i and j but also on certain of their *nonspatial* properties. This becomes even more clear when we attempt to relax the simple search model further. As one important illustration, recall that perhaps the strongest behavioral assumption in this simple model was that search *always* continues until a satisfactory opportunity is found (or until all opportunities have been searched). This implicitly assumes that the *costs* of search are sufficiently small to be ignored. Hence if search costs are indeed significant, then it is more realistic to postulate that actors can choose to discontinue searching after some point. In many search theories, this type of behavior is modeled by some form of 'stopping rule' for actors [as for example in Rogerson (1982) and Smith *et al.* (1992)]. If such stopping rules differ among actors, then one can treat such stopping rules as random variables (in a manner closely paralleling the 'threshold' random variables in Example 2 of Section 2.2.2 above). In particular, let the zero-one variable, $S_{ijc}^1(\omega)$, be defined to have value one if and only if actor, α_1, in search event ω starts from origin, i, under search scheme, c, and employs an (implicit) stopping rule which terminates his search at destination j whenever

j is reached and contains no satisfactory opportunity. With this definition, it follows that the probability, $P_c[j \mid i]$, in (2.195) that j is reached from i under search scheme c is now given by the joint probability that each destination prior to j has no satisfactory opportunity and is not a stopping point for the stopping rule employed by the actor, i.e., by

$$P_c[j \mid i] = P_c(\Delta_{1h} = 0, S_{ihc}^1 = 0 : h \in J_c[j \mid i]). \qquad (2.198)$$

In this more general context, mean interaction frequencies are still seen to be of the form given by the first line of (2.197) where $P_c[j \mid i]$ is now given by (2.198). Hence, one could again simply set $c_{ij} = -\log P_c[j \mid i]$ and represent $\mathrm{E}_c(N_{ij})$ by the 'exponential gravity model' in (2.197). But since each stopping-rule condition, $S_{ihc}^1 = 0$, implicitly depends on both the spatial and nonspatial properties of all opportunities beyond destination h, the notion of 'intervening-opportunity distance' now loses its meaning altogether. Moreover, it is difficult to give any compelling 'spatial separation' interpretation to such values of c_{ij}. Thus, within our present modeling framework, it seems more appropriate to regard this search model as one in which the spatial and nonspatial attributes of interaction opportunities are nonseparable, and hence in which the corresponding mean realized-interaction frequencies, $\mathrm{E}_c(N_{ij}^c)$, have no meaningful gravity-type representation.

2.5.2 Interaction Processes with Hierarchical Destinations

To motivate our second extension of gravity models, recall from the discussion of spatial aggregation in Section 1.3.2 that each *destination*, j, was implicitly assumed to include a population of interaction opportunities, B_j, which were sufficiently close in all relevant separation aspects to be represented by a common separation profile, c_{ij}, with respect to each origin population, A_i. In addition, recall from the discussion of structural independence in Section 1.3.3 that these populations were assumed to be *independent* of any changes in separation configurations, i.e., both the origin set, I, and destination set, J, were assumed to be the same for all separation configurations, $c \in C$. For example, while freeway improvements may alter the separation between city i and city j, they will generally not alter the clustering of actors and opportunities defined by these cities. Similarly, while improved political relations between countries i and j may effectively reduce the separation between them, such improved relations will generally not alter their individual political boundaries. Hence in such cases it is reasonable to focus only on configurational changes *between* these origins and destinations. However, in other cases the clustering of actors and opportunities can be significantly influenced by changes in the relevant dimensions of separation between them. For example, if all stores within 'walking distance' of each other are perceived by shoppers as a relevant cluster, then such clusters are clearly influenced by the relocation of stores (or the construction of pedestrian walkways, overpasses, etc.). Similarly, in a political voting context, changes of candidate positions on key issues can influence the perceived clustering of candidates by voters. With these types of examples in mind, we shall here treat the set, I, of actor populations as fixed, and shall concentrate on changes in the clustering of opportunities perceived by actors as the separation between opportunities changes. To model the behavioral effects of such changes, we shall postulate [following Fotheringham (1986)] that spatial interaction behavior involves a *hierarchical* interaction process in which a relevant opportunity cluster is first selected (or perceived) by an actor, and then a specific destination within this cluster is chosen. [For additional illustrations of spatial choice hierarchies in a transportation context, see for example Boyce *et al.* (1983)].

To formalize these ideas within our present framework, it is convenient to begin by equating destinations with opportunities (i.e., to let $J = B$), so that each opportunity is now treated as a distinct destination, $j \in J$. In addition, it is convenient to treat each destination as a possible origin (i.e., to assume that $J \subseteq I$), so that separation profiles, c_{jh}, between destination pairs, $j, h \in J$, are well defined. In this context, it is postulated that each spatial separation configuration, $c \in C$, gives rise to a partition, $\boldsymbol{D}_c$, of the destination set, J, into *destination clusters*, $D \in \boldsymbol{D}_c$, which are commonly perceived by all actors. Hence if no clustering effects are perceived under configuration, c, then the relevant partition is one in which each cluster, $D \in \boldsymbol{D}_c$, contains exactly one destination. [A more general formulation involving *random destination sets* is treated in Section 2.5.3 below.] Next, it is postulated that for any given separation configuration, $c \in C$, spatial interaction decisions by actors at each origin, $i \in I$, involve a two-stage process in which a destination cluster, $D \in \boldsymbol{D}_c$, is first selected, and then a specific destination, $j \in D$, is chosen. Hence each individual interaction event in the present context is representable by a triple, (iDj), and the class, Ω, of relevant population events consists of all finite *interaction patterns*, $\omega = (i_r D_r j_r : r = 1, \ldots, n)$, with $j_r \in D_r \subseteq J$ for each $r = 1, \ldots, n$. Population behavior of this type can then be modeled by an interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω satisfying the consistency condition that the only possible interaction patterns under distribution P_c involve cluster sets in $\boldsymbol{D}_c$. Hence we now say that [see also Definitions 4.10 and 4.11 in Chapter 4 below]:

Definition 2.25 An interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω is designated as a *hierarchical interaction process* if and only if for all $c \in C$ and $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$,

$$P_c(\omega) > 0 \Leftrightarrow D_r \in \boldsymbol{D}_c, \quad r = 1, \ldots, n. \tag{2.199}$$

To develop explicit model representations of hierarchical interaction processes within our present framework, we begin with the choice of destination clusters by actors. In particular, observe that if for any given separation configuration, $c \in C$, we simply substitute the *destination cluster set*, $\boldsymbol{D}_c$, for the destination set, J, in our basic model of spatial interaction processes, then the choice of a destination cluster, $D \in \boldsymbol{D}_c$, by an actor at origin, $i \in I$, can be treated in a manner which is formally identical to the choice of destinations, $j \in J$, in the basic model. In this case the appropriate class of *cluster-level interaction patterns*, $\boldsymbol{\Sigma}_c$, now consists of all finite populations of cluster choices, D, by actors at origins, i, with typical elements of the form, $\boldsymbol{\sigma} = (i_r D_r : r = 1, \ldots, n) \in \boldsymbol{\Sigma}_c$, where $i_r D_r \in I \times \boldsymbol{D}_c, r = 1, \ldots, n$. Hence, if $N_i(\boldsymbol{\sigma})$ and $N_D(\boldsymbol{\sigma})$ now denote the origin-activity at i and cluster-activity at D generated by pattern, $\boldsymbol{\sigma}$, respectively, and if we let the corresponding *cluster-level activity profile* be denoted by

$$A_c(\boldsymbol{\sigma}) = [(N_i(\boldsymbol{\sigma}) : i \in I), (N_D(\boldsymbol{\sigma}) : D \in \boldsymbol{D}_c)], \tag{2.200}$$

then we may again designate two patterns, $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \boldsymbol{\Sigma}_c$, to be *activity-equivalent* if and only if $A_c(\boldsymbol{\sigma}) = A_c(\boldsymbol{\sigma}')$. Moreover, if for each origin-cluster pair, $iD \in I \times \boldsymbol{D}_c$, we let

$$c_{iD} = (c_{ij} : j \in D) \tag{2.201}$$

denote the relevant *(iD)-separation profile* in c, then we can define an appropriate notion of a *cluster-level separation array* for each pattern, $\boldsymbol{\sigma} = (i_r D_r : r = 1, \ldots, n)$, by

$$c = (c_{i_r D_r} : r = 1, \ldots, n). \tag{2.202}$$

With these definitions, it follows that if for each for independent interaction processes, $\boldsymbol{P} = \{P_c : c \in C\}$ we let $P_c(\boldsymbol{\sigma})$ denote the probability of the spatial interaction event $\boldsymbol{\sigma}$ [as defined more fully in Section 4.6.1 below] then we are led to consider the following version of the separation dependence axiom (A3) at the cluster level:

A3.1. (Cluster-Level Separation Dependence) *For all separation configurations, $c \in C$, and activity-equivalent cluster-level interaction patterns, $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \boldsymbol{\Sigma}_c$,*

$$c_{\boldsymbol{\sigma}} = c_{\boldsymbol{\sigma}'} \Rightarrow P_c(\boldsymbol{\sigma}) = P_c(\boldsymbol{\sigma}'). \tag{2.203}$$

Given this behavioral axiom for the first-stage choice of a cluster, $D \in \boldsymbol{D}_c$, we next observe that the second-stage choice of a destination, $j \in D$, is again formally identical to our basic model of spatial interaction processes, with the full destination set, J, replaced by the cluster set, D. Hence, if S_D denotes the class of all spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n) \in S$, with $j_r \in D$ for each $r = 1, \ldots, n$, then we may formulate the following corresponding version of separation dependence (A3) at the destination level in terms of our original notation:

A3.2. (Destination-Level Separation Dependence) *For all separation configurations, $c \in C$, destination clusters, $D \in \boldsymbol{D}_c$, and activity-equivalent spatial interaction patterns, $s, s' \in S_D$,*

$$c_s = c_{s'} \Rightarrow P_c(s) = P_c(s'). \tag{2.204}$$

With this two-part version of separation dependence, we can now simply apply the results of Section 2.2.3 above to obtain a representation of mean interaction frequencies for independent interaction processes satisfying both these conditions. To do so, we begin by formulating the following version of the independence axioms A1 and A2 for hierarchical interaction processes. In particular, if for each $c \in C$ we denote the feasible class of individual interaction events in Ω_1 under c by $\Omega_{1c} = \{iDj \in \Omega_1 : D \in \boldsymbol{D}_c\}$, then a typical *frequency profile* of feasible interaction frequencies under c can be denoted by $\boldsymbol{n} = (n_{iDj} : iDj \in \Omega_{1c})$. Hence, if for each positive integer n we let $\Omega_{nc} = (\Omega_{1c})^n$ and as in (3.24) denote conditional probabilities given n by P_c^n, then we now have [see also Definition 4.12 in Chapter 4]:

Definition 2.26 A hierarchical interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ is said to be *independent* if and only if $\boldsymbol{P}$ satisfies the following two conditions for all separation configurations, $c \in C$, feasible interaction patterns, $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega_{nc}$, $n > 0$, and interaction frequency profiles, $\boldsymbol{n} = (n_{iDj} : iDj \in \Omega_{1c})$,

A1′. (Interaction Independence) $P_c^n(\omega) = \prod_{r=1}^n P_c^n(i_r D_r j_r)$.

A2′. (Frequency Independence) $P_c(\boldsymbol{n}) = \prod_{iDj \in \Omega_{1c}} P_c(n_{iDj})$.

Given this definition of independent hierarchical interaction processes, $\boldsymbol{P}$, if we now denote the *(iDj)-frequency* for processes $\boldsymbol{P}$ by the random variable, N_{iDj}, and similarly denote the *(iD)-frequency* by $N_{iD} = \Sigma_{j \in D} N_{iDj}$, then it can be shown [as in Proposition 4.11 of Chapter 4] that axiom A3.1 yields the following representational characterization of *mean (iD)-frequencies*, $\mathrm{E}_c(N_{iD})$, for such processes:

Representation of Cluster-Level Frequencies. *An independent hierarchical interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with cluster partitions, $(\boldsymbol{D}_c : c \in C)$, satisfies cluster-level separation dependence* (A3.1) *if and only if for each $c \in C$ there exists a positive origin function, A_c, destination-cluster function, B_c, and cluster-level deterrence function, F_c, such that for all $iDj \in \Omega_{1c}$,*

$$\mathrm{E}_c(N_{iD}) = A_c(i) B_c(D) F_c(c_{iD}). \tag{2.205}$$

Similarly, if for each possible cluster set, $D \subseteq J$, we now designate the subset of all interaction patterns in involving only interactions with D by $\Omega_D = \{\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega : D_1 = \ldots = D_n = D\}$, and denote the *conditional mean (ij)-interaction frequencies* given Ω_D by $\mathrm{E}_c(N_{ij}\,|\,\Omega_D)$ $[=\mathrm{E}_c(N_{iDj}\,|\,\Omega_D)]$, then it can be shown [Proposition 4.12 in Chapter 4] that axiom A3.2 yields the following representation of these conditional mean frequencies:

Representation of Destination-Level Frequencies. *An independent hierarchical interaction process, $\boldsymbol{P}=\{P_c : c\in C\}$ with cluster partitions, $(\boldsymbol{D}_c : c\in C)$, satisfies destination-level separation dependence* (A3.2) *if and only if for each $c\in C$ and $D\in\boldsymbol{D}_c$ there exists a positive origin function, A_{cD}, destination function, B_{cD}, and destination-level deterrence function, F_{cD}, such that for all $ij\in I\times D$,*

$$\mathrm{E}_c(N_{ij}\,|\,\Omega_D)=A_{cD}(i)B_{cD}(j)F_{cD}(c_{ij}). \tag{2.206}$$

Finally, it should be clear that one may construct a similar two-part version of the separation efficiency axiom (A4), which will imply that the cluster-level deterrence functions, F_c, in (2.205) and destination-level deterrence functions, F_{cD}, in (2.206) are all nonincreasing. Similarly, for the case of extensively measurable interaction costs, one may construct a similar two-part version of the cost dependency axiom (A8) and cost efficiency axiom (A9) to obtain exponential representations of these deterrence functions. However, it is important to note there are no analogous two-part versions of axioms A3′, A4′, A6, A7, A8′, and A9′. In particular, the dependency of cluster partitions, $\boldsymbol{D}_c$, on the underlying separation configuration, $c\in C$, implies that the resulting class of cluster-level interaction patterns, $\boldsymbol{\Sigma}_c$, can *differ* between configurations. Hence, there is no direct comparability among all configurations, as required by these axioms. [A possible approach to this problem is discussed below in the context of more general types of spatial interaction processes involving random destination sets.]

Given these two formal representations, we may now derive the desired representation of mean interaction levels, $\mathrm{E}_c(N_{ij})$ for the independent spatial interaction process generated by each independent hierarchical interaction process, $\boldsymbol{P}$, as follows. First observe that for each cluster set, $D\in\boldsymbol{D}_c$, under configuration c, and each destination, $j\in D$, the interaction event, (ij), occurs if and only if the interaction event, (iDj), occurs (since the occurrence of j implies the occurrence of D). Hence for any interaction event of size one, it follows by definition that

$$P_c^1(ij)=P_c^1(ijD)=P_c^1(j\,|\,iD)\cdot P_c^1(iD). \tag{2.207}$$

But by expression (1.30) in the Multinomial Sampling Theorem of Chapter 1 we must have

$$\begin{aligned}P_c^1(j\,|\,iD)&=\frac{P_c^1(ij)}{\sum_{h\in D}P_c^1(ih)}=\frac{\mathrm{E}_c(N_{ij}\,|\,\Omega_D)}{\sum_{h\in D}\mathrm{E}_c(N_{ih}\,|\,\Omega_D)}\\&=\frac{B_{cD}(j)F_{cD}(c_{ij})}{\sum_{h\in D}B_{cD}(h)F_{cD}(c_{ih})}\end{aligned} \tag{2.208}$$

and, similarly, must have

$$P_c^1(iD) = \frac{\mathrm{E}_c(N_{iD})}{\mathrm{E}_c(N)} = \frac{A_c(i)B_c(D)F_c(c_{iD})}{\mathrm{E}_c(N)}. \tag{2.209}$$

Hence, by combining (2.205) through (2.209), and applying (1.30) once more, we obtain the following representation of *mean spatial interaction frequencies*, $\mathrm{E}_c(N_{ij})$ for all such hierarchical choice processes [see also Proposition 4.13 in Chapter 4]:

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= P_c^1(ij) \cdot \mathrm{E}_c(N) = P_c^1(j \mid iD) \cdot P_c^1(iD) \cdot \mathrm{E}_c(N) \\ &= A_c(i)B_c(D)F_c(c_{iD})\frac{B_{cD}(j)F_{cD}(c_{ij})}{\sum_{h\in D} B_{cD}(h)F_{cD}(c_{ih})}. \end{aligned} \tag{2.210}$$

It should be clear from (2.210) that these representations do not generally correspond to simple gravity models. To gain additional insight, it is of interest to consider the *conditional interaction probabilities*, $p_c(j \mid i)$, derivable from (2.210). In particular, if for any fixed cluster, $D \in \boldsymbol{D}_c$, and destination, $j \in D$, we now let H denote an arbitrary destination cluster in $\boldsymbol{D}_c$ with destinations, $h, k \in H$ (and recall that $\boldsymbol{D}_c$ partitions J), then it follows from (1.30), (2.5), and (2.210) above that for each origin, $i \in I$,

$$\begin{aligned} p_c(j \mid i) &= \frac{P_c^1(ij)}{\sum_{H\in \boldsymbol{D}_c}\sum_{h\in H} P_c^1(ih)} = \frac{\mathrm{E}_c(N_{ij})}{\sum_{H\in \boldsymbol{D}_c}\sum_{h\in H} \mathrm{E}_c(N_{ih})} \\ &= \frac{B_c(D)F_c(c_{iD})}{\sum_{H\in \boldsymbol{D}_c} B_c(H)F_c(c_{iH})} \cdot \frac{B_{cD}(j)F_{cD}(c_{ij})}{\sum_{h\in D} B_{cD}(h)F_{cD}(c_{ih})}, \end{aligned} \tag{2.211}$$

since the last term on the first line equals

$$\frac{B_c(D)F_c(c_{iD})[B_{cD}(j)F_{cD}(c_{ij})/\sum_{k\in D} B_{cD}(k)F_{cD}(c_{ik})]}{\sum_{H\in \boldsymbol{D}_c} B_c(H)F_c(c_{iH})[\sum_{h\in H} B_{cH}(h)F_{cH}(c_{ih})/\sum_{k\in H} B_{cH}(k)F_{cH}(c_{ik})]}.$$

To interpret representation (2.211), observe that by definition, $P_c^1(iD) = P_c^1(D \mid i)P_c^1(i)$. Hence [again by expression (1.30)] it follows that if we define the conditional cluster probabilities for actors at i by

$$p_c(D \mid i) = P_c^1(iD)/P_c^1(i) \tag{2.212}$$

and let $p_c(j \mid iD)$ $[= P_c^1(j \mid iD)]$ denote the corresponding conditional destination probabilities given the choice of cluster D by an actor at i, then it follows from (2.207) that the conditional interaction probabilities, $p_c(j \mid i)$, in (2.211) can be equivalently written as

$$p_c(j \mid i) = P_c^1(ij)/P_c^1(i) = p_c(D \mid i) \cdot p_c(j \mid iD), \tag{2.213}$$

where the respective conditional probabilities, $p_c(D \mid i)$ and $p_c(j \mid iD)$, are seen to correspond to the first and second stages of the hierarchical interaction process. Hence a comparison of (2.211) and (2.213) shows that the conditional interaction probabilities, $p_c(j \mid i)$, are is this case representable as a *product* of the following two (conditional) gravity models, each corresponding to one stage of the choice process:

$$p_c(D \mid i) = \frac{B_c(D)F_c(c_{iD})}{\sum_{H \in \boldsymbol{D}_c} B_c(H)F_c(c_{iH})}, \tag{2.214}$$

$$p_c(j \mid iD) = \frac{B_{cD}(j)F_{cD}(c_{ij})}{\sum_{h \in D} B_{cD}(h)F_{cD}(c_{ih})}. \tag{2.215}$$

As an explicit illustration of this general class of hierarchical models, we now consider the following *exponential* model [see Fotheringham (1986)], in which it is postulated that opportunities differ only in terms of spatial location, and are otherwise homogeneous. Here it is convenient to begin with the *destination level* of the interaction process. In particular, within each cluster, $D \in \boldsymbol{D}_c$, it is postulated that the cluster-specific destination function, B_{cD}, and deterrence function, F_{cD}, are given respectively by

$$B_{cD}(j) = 1, \tag{2.216}$$

$$F_{cD}(c_{ij}) = \exp[-\theta^t c_{ij}]. \tag{2.217}$$

where the cost sensitivity vector, θ, is assumed to be positive, and where the unit value in (2.216) reflects the homogeneity property of all destinations (opportunities), $j \in J$. Hence the *conditional destination probabilities* in (2.215) now take the explicit form:

$$p_c(j \mid iD) = \frac{\exp[-\theta^t c_{ij}]}{\sum_{h \in D} \exp[-\theta^t c_{ih}]}. \tag{2.218}$$

Next, with respect to the *cluster level* of the interaction process, one may postulate that (in view of the homogeneity assumption) the appropriate attraction weight, $B_c(D)$, for each cluster, $D \in \boldsymbol{D}_c$, is a simple function of the number, $|D|$, of destinations in D, and in particular [following Fotheringham (1986)] that

$$B_c(D) = |D|^\beta, \tag{2.219}$$

where β is a positive scalar. Finally, the appropriate cluster-level deterrence function, $F_c(D)$, is postulated to depend both on separation factors common to all destinations in D, and on separation factors specific to each destination in D. With respect to the former, it is assumed that for each origin, $i \in I$, the interaction costs common to all destinations in cluster, $D \in \boldsymbol{D}_c$, are representable by the *average separation* between i and D, as defined by

$$s(c_{iD}) = |D|^{-1} \sum_{j \in D} c_{ij}. \tag{2.220}$$

In addition, there are postulated to exist separation factors specific to each destination in cluster D, which are taken to be (additively) representable by the Hansen-type accessibility of origin i to cluster D, as defined in terms of (2.216) and (2.217) by

$$a(c_{iD}) = \sum_{j \in D} B_{cD}(j) F_{cD}(c_{ij}) = \sum_{j \in D} \exp[-\theta^t c_{ij}]. \tag{2.221}$$

With these concepts, it is then postulated that the appropriate deterrence function, F_c, in (2.204) is given by the following composite function of accessibility and average separation:

$$F_c(c_{iD}) = a(c_{iD})^{1-\sigma} \exp[-\lambda^t s(c_{iD})], \tag{2.222}$$

where λ is a positive vector and where $0 \leq \sigma \leq 1$. [Note also that (2.222) is expressible as a multivariate exponential deterrence function (2.121) of the form, $F_c(c_{iD}) = \exp\{(1-\sigma)\log[a(c_{iD})] - \lambda^t s(c_{iD})\}$]. Hence, the *conditional cluster probabilities* in (2.214) now take the explicit form:

$$p_c(D \mid i) = \frac{|D|^\beta a(c_{iD})^{1-\sigma} \exp[-\lambda^t s(c_{iD})]}{\sum_{H \in \boldsymbol{D}_c} |H|^\beta a(c_{iH})^{1-\sigma} \exp[-\lambda^t s(c_{iH})]}. \tag{2.223}$$

Finally, observe from (2.218) and (2.223), together with the definition of $a(c_{iD})$ in (2.221), that the conditional interaction probabilities in (2.211) are in this case given by:

$$p_c(j \mid i) = \frac{|D|^\beta a(c_{iD})^{1-\sigma} \exp[-\lambda^t s(c_{iD})]}{\sum_{H \in \boldsymbol{D}_c} |H|^\beta a(c_{iH})^{1-\sigma} \exp[-\lambda^t s(c_{iH})]} \cdot \frac{\exp[-\theta^t c_{ij}]}{a(c_{iD})}. \tag{2.224}$$

Hence, with these assumptions, the conditional interaction probabilities in (2.224) are seen to be representable by a *nested logit model* with an 'inclusive value' given by the accessibility measure, $a(c_{iD})$, appearing in both factors of (2.224). This accessibility measure thus constitutes a type of linkage between the two stages of the hierarchical choice process [as discussed in more detail in Fotheringham (1986) and in Fotheringham and O'Kelly (1989, Section 4.8].

2.5.3 Interaction Processes with Random Destination Sets

The hierarchical model above represents a simple type of interaction behavior in which the interaction process is decomposed into the selection of a relevant set of destinations, followed by the choice of a specific destination within that set. This basic concept is far more general however. Even within the 'clustering' framework above, one can consider more general types of behavior in which the perception of destination clusters need not result in well-defined partitions of the basic destination set, J. For example, while the notion of an urban 'central business district' (CBD) is based loosely on a perceived clustering of business establishments, the actual boundary of a given CBD is often quite arbitrary, and hence might more accurately be viewed as a 'fuzzy set' in which the degree of perceived membership of any establishment varies across both establishments and actors [Fotheringham (1986) and Fotheringham and O'Kelly (1989, p.77)].

These observations suggest that a more powerful modeling approach would be to treat each possible subset of destinations as a potentially relevant 'destination set' in the first stage. With this in mind we now define the relevant class, $\boldsymbol{D}$, of potential *destination sets* to include all nonempty subsets of J, i.e.,

$$\boldsymbol{D} = \{D \subseteq J : D \neq \emptyset\}. \tag{2.225}$$

In this more general setting, one can still view interactions as involving an implicit two-stage process in which a relevant destination set, $D \in \boldsymbol{D}$, is first selected (or perceived) by an actor at i, and then a specific destination, $j \in D$, is chosen. To model such two-stage processes in a simple way, we shall assume that J contains at least three distinct destinations (i.e., $|J| \geq 3$) and, in addition, shall assume that all destination sets in $\boldsymbol{D}$ are possible. More precisely, if Ω is defined as in Section 2.5.2 above, then this class of two-stage interaction processes can be formalized as follows [see also Definition 4.13(i) in Chapter 4]:

Definition 2.27 An interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω is designated as a *regular two-stage interaction process* if and only if $|J| \geq 3$ and $P_c(\omega) > 0$ for all $\omega \in \Omega$.

Note that each spatial interaction pattern, $s = (i_r j_r : r = 1, \ldots, n) \in S$, corresponds to a measurable event in Ω, and hence that the spatial interaction process generated by $\boldsymbol{P}$ on S [i.e., with spatial interaction probabilities $P_c(s)$] is well defined. With this observation, we now have the following notion of *independence* for regular two-stage processes, which is somewhat weaker than that for the hierarchical processes above [see also Definition 4.13(ii) in Chapter 4]:

Definition 2.28 A regular two-stage process, $\boldsymbol{P}$, on Ω is designated as an *independent two-stage interaction process* if and only if the spatial interaction process generated by $\boldsymbol{P}$ on S satisfies axioms A1 and A2.

Given this class of independent two-stage processes, we next consider explicit representational axioms for such processes. To do so, observe first that the set-level stage of these processes is formally parallel to the cluster-level stage of the hierarchical interaction processes above, with the cluster partitions, $\boldsymbol{D}_c$, for each separation configuration, $c \in C$, now replaced by the more general domain, $\boldsymbol{D}$, of possible destination sets. Thus one can easily define appropriate versions of separation dependence (A3) both with respect to the choice of destination sets, $D \in \boldsymbol{D}$, and individual destinations, $j \in D$. However, since the relevant domain, $\boldsymbol{D}$, of destination sets is now *independent* of the underlying separation configuration, $c \in C$, the types of limitations observed above for cluster partitions, $\boldsymbol{D}_c$, are no longer present. Hence it is possible to consider stronger behavioral theories in this setting which ensure the existence of *relatively-invariant* representations of mean interaction frequencies. With this in mind, we now develop a class of behavioral theories based on the types of local proportionality and separability properties in axioms A6 and A8 above. To do so, let us begin by observing that if the j-domain of all destination sets in $\boldsymbol{D}$ containing j is denoted by

$$\boldsymbol{D}_j = (D \in \boldsymbol{D} : j \in D), \tag{2.226}$$

then for any two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ on Ω it follows by definition that for all $ij \in I \times J$ and $c \in C$,

$$P_c^1(ij) = \sum_{D \in \boldsymbol{D}_j} P_c^1(iD) P_c^1(j \mid iD). \tag{2.227}$$

[Note in particular that, in contrast to (2.207) above, it is no longer true that each destination-level event, (ij), implies a unique set-level event (iD), since the j-domain, $\boldsymbol{D}_j$, contains many different sets.] Hence for each independent two-stage interaction process, $\boldsymbol{P}$, it follows at once from Definition 2.28 together with (1.30) that [as a parallel to (2.209) above] all *mean spatial interaction frequencies*, $\mathrm{E}_c(N_{ij})$ have the general representation [see also expression (4.522) in Chapter 4]:

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \mathrm{E}_c(N) P_c^1(ij) = \mathrm{E}_c(N) P_c^1(i) P_c^1(j \mid i) = \mathrm{E}_c(N_i) P_c^1(j \mid i) \\ &= \mathrm{E}_c(N_i) \sum_{D \in \boldsymbol{D}_j} P_c^1(D \mid i) P_c^1(j \mid iD). \end{aligned} \tag{2.228}$$

Given this representation of $\mathrm{E}_c(N_{ij})$ in terms of the conditional probabilities, $P_c^1(D \mid i)$ and $P_c^1(j \mid iD)$, it is again convenient to adopt the following simpler notation paralleling Section 2.5.2 above. In particular, if for each process, $\boldsymbol{P}$, we denote the corresponding *conditional set probabilities* by $p_c(D \mid i) = P_c^1(D \mid i)$, and denote the *conditional destination probabilities* by $p_c(j \mid iD) = P_c^1(j \mid iD)$, then each mean spatial interaction frequency, $\mathrm{E}_c(N_{ij})$ is equivalently representable in terms of these conditional probabilities as

$$\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(N_i) p_c(j \mid i) = \mathrm{E}_c(N_i) \sum_{D \in \boldsymbol{D}_j} p_c(D \mid i) p_c(j \mid iD). \tag{2.229}$$

Given this representation of mean interaction frequencies, it thus suffices to develop appropriate local axioms for conditional set probabilities and conditional destination probabilities, respectively. Turning first to the *conditional set probabilities*, we begin by observing that in many behavioral processes of set selection, it is not possible to separate the spatial and nonspatial attributes of destination sets in a simple multiplicative way [as in

(2.205) above]. In particular, recall from expressions (2.221) and (2.223) in the hierarchical spatial interaction process illustrated above, that this separation at the cluster level was achieved only by postulating that all opportunities were ***homogeneous*** with respect all nonspatial attributes. Indeed, if $B_{cD}(j)$ and $F_{cD}(c_{ij})$ in (2.216) and (2.217), respectively, are not independent of D, then it is clear from the definition of accessibility in (2.221) that the cluster-level deterrence function, F_c, in (2.222) is no longer solely a function of separation attributes, c_{iD}, but also depends on nonspatial attributes of D [as reflected, for example, by the values of $B_{cD}(j)$, $j \in D$]. Hence in this case, F_c is seen to be a *D-dependent function* of the form, F_{cD}, and it is clear that no simple multiplicative separation of spatial and nonspatial properties of D is possible. [An additional example of this type of nonseparability is illustrated by the 'limited information' example given below.] With these observations in mind, it is of interest to consider more general types of behavioral axioms at the set level which allow for such nonseparabilities. To motivate one such axiom, suppose that the conditional set probabilities, $p_c(D \mid i)$, in (2.229) are of the general form

$$p_c(D \mid i) = \frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})}, \tag{2.230}$$

where the *D-deterrence function*, F_D, now reflects the types of nonseparability effects discussed above. Then observe that for any origins, $i, g \in I$, destination sets, $D, H \in \boldsymbol{D}$, and separation configurations, $c, c' \in C$, with $c_{iD} = c'_{gD}$ and $c_{iH} = c'_{gH}$, we must have

$$\frac{p_c(D \mid i)}{p_c(H \mid i)} = \frac{F_D(c_{iD})}{F_H(c_{iH})} = \frac{F_D(c'_{gD})}{F_H(c'_{gH})} = \frac{p_{c'}(D \mid g)}{p_{c'}(H \mid g)}. \tag{2.231}$$

Hence the conditional set probabilities, $p_c(D \mid i)$, are seen to satisfy the following version of the *destination proportionality axiom* (A6):

A6.1. (Conditional Set Proportionality) *For all origins, $i, g \in I$, destination sets, $D, H \in \boldsymbol{D}$, and separation configurations, $c, c' \in C$,*

$$(c_{iD} = c'_{gD}, c_{iH} = c'_{gH}) \Rightarrow \frac{p_c(D \mid i)}{p_c(H \mid i)} = \frac{p_{c'}(D \mid g)}{p_{c'}(H \mid g)}. \tag{2.232}$$

Conversely, this conditional set proportionality condition can be shown to imply the existence of *D-deterrence functions*, F_D, satisfying (2.230) above [as in Proposition 4.14 of Chapter 4 below]. Hence we obtain the following characterization of this type of nonseparable representation:

Representation of Conditional Set Probabilities. *For any independent two-stage interaction process, $\boldsymbol{P}$, there exists a family of D-deterrence functions, $\{F_D : D \in \boldsymbol{D}\}$, satisfying* (2.230) *for all $iD \in I \times \boldsymbol{D}$ and $c \in C$ if and only if $\boldsymbol{P}$ satisfies conditional set proportionality* (A6.1).

Turning next to the *conditional destination probabilities* for process $\boldsymbol{P}$, we again start by considering the destination level of the hierarchical spatial interaction process developed in expressions (2.216) through (2.224). In particular, observe that [unlike the general representation in (2.211)] the destination function, B_{cD}, and the destination-level deterrence function, F_{cD}, are not only independent of the separation configuration, c, but are also independent of the particular destination set, D. [A second illustration of this type of independence property is given in the 'limited information' example below.] To characterize this type of independence, we begin by observing that if conditional destination probabilities are representable in terms of a *single destination function*, B, and *deterrence function*, F, by

$$p_c(j \mid iD) = \frac{B(j)F(c_{ij})}{\sum_{h \in D} B(h)F(c_{ih})}, \tag{2.233}$$

then it follows at once that for all $i, g \in I$, $D, H \in \boldsymbol{D}$, $j, h \in D \cap H$, and $c, c' \in C$ with $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$,

$$\frac{p_c(j \mid iD)}{p_c(h \mid iD)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{p_{c'}(j \mid gH)}{p_{c'}(h \mid gH)}. \tag{2.234}$$

Hence the conditional destination probabilities, $p_c(j \mid iD)$, are seen to satisfy an appropriate version of the *destination proportionality axiom* (A6). In addition, it also follows from (2.231) that for any configurations, c and c', with $c_{ij} = c_{ih}$ and $c'_{ij} = c'_{ih}$, we must have

$$\frac{p_c(j \mid iD)}{p_c(h \mid iD)} = \frac{B(j)}{B(h)} = \frac{p_{c'}(j \mid iH)}{p_{c'}(h \mid iH)}, \tag{2.235}$$

for all origins, $i, g \in I$, and destination sets, $D, H \in \boldsymbol{D}$ with $j, h \in D \cap H$. Hence, conditional destination probabilities are also seen to satisfy an appropriate version of the *destination separability axiom* (A8). More formally, destination choice probabilities must satisfy the following conditional versions of these two axioms:

A6.2. (Conditional Destination Proportionality) *For all origins, $i, g \in I$, separation configurations, $c, c' \in C$, destination sets, $D, H \in \boldsymbol{D}$, and destinations, $j, h \in D \cap H$,*

$$(c_{ij} = c'_{gj}, c_{ih} = c'_{gh}) \Rightarrow \frac{p_c(j \mid iD)}{p_c(h \mid iD)} = \frac{p_{c'}(j \mid gH)}{p_{c'}(h \mid gH)}. \tag{2.236}$$

A8.2. (Conditional Destination Separability) *For all origins, $i \in I$, separation configurations, $c, c' \in C$, destination sets, $D \in \boldsymbol{D}$, and destinations, $j, h \in D$,*

$$(c_{ij} = c_{ih}, c'_{ij} = c'_{ih}) \Rightarrow \frac{p_c(j \mid iD)}{p_c(h \mid iD)} = \frac{p_{c'}(j \mid iD)}{p_{c'}(h \mid iD)}. \tag{2.237}$$

Conversely, these axioms can be shown to imply the existence of a destination function and deterrence function satisfying (2.233) [as in Proposition 4.15 of Chapter 4]. Hence we now have the following characterization of conditional destination probabilities for independent two-stage interaction processes:

Representation of Conditional Destination Probabilities. *For any independent two-stage interaction process, $\boldsymbol{P}$, there exists a positive destination function, B, and deterrence function, F, satisfying* (2.233) *for all $i \in I$, $D \in \boldsymbol{D}$, $j \in D$, and $c \in C$, if and only if $\boldsymbol{P}$ satisfies both conditional destination proportionality* (A6.2) *and conditional destination separability* (A8.2).

Finally, by combining (2.229), (2.230) and (2.233) we may conclude that the desired representation of *mean interaction frequencies*, $\mathrm{E}_c(N_{ij})$ for independent spatial interaction processes, $\boldsymbol{P}$, satisfying axioms A6.1, A6.2, and A8.2 is given by

$$\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(N_i) \sum_{D \in \boldsymbol{D}_j} [\frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})} \cdot \frac{B(j)F(c_{ij})}{\sum_{h \in D} B(h)F(c_{ih})}]. \quad (2.238)$$

This general representation provides framework within which a wide variety of behavioral models can be constructed involving *random destination sets*. As one immediate illustration, it is of interest to observe that the hierarchical decision process in Section 2.5.2 above can be derived as a limiting case within this framework. In particular, a comparison of (2.207) and (2.227) shows that each hierarchical interaction process with cluster partitions, $\{\boldsymbol{D}_c : c \in C\}$, is the limiting case of (2.227) in which all probability mass for the set-level interaction probabilities, $p_c(iD) = P_c^1(iD)$, on $I \times \boldsymbol{D}$ is concentrated on the subset, $I \times \boldsymbol{D}_c$, for each $c \in C$. Since all probabilities, $p_c(iD)$, are assumed to *positive* for regular two-stage processes, such hierarchical processes are not properly within this class. However, it should be clear that hierarchical processes can be approximated to any degree of precision by processes within this class [by simply treating the positive event probabilities $p_c(iD)$ with $D \notin \boldsymbol{D}_c$ as very small]. The advantage of this approach to hierarchical interaction processes is that one can develop behavioral axioms characterizing *configuration-free* representations, as in (2.238) above. In particular, if axioms A6.1, A6.2, and A8.2 are postulated to hold for regular two-stage processes 'very close' to a given hierarchical process, then simple continuity hypotheses are sufficient to insure that the resulting configuration-free property of the representations in (2.238) will continue to hold for this limiting process.

To illustrate additional types of explicit behavioral models which can be developed within this framework, it is convenient [as in (2.211) above] to focus on the corresponding representation of conditional interaction probabilities, $p_c(j \mid i)$, which are seen from (2.228) and (2.238) to be of the form

$$\begin{aligned} p_c(j \mid i) &= \mathrm{E}_c(N_{ij})/\mathrm{E}_c(N_i) \\ &= \sum_{D \in \boldsymbol{D}_j} \left[\frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})} \cdot \frac{B(j)F(c_{ij})}{\sum_{h \in D} B(h)F(c_{ih})} \right]. \end{aligned} \tag{2.239}$$

Given this general representation of conditional interaction probabilities, we now illustrate one explicit type of interaction behavior with random destination sets, based on limited information considerations:

(A) A Limited-Information Model of Set Formation

For each origin, $i \in I$, the set-level choice probability, $p_c(D \mid i)$, can be interpreted as the probability that D constitutes the relevant set of destination choice alternatives considered by an actor at origin i. Hence the conditional distribution, $p_c(\cdot \mid i)$, on the domain, $\boldsymbol{D}$, is often designated as a *choice-set generating process* [see for example Manski (1977)]. Explicit models of such processes have been developed in a variety of choice contexts. For example, Ben-Akiva and Swait (1987) have developed a simple model of choice-set generation for mode choice by commuters. In the context of shopping behavior, a model involving variable choice sets has been developed by Fotheringham (1989). For our present purposes, however, it is convenient to illustrate choice-set generation in terms of a 'limited-information' model which is based on the labor-migration model constructed by Smith and Slater (1981).

In particular, we now consider a *labor-migration process* in which *workers*, $\alpha \in A_i$, at origin locations, $i \in I$, receive information about *job vacancies*, $\beta \in B_j$, at various destination locations, $j \in J$, and base their labor migration decisions on this information. Hence the relevant *destination set*, D, for any worker, α, is taken to include all those destinations where α has obtained information about job opportunities. In this context, the relevant separation profile, c_{ij}, between each origin, i, and destination, j, is assumed to include (positive, extensively measurable) *information costs*, c_{ij}^1, and *migration costs*, c_{ij}^2, which may influence these migration decisions. In particular, it is assumed that probable information flows between origins and destinations are influenced by the information-cost components, c_{ij}^1, and that probable migration decisions are influenced by the migration-cost components, c_{ij}^2. Turning first to information flows, it is postulated [following Smith and Slater (1981)] that the probability, $\rho_c(ij)$, that information about a job vacancy at j will be received by worker at i is of the exponential form,

$$\rho_c(ij) = p_j \exp[-\beta^t c_{ij}^1], \tag{2.240}$$

where p_j denotes a positive *job-information emission probability* at each destination j, and where the *information-cost sensitivity* parameter vector, β, is also assumed to be positive, so that $0 < \rho_c(ij) < 1$. In addition, it is postulated that information flows are statistically independent, so that for any destination set, $D \in \boldsymbol{D}$, the conditional probability, $p_c(D \mid i)$, that information from exactly those destinations in D will be received at origin i is given [as in Smith and Slater (1981)] by the product

$$p_c(D \mid i) = \prod_{j \in D} \rho_c(ij) \cdot \prod_{j \in J-D} [1 - \rho_c(ij)]. \tag{2.241}$$

This probability can be rewritten in a more convenient form for our purposes as follows. If for each (iD)-separation profile, $c_{iD} = (c_{ij} : j \in D)$, we define the *D-deterrence function*, F_D, by

$$\begin{aligned} F_D(c_{iD}) &= \prod_{j \in D} (\rho_c(ij)/[1 - \rho_c(ij)]) \\ &= \prod_{j \in D} (p_j \exp(-\beta^t c_{ij}^1)/[1 - p_j \exp(-\beta^t c_{ij}^1)]), \end{aligned} \tag{2.242}$$

then $p_c(D \mid i)$ can be written as

$$\begin{aligned} p_c(D \mid i) &= \prod_{j \in J} [1 - \rho_c(ij)] \cdot \prod_{j \in D} (\rho_c(ij)/[1 - \rho_c(ij)]) \\ &= \lambda_c(i) F_D(c_{iD}), \end{aligned} \tag{2.243}$$

where $\lambda_c(i) = \prod_{j \in J}[1 - \rho_c(ij)]$. Hence (2.243) together with the summation identity

$$1 = \sum_{D \in \boldsymbol{D}} p_c(D \mid i) = \lambda_c(i) \sum_{D \in \boldsymbol{D}} F_D(c_{iD})$$

implies that

$$p_c(D \mid i) = \frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})} \tag{2.244}$$

and we may conclude that the *conditional set probabilities*, $p_c(D \mid i)$, are of the form (2.230).

Turning next to the actual migration decision of workers at origin, i, with information about job vacancies at some nonempty set of destinations, D, we focus on the destination choices of such workers *given the decision to migrate*, i.e., we focus on the *conditional interaction probabilities*, $p_c(j \mid i)$, rather than the mean migration frequencies, $\mathrm{E}_c(N_{ij})$, $j \in D$. In this context, it is postulated [as in Smith and Slater (1981)] that the *conditional destination probabilities*, $p_c(j \mid iD)$, depend exponentially on the migration-cost component, c_{ij}^2, of c_{ij}, and in particular, are of the form

$$p_c(j \mid iD) = \frac{w_j \exp[-\theta^t c_{ij}^2]}{\sum_{h \in D} w_h \exp[-\theta^t c_{ih}^2]}, \tag{2.245}$$

where the *migration-cost sensitivity vector,* θ, and *destination-attractiveness weights,* $w_j > 0$, for each $j \in D$, are parameters to be estimated. Under these assumptions, the resulting conditional destination probabilities, $p_c(j|iD)$, are seen to be of the form (2.233) with $B(j) = w_j$ and $F(c_{ij}) = \exp[-\theta^t c_{ij}^2]$. Hence the resulting *conditional interaction probabilities,*

$$p_c(j \mid i) = \sum_{D \in \boldsymbol{D}} [\frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})} \cdot \frac{w_j \exp[-\theta^t c_{ij}]}{\sum_{h \in D} w_h \exp[-\theta^t c_{ih}]}], \qquad (2.246)$$

together with (2.242) provide an explicit illustration of an independent two-stage interaction process.

(B) A Prominence Model of Destination Choice

To motivate our final behavioral illustration, we begin by observing that (except in degenerate cases) the interaction probabilities, $p_c(ij)[\equiv P_c^1(ij)]$, arising from independent two-stage interaction processes always violate the independence-of-irrelevant-alternatives (IIA) property of gravity models [i.e., violate (2.94) above, where ratios of (size one) pattern probabilities, $P_c(ij)$, are by definition equal the corresponding ratios of interaction probabilities, $p_c(ij)$, for independent interaction processes]. Indeed, the restrictive nature of this IIA-property has served as the motivating force for the construction of a wide variety of probabilistic choice models [as discussed for example in Luce (1977)]. From a conceptual viewpoint, the IIA-property implies a strong form of *context-insensitive* behavior by choice makers, in the sense that relative likelihoods of choices among pairs of alternatives are independent of the presence or absence of any other alternatives. Hence the variety of models which have been developed to overcome this restriction can each be said to involve some type of *context-sensitive* behavior. With these general observations in mind, our final example illustrates a type of context-sensitive behavior which may be present even in cases where no variability of choice sets is involved. Hence, within the framework above, it is appropriate to develop this behavioral model in terms of *conditional destination probabilities.*

The present notion of context sensitivity focuses on dissimilarity relationships among the alternative destinations within any given destination set, $D \subseteq J(= B)$. In particular, it is postulated that the dissimilarity of a given interaction opportunity, $j \in D$, from all other opportunities in D may enhance its *prominence* (as perceived by spatial actors), and thereby influence the probability that j is chosen. The most common examples of such effects involve the degree of 'substitutability' among possible interaction opportunities. For example, if there is only one chinese restaurant in town, then this restaurant may be said to exhibit certain distinguishing features which enhance its prominence relative to all other dining opportunities. Hence (other things being equal), one may expect this restaurant to be more popular than it would be if other chinese restaurants were also present. Such prominence effects have long been recognized to be a major factor in consumer choice behavior, and have been incorporated in a variety of spatial choice models [as discussed for example in Borgers and Timmermans (1987) and Fotheringham (1988)]. In particular, the notion of perceived prominence is usually taken to be reflected by some measure of the *dissimilarity* between alternatives with respect to some relevant set of attributes. More formally, if for each opportunity, $j \in J$, the vector of (extensively measurable) attributes which are relevant for similarity comparisons is denoted by $z_j = (z_{jw} : w \in W)$, and if the absolute deviations between these vectors are denoted by $|z_j - z_h| = (|z_{jw} - z_{hw}| : w \in W)$, then the *prominence* of j in D can be taken to be a function of its deviations from all other opportunities in D, i.e., a function of the form $\Phi_D(|z_j - z_h| : h \in D/j)$, where again $D/j = D - \{j\}$. One example of such a function, which focuses on the *average absolute deviations* between opportunities in D (with $|D| \geq 2$), is the following [based on Batsell (1981)]:

$$\Phi_D(|z_j - z_h| : h \in D/j) = \exp[\sum_{w \in W} \theta_w (\frac{1}{|D| - 1} \sum_{h \in D/j} |z_{jw} - z_{hw}|)], \quad (2.247)$$

where $\theta_w > 0$, $w \in W$. A variety of other functions involving alternative distance or deviation concepts with respect to general destination attributes are discussed in Borgers and Timmermans (1987) and Fotheringham (1988). Within our present formal framework, all such dissimilarity measures are here taken to reflect some implicit notion of *separation* between destinations. Hence each attribute-deviation value involved in such measures is here treated as a component of the relevant *separation profile* between destination pairs. For example, if for each destination attribute, $w \in W$, in (2.247) we let $c_{jh}^w = |z_{jw} - z_{hw}|$, then the relevant *(jh)-separation profile* for this case is given by $c_{jh} = (c_{jh}^w : w \in W)$. Hence if we designate the array of separation profiles

$$c_{jD} = (c_{jh} : h \in D/j) \quad (2.248)$$

as the *(jD)-separation profile*, then (2.247) can be rewritten in terms of this notation as follows:

$$\Phi_D(c_{jD}) = \exp[\sum_{w \in W} \theta_w (\frac{1}{|D| - 1} \sum_{h \in D/j} c_{jh}^w)]. \quad (2.249)$$

Note in particular that such measures of prominence are meaningless for destination sets with only one element. Hence if for each positive integer, $k \leq |J|$, we denote the class of destination sets with at least k elements by

$$\boldsymbol{D}_k = (D \in \boldsymbol{D} : |D| \geq k), \quad (2.250)$$

then we shall henceforth only be concerned with destination sets, $D \in \boldsymbol{D}_2$.

In this context, it is typically postulated that such prominence effects tend to enhance the likelihood that opportunity j will be chosen in context D. Hence, as a modification of the basic gravity model form in (2.233), we now consider the following class of context-sensitive gravity models, which we designate simply as *prominence models* of conditional destination probabilities:

$$p_c(j \mid iD) = \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\sum_{h \in D} \Phi_D(c_{hD})B(h)F(c_{ih})}, \qquad j \in D \in \boldsymbol{D}_2. \qquad (2.251)$$

Given this class of prominence models [which includes the spatial choice models of Batsell (1981), Meyer and Eagle (1981,1982), Borgers and Timmermans (1987), and Fotheringham (1983a,1986,1988) among others] we next explore the broader question of characterizing the types of spatial interaction behavior representable by such models. In doing so, we shall employ the basic elements of a *prominence theory of choice* developed by Smith and Yu (1982), which focuses precisely on those types of prominence effects which are attributable to dissimilarities between choice alternatives. To formalize the main concepts within our present framework, we again identify opportunities with distinct destinations, i.e., set $J = B$, and observe that in the present setting the relevant measures of separation between actors at origins, $i \in I$, and opportunities, $j \in J$, may in general be quite different than those involved in dissimilarity comparisons between opportunities, $j, h \in J$. To allow for this possibility, we again let $(c^k : k \in K)$ denote the relevant separation measures between actors and opportunities, but now let $(c^w : w \in W)$ denote the relevant separation measures for dissimilarity comparisons (where possibly $K = W$). Hence each *separation configuration*, is now assumed to be of the form, $c = [(c_{ij} : ij \in I \times J), (c_{jh} : j, h \in J)]$, where *$(ij)$-separation profiles* are again of the form, $c_{ij} = (c_{ij}^k : k \in K)$, but *$(jh)$-separation profiles* are of the form, $c_{jh} = (c_{jh}^w : w \in W)$, as for example in (2.249) above. In this context, we impose the additional symmetry condition on each separation measure, c_w, that $c_{jh}^w = c_{hj}^w$ for all $w \in W$ and $j, h \in J$. [Observe in particular that symmetry is satisfied by all absolute deviation measures, as in (2.248), as well as most standard measures of distance]. Hence if the class of separation profile values for origin-destination pairs is again denoted by V, and if we now denote the relevant class of dissimilarity profile values for destination pairs by U, then the relevant *configuration class*, C, in the present context is taken to be of the form

$$C = \{c \in V^{I \times J} \times U^{J \times J} : c_{jh} = c_{hj}, j, h \in J\}. \qquad (2.252)$$

With these conventions, it is next hypothesized that the prominence of each destination, $j \in D$, varies directly with its degree of separation (dissimilarity) from all other destinations in D, as reflected by the separation values in c_{jD}. Moreover, it is hypothesized that prominence of j in D does not depend on the particular labeling of other destinations in D, and hence is a symmetric function of the separation profiles in c_{jD}. To formalize these hypotheses, observe first from the definition of C in (2.252) above that for any destination set $D \in \boldsymbol{D}_2$, the set of possible values of (jD)-separation profiles for destinations $j \in D$ is given by $U^{|D|-1}$. Hence for any two arrays of dissimilarity profiles, $x = (x_1, \ldots, x_{|D|-1})$ and $y = (y_1, \ldots, y_{|D|-1})$ in $U^{|D|-1}$, we now say that x and y are *symmetrically equivalent* (written $x =_s y$) when they differ only by the ordering of their component value [for example, $(1,2,3) =_s (3,1,2)$]. Next, x is said to *symmetrically dominate* y (written $x \geq_s y$) whenever x is greater than or equal to some reordering of y, i.e., whenever $x \geq y'$ holds for some $y' =_s y$ [so that for example, $(4,5,6) \geq_s (2,6,4)$]. Similarly, if we also write $x \leq_s y \Leftrightarrow y \geq_s x$, then one may readily verify that

$$x =_s y \Leftrightarrow (x \geq_s y, x \geq_s y). \tag{2.253}$$

With these notational conventions, we now designate a positive function, Φ_D, of (jD)-separation profiles as a possible *D-prominence function* if and only if Φ_D satisfies the following *symmetric monotonicity* condition for all for profile values, $x, y \in U^{|D|-1}$:

$$x \geq_s y \Rightarrow \Phi_D(x) \geq \Phi_D(y). \tag{2.254}$$

Hence, whenever the set of dissimilarity profiles for j symmetrically dominate those for h in D, destination j is hypothesized to be at least as prominent as h in D. As one illustration, observe that the positive function in (2.249) satisfies (2.254), and hence constitutes an admissible D-prominence function in the present framework.

Given this class of prominence functions, we now seek to characterize the associated class of *prominence models* in (2.251) above. To do so, we begin by observing that for all configurations, $c \in C$, and all *binary destination sets*, $\{jh\} \in \boldsymbol{D}_2$, since $c_{j\{jh\}} = c_{jh}$ and $c_{h\{jh\}} = c_{hj}$, it follows from (2.254) together with the symmetry assumption on dissimilarity measures that

$$c_{jh} = c_{hj} \Rightarrow \Phi_{\{jh\}}(c_{jh}) = \Phi_{\{jh\}}(c_{hj}). \tag{2.255}$$

Hence in all *binary* interaction situations, our hypotheses imply that neither opportunity is more prominent than the other. In particular, this implies that if conditional destination probabilities are representable as in (2.251) then the *relative* likelihoods of interactions with either opportunity are always of the simple gravity form in (2.233). Thus for any origins, $i, g \in I$, and separation configurations, $c, c' \in C$, with $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$, we must have,

$$\frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{p_{c'}(j \mid g\{jh\})}{p_{c'}(h \mid g\{jh\})}. \tag{2.256}$$

Similarly, if $c_{ij} = c_{ih}$ and $c'_{ij} = c'_{ih}$, then we must also have

$$\frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} = \frac{B(j)}{B(h)} = \frac{p_{c'}(j \mid i\{jh\})}{p_{c'}(h \mid i\{jh\})}. \tag{2.257}$$

Hence, we see that conditional destination probabilities continue to satisfy the following *pairwise* version of the destination proportionality conditions in (2.236) and (2.237) above:

P1. (Pairwise Destination Proportionality) *For all origins, $i, g \in I$, distinct destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ih} = c'_{gj}, c_{ih} = c'_{gh}) \Rightarrow \frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} = \frac{p_{c'}(j \mid i\{jh\})}{p_{c'}(h \mid i\{jh\})}. \tag{2.258}$$

P2. (Pairwise Destination Separability) *For all origins, $i \in I$, distinct destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c_{ih}, c'_{ij} = c'_{ih}) \Rightarrow \frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} = \frac{p_{c'}(j \mid i\{jh\})}{p_{c'}(h \mid i\{jh\})}. \tag{2.259}$$

As one additional consequence of (2.255), observe that for any given origin, $i \in I$, separation configuration, $c \in C$, and distinct destinations, $j, h, d \in J$ (where we again assume that $|J| \geq 3$) it must be true that

$$\begin{aligned} \frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} &= \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c_{ij})/B(d)F(c_{id})}{B(h)F(c_{ih})/B(d)F(c_{id})} \\ &= \frac{p_c(j \mid i\{jd\}/p_c(d|i\{jd\})}{p_c(h \mid i\{hd\})/p_c(d \mid i\{hd\})}. \end{aligned} \tag{2.260}$$

Thus conditional destination probabilities representable as in (2.251) also satisfy the following *pairwise* version of independence-of-irrelevant-alternatives [also designated as the 'product rule' (Luce and Suppes 1965, Definition 25)]:

P3. (Pairwise Destination Independence) *For all origins, $i \in I$, distinct destinations, $j, h, d \in J$, and separation configurations, $c \in C$,*

$$\frac{p_c(j \mid i\{jh\})}{p_c(h \mid i\{jh\})} = \frac{p_c(j \mid i\{jd\})/p_c(d \mid i\{jd\})}{p_c(h \mid i\{hd\})/p_c(d \mid i\{hd\})}. \tag{2.261}$$

Next observe that for all larger destination sets, $D \in \boldsymbol{D}_3$, (2.251) implies that for all $c \in C$, $i \in I$, and $j, h \in D$,

$$\begin{aligned} \frac{p_c(j \mid iD)/p_c(h \mid iD)}{p_c(j \mid i\{jh\})/p_c(h \mid i\{jh\})} &= \frac{\Phi_D(c_{jD})B(j)F(c_{ij})/\Phi_D(c_{hD})B(h)F(c_{ih})}{B(j)F(c_{ij})/B(h)F(c_{ih})} \\ &= \frac{\Phi_D(c_{jD})}{\Phi_D(c_{hD})}. \end{aligned} \tag{2.262}$$

In other words, if any new alternatives are added to the binary destination set, $\{jh\}$, then the change in the relative likelihood of choosing j over h is determined entirely by the *relative prominence* of j versus h in the larger set D. As a direct consequence of (2.262), observe that for all destination pairs, $\{jh\},\{bd\} \subseteq D$, and all separation configurations, $c, c' \in C$, with $c_{jD} \geq_s c'_{bD}$ and $c_{hD} \leq_s c'_{dD}$, the symmetric monotonicity condition (2.254) implies that $\Phi_D(c_{jD})/\Phi_D(c_{hD}) \geq \Phi_D(c'_{bD})/\Phi_D(c'_{dD})$, and hence that

$$\begin{aligned}\frac{p_c(j \mid iD)/p_c(h \mid iD)}{p_c(j \mid i\{jh\})/p_c(h \mid i\{jh\})} &= \frac{\Phi_D(c_{jD})}{\Phi_D(c_{hD})} \geq \frac{\Phi_D(c'_{bD})}{\Phi_D(c'_{dD})} \\ &= \frac{p_{c'}(b \mid gD)/p_{c'}(d \mid gD)}{p_{c'}(b \mid g\{bd\})/p_{c'}(d \mid g\{bd\})}.\end{aligned} \tag{2.263}$$

Thus we see that conditional destination probabilities representable by (2.251) must satisfy the following monotonicity property with respect to prominence:

P4. (Prominence Monotonicity) *For all origins, $i, g \in I$, destination pairs, $\{jh\},\{bd\} \subseteq D \in \boldsymbol{D}_3$, and separation configurations, $c, c' \in C$, if both $c_{jD} \geq_s c'_{bD}$ and $c_{hD} \leq_s c'_{dD}$ then*

$$\frac{p_c(j \mid iD)/p_c(h \mid iD)}{p_c(j \mid i\{jh\})/p_c(h \mid i\{jh\})} \geq \frac{p_{c'}(b \mid gD)/p_{c'}(d \mid gD)}{p_{c'}(b \mid g\{bd\})/p_c(d \mid g\{bd\})}. \tag{2.264}$$

To establish a final consequence of (2.251), observe first from (2.253) and (2.254) that for all dissimilarity profile values, $x, y \in U^{|D|-1}$,

$$\begin{aligned}x =_s y &\Rightarrow (x \geq_s y, y \geq_s x) \Rightarrow [\Phi_D(x) \geq \Phi_D(y), \Phi_D(y) \geq \Phi_D(x)] \\ &\Rightarrow \Phi_D(x) = \Phi_D(y),\end{aligned} \tag{2.265}$$

so that for any destinations, $j, h, d \in D \in \boldsymbol{D}_3$, and separation configurations, $c, c', c'' \in C$, with $c_{jD} =_s c'_{jD}$, $c_{hD} =_s c''_{hD}$, and $c'_{dD} =_s c''_{dD}$, it must be true that $\Phi_D(c_{jD}) = \Phi_D(c'_{jD})$, $\Phi_D(c_{hD}) = \Phi_D(c''_{hD})$, and $\Phi_D(c'_{dD}) = \Phi_D(c''_{dD})$. Hence if for any origin, $i \in I$, these three profiles also satisfy $c_{iD} = c'_{iD} = c''_{iD}$, [so that $F(c_{ik}) = F(c'_{ik}) = F(c''_{ik})$ for all $k \in D$] then it must follow from (2.251) that

$$\begin{aligned}\frac{p_c(j \mid iD)}{p_c(h \mid iD)} &= \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\Phi_D(c_{hD})B(h)F(c_{ih})} = \frac{\Phi_D(c'_{jD})B(j)F(c'_{ij})}{\Phi_D(c''_{hD})B(h)F(c''_{ih})} \\ &= \frac{\Phi_D(c'_{jD})B(j)F(c'_{ij})}{\Phi_D(c''_{hD})B(h)F(c''_{ih})} \cdot \frac{\Phi_D(c''_{dD})B(d)F(c''_{id})}{\Phi_D(c'_{dD})B(d)F(c'_{id})} \\ &= \frac{\Phi_D(c'_{jD})B(j)F(c'_{ij})/\Phi_D(c'_{dD})B(d)F(c'_{id})}{\Phi_D(c''_{hD})B(h)F(c''_{ih})/\Phi_D(c''_{dD})B(d)F(c''_{id})} \\ &= \frac{p_{c'}(j \mid iD)/P_{c'}(d \mid iD)}{p_{c''}(h \mid iD)/P_{c''}(d \mid iD)}.\end{aligned} \tag{2.266}$$

Hence, as a parallel to the pairwise destination independence condition (P3) for binary destination sets, we see that destinations in all larger sets must satisfy the following pairwise independence property with respect to prominence effects:

P5. (Pairwise Prominence Independence) *For all* $i \in I$, $D \in \boldsymbol{D}_3$, $j, h, d \in D$, *and all separation configurations*, $c, c', c'' \in C$ *with* $c_{iD} = c'_{iD} = c''_{iD}$, $c_{jD} =_s c'_{jD}$, $c_{hD} =_s c''_{hD}$, *and* $c'_{dD} =_s c''_{dD}$,

$$\frac{p_c(j \mid iD)}{p_c(h \mid iD)} = \frac{p_{c'}(j \mid iD)/p_{c'}(d \mid iD)}{p_{c''}(h \mid iD)/p_{c''}(d \mid iD)}. \tag{2.267}$$

Given these five properties, it can be shown [as in Proposition 4.16 in Chapter 4 below] that there must exist D-prominence functions, Φ_D, satisfying (2.251) for some choice of destination function, B, and deterrence function, F. To state this result formally, we now take these properties to define the following *prominence theory* of conditional interaction behavior:

Definition 2.29 A regular two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$ is said to satisfy the *prominence theory* of conditional destination interactions if and only if $\boldsymbol{P}$ satisfies (P1, P2, P3, P4, P5).

In terms of this definition, we then have the following characterization of such prominence models for all regular two-stage processes:

Characterization of Prominence Models. *A regular two-stage interaction process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *satisfies the prominence theory of conditional destination interactions if and only if there exists a positive destination function, B, deterrence function, F, and family of D-prominence functions,* Φ_D, *for each* $D \in \boldsymbol{D}_2$ *such that* (2.251) *holds for all* $i \in I$, $c \in C$, $D \in \boldsymbol{D}_2$ *and* $j \in D$.

Note finally that the simple gravity-type conditional destination probabilities in (2.233) are also consistent with this general form of the prominence theory (since P4 is an equality for this case, and since P1, P2, P3 and P5 hold identically for all gravity models of conditional destination probabilities). With this in mind, observe that if for any destination set, $D \in \boldsymbol{D}_2$, and dissimilarity profile values, $x, y \in U^{|D|-1}$, we write $x >_s y$ whenever $x \geq y' \neq x$ for some $y' =_s y$, then as a natural strengthening of this theory, we may include in P4 the additional *strict prominence monotonicity* condition that

$$\frac{p_c(j \mid iD)/p_c(h \mid iD)}{p_c(j \mid i\{jh\})/p_c(h \mid i\{jh\})} > \frac{p_{c'}(j \mid gD)/p_{c'}(h \mid gD)}{p_{c'}(j \mid g\{jh\})/p_{c'}(h \mid i\{jh\})} \tag{2.268}$$

whenever $c_{jD} >_s c'_{bD}$ and $c_{hD} \leq_s c'_{dD}$. Under this stronger version of the prominence theory, it should be clear that the corresponding prominence functions in (2.251) must satisfy the additional *strict monotonicity* property that

$$x >_s y \Rightarrow \Phi_D(x) > \Phi_D(y) \tag{2.269}$$

for all $x, y \in U^{|D|-1}$ [as exemplified by the D-prominence function in (2.249)]. Hence we see that all interaction behavior consistent with this type of strict context sensitivity must again fail to be representable by any gravity model.

CHAPTER 3

Spatial Interaction Processes: Formal Development

3.1 Introduction

In this chapter, the basic notations of interaction processes and their associated frequency processes are developed in a formal way. We begin in Section 3.2 below with a development of certain mathematical concepts which will be employed throughout the analysis. This is followed in Section 3.3 by a development of a general probability model of interactions which focuses on the locational and frequency attributes of interaction patterns. In Section 3.4, the class of *interaction processes* discussed in Section 1.4.2 is then formalized, and the important subclass of *independent interaction processes* discussed in Section 1.4.3 is developed in detail. In Section 3.5, the corresponding concepts of *frequency processes* and *independent frequency processes* are developed, and the fundamental *Poisson characterization theorem* discussed in Section 3.6.2 is established in terms of these concepts. Finally, the application of this general result to the classes of threshold interaction processes and search processes discussed in Examples 2 and 4 of Chapter 2 are developed formally in Sections 3.7 and 3.8, respectively.

3.2 Analytical Preliminaries

Throughout this development, the following basic notation and concepts will be employed. Let Z and R denote the *integers* and *real numbers*, respectively, and let Z_+ and R_+ (Z_{++} and R_{++}) denote the corresponding subsets of nonnegative (positive) numbers. For any nonempty sets $A_1, \ldots, A_n$, let $\prod_{i=1}^n A_i = A_1 \times \ldots \times A_n$ denote the corresponding Cartesian product set, and for each integer, $k \in Z_{++}$, let A^k denote the k-fold product of a set, A, with itself. For notational simplicity, we adopt the convention that $A^0 \times B = B \times A^0 = B$ for all sets A and B. If $|A|$ denotes the *cardinality* of a set, A, then A is designated as a *singleton set* iff $|A| = 1$. Similarly, A is said to be *finite* iff $|A| < |Z| (= \infty)$, *countable* iff $|A| \leq |Z|$, and *uncountable* iff $|A| > |Z|$. For any finite set with cardinality, $|A| = k$, the class R^A, of all real-valued functions, $f : A \rightarrow R$, will always be identified with the k-dimensional Euclidean space, R^k. If $||x||$ denotes the *Euclidean norm* of each $x \in R^k$, then for each $x \in R^k$ and sequence

$(x_n : n \in Z_+)$ on R^k, we write $x_n \to x$ iff $\lim_{n\to\infty} ||x_n - x|| = 0$. For any elements $a = (a_1, \ldots, a_k)$, $b = (b_1, \ldots, b_n) \in R^k$, we write $a \leq b (a < b)$ iff $a_i \leq b_i (a_i < b_i)$ for all $i = 1, \ldots, n$, and let $(a, b) = \{x \in R^k : a < x < b\}$ and $[a, b] = \{x \in R^k : a \leq x \leq b\}$ denote the (possibly empty) *open interval* and *closed interval*, respectively, generated by a and b in R^k. Similarly, let $[a, \infty) = \{x \in R^k : a \leq x\}$ and $(-\infty, a] = \{x \in R^k : x \leq a\}$ denote the corresponding *upper half-interval* and *lower half-interval* generated by a in R^k.

3.2.1 Measurable Spaces

If for any set, X, we denote the family of all subsets of X (i.e., the *power set* of X) by $\boldsymbol{S}(X) = \{A : A \subseteq X\}$, and let $\emptyset$ denote the *empty set*, then a family of subsets, $\boldsymbol{X} \subseteq \boldsymbol{S}(X)$, is designated as a *σ-field* (or *σ-algebra*) on X iff $\boldsymbol{X}$ satisfies the following three conditions: (i) $\emptyset \in \boldsymbol{X}$, (ii) $A \in \boldsymbol{X} \Rightarrow X - A \in \boldsymbol{X}$, and (iii) $\{A_n : n \in Z_+\} \subseteq \boldsymbol{X} \Rightarrow \cup_n A_n \in \boldsymbol{X}$. Each σ-field, $\boldsymbol{X}$, on X by definition satisfies the additional conditions that (iv) $X \in \boldsymbol{X}$, (v) $\{A_n : n \in Z_+\} \subseteq \boldsymbol{X} \Rightarrow \cap_n A_n \in \boldsymbol{X}$ and (vi) $A, B \in \boldsymbol{X} \Rightarrow A - B \in \boldsymbol{X}$. Next observe that since the power set, $\boldsymbol{S}(X)$, automatically satisfies (i), (ii), and (iii), it follows that for any family of sets, $\boldsymbol{S} \subseteq \boldsymbol{S}(X)$, there always exists at least one σ-field on X containing $\boldsymbol{S}$. Hence the intersection, $\sigma(\boldsymbol{S})$, of all σ-fields on X containing $\boldsymbol{S}$ is nonempty, and is easily seen to be a σ-field [Halmos, 1950, Theorem 5.A]. By definition, $\sigma(\boldsymbol{S})$ is thus the unique smallest σ-field on X containing $\boldsymbol{S}$, and is designated as the σ-field *generated* by $\boldsymbol{S}$. For example, if $\boldsymbol{H}^k \subseteq \boldsymbol{S}(R^k)$ denotes the family of all upper half-intervals (or lower half-intervals) in R^k, then the usual *Borel σ-field*, $\boldsymbol{B}(R^k)$, on R^k is precisely the σ-field generated by $\boldsymbol{H}^k$, i.e., $\boldsymbol{B}(R^k) = \sigma(\boldsymbol{H}^k)$. Each set $B \in \boldsymbol{B}(R^k)$ is designated as a *Borel set* in R^k.

For any σ-field, $\boldsymbol{X}$, on a nonempty set, X, the pair $\langle X, \boldsymbol{X}\rangle$ is designated as a *measurable space* and each subset $A \in \boldsymbol{X}$ is designated as a *measurable set* in $\langle X, \boldsymbol{X}\rangle$. If $\boldsymbol{X}$ contains all singleton sets in X, i.e., if $\{x\} \in \boldsymbol{X}$ for all $x \in X$, then $\langle X, \boldsymbol{X}\rangle$ is said to be a *fully measurable space.* Certain spaces of this type are of special interest. If For any countable set, X, the (maximal) fully measurable space, $\langle X, \boldsymbol{S}(X)\rangle$, will always be associated with X unless otherwise stated. Similarly, the *Borel space,* $\langle R^k, \boldsymbol{B}(R^k)\rangle$, will always be associated with R^k unless otherwise stated. Next, for any measurable space, $\langle X, \boldsymbol{X}\rangle$, and index set, T, a family of subsets, $\{A_t : t \in T\} \subseteq \boldsymbol{X}$, is designated as an *$\boldsymbol{X}$-measurable partition* of X iff $\cup_{t\in T} A_t = X$ and $A_t \cap A_s = \emptyset$ for all distinct $t, s \in T$. In addition, if the index set, T, is finite (countable), then $\{A_t : t \in T\}$ is said to be a *finite (countable) partition* of X.

3.2.2 Measurable Functions

For any function, $f : X \to Y$, and subset $A \subseteq X$, the set $f(A) = \{f(x) \in Y : x \in A\}$ denotes the *image* of A under f. Similarly, for any subset $B \subseteq Y$, the (possibly empty) set $f^{-1}(B) = \{x \in X : f(x) \in B\}$ denotes the *inverse image* of B under f. If $f^{-1}(\{y\})$ is a singleton set for each $y \in Y$, then f is designated as a *bijection* from X to Y (i.e., one-to-one map of X onto Y), and the unique function, $f^{-1} : Y \to X$, defined for all $y \in Y$ by $f^{-1}(y) = x \Leftrightarrow y = f(x)$ is designated as the *inverse* of f. For any measurable spaces, $\langle X, \boldsymbol{X}\rangle$ and $\langle Y, \boldsymbol{Y}\rangle$, a function, $f : X \to Y$, is said to be $(\boldsymbol{X}, \boldsymbol{Y})$-*measurable* iff $f^{-1}(A) \in \boldsymbol{X}$ for all $A \in \boldsymbol{Y}$. Where the σ-fields $\boldsymbol{X}$ and $\boldsymbol{Y}$ are understood, we shall simply designate f as *measurable*. If Y is countable [so that $\boldsymbol{Y} = \boldsymbol{S}(Y)$ by convention], then $\boldsymbol{X}$-measurability of f is equivalent to the condition that $f^{-1}(\{y\}) \in \boldsymbol{X}$ hold for all $y \in Y$. Finally, if $Y = R^k$ [so that $\boldsymbol{Y} = \boldsymbol{B}(R^k)$], then measurability of f is equivalent to the condition that $f^{-1}([a, \infty)) \in \boldsymbol{X}$ hold for all upper half-intervals in R^k (or all lower half-intervals in R^k).

Observe next that for any pairs of σ-fields, $(\boldsymbol{X}, \boldsymbol{X}')$ on X and $(\boldsymbol{Y}, \boldsymbol{Y}')$ on Y, if $\boldsymbol{X} \subseteq \boldsymbol{X}'$ and $\boldsymbol{Y}' \subseteq \boldsymbol{Y}$, then each $(\boldsymbol{X}, \boldsymbol{Y})$-measurable function, $f : X \to Y$, must also be $(\boldsymbol{X}', \boldsymbol{Y}')$-measurable. In addition, if for any set, $A \in \boldsymbol{X}$, we designate the σ-field, $\boldsymbol{X}_A = \{A \cap B : B \in \boldsymbol{X}\}$, as the *A-restriction* of $\boldsymbol{X}$, and designate the function, $f_A : A \to Y$, defined for all $x \in A$ by $f_A(x) = f(x)$ as the *A-restriction* of f, then it follows that $(\boldsymbol{X}, \boldsymbol{Y})$-measurability of f implies $(\boldsymbol{X}_A, \boldsymbol{Y})$-measurability of f_A.

Finally, for any nonempty set, X, together with a family of measurable spaces, $\{\langle Y_t, \boldsymbol{Y}_t\rangle : t \in T\}$, and an associated family of functions, $F = \{f : X \to Y_t | t \in T\}$, it follows by definition that the σ-field, $\sigma(F) = \{f_t^{-1}(A_t) : A_t \in \boldsymbol{Y}_t, t \in T\}$, is the unique smallest σ-field on X with respect to which all functions in F are measurable [i.e., for which each function, $f_t \in F$, is $(\sigma(F), \boldsymbol{Y}_t)$-measurable]. Hence, we shall designate $\sigma(F)$ as the σ-field *generated by* F (together with the implicit σ-fields, $\boldsymbol{Y}_t$, for the range space of each function $f_t \in F$).

3.2.3 Probability Spaces

For any measurable space, $\langle X, \boldsymbol{X}\rangle$, each function, $\mu : \boldsymbol{X} \to R_+$, satisfying (i) $\mu(\emptyset) = 0$ and (ii) $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ for every countable $\boldsymbol{X}$-measurable partition, $\{A_n : n \in Z_+\}$, of X is designated as a *measure* on X, and $\langle X, \boldsymbol{X}, \mu\rangle$ is designated as the associated *measure space*. If in addition, $\mu(X) < \infty$, then μ is said to be a *finite measure*. Each finite measure, $P : \boldsymbol{X} \to [0, 1]$, with $P(X) = 1$ is designated as a *probability measure* (or *normed measure*) on X, with associated *probability space*, $\langle X, \boldsymbol{X}, P\rangle$. Each measurable set, $A \in \boldsymbol{X}$, is then designated as a *probabilizable event* in X, or simply as an *event* in X. If X is countable [so that $\boldsymbol{X} = \boldsymbol{S}(X)$], then each probability measure, P, on X is uniquely specified by its associated *probability function*, $p: X \to [0,1]$, defined for all $x \in X$ by $p(x) = P(\{x\})$ [and satisfying $\sum_{x \in X} P(x) = 1$]. In this case, the associated probability space can be written simply as $\langle X, p \rangle$.

For any probability space, $\langle X, \boldsymbol{X}, P \rangle$, and measurable space, $\langle Y, \boldsymbol{Y} \rangle$, each $(\boldsymbol{X}, \boldsymbol{Y})$-measurable function, $H : X \to Y$, is designated as a (Y-*valued*) *random variable* on $\langle X, \boldsymbol{X}, P \rangle$. Each random variable, $H : X \to Y$, induces a probability measure, $P_H : \boldsymbol{Y} \to [0,1]$, on $\langle Y, \boldsymbol{Y} \rangle$ defined for all $A \in \boldsymbol{Y}$ by $P_H(A) = P[H^{-1}(A)]$. Typically, the probability of each H-event, $A \in \boldsymbol{Y}$, is written as $P(H \in A)$. If X and Y are both countable sets, then every function $H : X \to Y$, is automatically a random variable. If $Y = R^k$, then $H : X \to R^k$ is said to be a (k-dimensional) *random vector* on $\langle X, \boldsymbol{X}, P \rangle$, and the probability measure P_H on $\langle R^K, \boldsymbol{B}(R^K) \rangle$ is uniquely determined by the family of event probabilities $\{P_H(H \geq h : h \in R^K)\}$. If $Y = Z_+$ and $P(H \neq 0) > 0$, then H is said to be a *nondegenerate* Z_+-valued random variable.

For any measurable spaces $\langle Y_r, \boldsymbol{Y}_r \rangle$, $r = 1, \ldots, m$, and Y_r-valued random variables, $H_r : X \to Y_r$, on $\langle X, \boldsymbol{X}, P \rangle$, the family of random variables, $(H_1, \ldots, H_m)$, is said to be *independent* iff $P(H_r \in A_r, r = 1, \ldots, m) = \prod_{r=1}^{m} P(H_r \in A_r)$ for all events $A_r \in \boldsymbol{Y}_r, r = 1, \ldots, m$. If $Y_r = R^k$ for $r = 1, \ldots, g$, and Y_r is countable for $r = g+1, \ldots, m$, then independence of $(H_1, \ldots, H_m)$ is equivalent to the condition that

$$\begin{aligned} P(H_1 \geq h_1, \ldots, H_g \geq h_g, H_{g+1} = h_{g+1}, \ldots, H_m = h_m) \\ = \prod_{r=1}^{g} P(H_r \geq h_r) \prod_{r=g+1}^{m} P(H_r = h_r), \end{aligned}$$

for all $h_r \in Y_r, r = 1, \ldots, m$.

For any event, $A \in \boldsymbol{X}$, with $P(A) > 0$, the probability measure, $P(\cdot|A)$: $\boldsymbol{X} \to [0,1]$, defined for all $B \in \boldsymbol{X}$ by $P(B|A) = P(A \cap B)/P(A)$ is designated as the *conditional probability measure* on $\langle X, \boldsymbol{X} \rangle$ given A. Equivalently, if the function, P_A, denotes the $\boldsymbol{X}_A$-restriction of $P(\cdot|A)$, then each conditional probability measure is representable by its associated *conditional probability space*, $\langle A, \boldsymbol{X}_A, P_A \rangle$. Finally, since $\boldsymbol{X}_A \subseteq \boldsymbol{X}$ by definition, it follows that for each Y-valued random variable, $H : X \to Y$, on $\langle X, \boldsymbol{X}, P \rangle$, the A-restriction, H_A, of H yields a well defined Y-valued random variable on $\langle A, \boldsymbol{X}_A, P_A \rangle$.

3.3 Interaction Probability Spaces

The behavioral phenomenon of interest for our purposes consist of patterns of interaction activity between spatially distributed populations of actors and opportunities occurring during some relevant time period. To allow flexibility in analyzing any attributes of these phenomena, it is convenient to treat *interaction patterns*, ω, as abstract elements of some universe,

Ω, of possible patterns, and to define *attributes* of interaction patterns as functions on Ω. In this context, each given family, F, of possible *attribute functions*, f, generates a unique σ-field, $\sigma(F)$, on Ω (as in Section 3.2.2 above), which may be taken to define the class of relevant *interaction events* for analyzing interaction behavior in terms of attributes in F. In this way, each probability measure, P, on $\langle\Omega, \sigma(F)\rangle$ generates a possible probability space, $\langle\Omega, \sigma(F), P\rangle$, on which every attribute, $f \in F$, is by definition a random variable. Hence, all statistical analyses of such attributes can be carried out within this probabilistic framework.

3.3.1 Interaction Patterns

To begin with, let Ω_1 denote an abstract *individual interaction space* consisting of all possible interactions, ω, between individual actors and opportunities (which may include other actors). In terms of the notation in Chapter 1, we shall usually take the individual interaction space to be the product space, $\Omega_1 = A \times B$. [Exceptions are illustrated in Sections 3.7 and 3.8 below.] The subscript '1' indicates that each element, $\omega \in \Omega_1$, consists of exactly one interaction. To include the possibility of larger numbers of interactions, let the associated *n-interaction space,* Ω_n, be defined as the n-fold product of Ω_1, i.e., $\Omega_n = (\Omega_1)^n$ for each $n \in Z_{++}$. The elements, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$, are then designated as *n-interaction patterns,* or equivalently, as interaction patterns of size n. To allow for the possible occurrence of no interactions at all (i.e., interaction patterns of size zero), it is convenient to introduce an abstract *null element*, o, and to let $\Omega_0 = \{o\}$ denote the 0-*interaction space*. With this convention, the desired *interaction space,* Ω, is given by

$$\Omega = \Omega_0 \cup [\cup_{n>0}\Omega_n] = \cup_{n\geq 0}\Omega_n \tag{3.1}$$

with elements, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n \subseteq \Omega$, designated as *interaction patterns.*

3.3.2 Locational Attributes of Interactions

The actors and opportunities involved in interactions are assumed to occupy locations in (possibly identical) finite state spaces, I and J, each containing of at least two elements, i.e., with $2 \leq |I| < \infty$ and $2 \leq |J| < \infty$. (An example involving continuous state spaces is given in Section 3.7 below.) To distinguish the *locational* attributes of actors and opportunities involved in interactions, it is convenient to designate I and J as, respectively, the *origin state space* and *destination state space* for each interaction. Individual states, $i \in I$ and $j \in J$, are then said to represent, respectively, the possible *origin attributes* and *destination attributes* of interactions. [Recall from Section 1.4.2 that the elements, $i \in I$ and $j \in J$, may implicitly be taken to represent the subsets of actors, A_i, and opportunities, B_j, in underlying spatial aggregation schemes, $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$, respectively.] With these conventions, all locational attributes of actors involved in interactions can be summarized by an *origin-attribute function*, $\boldsymbol{i} : \Omega_1 \longrightarrow I$, in which each value, $\boldsymbol{i}(\omega) \in I$, represents the locational attributes of the actor involved in interaction ω. Similarly, all locational attributes of opportunities involved in interactions are summarized by a *destination-attribute function*, $\boldsymbol{j} : \Omega_1 \longrightarrow J$, with values, $\boldsymbol{j}(\omega) \in J$, representing the locational attributes of the opportunity involved in interaction ω.

To represent the relevant locational attributes of interaction patterns, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$, we begin with the origin attributes of individual interactions, ω_r, in ω. In particular, if the null element, o, is adjoined to the origin state space, I, to form the set $I_o = I \cup \{o\}$, then we may define the following family of *origin functions*, $\boldsymbol{I} = \{I_r : \Omega \longrightarrow I_o | r \in Z_{++}\}$, for all $r \in Z_{++}$, $n \in Z_+$, and $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ by

$$I_r(\omega) = \begin{cases} \boldsymbol{i}(\omega_r), & r \leq n \\ o, & r > n. \end{cases} \tag{3.2}$$

Each value, $I_r(\omega)$, denotes the origin attribute of the rth interaction in all those patterns, $\omega \in \Omega$, with at least r interactions, and is null otherwise. Similarly, to represent the destination attributes of individual interactions in $\omega = (\omega_1, \ldots, \omega_n)$, we let $J_o = J \cup \{o\}$ and define the family of *destination functions*, $\boldsymbol{J} = \{J_r : \Omega \longrightarrow J_o | r \in Z_{++}\}$, for all $r \in Z_{++}$, $n \in Z_+$, and $\omega = (\omega_1, \ldots, \omega_n)$ by

$$J_r(\omega) = \begin{cases} \boldsymbol{j}(\omega_r), & r \leq n \\ o, & r > n. \end{cases} \tag{3.3}$$

The combined family, $\boldsymbol{I} \cup \boldsymbol{J}$, is designated as the set of *locational attributes* of interaction patterns.

3.3.3 Interaction Events

Throughout most of the analysis to follow, all relevant attributes of interaction patterns will be constructed as functions of locational attributes in $\boldsymbol{I} \cup \boldsymbol{J}$. Hence, to construct the simplest possible probability model of interactions, it suffices to require that all functions, $I_r \in \boldsymbol{I}$ and $J_r \in \boldsymbol{J}$, be measurable. With this in mind, we now define the relevant class, M, of measurable *interaction events* in Ω to be the smallest σ-field on Ω for which all functions in $\boldsymbol{I} \cup \boldsymbol{J}$ are measurable, i.e.,

$$M = \sigma(\boldsymbol{I} \cup \boldsymbol{J}) \tag{3.4}$$

(where the appropriate σ-fields for the range spaces of the functions, $I_r : \Omega \longrightarrow I_o$ and $J_r : \Omega \longrightarrow J_o$, are taken to be the power sets of I_o and J_o, respectively, as in Section 3.2.2 above). Given this choice of σ-fields, each possible probability measure, P, on the measurable space, $\langle \Omega, M \rangle$, defines a possible *interaction probability space*, $\langle \Omega, M, P \rangle$, on which every function, $I_r \in \boldsymbol{I}$ is an I_o-valued random variable, and every function, $J_r \in \boldsymbol{J}$, is a J_o-valued random variable.

While (3.4) yields a precise definition of measurable interaction events which is readily extendable to more general classes of attribute functions (as illustrated in Section 3.7 and 3.8 below), it is convenient for our present purposes to give a more explicit construction of M in terms of the *origin-attribute function*, $\boldsymbol{i}:\Omega_1 \to I$, and *destination-attribute function*, $\boldsymbol{j}:\Omega_1 \to J$, defined in Section 3.3.2 above. To do so, we begin by letting $M_1 = \sigma(\boldsymbol{i}, \boldsymbol{j})$ denote the σ-field of *individual interaction events* generated by the family of functions, $\{\boldsymbol{i}, \boldsymbol{j}\}$, on Ω_1. Equivalently, M_1 is the smallest σ-field on Ω_1 containing all inverse image sets, $\boldsymbol{i}^{-1}(i) = \{\omega \in \Omega_1 : \boldsymbol{i}(\omega) = i\}$ and $\boldsymbol{j}^{-1}(j) = \{\omega \in \Omega_1 : \boldsymbol{j}(\omega) = j\}$, and hence can also be written as

$$M_1 = \sigma[\{\boldsymbol{i}^{-1}(i) : i \in I\} \cup \{\boldsymbol{j}^{-1}(j) : j \in J\}]. \tag{3.5}$$

Next, for each $n \in Z_{++}$, let $M_n \subseteq \sigma[(M_1)^n] \subseteq \boldsymbol{S}(\Omega_n)$ denote the relevant σ-field of *n-interaction events*, defined by

$$\begin{aligned} M_n = \sigma[\{\prod_{r=1}^{n} \boldsymbol{i}^{-1}(i_r) : (i_1, \ldots, i_n) \in I^n\} \\ \cup \{\prod_{r=1}^{n} \boldsymbol{j}^{-1}(j_r) : (j_1, \ldots, j_n) \in J^n\}]. \end{aligned} \tag{3.6}$$

Finally, letting $M_0 = \sigma(\Omega_0) = \{\emptyset, \Omega_0\}$ denote the *null interaction events*, we have the following equivalent characterization of M [as in Moyal (1962) and Smith (1986a, Section 3.4)]:

Proposition 3.1 *The σ-field, $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J})$, can be equivalently written as*

$$M = \sigma(\cup_{n \geq 0} M_n). \tag{3.7}$$

PROOF: (i) To establish that $M \subseteq \sigma(\cup_{n \geq 0} M_n)$, it suffices to show that all functions in $\boldsymbol{I} \cup \boldsymbol{J}$ are $\sigma(\cup_{n \geq 0} M_n)$-measurable. But for all $r \in Z_{++}$ and $i \in I$,

$$\begin{aligned} I_r^{-1}(i) &= \cup_{n \geq r} \{\omega \in \Omega_n : I_r(\omega) = i\} \\ &= \cup_{n \geq r} [\cup \{\prod_{s=1}^{n} \boldsymbol{i}^{-1}(i_s) : i_r = i, i_s \in I, s \neq r\}], \end{aligned} \tag{3.8}$$

which, together with $\cup\{\prod_{s=1}^{n} \boldsymbol{i}^{-1}(i_s) : i_r = i, i_s \in I, s \neq r\} \in M_n$, implies that each set, $I_r^{-1}(i)$, is a countable union of sets in $\cup_{n \geq 0} M_n$, and hence is an element of $\sigma(\cup_{n \geq 0} M_n)$. Moreover, since $I_r^{-1}(o) = \cup_{n < r} \Omega_n \in \sigma(\cup_{n \geq 0} M_n)$ for all $r \in Z_{+}$, it follows that each function, $I_r : \Omega \to I_o$, is

$\sigma(\cup_{n\geq 0} M_n)$-measurable. Finally, since a parallel argument is easily seen to hold for each $J_r \in \boldsymbol{J}$, it follows that $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J}) \subseteq \sigma(\cup_{n\geq 0} M_n)$.

(ii) Conversely, to establish that $\sigma(\cup_{n\geq 0} M_n) \subseteq M$, it suffices to show that $M_n \subseteq M$ for each $n \in Z_+$, since this will imply at once that $\cup_{n\geq 0} M_n \subseteq M$, and hence that $\sigma(\cup_{n\geq 0} M_n) \subseteq \sigma(M) = M$. To do so, observe first that $I_1 \in \boldsymbol{I} \Rightarrow \Omega_0 = I_1^{-1}(o) \in M$, so that $\emptyset \in M$ implies $M_0 = \{\emptyset, \Omega_0\} \subseteq M$. Next, observe that for any $n \in Z_{++}$, $I_{n+1}^{-1}(o) - I_n^{-1}(o) = \cup_{m\leq n}\Omega_m - \cup_{m<n}\Omega_m = \Omega_n$ implies from the M-measurability of I_{n+1} and I_n together with property (vi) of σ-fields (Section 3.2.1 above) that $\Omega_n \in M$. Also we see from (3.2) that for all $(i_1, \ldots, i_n) \in I^n$,

$$\begin{aligned}\prod_{r=1}^{n} \boldsymbol{i}^{-1}(i_r) &= \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n : \boldsymbol{i}(\omega_r) = i_r, r = 1, \ldots, n\} \\ &= \{\omega \in \Omega_n : I_r(\omega) = i_r, r = 1, \ldots, n\} \\ &= \left[\cap_{r=1}^{n} I_r^{-1}(i_r)\right] \cap \Omega_n\end{aligned} \tag{3.9}$$

and hence from the M-measurability of $I_1, \ldots, I_n \in \boldsymbol{I}$ together with property (v) of σ-fields that $\prod_{r=1}^{n} \boldsymbol{i}^{-1}(i_r) \in M$ for all $(i_1, \ldots, i_n) \in I^n$. Similarly, (3.3) together with the M-measurability of $J_1, \ldots, J_n \in \boldsymbol{J}$ implies that

$$\prod_{r=1}^{n} \boldsymbol{j}^{-1}(j_r) = \left[\cap_{r=1}^{n} J_r^{-1}(j_r)\right] \cap \Omega_n \in M.$$

Hence,

$$\begin{aligned}&\{\prod_{r=1}^{n} \boldsymbol{i}^{-1}(i_r) : (i_1, \ldots, i_n) \in I^n\} \\ &\quad \cup \{\prod_{r=1}^{n} \boldsymbol{j}^{-1}(j_r) : (j_1, \ldots, j_n) \in J^n\} \subseteq M,\end{aligned}$$

and it follows from (3.6) that $M_n \subseteq \sigma(M) = M$ for all $n \in Z_{++}$. □

Finally, it is of interest to note that $\langle \Omega, M \rangle$ need not be a *fully* measurable space, i.e., that the singleton sets, $\{\omega\} \subseteq \Omega$, need not be M-measurable. In the present context, this simply means that the relevant interaction patterns, $\omega \in \Omega$, may not be fully describable in terms of their locational attributes, $\boldsymbol{I} \cup \boldsymbol{J}$, and hence that the corresponding σ-field, M, is not sufficiently rich to discriminate between individual interaction patterns in Ω. However, $\langle \Omega, M \rangle$ is always a fully measurable space with respect to the attributes in $\boldsymbol{I} \cup \boldsymbol{J}$. More precisely, the inverse image sets of these functions induce a (countable) M-measurable partition, M_S, of Ω containing all locational-attribute information, as defined in Section 3.6.2 below. Moreover, it will be shown there (see Lemma 3.2 below) that if the elements of M_S are regarded as 'atoms' in Ω, then M yields a fully measurable space with respect to this (countable) set of atoms.

3.3.4 FREQUENCY ATTRIBUTES OF INTERACTIONS

Given the class of interaction probability spaces above, we next define certain *frequency* attributes of interaction patterns, and demonstrate that these attributes yield well-defined random variables on each possible interaction probability space, $\langle \Omega, M, P \rangle$. To begin with, let the family of *indicator functions*, $\delta_{ij}^r : \Omega \rightarrow \{0,1\}$, be defined for all $r \in Z_+, ij \in I \times J$, and $\omega \in \Omega$ by

$$\delta_{ij}^r(\omega) = \begin{cases} 1, & I_r(\omega) = i \text{ and } J_r(\omega) = j \\ 0. & \text{otherwise,} \end{cases} \tag{3.10}$$

[Note by definition that $\delta_{ij}^r(\omega) = 0$ for all $\omega \in \Omega_n$ with $n < r$.] The associated function, $N_{ij} : \Omega \rightarrow Z_+$, defined for all $\omega \in \Omega$ by

$$N_{ij}(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^n \delta_{ij}^r(\omega), & \omega \in \Omega_n, n \in Z_{++}, \end{cases} \tag{3.11}$$

is designated as the (ij)*-frequency function* on Ω, with values, $N_{ij}(\omega)$, denoting the number of occurrences of individual interactions in ω involving both origin, i, and destination, j. A number of more aggregate frequency functions are derivable from (ij)-frequencies. First, for each origin, $i \in I$, the *i-frequency function,* $N_i : \Omega \rightarrow Z_+$, is defined for all $\omega \in \Omega$ by

$$N_i(\omega) = \sum_{j \in J} N_{ij}(\omega). \tag{3.12}$$

Next, for each destination, $j \in J$, the *j-frequency function,* $N_j : \Omega \rightarrow Z_+$, is defined for all $\omega \in \Omega$ by

$$N_j(\omega) = \sum_{i \in I} N_{ij}(\omega). \tag{3.13}$$

Finally, the *total frequency function,* $N : \Omega \rightarrow Z_+$, is defined for all $\omega \in \Omega$ by

$$N(\omega) = \sum_{i \in I} N_i(\omega) = \sum_{j \in J} N_j(\omega). \tag{3.14}$$

With these definitions, we now verify that [see also Moyal (1962, Section 3)]:

Proposition 3.2 *All frequency functions,* N, N_i, N_j, *and* N_{ij}, $ij \in I \times J$, *are* M*-measurable.*

PROOF: Observe first that since finite sums of real-valued M-measurable functions are also M-measurable [Halmos (1950, Theorem 19.C)], it follows from (3.12), (3.13) and (3.14) that we need only establish M-measurability of each (ij)-frequency functions, $N_{ij} : \Omega \rightarrow Z_+$. To do so, observe next that since $\Omega_n \cap \Omega_m = \emptyset$ for all distinct $n, m \in Z_+$, it follows from Proposition 3.1

above that $\{\Omega_n : n \in Z_+\}$ defines a countable M-measurable partition of Ω. Hence, denoting the Ω_n-restriction of N_{ij} by $N_{ij}^n : \Omega_n \to Z_+$, and observing that by definition, $N_{ij}^{-1}(z) = \cup_{n \geq 0}[N_{ij}^n]^{-1}(z)$ for all $z \in Z_+$, it follows from the definition of M-measurability for countable-valued functions (Section 3.2.2 above) together with property (iii) of σ-fields (Section 3.2.1 above) that M-measurability of each restriction, N_{ij}^n, will imply M-measurability of N_{ij}. But since N_{ij}^0 is trivially M_0-measurable, we need only consider $n \in Z_{++}$. Moreover, since sums of M_n-measurable functions are again M_n-measurable [Halmos (1950, Theorem 19.C)], it follows from (3.11) that N_{ij}^n will be M_n-measurable if the Ω_n-restriction, $\delta_{ij}^{rn} : \Omega_n \to \{0, 1\}$, of the indicator function, δ_{ij}^r, in (3.10) is M_n-measurable for each $n \in Z_{++}$ and $r = 1, \ldots, n$. Finally, since $[\delta_{ij}^{rn}]^{-1}(1) = [\Omega_n \cap I_r^{-1}(i) \cap J_r^{-1}(j)] \in M_n$ and $[\delta_{ij}^{rn}]^{-1}(0) = \Omega_n - [\delta_{ij}^{rn}]^{-1}(1) \in M_n$ imply the M_n-measurability of δ_{ij}^{rn} for all $ij \in I \times J$ and $r = 1, \ldots, n \in Z_{++}$, the result follows from Proposition 3.1. □

Hence the frequency functions $(N_{ij} : ij \in I \times J)$, $(N_i : i \in I)$, $(N_j : j \in J)$, and N all yield well defined Z_+-valued random variates on $\langle \Omega, M, P \rangle$. In particular, the distribution of the *total frequency variate*, N, is given by

$$P(N = n) = P(\Omega_n), \quad n \in Z_+. \tag{3.15}$$

The distributions of the *(ij)-frequency variates, N_{ij}, i-frequency variates, N_i,* and *j-frequency variates, N_j*, are all derivable from the joint distribution, $P(N_{ij} = n_{ij} : ij \in I \times J)$, of the origin-destination frequencies. The exact form of this distribution is somewhat more complex than (3.15), and is deferred to Section 3.6 below.

3.4 Interaction Processes

Within the general probabilistic framework above, we are now ready to formulate the class of spatial interaction models which will be employed throughout most of the book. The essential focus of these models is on how spatial separation influences the probabilities of interaction between actors and opportunities. Hence, we first define a general class of spatial separation measures in Section 3.4.1 below, and then construct an appropriate class of interaction probability models as functions of these measures.

3.4.1 Separation Configurations

Spatial separation between actors and opportunities is assumed to depend only on their locational attributes, i.e., on their locations in I and J, respectively. In particular, these spatial relationships are assumed to be representable by a finite family of *separation functions*, $\{c^k : I \times J \to R \mid k \in K\}$, where $0 < |K| < \infty$, and where each value, c_{ij}^k, denotes the *k-type separation* between origin i, and destination j. Of particular interest are *nonnegative* measures of separation (including various types of interaction costs, such as travel time). These nonnegative measures are taken to correspond to a (possibly empty) subset, $K_+ \subseteq K$, where $c^k : I \times J \to R_+$ for all $k \in K_+$. Each vector of separation values, $c_{ij} = (c_{ij}^k : k \in K)$, is designated as a possible *separation profile* for origin-destination pair, $ij \in I \times J$, and the set of *profile values* is denoted [as in (1.6)] by $V = R_+^{K+} \times R^{K-K_+}$. Each possible array of profile values, $c = (c_{ij} : ij \in I \times J)$, is designated as a *separation configuration* on $I \times J$, and the set of all such configurations is designated as the *configuration class*, $C = V^{I \times J}$.

In terms of these definitions, all spatial structure relevant for interaction behavior is now assumed to be specified by the prevailing separation configuration, $c \in C$. Hence, each probabilistic description of interaction behavior within this spatial structure is representable by some probability measure, P_c, on $\langle \Omega, M \rangle$, which may in general depend on c. In order to analyze possible changes in behavior resulting from changes in c, one must thus specify a family, $\{P_c : c \in C\}$, of interaction probability measures for each possible configuration in C. With this in mind, we turn now to the study of such families.

3.4.2 General Interaction Processes

For purposes of analysis, it is convenient to impose certain regularity conditions on the families of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, which are appropriate for the types of interaction behavior to be modeled. These three conditions, designated respectively as *positivity, symmetry*, and *continuity*, will now be developed in turn.

Positivity Condition

First, it is assumed that each origin-destination pair, $ij \in I \times J$, is potentially relevant in the sense that it can actually occur as part of some interaction. This regularity condition (which may alternatively be viewed as an implicit definition of the relevant sets of origins and destinations, I and J), can be formalized in terms of (ij)-frequency variates, N_{ij}, defined in (3.11) above. In particular, for any prevailing separation configuration, $c \in C$, a given origin-destination pair, ij, can occur iff the event, $\{\omega \in \Omega : N_{ij}(\omega) \neq 0\}$, has *positive* probability under c [i.e., iff the Z_+-valued random variate, N_{ij}, on $\langle \Omega, M, P_c \rangle$ is *nondegenerate* in the sense of Section 3.2.3 above]. Hence our first regularity condition on $\boldsymbol{P} = \{P_c : c \in C\}$ can be stated as follows:

R1. (Positivity) *For all* $c \in C$ *and* $ij \in I \times J$,

$$P_c(N_{ij} \neq 0) > 0. \tag{3.16}$$

Symmetry Condition

Our second regularity condition relates to the basic meaning of an interaction pattern, $\omega = (\omega_1, \ldots, \omega_n)$. In essence, each such pattern is regarded as a set of interactions which may occur during some relevant period. Hence the particular labeling (or ordering) of individual interactions, ω_r, in ω is assumed to have no effect on the probable occurrence of ω. To state this assumption precisely, we begin by formalizing each relabeling of the individual interactions in $\omega \in \Omega_n$ as a *permutation*, $\pi = (\pi_1, \ldots, \pi_n)$, of the integers $(1, \ldots, n)$ [i.e., a bijection, $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$, with images, $\pi(r) = \pi_r, r = 1, \ldots, n$]. If for any $n \in Z_{++}$, we let Π_n denote the set of all such permutations, then each $\pi \in \Pi_n$ induces a relabeling of every n-interaction pattern, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$, denoted by

$$\omega_\pi = (\omega_{\pi_1}, \ldots, \omega_{\pi_n}), \quad \pi \in \Pi_n. \tag{3.17}$$

Similarly, each n-interaction event, $A \in M_n, n \in Z_{++}$, has a corresponding relabeling defined by

$$A_\pi = \{\omega_\pi : \omega \in A\}. \tag{3.18}$$

With this notation, our basic assumption is that such relabelings of events do not affect their probability of occurrence. To make this assumption precise, it is first appropriate to verify that such relabelings always yield probabilizable events, i.e., that,

Proposition 3.3 *For all $n \in Z_{++}, \pi \in \Pi_n$, and $A \subseteq \Omega_n$,*

$$A \in M_n \Rightarrow A_\pi \in M_n. \tag{3.19}$$

PROOF: If for each $n \in Z_{++}$ and $\pi \in \Pi_n$ we denote the class of all n-interaction events with M_n-measurable relabelings by

$$M_n^\pi = \{A \in M_n : A_\pi \in M_n\} \subseteq M_n, \tag{3.20}$$

then by definition, it suffices to show that $M_n^\pi = M_n$. To do so, observe first that for any $A = (A_1, \ldots, A_n) \in (M_1)^n$, it follows from (3.17) and (3.18) that $A_\pi = (A_{\pi_1}, \ldots, A_{\pi_n}) \in (M_1)^n$ so that $(M_1)^n \subseteq M_n^\pi$. Hence, if it can be shown that M_n^π is a σ-field on Ω_n, then we must have $M_n \subseteq \sigma[(M_1)^n] \subseteq M_n^\pi$, which together with $M_n^\pi \subseteq M_n$ in (3.20) will yield the desired result. To verify that M_n^π is a σ-field on Ω_n, observe first that $\emptyset_\pi = \emptyset \in M_n$ implies from (3.20) that $\emptyset \in M_n^\pi$. Next observe (from the bijective properties of permutations) that both $A_\pi \cap (\Omega_n - A)_\pi = \emptyset$ and $A_\pi \cup (\Omega_n - A)_\pi = \Omega_n$ hold identically for all $\pi \in \Pi_n$ and $A \subseteq \Omega_n$. Hence $(\Omega_n - A)_\pi = \Omega_n - A_\pi$, and we may conclude that $A \in M_n^\pi \Rightarrow A_\pi \in M_n \Rightarrow \Omega_n - A_\pi \in M_n \Rightarrow (\Omega_n - A)_\pi \in M_n \Rightarrow \Omega_n - A \in M_n^\pi$. Finally, since $(\cup_m A_m)_\pi = \cup_m (A_m)_\pi$ holds identically for all $\{A_m : m \in Z_+\} \subseteq \boldsymbol{S}(\Omega_n)$, it also follows that $\{A_m : m \in Z_+\} \subseteq M_n^\pi \Rightarrow \{(A_m)_\pi : m \in Z_+\} \subseteq M_n \Rightarrow \cup_m (A_m)_\pi \in M_n \Rightarrow$

$(\cup_m A_m)_\pi \in M_n \Rightarrow \cup_m A_m \in M_n^\pi$. Hence M_n^π must be a σ-field on Ω_n, and the result is established. □

Given this measurability property of relabelings, we now require that each interaction probability measure, P_c, in $\boldsymbol{P}$ exhibit the following symmetry condition with respect to relabelings of interaction events:

R2. (Symmetry) *For all* $c \in C$, $n \in Z_{++}$, $\pi \in \Pi_n$, *and* $A \in M_n$,

$$P_c(A) = P_c(A_\pi). \tag{3.21}$$

Equivalently, condition R2 can be seen as an 'exchangeability' condition on the n individual interaction events defining any realized n-interaction pattern [as discussed, for example, in Kingman (1978)].

Continuity Condition

Finally, in analyzing the effects of changes in spatial separation on interaction behavior, our only prior assumption is that small changes in separation levels yield small changes in mean interaction frequency levels. More precisely, if for each separation configuration, $c \in C$, and origin-destination pair, $ij \in I \times J$, we denote the *mean* (ij)*-frequency* under configuration, c, by

$$\mathrm{E}_c(N_{ij}) = \sum_{m \geq 0} m P_c(N_{ij} = m). \tag{3.22}$$

then $\boldsymbol{P} = \{P_c : c \in C\}$ is assumed to satisfy the following continuity condition:

R3. (Continuity)*For all* $ij \in I \times J, c \in C$ *and sequences* $\{c_m : m \in Z_+\}$ *in* C *with* $\mathrm{E}_{c_m}(N_{ij}) < \infty, m \in Z_+$,

$$c_m \to c \Rightarrow \mathrm{E}_{c_m}(N_{ij}) \to \mathrm{E}_c(N_{ij}) < \infty. \tag{3.23}$$

Given these regularity conditions, the desired families of interaction probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, can now be formalized as follows:

Definition 3.1 For any σ-field, $\overline{M}$, on Ω with $M \subseteq \overline{M}$, family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is designated as an *interaction process* on Ω iff $\boldsymbol{P}$ satisfies regularity conditions (R1, R2, R3).

Note in particular that condition R1 implies that for each $ij \in I \times J$ and $c \in C$ there is some $m > 0$ with $P_c(N_{ij} = m) > 0$, so that $\mathrm{E}_c(N_{ij}) \geq m P_c(N_{ij} = m) > 0$. Hence all mean interaction frequencies, $\mathrm{E}_c(N_{ij})$, are *positive* for interaction processes $\boldsymbol{P} = \{P_c : c \in C\}$ on Ω.

3.4.3 Independent Interaction Processes

Within this general framework, the notion of *independent* interaction processes developed in Chapter 1 can now be formalized as follows. Recall first from Section 3.2.3 [together with (3.15) above] that for each $c \in C$ and $n \in Z_{++}$ with $P_c(\Omega_n) > 0$, the *conditional probability measure*, P_c^n, on $\langle \Omega, \overline{M} \rangle$ given for all $A \in \overline{M}$ by

$$P_c^n(A) = \frac{P_c(A \cap \Omega_n)}{P_c(\Omega_n)} \tag{3.24}$$

is well defined and specifies the conditional probability of interaction events, $A \in \overline{M}$, given that an interaction pattern of size n occurs. With respect to this conditional distribution, the *conditional joint-location probabilities*, $P_c^n(I_r = i_r, J_r = j_r)$ and $P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n)$, are well defined for all locations $(i_r : r = 1, \ldots, n) \in I^n$ and $(j_r : r = 1, \ldots, n) \in J^n$. In terms of these probabilities, the desired class of independent interaction processes can now be defined as follows:

Definition 3.2 An interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on Ω is designated as an *independent interaction process* iff $\boldsymbol{P}$ satisfies the following two independence conditions for all $c \in C$ and $n \in Z_{++}$ with $P_c(\Omega_n) > 0$,

A1. (Locational Independence) *For all* $i_1, \ldots, i_n \in I$ *and* $j_1, \ldots, j_n \in J$,

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) = \prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r). \tag{3.25}$$

A2. (Frequency Independence) *For all* $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$,

$$P_c(N_{ij} = n_{ij} : ij \in I \times J) = \prod_{ij} P_c(N_{ij} = n_{ij}). \tag{3.26}$$

3.5 Frequency Processes

Since all the gravity models in Chapter 2 were defined solely in terms of *interaction frequencies*, $(N_{ij} : ij \in I \times J)$, it is clear that these random variables are of central importance for our purposes. With this in mind, it is convenient to begin by defining general processes involving such frequency variates, and then to focus on those frequencies which are generated by the interaction processes defined above.

For any given separation configuration, $c \in C$, let each vector, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, of nondegenerate Z_+-valued random variables with associated probability function, $p_c : Z_+^{I \times J} \to [0, 1]$, be designated as a possible *frequency profile* on $I \times J$ under configuration, c. By definition, the joint distribution of the random variates, $(N_{ij}^c : ij \in I \times J)$, is given for all realizations, $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$, by

$$P(N_{ij}^c = n_{ij} : ij \in I \times J) = p_c(n_{ij} : ij \in I \times J). \tag{3.27}$$

The corresponding marginal distribution of each (ij)*-frequency,* N_{ij}^c, is denoted by $P(N_{ij}^c = n_{ij})$, where the underlying probability function, p_c, is understood. Similarly, the distributions of the associated *total frequency,* N^c, *i-frequencies,* N_i^c, and *j-frequencies,* N_j^c, are written as $P(N^c = n)$, $P(N_i^c = n_i)$, and $P(N_j^c = n_j)$, respectively. If the *mean* (ij)*-frequency* is denoted by

$$\mathrm{E}(N_{ij}^c) = \sum_{m \geq 0} m P(N_{ij}^c = m), \tag{3.28}$$

then we now say that:

Definition 3.3
(i) A family, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, of frequency profiles on $I \times J$ is designated as a *frequency process* on $I \times J$ iff $\boldsymbol{N}$ satisfies the following continuity condition for all $ij \in I \times J$, $c \in C$ and all sequences $\{c_m : m \in Z_+\}$ on C with $\mathrm{E}(N_{ij}^{c_m}) < \infty, m \in Z_+$,

$$c_m \to c \Rightarrow \mathrm{E}(N_{ij}^{c_m}) \to \mathrm{E}(N_{ij}^c). \tag{3.29}$$

(ii) Let $\langle I \times J \rangle$ denote the class of all *frequency processes* on $I \times J$.

3.6 Generated Frequency Processes

Each interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, generates a frequency process on $I \times J$ in the obvious way. In particular, if for each configuration, $c \in C$, the joint frequencies, $(N_{ij}(\omega) : ij \in I \times J)$, associated with each realized interaction pattern, $\omega \in \Omega$, are taken to define the realizations of a random vector, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, with distribution generated by the probability measure, P_c, on $\langle \Omega, M \rangle$, then it follows from the positivity condition, R1, on $\boldsymbol{P}$ that each random vector, $\boldsymbol{N}_c$, is a frequency profile on $I \times J$. Moreover, it also follows from the continuity condition, R3, on $\boldsymbol{P}$ that the resulting family of frequency profiles yields a well defined frequency process on $I \times J$, which is now designated as the $\boldsymbol{P}$*-frequency process,* $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\} \in \langle I \times J \rangle$. More generally, for any interaction process, $\boldsymbol{P}$, and frequency process, $\boldsymbol{N} \in \langle I \times J \rangle$, if the distribution of each frequency profile, $\boldsymbol{N}_c$, in $\boldsymbol{N}$ (given by the probability function, p_c) is the same as the distribution of the corresponding frequency profile in $\boldsymbol{N_P}$ (given the probability measure, P_c), i.e., if

$$p_c(n_{ij} : ij \in I \times J) = P_c(N_{ij} = n_{ij} : ij \in I \times J), \tag{3.30}$$

for all $c \in C$ and $(n_{ij} : ij \in I \times J) \in Z_+^{I\times J}$, then $\boldsymbol{N}$ is now said to be *generated* by $\boldsymbol{P}$, and $\boldsymbol{P}$ is designated as a *generator* of $\boldsymbol{N}$.

Given these definitions, we next develop an explicit representation of the frequency profile distributions for $\boldsymbol{P}$-frequency processes in terms of the underlying probability measures, $P_c \in \boldsymbol{P}$. To so do, we begin by letting

$$\Omega_{ij} = \boldsymbol{i}^{-1}(i) \cap \boldsymbol{j}^{-1}(j) \in M_1, \quad ij \in I \times J \tag{3.31}$$

denote the (ij)-*interaction event* consisting of all those individual interactions in Ω_1 which involve both origin, i, and destination, j. Then for any nonnegative integer vector $(n_{ij} : ij \in I \times J) \in Z_+^{I\times J}$ with $n = \sum_{ij} n_{ij}$, the event $\{\omega \in \Omega : N_{ij}(\omega) = n_{ij}, ij \in I \times J\}$ occurs iff some n-interaction pattern, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$ occurs with $\omega_r \in \Omega_{ij}$ for exactly n_{ij} values of $r \in \{1, \ldots, n\}$ for each $ij \in I \times J$. This in turn is equivalent to the occurrence of some permutation (relabeling) of the n-interaction event, $\prod_{ij} \Omega_{ij}^{n_{ij}} \in M_n$. Hence, recalling the definition of A_π [from (3.18) above] for each event, $A \in \Omega_n$, and permutation, $\pi \in \Pi_n$, it follows that the event $\{\omega \in \Omega : N_{ij}(\omega) = n_{ij}, ij \in I \times J\}$ is precisely the n-interaction event, $\cup_\pi (\prod_{ij} \Omega_{ij}^{n_{ij}})_\pi \in M_n$. Hence, the probability function, p_c, for each frequency profile, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, in $\boldsymbol{N_P}$ is given by $p_c(0, \ldots, 0) = P_c(N = 0) = P_c(\Omega_0)$ and by

$$p_c(n_{ij} : ij \in I \times J) = P_c\big[\cup_{\pi \in \Pi_n} \big(\prod_{ij} \Omega_{ij}^{n_{ij}}\big)_\pi\big], \tag{3.32}$$

for all $(n_{ij} : ij \in I \times J) \in Z_+^{I\times J}$ with $\sum_{ij} n_{ij} = n \in Z_{++}$. The distributions of the corresponding i-frequencies, N_i^c, in (3.12) above and j-frequencies, N_j^c, in (3.13) above are constructable as appropriate sums of the probabilities in (3.32). In particular, if for each $i \in I$, the marginal distribution of the random vector, $(N_{ij} : j \in J)$, is denoted by $P_c(N_{ij} = n_{ij} : j \in J)$, and if for each $n_i \in Z_+$ we let $Z_J(n_i) = \{(n_{ij} : j \in J) \in Z_+^J : \sum_j n_{ij} = n_i\}$ denote the set of integer vectors, $(n_{ij} : j \in J)$, summing to n_i, then the distribution of each *i-frequency variate*, N_i^c, is given for all $n_i \in Z_+$ by

$$P(N_i^c = n_i) = \sum_{(n_{ij}:j\in J)\in Z_J(n_i)} P_c(N_{ij} = n_{ij} : j \in J). \tag{3.33}$$

Similarly, letting $P_c(N_{ij} = n_{ij} : i \in I)$ denote the marginal distribution of $(N_{ij} : i \in I)$ in (3.32), and setting $Z_I(n_j) = \{(n_{ij} : i \in I) \in Z_+^I : \sum_{ij} n_{ij} = n_j\}$, the distribution of each *j-frequency variate*, N_j^c, is given for all $n_j \in Z_+, j \in J$, and $c \in C$ by

$$P(N_j^c = n_j) = \sum_{(n_{ij}:i\in I)\in Z_I(n_j)} P_c(N_{ij} = n_{ij} : i \in I). \tag{3.34}$$

3.6.1 Poisson Frequency Processes

The single most important frequency processes employed in this book are *Poisson frequency processes*, i.e., those frequency processes, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, in which each frequency profile, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, consists of independently distributed Poisson variates, N_{ij}^c, with finite means, $\mathrm{E}(N_{ij}^c)$. To analyze these processes, it is convenient to begin by characterizing certain more general families of Poisson variates, which are of interest beyond the present context. In particular, for any nonempty finite index set, T, a family of Z_+-valued random variates, $\boldsymbol{X} = (X_t : t \in T)$, with joint distribution, P, and finite means, $\mathrm{E}(X_t), t \in T$, is designated as *Poisson family* on T iff for all $(n_t : t \in T) \in Z_+^T$,

$$P(X_t = n_t : t \in T) = \prod_{t \in T} \frac{\mathrm{E}(X_t)^{n_t}}{n_t!} \exp[-\mathrm{E}(X_t)]. \tag{3.35}$$

The following characterization of Poisson families is based on the results of Moran (1952) and Chatterji (1963) [see also Smith (1986, Lemmas 3.1 and 3.2), (1985b, Lemma 3.1)]. For any finite family of Z_+-valued random variates, $\boldsymbol{X} = (X_t : t \in T)$, with joint distribution, P, if we denote by $Z_{++}(\boldsymbol{X}) = \{n \in Z_{++} : P(\sum_t X_t = n) > 0\}$ the set of positive total values which can actually occur in $\boldsymbol{X}$, [and recall that X_t is *nondegenerate* iff $P(X_t \neq 0) > 0$], then we now have:

Theorem 3.1 (Poisson Families)
(i) *A finite family of independent Z_+-valued random variates, $\boldsymbol{X} = \{X_t : t \in T\}$, containing at least two nondegenerate variates is a Poisson family on T iff for each $n \in Z_{++}(\boldsymbol{X})$ there exists a probability function, $p_n : T \to [0, 1]$, such that for all $(n_t : t \in T) \in Z_+^T$ with $\sum_{t \in T} X_t = n$,*

$$P[(X_t = n_t : t \in T) | \sum_{t \in T} X_t = n] = n! \prod_{t \in T} \frac{p_n(t)^{n_t}}{n_t!}. \tag{3.36}$$

(ii) *In addition, the equivalence in* (i) *necessarily implies that $Z_{++}(\boldsymbol{X}) = Z_{++}$, and that $p_n(t) = p_1(t) = \mathrm{E}(X_t)/\mathrm{E}(\sum_{s \in T} X_s)$ for all $t \in T$ and $n \in Z_{++}$.*

Proof: **(i)** Necessity of condition (3.36) is immediate, since every Poisson family (with at least one nondegenerate variate) automatically satisfies the conditional multinomial property in (3.36). To establish the sufficiency of condition (3.36), suppose that the family of probability functions, $\{p_n : T \to [0, 1] | n \in Z_{++}(\boldsymbol{X})\}$, satisfies (3.36), and consider any $t^\circ \in T$ with X_{t° nondegenerate. Then by setting $X_1 = X_{t^\circ}$ and $X_2 = \sum_{t \in T - \{t^\circ\}} X_t$, it follows by hypothesis that both X_1 and X_2 are nondegenerate and independent. Next consider any $n \in Z_{++}$ with $P(X_1 + X_2 = n) = P(\sum_t X_t = n) > 0$. Then since (3.36) must hold for n, it follows from well known properties of the multinomial distribution in (3.36) [see for example Johnson and Kotz (1969, Section 11.2)] that X_1 and X_2 must be conditionally binomially distributed, given $X_1 + X_2 = n$, i.e., that for all $k = 0, 1, \ldots, n$,

$$P(X_1 = k | X_1 + X_2 = n) = \frac{n!}{k!(n-k)!} q_n^k (1 - q_n)^{n-k}, \tag{3.37}$$

where $q_n = p_n(t^\circ)$ in (3.36). We now show that for this choice of X_1 and X_2, (3.37) holds for *all* $n \in Z_{++}$, and in addition that $0 < q_n < 1$ for all $n \in Z_{++}$. To do so, observe first from the nondegeneracy of X_1 and X_2 that $P(X_1 = n_1) > 0$ and $P(X_2 = n_2) > 0$ for some $n_1, n_2 \in Z_{++}$. Hence, setting $m = n_1 + n_2 \in Z_{++}$, and observing from independence that $P(X_1 + X_2 = m) \geq P(X_1 = n_1, X_2 = n_2) = P(X_1 = n_1)P(X_2 = n_2) > 0$, it follows that (3.37) is satisfied for $n = m$. Moreover, setting $k = n_1$ and $n = m$ in (3.37), and observing that $P(X_1 = n_1 | X_1 + X_2 = m) \geq P(X_1 = n_1, X_2 = n_2)/P(X_1 + X_2 = m) > 0$, it follows from the positivity of n_1 and $n_2 (= m - k)$ that $0 < q_m < 1$. Hence the above assertion holds in particular for $n = m$. Moreover, since this in turn implies from (3.37) that $P(X_i = k) \geq P(X_i = k | X_1 + X_2 = m)P(X_1 + X_2 = m) > 0$ for all $i = 1, 2$ and $k = 0, 1, \ldots, m$, it follows by replacing n_1 and n_2 above with any $k_1, k_2 \leq m$ that (3.37) holds with $0 < q_n < 1$ for all $n = k_1 + k_2 \leq 2m$. Hence we have found a positive integer, $2m$, such that the assertion holds for all $n \leq 2m$. Given this result, it remains only to show that there is no largest integer with this property. To do so, suppose to the contrary that $n^* (\geq 2m)$ is the largest integer such that (3.37) holds with $0 < q_n < 1$ for all $n \leq n^*$. Then in particular, $P(X_i = n^*) = P(X_i = n^* | X_1 + X_2 = n^*)P(X_1 + X_2 = n^*)$ for $i = 1, 2$, so that by replacing n_1 and n_2 above with any $k_1, k_2 \leq n^*$, it now follows that (3.37) holds with $0 < q_n < 1$ for all $n \leq 2n^*$. But since this contradicts the maximality of n^*, we may conclude that no largest integer with this property exists.

Given this result, we can now apply the argument of Chatterji (1963) to show that both X_1 and X_2 are necessarily Poisson distributed with positive finite means. To do so, let us begin by observing that since (3.37) holds with $0 < q_n < 1$ for all $n \in Z_{++}$, it follows that the probabilities, $f(n) = P(X_1 = n)$ and $g(n) = P(X_2 = n)$, are both positive for all $n \in Z_{++}$, and satisfy the following identity for all $k, n \in Z_+$ with $k \leq n$,

$$\begin{aligned} f(k)g(n-k) &= P(X_1 = k, X_2 = n-k) \\ &= P(X_1 = k | X_1 + X_2 = n)P(X_1 + X_2 = n) \\ &= \frac{n!}{k!(n-k)!} q_n^k (1 - q_n)^{n-k} P(X_1 + X_2 = n). \end{aligned} \tag{3.38}$$

Hence, taking the ratio of $f(k)g(n-k)$ and $f(k-1)g(n-k+1)$, we obtain

$$\frac{f(k)g(n-k)}{f(k-1)g(n-k+1)} = a_n \frac{(n-k+1)}{k}, \quad k = 1, \ldots, n, \tag{3.39}$$

where $a_n = q_n/(1-q_n) > 0, n \in Z_+$. Next, by setting $k = n$ in (3.39) and letting $\theta = g(1)/g(0)$, we are led to the following recursion relation for f:

$$f(n) = [a_n\theta/n]f(n-1), \quad n \in Z_+. \tag{3.40}$$

Similarly, by setting $k = 1$ in (3.39) and noting from (3.40) that $f(1)/f(0) = a_1\theta$, we also have the following recursion relation for g:

$$g(n) = [a_1\theta/na_n]g(n-1), \quad n \in Z_+. \tag{3.41}$$

Next, by evaluating (3.39) at $2 = k \leq n$ and employing (3.40) and (3.41), we obtain the following recursion relation for a_n:

$$\begin{aligned} a_n\frac{(n-1)}{2} &= \frac{f(2)}{f(1)} \cdot \frac{g(n-2)}{g(n-1)} = (a_2\theta/2)[a_{n-1}(n-1)/a_1\theta] \\ &\Rightarrow a_n = (a_2/a_1)a_{n-1}, \quad n \geq 2, \end{aligned} \tag{3.42}$$

which is easily seen to yield the following solution for a_n:

$$a_n = (a_2/a_1)^{n-1}a_1, \quad n \in Z_+. \tag{3.43}$$

But if $a_2 > a_1$, then by (3.40) and (3.43), we would have $f(n)/f(n-1) = (a_1\theta/n)(a_2/a_1)^{n-1} \to \infty$, which contradicts the condition that $1 = \sum_n P(X_1 = n) = \sum_n f(n)$. Similarly if $a_1 > a_2$, then by (3.41) and (3.43) we would have $g(n)/g(n-1) = (\theta/n) \cdot (a_1/a_2)^{n-1} \to \infty$, which contradicts the condition that $1 = \sum_n P(X_2 = n) = \sum_n g(n)$. Hence we must have $a_2 = a_1$, and may conclude from (3.43) that $a_n = a_1$ (and thus that $q_n = q_1$) holds identically for all $n \in Z_+$. Finally, by substituting this identity into the recursion relation (3.40), we obtain the following solution for f:

$$f(n) = \frac{(a_1\theta)^n}{n!}f(0), \quad n \in Z_+ \tag{3.44}$$

which together with $\sum_n f(n) = 1$ implies that $f(0) = \exp[-a_1\theta]$. Hence, $X_{t^\circ}(= X_1)$ is seen to be Poisson distributed with positive finite mean, $\mathrm{E}(X_{t^\circ}) = a_1\theta$.

To complete the proof of part (i), observe that since the choice of the nondegenerate variate, X_{t°, was arbitrary, it follows that every nondegenerate variate in $\{X_t : t \in T\}$ must be Poisson distributed with positive finite mean. Moreover, since each degenerate variate in $\{X_t : t \in T\}$ is trivially Poisson with mean zero, it follows from the hypothesized independence of all random variates in $\{X_t : t \in T\}$ that (3.35) holds identically for all $(n_t : t \in T) \in Z_+^T$, and that $\mathrm{E}(X_t) < \infty$ for all $t \in T$.

(ii) Finally, to establish the assertion in part (ii), observe from (3.35) together with the independence hypothesis that for any nondegenerate variate, $X_{t^\circ} \in \boldsymbol{X}$, and any $n \in Z_{++}$,

$$P(\sum_t X_t = n) \geq P[X_{t^\circ} = n, (X_t = 0 : t \in T - \{t^\circ\})]$$
$$= P(X_{t^\circ} = n) \prod_{t \in T - \{t^o\}} P(X_t = 0) > 0$$

implies that $Z_{++}(\boldsymbol{X}) = Z_{++}$. Moreover, since (3.35) implies that

$$P(\sum_t X_t = n) = \mathrm{E}(\sum_t X_t)^n \exp[-\mathrm{E}(\sum_t X_t)]/n!,$$

it follows by substituting this relation and (3.35) and (3.36) into the identity,

$$P[X_t = n,\ (X_s = 0 : s \neq t)]$$
$$= P[X_t = n,\ (X_s = 0 : s \neq t) | \sum_t X_t = n] P(\sum_t X_t = n),$$

that $p_n(t)^n = [\mathrm{E}(X_t)/\mathrm{E}(\sum_s X_s)]^n$, and hence that

$$p_n(t) = \mathrm{E}(X_t)/\mathrm{E}(\sum_s X_s) = p_1(t),$$

for all $n \in Z_{++}$. □

Given this general result, we now consider the following special class of *Poisson processes* defined on the origin-destination set, $I \times J$:

Definition 3.4
(i) A frequency process, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\} \in \langle I \times J \rangle$, with positive finite mean frequencies, $0 < \mathrm{E}(N_{ij}^c) < \infty$, for each $c \in C$ and $ij \in I \times J$ is designated as a *Poisson process* on $I \times J$ iff for all configurations, $c \in C$, and all realized frequencies, $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$,

$$P(N_{ij}^c = n_{ij} : ij \in I \times J) = \prod_{ij} \frac{\mathrm{E}(N_{ij}^c)^{n_{ij}}}{n_{ij}!} \exp[-\mathrm{E}(N_{ij}^c)]. \tag{3.45}$$

(ii) The class of all Poisson processes on $I \times J$ is denoted by $\langle$POISSON$\rangle$.

To characterize this class of frequency processes, we begin by designating a frequency process, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, as *independent* iff

$$P(N_{ij}^c = n_{ij} : ij \in I \times J) = \prod_{ij} P(N_{ij}^c = n_{ij}), \tag{3.46}$$

for all $c \in C$ and $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$. Then, denoting by

$$Z_{++}(\boldsymbol{N}_c) = \{n \in Z_{++} : P(N^c = n) > 0\} \tag{3.47}$$

the set of *positive frequency totals* which are possible in each frequency profile, $\boldsymbol{N}_c \in \boldsymbol{N}$, we can now employ Theorem 3.1 to obtain the following characterization of Poisson processes:

Corollary 3.1 (Poisson Processes)
(i) *If* $\boldsymbol{N}=\{\boldsymbol{N}_c : c \in C\}$ *is an independent frequency process on* $I \times J$, *then* $\boldsymbol{N} \in \langle \text{POISSON} \rangle$ *iff for each* $c \in C$ *and* $n \in Z_{++}(\boldsymbol{N})$ *there exists a probability function,* $p_c^n : I \times J \to [0,1]$, *such that for all frequency realizations,* $(n_{ij} : ij \in I \times J) \in Z_+^{I\times J}$, *with* $\sum_{ij} n_{ij} = n$,

$$P(N_{ij}^c = n_{ij} : ij \in I \times J | N^c = n) = n! \prod_{ij} \frac{p_c^n(ij)^{n_{ij}}}{n_{ij}!}. \tag{3.48}$$

(ii) *In addition, the equivalence in* (i) *implies that* $Z_{++}(\boldsymbol{N}_c) = Z_{++}$ *and*

$$p_c^n(ij) = p_c^1(ij) = \mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c) > 0, n \in Z_{++} \tag{3.49}$$

for all $c \in C$ *and* $ij \in I \times J$.

PROOF: **(i)** By setting $T = I \times J$ in Theorem 3.1 and observing that by hypothesis, $|I \times J| = |I| \cdot |J| > 2$, it follows from (3.46) that each frequency profile, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J) \in \boldsymbol{N}$, is a finite family of independent, nondegenerate Z_+-valued random variates (containing more than two members). Hence it follows at once from part (i) of Theorem 3.1 that the family of random variates, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, satisfies (3.45) iff for each $n \in Z_{++}$ there exists a probability function, $p_c^n : I \times J \to [0,1]$, satisfying (3.48).

(ii) Finally, the equalities, $Z_{++}(\boldsymbol{N}_c) = Z_{++}, c \in C$, follow from part (i) above, and condition (3.49) follows directly from part (ii) of Theorem 3.1 together with the positivity of each mean value, $\mathrm{E}_c(N_{ij})$. □

Several additional properties of Poisson processes are worth noting. In particular, since sums of independent Poisson variates are also Poisson [see for example Johnson and Kotz (1969, Section 11)], it follows at once that all origin-frequencies, destination-frequencies, and total frequencies for Poisson processes are necessarily Poisson. Hence the distributions defined in (3.15), (3.33) and (3.34) above have the following explicit forms for each Poisson process, $\boldsymbol{N} \in \langle \text{POISSON} \rangle$:

$$P(N_i^c = n_i) = \frac{\mathrm{E}(N_i^c)^{n_i}}{n_i!} \exp[-\mathrm{E}(N_i^c)], \quad n_i \in Z_+, i \in I, \tag{3.50}$$

$$P(N_j^c = n_j) = \frac{\mathrm{E}(N_j^c)^{n_j}}{n_j!} \exp[-\mathrm{E}(N_j^c)], \quad n_j \in Z_+, j \in J, \tag{3.51}$$

$$P(N^c = n) = \frac{\mathrm{E}(N^c)^{n}}{n!} \exp[-\mathrm{E}(N^c)], \quad n \in Z_+. \tag{3.52}$$

3.6.2 Poisson Characterization Theorem

Let $\langle A1,A2\rangle$ denote the class of all frequency processes generated by independent interaction processes, i.e., the subclass of all frequency processes, $\boldsymbol{N}$ on $I \times J$ which are generated by interaction processes, $\boldsymbol{P}$, on Ω satisfying axioms A1 and A2 in Definition 3.2 above. With this notation, our main objective is to show that $\langle \text{POISSON}\rangle = \langle A1,A2\rangle$, i.e., that the Poisson processes on $I \times J$ are precisely those frequency processes generated by independent interaction processes on Ω. To do so, we begin by establishing the following multinomial property of (ij)-frequencies for interaction processes satisfying the locational independence axiom:

Lemma 3.1 *If $\boldsymbol{P} = \{P_c : c \in C\}$ is an interaction process on Ω satisfying* A1 *and if for each $c \in C$ and $n \in Z_{++}$ with $P(\Omega_n) > 0$, the probability function, $p_c^n : I \times J \to [0,1]$, is defined for all $ij \in I \times J$ by*

$$p_c^n(ij) = P_c^n(I_1 = i, J_1 = j), \tag{3.53}$$

then for all $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ with $\sum_{ij} n_{ij} = n$,

$$P_c^n(N_{ij} = n_{ij} : ij \in I \times J) = n! \prod_{ij} \frac{p_c^n(ij)^{n_{ij}}}{n_{ij}!}. \tag{3.54}$$

Proof: First recall from (3.24) and (3.32) that for all $c \in C$, $n \in Z_{++}$ with $P_c(\Omega_n) > 0$, and $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ with $\sum_{ij} n_{ij} = n$,

$$\begin{aligned} P_c^n(N_{ij} = n_{ij} : ij \in I \times J) &= P_c(N_{ij} = n_{ij} : ij \in I \times J)/P_c(\Omega_n) \\ &= P_c\big[\cup_{\pi \in \Pi_n} \big(\prod_{ij} \Omega_{ij}^{n_{ij}}\big)_\pi\big]/P_c(\Omega_n). \end{aligned} \tag{3.55}$$

Next from symmetry condition, R2, for interaction processes, it follows that $P_c[(\prod_{ij} \Omega_{ij}^{n_{ij}})_\pi] = P_c(\prod_{ij} \Omega_{ij}^{n_{ij}})$ for all $\pi \in \Pi_n$. Moreover, since there are exactly $n!/\prod_{ij} n_{ij}!$ distinct events, $(\prod_{ij} \Omega_{ij}^{n_{ij}})_\pi$, $\pi \in \Pi_n$, and since these events are pairwise disjoint, it follows that

$$P_c\big[\cup_{\pi \in \Pi_n} \big(\prod_{ij} \Omega_{ij}^{n_{ij}}\big)_\pi\big] = \frac{n!}{\prod_{ij} n_{ij}!} P_c\big(\prod_{ij} \Omega_{ij}^{n_{ij}}\big). \tag{3.56}$$

To simplify this expression further, choose any elements, $(i_1, \ldots, i_n) \in I^n$ and $(j_1, \ldots, j_n) \in J^n$, such that for all $ij \in I \times J, (i_r, j_r) = (i,j)$ holds for exactly n_{ij} values of $r = 1, \ldots, n$. Then by construction, the associated event, $(I_r = i_r, J_r = j_r : r = 1, \ldots, n)$ must correspond to a relabeling of the event, $\prod_{ij} \Omega_{ij}^{n_{ij}}$. Hence it follows from R2 together with independence axiom, A1, for $\boldsymbol{P}$ that

$$\begin{aligned} P_c\Big(\prod_{ij} \Omega_{ij}^{n_{ij}}\Big) &= P_c(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ &= P_c(\Omega_n) P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ &= P_c(\Omega_n) \prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r). \end{aligned} \tag{3.57}$$

But since R2 also implies that for all $ij \in I \times J$ and $r = 1, \ldots, n$, the permutation, $\pi \in \Pi_n$, defined by $\pi_r = 1, \pi_1 = r$, and $\pi_s = s$ for all $s \notin \{1, r\}$ yields the identity

$$\begin{aligned} &P_c^n(I_r = i, J_r = j) = P_c^n[I_r = i, J_r = j, (I_s \in I, J_s \in J : s \neq r)] \\ &= P_c[(I \times J)^{r-1} \times \{(i,j)\} \times (I \times J)^{n-r}]/P_c(\Omega_n) \\ &= P_c[(\{(i,j)\} \times (I \times J)^{n-1})_\pi]/P_c(\Omega_n) \\ &= P_c(\{(i,j)\} \times (I \times J)^{n-1})/P_c(\Omega_n) \\ &= P_c^n(I_1 = i, J_1 = j). \end{aligned} \tag{3.58}$$

It may be concluded by successive backward substitution from (3.58) to (3.55) that

$$\begin{aligned} P_c^n(N_{ij} = n_{ij} : ij \in I \times J) &= \frac{n!}{\prod_{ij} n_{ij}!} \prod_{r=1}^{n} P_c^n(I_1 = i_r, J_1 = j_r) \\ &= \frac{n!}{\prod_{ij} n_{ij}!} \prod_{ij} [P_c^n(I_1 = i, J_1 = j)]^{n_{ij}}, \end{aligned} \tag{3.59}$$

which, together with (3.53), yields the desired result. □

Next we establish a sharper characterization of the σ-field, M, in terms of the following class of interaction events. First observe that each realized interaction pattern, $\omega \in \Omega_n \subseteq \Omega, n \in Z_{++}$, generates a pattern of origin-destination pairs, $[I_1(\omega), J_1(\omega), \ldots, I_n(\omega), J_n(\omega)] \in (I \times J)^n$, which we designate as the associated *spatial interaction pattern.* If for each $n \in Z_{++}$, the set of possible *spatial n-interaction patterns* is denoted by $S_n = (I \times J)^n$, then as in Chapter 1, the (countable) set of spatial interaction patterns is given by $S = \cup_{n \geq 0} S_n$, where the singleton set, $S_0 = \{s_o\}$, contains the *null interaction pattern*, s_o, corresponding to the null event, $\Omega_0 = \{o\}$. Next, if for each spatial interaction pattern, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$, we let

$$\Omega(s) = \{\omega \in \Omega_n : I_r(\omega) = i_r, J_r(\omega) = j_r, r = 1, \ldots, n\} \in M \tag{3.60}$$

denote the corresponding *spatial interaction event* in M, and write $\Omega(s_o) = \Omega_0$, then the family of sets,

$$M_S = \{\Omega(s) : s \in S\} \cup \{\emptyset\} \subseteq M, \tag{3.61}$$

contains all spatial interaction events in M, together with the null interaction event, Ω_0, and the empty set, $\emptyset$. Finally, if for any interaction event, $A \in M$, we denote the subfamily of events in M_S contained in A by

$$M_S(A) = \{A' \in M_S : A' \subseteq A\} \tag{3.62}$$

and let $\cup[M_S] = \{\cup_k A_k : A_k \in M_S, k \in Z_+\}$ denote the family of all *countable unions* of sets in M_S, then as an alternative to Proposition 3.1 above, we now have the following characterization of M, which is more useful for the analysis and construction of measures, μ, on $\langle \Omega, M \rangle$:

Lemma 3.2 *The σ-field, $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J})$, is equivalently representable as*

$$M = \cup[M_S] \tag{3.63}$$

and each measure, μ, on $\langle \Omega, M \rangle$ is uniquely defined for all nonempty sets, $A \in M$, by

$$\mu(A) = \sum_{A' \in M_S(A)} \mu(A'). \tag{3.64}$$

PROOF: **(i)** To establish (3.63), observe first that since each set, $A \in M_S$, is M-measurable by construction, and since σ-fields are closed under countable unions, it follows at once that $\cup[M_S] \subseteq M$. Moreover, since $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J})$ by definition, it will also follow that $M \subseteq \cup[M_S]$ if $\cup[M_S]$ can be shown to be a σ-field on which all functions in $\boldsymbol{I} \cup \boldsymbol{J}$ are measurable. To do so, observe first that by definition, $\cup[M_S]$ is closed under countable unions and contains $\emptyset$. Hence, to establish that $\cup[M_S]$ is a σ-field, it suffices to show that $\cup[M_S]$ is closed under complements. To do so, observe that since $A \in M_S \Rightarrow A \subseteq \Omega \Rightarrow \cup\{A : A \in M_S\} \subseteq \Omega$ and since $\Omega_n = \cup\{\Omega(s) : s \in S_n\} \in \cup[M_S]$ for each $n \in Z_+$ implies $\Omega \subseteq \cup\{A : A \in M_S\}$, we must have $\Omega = \cup\{A : A \in M_S\}$. Moreover, since $A \cap A' = \emptyset$ for all distinct $A, A' \in M_S$, this in turn implies that M_S must be a countable M-measurable partition of Ω. Thus for any $A \in \cup[M_S]$, the complement, $\Omega - A$, must also be a countable union of sets in M_S, and hence in $\cup[M_S]$. Finally, to see that each function in $\boldsymbol{I} \cup \boldsymbol{J}$ is $\cup[M_S]$-measurable, consider any $I_r \in \boldsymbol{I}$ and observe that if for each spatial interaction pattern, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$, we write $i_r(s) = i_r$ and $j_r(s) = j_r, r = 1, \ldots, n$, then for every $i \in I$ and $r \in Z_{++}$, the inverse image set, $I_r^{-1}(i) = \cup_{n \geq r}[\cup\{\Omega(s) : s \in S_n, i_r(s) = i\}]$, is a countable union of sets in M_S, and hence in $\cup[M_S]$. Moreover, since $\Omega_n \in \cup[M_S]$ for each $n \in Z_+$ implies that $I_r^{-1}(o) = \cup_{n<r}\Omega_n \in \cup[M_S]$, it follows that I_r is $\cup[M_S]$-measurable for all $r \in Z_{++}$. Similarly, $J_r^{-1}(j) = \cup_{n \geq r}[\cup\{\Omega(s) : s \in S_n, j_r(s) = j\}] \in \cup[M_S]$ for all $j \in J$, which together with $J_r^{-1}(o) = I_r^{-1}(o) \in \cup[M_S]$, implies that each $J_r \in \boldsymbol{J}$ is also $\cup[M_S]$-measurable. Hence (3.63) must hold.

(ii) To establish (3.64), observe simply that since $A \cap A' = \emptyset$ for all distinct sets, $A, A' \in M_S$, and since (3.63) implies that $A = \cup\{A' : A' \in M_S(A)\}$

holds identically for all $A \in M$, the result follows at once from the countable additivity property of measures. □

Given these preliminary results, we are now ready to establish the desired correspondence between independent interaction processes and Poisson frequency processes:

Theorem 3.2 (Poisson Characterization Theorem) *An interaction process, $\boldsymbol{P}$, generates a Poisson frequency process, $\boldsymbol{N_P} \in \langle \text{POISSON} \rangle$, iff $\boldsymbol{P}$ is independent.*

PROOF: **(i)** Suppose that $\boldsymbol{P}$ is independent, i.e., that $\boldsymbol{P}$ satisfies A1 and A2. To verify that $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$, observe first from A2 that $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\}$ is an independent frequency process on $I \times J$. Moreover, by Lemma 3.1 it follows that for each $c \in C$ and $n \in Z_{++}(\boldsymbol{N}_c)$, the probability function, $p_c^n : I \times J \to [0,1]$, in (3.53) satisfies condition (3.48). Hence $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$ by Corollary 3.1.

(ii) To establish the converse, suppose that $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$, for some interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$. Then in particular, $\boldsymbol{N_P}$ is an independent frequency process, and it follows at once from (3.27), (3.30) and (3.46) that $\boldsymbol{P}$ satisfies A2. To establish that $\boldsymbol{P}$ also satisfies A1, we begin by constructing an explicit representation of $\boldsymbol{P}$ as follows. Observe first from Lemma 3.2 that each probability measure, P_c, is completely defined by its values on each set, $A \in M_S$. Moreover, since $P_c(\emptyset) = 0$ by definition, it follows from (3.61) that P_c is completely defined by its value on each spatial interaction event, $\Omega(s), s \in S_n, n \in Z_{++}$. To determine these values, observe that by definition, each (ij)-frequency function, N_{ij}, is constant on the set, $\Omega(s)$, i.e., that $N_{ij}(\omega) = N_{ij}(\omega')$ for all ω, $\omega' \in \Omega(s)$ and all $ij \in I \times J$. Hence, without loss of generality, we may write $N_{ij}(s) = N_{ij}(\omega)$ for any $\omega \in \Omega(s)$. Similarly, if the associated frequency profile for ω is denoted by $\boldsymbol{N}(\omega) = (N_{ij}(\omega) : ij \in I \times J)$, then we may now write $\boldsymbol{N}(s) = \boldsymbol{N}(\omega)$ for any $\omega \in \Omega(s)$. Next, if we denote typical frequency-profile realizations of $\boldsymbol{N}(\omega)$ by $\boldsymbol{n} = (n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$, and let

$$\Omega_{\boldsymbol{n}} = \{\omega \in \Omega : \boldsymbol{N}(\omega) = \boldsymbol{n}\} \tag{3.65}$$

denote the corresponding $\boldsymbol{n}$-event in M (i.e., the occurrence of an interaction pattern with frequency profile, $\boldsymbol{n}$) then the constancy of (ij)-frequencies on each set, $\Omega(s)$, implies from (3.62) that

$$M_S(\Omega_{\boldsymbol{n}}) = \{\Omega(s) \in M_S : \boldsymbol{N}(s) = \boldsymbol{n}\}. \tag{3.66}$$

Hence, if for any $s \in S_n$ with $\boldsymbol{N}(s) = \boldsymbol{n}$, we let $P_c[\Omega(s)|\Omega_{\boldsymbol{n}}]$ denote the conditional probability of $\Omega(s)$ given $\Omega_{\boldsymbol{n}}$, and observe from (3.66) that $\Omega(s) \cap \Omega_{\boldsymbol{n}} = \Omega(s)$, it then follows by definition that,

$$P_c[\Omega(s)] = P_c[\Omega(s) \cap \Omega_{\boldsymbol{n}}] = P_c[\Omega(s)|\Omega_{\boldsymbol{n}}]P_c(\Omega_{\boldsymbol{n}}). \tag{3.67}$$

But if we let $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\}$ denote the frequency process generated by $\boldsymbol{P}$, [where by definition, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J),\ c \in C$], and recall that by hypothesis, $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$, it then follows at once from (3.45) that for each realization, $\boldsymbol{n} = (n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$,

$$\begin{aligned} P_c(\Omega_{\boldsymbol{n}}) &= P_c(N_{ij} = n_{ij} : ij \in I \times J) \\ &= \prod_{ij} \frac{\mathrm{E}(N_{ij}^c)^{n_{ij}}}{n_{ij}!} \exp[-\mathrm{E}(N_{ij}^c)], \end{aligned} \tag{3.68}$$

where again by (3.27) and (3.30),

$$\mathrm{E}(N_{ij}^c) = \mathrm{E}_c(N_{ij}), \quad c \in C,\ ij \in I \times J. \tag{3.69}$$

Finally, to evaluate $P_c[\Omega(s)|\Omega_{\boldsymbol{n}}]$, observe that for any interaction patterns, $s = (i_r, j_r : r = 1, \ldots, n)$, $s' = (i'_r, j'_r : r = 1, \ldots, n) \in S_n$ with $\boldsymbol{N}(s) = \boldsymbol{N}(s')$, there is some permutation, $\pi \in \Pi_n$, such that

$$(i'_r, j'_r) = (i_{\pi_r}, j_{\pi_r}),\ r = 1, \ldots, n.$$

Hence, for any n-interaction pattern, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$, we must have

$$\begin{aligned} \omega \in \Omega(s) &\Leftrightarrow [\boldsymbol{i}(\omega_r) = i_r, \boldsymbol{j}(\omega_r) = j_r : r = 1, \ldots, n] \\ &\Leftrightarrow [\boldsymbol{i}(\omega_{\pi_r}) = i_{\pi_r}, \boldsymbol{j}(\omega_{\pi_r}) = j_{\pi_r} : r = 1, \ldots, n] \\ &\Leftrightarrow \omega_\pi = (\omega_{\pi_1}, \ldots, \omega_{\pi_n}) \in \Omega(s'), \end{aligned} \tag{3.70}$$

which implies that $\Omega(s') = \Omega(s)_\pi$ for this choice of $\pi \in \Pi_n$. But by the symmetry property (R2) of P_c, it then follows that

$$P_c[\Omega(s)] = P_c[\Omega(s)_\pi] = P_c[\Omega(s')], \tag{3.71}$$

for all $s, s' \in S$ with $\boldsymbol{N}(s) = \boldsymbol{N}(s')$. Hence for each $\Omega(s) \in M_S(\Omega_{\boldsymbol{n}})$, it follows at once from (3.64), (3.66), and (3.71) that

$$P_c[\Omega(s)|\Omega_{\boldsymbol{n}}] = \frac{P_c[\Omega(s)]}{\sum_{\Omega(s') \in M_S(\Omega_{\boldsymbol{n}})} P_c[\Omega(s')]} = \frac{1}{|M_S(\Omega_{\boldsymbol{n}})|}. \tag{3.72}$$

Finally, observing that the number of *distinct* elements, $s \in S_n = (I \times J)^n$ with $\boldsymbol{N}(s) = \boldsymbol{n}$ is given by $|M_S(\Omega_{\boldsymbol{n}})| = n!/\prod_{ij}(n_{ij})!$, it follows by combining (3.67), (3.68), and (3.72) that for all $s \in S_n$ with $\boldsymbol{N}(s) = \boldsymbol{n}$,

$$\begin{aligned} P_c[\Omega(s)] &= \left[\frac{n!}{\prod_{ij} n_{ij}!}\right]^{-1} \prod_{ij} \left[\frac{\mathrm{E}(N_{ij}^c)^{n_{ij}}}{n_{ij}!} \exp[-\mathrm{E}(N_{ij}^c)]\right] \\ &= \frac{\prod_{ij} n_{ij}!}{n!} \exp[-\sum_{ij} \mathrm{E}(N_{ij}^c)] \prod_{ij} \frac{\mathrm{E}(N_{ij}^c)^{n_{ij}}}{n_{ij}!} \\ &= \frac{\mathrm{E}(N^c)^n}{n!} \exp[-\mathrm{E}(N^c)] \prod_{ij} \left[\frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N^c)}\right]^{n_{ij}}. \end{aligned} \tag{3.73}$$

Given this representation of $\boldsymbol{P}$, we can now verify that $\boldsymbol{P}$ satisfies A1. To do so, observe first from (3.73) together with (3.52) that for all spatial interaction patterns, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$,

$$\begin{aligned} P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) &= P_c^n[\Omega(s)] \\ &= P_c[\Omega(s)]/P_c(N = n) = \prod_{ij} [\mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c)]^{n_{ij}} \\ &= \prod_{r=1}^{n} [\mathrm{E}(N_{i_r j_r}^c)/\mathrm{E}(N^c)]. \end{aligned} \tag{3.74}$$

Finally, if for any $r \in \{1, \ldots, n\}$ we sum (3.74) over all values of $(I_{r'}, J_{r'})$ with $r' \neq r$, then it follows from the identity, $\mathrm{E}(N^c) = \sum_{ij} \mathrm{E}(N_{ij}^c)$, that the joint distribution of (I_r, J_r) is given for all $(i_r, j_r) \in I \times J$ by

$$P_c^n(I_r = i_r, J_r = j_r) = \mathrm{E}(N_{i_r j_r}^c)/\mathrm{E}(N^c). \tag{3.75}$$

Hence by substituting (3.75) into (3.74), we may conclude that $\boldsymbol{P}$ satisfies A1 as well. □

The representation in (3.73) of each $\boldsymbol{P}$ with $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$ has a number of additional consequences. First of all, it shows that each Poisson frequency process, $\boldsymbol{N}$, is generated by *exactly one* interaction process on $\langle\Omega, M\rangle$. More precisely, if for each $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ with positive mean frequencies, $\mathrm{E}(N_{ij}^c), ij \in I \times J$ [satisfying the continuity condition (3.29)], we let $\boldsymbol{P_N} = \{P_c : c \in C\}$ denote the family of probability measures on $\langle\Omega, M\rangle$ generated by (3.73) together with (3.61) and (3.64), then it follows at once from Definition 3.1 that $\boldsymbol{P_N}$ is the *unique* interaction process which generates $\boldsymbol{N}$. Thus, if for each spatial interaction pattern, $s \in S_n$, we now set $N(s) = n = \sum_{ij} N_{ij}(s)$, then this observation motivates the following definition:

Definition 3.5 For each $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ with mean frequencies, $\mathrm{E}(N_{ij}^c)$, $ij \in I \times J$, and $\mathrm{E}(N^c) = \sum_{ij} \mathrm{E}(N_{ij}^c), c \in C$, the interaction process $\boldsymbol{P_N} = \{P_c : c \in C\}$ on $\langle\Omega, M\rangle$, defined for all separation configurations, $c \in C$, and spatial interaction patterns, $s \in S$, by

$$P_c(s) = \frac{\mathrm{E}(N^c)^{N(s)}}{N(s)!} \exp[-\mathrm{E}(N^c)] \prod_{ij} \left[\frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N^c)}\right]^{N_{ij}(s)} \tag{3.76}$$

is designated as the *generator* of $\boldsymbol{N}$.

As a second consequence of the representation in (3.73) [or (3.76)], it also follows that the class of Poisson frequency processes must be identical with the class of frequency processes generated by independent interaction processes. More precisely, if we now let $\langle\text{A1, A2}\rangle$ denote the class of frequency processes generated by interaction processes satisfying A1 and A2, then:

Corollary 3.2 $\langle \text{POISSON} \rangle = \langle \text{A1}, \text{A2} \rangle$.

PROOF: If $\boldsymbol{N} \in \langle \text{A1}, \text{A2} \rangle$, then $\boldsymbol{N} = \boldsymbol{N_P}$ for some independent interaction process, $\boldsymbol{P}$. Hence, $\boldsymbol{N} \in \langle \text{POISSON} \rangle$ by Theorem 3.2, so that $\langle \text{A1}, \text{A2} \rangle \subseteq \langle \text{POISSON} \rangle$. Conversely, if $\boldsymbol{N} \in \langle \text{POISSON} \rangle$, then again by Theorem 3.2, the generator, $\boldsymbol{P_N}$, of $\boldsymbol{N}$ satisfies A1 and A2, so that by definition, $\boldsymbol{N} \in \langle \text{A1}, \text{A2} \rangle$. Hence we see that $\langle \text{POISSON} \rangle \subseteq \langle \text{A1}, \text{A2} \rangle$ and the result is established. □

Finally, it is convenient to restate Corollary 3.1 above in a form which is more useful for analysis. If for any Poisson frequency process, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, we now define the associated (positive) *interaction probability functions,* $p_c : I \times J \to (0,1)$, for each $c \in C$ by

$$p_c(ij) = \mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c), \quad ij \in I \times J, \tag{3.77}$$

then it follows that once that:

Corollary 3.3 (Multinomial Sampling)
(i) *If* $\boldsymbol{N} \in \langle \text{POISSON} \rangle$, *then for each* $c \in C, n \in Z_{++}$, *and* $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ *with* $\sum_{ij} n_{ij} = n$,

$$P(N_{ij}^c = n_{ij} : \ ij \in I \times J | N^c = n) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}. \tag{3.78}$$

(ii) *In particular, if* $\boldsymbol{N}$ *is generated by a given independent interaction process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *then the interaction probability functions,* p_c, *in* (3.77) *are related to the corresponding probability measures,* P_c, *by the following identity for all* $c \in C, ij \in I \times J$, *and* $r \le n \in Z_{++}$,

$$P_c^n(I_r = i, J_r = j) = p_c(ij) = \mathrm{E}_c[N_{ij}/N | N > 0]. \tag{3.79}$$

PROOF: **(i)** Relation (3.78) follows immediately from parts (i) and (ii) of Corollary 3.1 together with (3.77).

(ii) The first equality in (3.79) follows from (3.75) and (3.77). To establish the second equality in (3.79), observe first from the properties of the multinomial distribution in (3.78) that

$$\mathrm{E}(N_{ij}^c/N^c | N^c = n) = (1/n) \cdot \mathrm{E}(N_{ij}^c | N^c = n) = p_c(ij),$$

and hence from the definition of generated frequency processes that

$$\begin{aligned} \mathrm{E}_c(N_{ij}/N | N > 0) &= \mathrm{E}(N_{ij}^c/N^c | N^c > 0) \\ &= \sum_{n>0} \mathrm{E}(N_{ij}^c/N^c | N^c = n) P(N^c = n | N^c > 0) = p_c(ij). \end{aligned}$$

□

3.7 Threshold Interaction Processes

As a first application of the general probabilistic framework above, we now develop a general class of *threshold interaction processes* which extend those developed in Example 2 of Chapter 2 above. [The results established here will be employed in Section 4.3.2 below to obtain gravity model representations of the special class of *fully independent* threshold interaction processes developed in Chapter 2.] From a formal viewpoint, the present class of processes involves an extension of the basic set of interaction attributes to include a set of 'threshold attributes' describing certain behavioral attitudes of actors toward spatial separation. Hence we first formalize an appropriate probability space for analyzing such attributes, and then establish the distributional consequences of certain independence assumptions on such attributes.

3.7.1 Potential Interactions

In the present model each element, ω, of the individual interaction space, Ω_1, in Section 3.3.1 is now interpreted as a *potential interaction situation* in which the actor is currently considering the possibility of an interaction with some spatial opportunity. If the relevant locational attributes of this actor and opportunity are denoted respectively by $i = \boldsymbol{i}(\omega)$ and $j = \boldsymbol{j}(\omega)$, then for any prevailing separation configuration, $c \in C$, the relevant separation profile for situation, ω, is $c_{ij} = (c_{ij}^k : k \in K) \in V \subseteq R^K$, as in Section 3.4.1. In this context, it is postulated that interaction situation, ω, results in a realized interaction between i and j iff no separation value, c_{ij}^k, exceeds the actor's current maximum tolerable level, t_k, for separation of type k. Hence, along with the locational attribute functions, $\boldsymbol{i} : \Omega_1 \to I$ and $\boldsymbol{j} : \Omega_1 \to J$, we now consider a *threshold attribute function,* $\boldsymbol{t} : \Omega_1 \to R^K$, in which each *threshold vector,* $\boldsymbol{t}(\omega) = (\boldsymbol{t}_k(\omega) : k \in K)$, represents the current maximum tolerable levels of each k-type separation for the actor in potential interaction situation, $\omega \in \Omega_1$.

If the set, Ω, generated by Ω_1 in (3.1) is now taken to represent the space of all finite *potential interaction patterns,* $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n \subseteq \Omega$, then the threshold attribute function, $\boldsymbol{t}$, may be extended to all of Ω in a manner paralleling (3.2) and (3.3). In particular, if the zero vector in R^K is denoted by $0 = (0, \ldots, 0)$, then the desired family of *threshold functions,* $\boldsymbol{T} = \{T_r : \Omega \to R^K | r \in Z_{++}\}$ can now be defined for all $r \in Z_{++}, n \in Z_+$, and $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ by

$$T_r(\omega) = \begin{cases} \boldsymbol{t}(\omega_r), & r \leq n \\ 0, & r > n. \end{cases} \tag{3.80}$$

[Note that the convention, $T_r(\omega) = 0$ for $r > n$, serves only to complete the formal definition of each threshold function, T_r, on Ω and plays no role in the analysis to follow. In particular, the value of $T_r(\omega)$ only has behavioral meaning for interaction patterns, $\omega \in \Omega_n$, with $r \leq n$.]

Given this extended family of attributes, $\boldsymbol{I} \cup \boldsymbol{J} \cup \boldsymbol{T}$, for potential interaction situations in Ω, it should be clear from the introductory discussion to Section 3.3 above that the appropriate σ-field of potential interaction events, $\overline{M}$, is now given by

$$\overline{M} = \sigma(\boldsymbol{I} \cup \boldsymbol{J} \cup \boldsymbol{T}), \tag{3.81}$$

where the σ-field for the range space of each function, $T_r : \Omega \to R^K$, is implicitly taken to be the *Borel σ-field*, $\boldsymbol{B}(R^K)$. Moreover, it should be clear from (3.4) together with the definition of minimal σ-fields that $M \subseteq \overline{M}$, and hence that every M-measurable function is by definition $\overline{M}$-measurable. As a parallel to Proposition 3.1 and Lemma 3.2 above, the σ-field, $\overline{M}$, can be given a more explicit characterization as follows. First recall from Section 3.2.2 that T_r is $\overline{M}$-measurable iff $T_r^{-1}([t, \infty)) \in \overline{M}$ for all $t \in R^K$. Hence if for each $r \in Z_{++}, n \in Z_+$ and $t \in R^K$ we set

$$\Omega_n^r(t) = \{\omega \in \Omega_n : T_r(\omega) \geq t\} \tag{3.82}$$

and observe that $T_r^{-1}([t, \infty)) = \cup_{n \geq 0} \Omega_n^r(t)$ by definition, it follows (from the mutual disjointness of the sets Ω_n) that T_r is $\overline{M}$-measurable iff $\Omega_n^r(t) \in \overline{M}$ for all $n \in Z_+$ and $t \in R^K$. But if $n < r$ then $\Omega_n^r(t) = \Omega_n$ for $t \leq 0$ and $\Omega_n^r(t) = \emptyset$ otherwise. Moreover since $\Omega_n = \cup_{z \in Z^K} \Omega_n^n(z)$ for all $n > 0$, it then follows from the σ-field properties of $\overline{M}$ that each function T_r is $\overline{M}$-measurable iff

$$\{\Omega_0\} \cup \{\Omega_n^r(t) : r \leq n, t \in R^K\} \subset \overline{M}. \tag{3.83}$$

Thus, if for each $n \in Z_{++}$, $i^n = (i_1, \ldots, i_n) \in I^n$, $j^n = (j_1, \ldots, j_n) \in J^n$ and $t^n = (t_1, \ldots, t_n) \in (R^K)^n$ we let

$$\Omega(i^n, j^n) = \{\omega \in \Omega : I_r(\omega) = i_r, J_r(\omega) = j_r, r = 1, \ldots, n\} \tag{3.84}$$

and let

$$\Omega(i^n, j^n, t^n) = \Omega(i^n, j^n) \cap [\cap_{r=1}^n \Omega_n^r(t_r)], \tag{3.85}$$

then by setting $\overline{M}_0 = M_0 = \{\emptyset, \Omega_0\}$ and

$$\overline{M}_n = \sigma\{\Omega(i^n, j^n, t^n) : i^n \in I^n, j^n \in J^n, t^n \in (R^K)^n\} \tag{3.86}$$

for each $n \in Z_{++}$, we obtain the following parallel to Proposition 3.1:

Proposition 3.4 *The σ-field, $\overline{M} = \sigma(\boldsymbol{I} \cup \boldsymbol{J} \cup \mathbf{T})$ is equivalently representable as*

$$\overline{M} = \sigma(\cup_{n \geq 0} \overline{M}_n). \tag{3.87}$$

PROOF: To establish that $\overline{M} \subseteq \sigma(\cup_n \overline{M}_n)$, observe first from (3.82) and (3.85) that for all $r \leq n \in Z_{++}$ and $t \in R^K$,

$$\Omega_n^r(t) = \cup_{i^n \in I^n} \cup_{j^n \in J^n} \cup_{z \in Z^K} \Omega[i^n, j^n, (z, \ldots, t, \ldots, z)] \in \sigma(\cup_n \overline{M}_n),$$

which together with $\Omega_0 \in \overline{M_0}$ implies from (3.83) that all attribute functions in $\boldsymbol{T}$ are $\sigma(\cup_n \overline{M}_n)$-measurable. Next observe that

$$\Omega(i^n, j^n) = \cup_{z \in Z^K} \Omega[i^n, j^n, (z, ..., z)] \in \sigma(\cup_n \overline{M}_n),$$

which together with (3.84) and (3.61) through (3.64) implies that all attribute functions in $\boldsymbol{I} \cup \boldsymbol{J}$ are also $\sigma(\cup_n \overline{M}_n)$-measurable, and hence by definition that $\overline{M} \subseteq \sigma(\cup_n \overline{M}_n)$. Conversely, since $\Omega(i^n, j^n, t^n) = \cap_{r=1}^n [I_r^{-1}(i_r) \cap J_r^{-1}(j_r) \cap T_r^{-1}[t_r, \infty)]$ for each $i^n = (i_1, \ldots, i_n) \in I^n, j^n = (j_1, \ldots, j_n) \in J^n$, and $t^n = (t_1, \ldots, t_n) \in (R^K)^n$, the $\overline{M}$-measurability of all attributes in $\boldsymbol{I} \cup \boldsymbol{J} \cup \boldsymbol{T}$ implies from (3.86) that $\sigma(\cup_n \overline{M}_n) \subseteq \overline{M}$. □

Given the measurable space, $\langle \Omega, \overline{M} \rangle$, we next specify regularity conditions on admissible probability measures, P, for $\langle \Omega, \overline{M} \rangle$ in a manner paralleling Definition 3.1 above. To do so, let us begin by observing from the $\overline{M}$-measurability of all locational attributes in $\boldsymbol{I} \cup \boldsymbol{J}$ that each (ij)-frequency function, $N_{ij} : \Omega \to Z_+$, in (3.11) is also $\overline{M}$-measurable, and hence yields a well-defined random variable on every probability space, $\langle \Omega, \overline{M}, P \rangle$. Thus the positivity condition, R1, for probability measures on $\langle \Omega, M \rangle$ is also meaningful for probability measures on $\langle \Omega, \overline{M} \rangle$. Similarly, the symmetry condition, R2, is also meaningful for probability measures on $\langle \Omega, \overline{M} \rangle$, since Proposition 3.3 continues to hold for $\overline{M}$, as we now show:

Proposition 3.5 *For all $n \in Z_{++}, \pi \in \Pi_n$, and $A \in \Omega_n$,*

$$A \in \overline{M}_n \Rightarrow A_\pi \in \overline{M}_n. \tag{3.88}$$

PROOF: To establish (3.88), observe that if we set $\overline{M}_n^\pi = \{A \in \overline{M}_n : A_\pi \in \overline{M}_n\} \subseteq \overline{M}_n$, then the same argument as in Proposition 3.3 shows that $\overline{M}_n^\pi$ is a σ-field on Ω_n. Hence if it can be shown that $\Omega(i^n, j^n, t^n) \in \overline{M}_n^\pi$ for each $(i^n, j^n, t^n) \in I^n \times J^n \times (R^K)^n$, then by (3.86) we will have $\overline{M}_n \subseteq \sigma(\overline{M}_n^\pi) = \overline{M}_n^\pi$, and the result will follow. But if for each $i^n = (i_1, \ldots, i_n) \in I^n$ we set $i_\pi^n = (i_{\pi_1}, \ldots, i_{\pi_n}) \in I^n$, and similarly, define $j_\pi^n \in J^n$ and $t_\pi^n \in (R^K)^n$, then it follows at once from the symmetry of the product set, $I^n \times J^n \times (R^K)^n$, that $\Omega(i^n, j^n, t^n)_\pi = \Omega(i_\pi^n, j_\pi^n, t_\pi^n) \in \overline{M}_n$, and hence that $\Omega(i^n, j^n, t^n) \in \overline{M}_n^\pi$ for all $(i^n, j^n, t^n) \in I^n \times J^n \times (R^K)^n$. □

Hence, if P_c^n again denotes the conditional probability measure defined by (3.24) above, then in a manner paralleling Definition 3.3, we now say that:

Definition 3.6 A family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle\Omega, \overline{M}\rangle$ is designated as a *threshold interaction process (TI-process)* on $\langle\Omega, \overline{M}\rangle$ iff $\boldsymbol{P}$ satisfies the following four conditions for all $c \in C$, all $r \leq n \in Z_{++}$ with $P_c(\Omega_n) > 0$, all $\pi \in \Pi_n$, $A \in \overline{M}_n$, $t \in R^K$, $ij \in I \times J$, and all sequences, $\{c_m : m \in Z_+\}$ on C:

P1. (Frequency Positivity) $P_c(N_{ij} \neq 0) > 0$.

P2. (Threshold Positivity) $P_c^n(T_r \geq t) > 0$.

P3. (Symmetry) $P_c(A) = P_c(A_\pi)$.

P4. (Continuity) $c_m \to c \Rightarrow \mathrm{E}_{c_m}(N_{ij}) \to \mathrm{E}_c(N_{ij})$.

Conditions P1, P3, and P4 are completely parallel to conditions R1, R2, and R3 of Definition 3.1, respectively [and hence $\boldsymbol{P}$ is an *interaction process* in the sense of Definition 3.1]. The threshold positivity condition, P2, ensures that realized interactions between origin-destination pair are always possible. [This condition is not essential from a behavioral viewpoint, but serves rather to simplify the analysis by avoiding the need for special analysis of degenerate cases.]

3.7.2 Independent Threshold Interaction Processes

The locational-independence and frequency-independence axioms for interaction processes in Section 3.4.3 above continue to be meaningful for TI-processes. In addition, we now postulate that threshold values of actors are independent of both locational considerations and the properties of other actors. More precisely:

Definition 3.7 A TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle\Omega, \overline{M}\rangle$ is designated as an *independent TI-process* iff $\boldsymbol{P}$ satisfies the following three independence conditions for all $c \in C$ and $n \in Z_{++}$ with $P_c(\Omega_n) > 0$,

C1. (Locational Independence)*For all* $i_1, \ldots, i_n \in I$ *and* $j_1, \ldots, j_n \in J$,

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) = \prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r). \quad (3.89)$$

C2. (Frequency Independence)*For all* $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$,

$$P_c(N_{ij} = n_{ij} : ij \in I \times J) = \prod_{ij} P_c(N_{ij} = n_{ij}). \quad (3.90)$$

C3. (Threshold Independence) *For all* $t_1, \ldots, t_n \in R^K, i_1, \ldots, i_n \in I$, *and* $j_1, \ldots, j_n \in J$ *with* $P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) > 0$,

$$\begin{aligned} P_c^n(T_1 \geq t_1, \ldots, T_n \geq t_n | I_r = i_r, J_r = j_r, r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^1(T_1 \geq t_r). \end{aligned} \tag{3.91}$$

[Note in particular that C3 asserts that the threshold values of actors are independent of the *number* of other actors present]. To analyze these processes, we first establish the following consequences of conditions C1 and C3:

Lemma 3.3

(i) *If a TI-process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *satisfies* C1, *then for all* $i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J$, *and all* $c \in C$ *and* $n \in Z_{++}$ *with* $P_c(\Omega_n) > 0$,

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) = \prod_{r=1}^{n} P_c^n(I_1 = i_r, J_1 = j_r). \tag{3.92}$$

(ii) *If, in addition,* $\boldsymbol{P}$, *satisfies* C3, *then for all Borel sets,* $B_1, \ldots, B_n \in \boldsymbol{B}(R^K)$,

$$\begin{aligned} P_c^n(I_r = i_r, J_r = j_r, T_r \in B_r : r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^n(I_1 = i_r, J_1 = j_r) P_c^1(T_1 \in B_r). \end{aligned} \tag{3.93}$$

PROOF: **(i)** By Proposition 3.5, the argument in (3.58) is valid for each probability space, $\langle \Omega_n, \overline{M}_n, P_c^n \rangle$. Hence (3.92) follows at once by combining (3.58) and condition C1.

(ii) To establish (3.93), observe first from C1 and C3 that since

$$\begin{aligned} P_c^n(T_1 \geq t_1, \ldots, T_n \geq t_n, I_1 = i_1, \ldots, I_n = i_n, J_1 = j_1, \ldots J_n = j_n) \\ = [\prod_{r=1}^{n} P_c^1(T_1 \geq t_r)] \cdot [\prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r)], \end{aligned}$$

for all $t_1, \ldots, t_n \in R^K$, $i_1, \ldots, i_n \in I$, $j_1, \ldots, j_n \in J$ and $n \in Z_{++}$ with $P_c(\Omega_n) > 0$, it follows at once from the definitions in Section 3.2.3 above that the Ω_n-restrictions of the functions, $T_r : \Omega \to R^K$ and $(I_r, J_r) : \Omega \to I_o \times J_o$, yield a family of independent random variables,

$$\{T_1, \ldots, T_n, (I_1, J_1), \ldots, (I_n, J_n)\},$$

on the (conditional) probability space, $\langle \Omega_n, \overline{M}_n, P_c^n \rangle$ [where in particular, each conditional random variable, (I_r, J_r), is seen from (3.2) and (3.3) to be

$I \times J$-valued]. Hence, the general definition of independence in Section 3.2.3 together with (3.92) implies that

$$\begin{aligned} P_c^n(I_r = i_r, J_r = j_r, T_r \in B_r : r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^n(I_1 = i_r, J_1 = j_r) P_c^n(T_r \in B_r), \end{aligned} \tag{3.94}$$

for all $i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J$, and Borel sets $B_1, \ldots, B_n \in \boldsymbol{B}(R^K)$. Finally, since all probabilities $P_c^n(T_r \in B), B \in \boldsymbol{B}(R^K)$, are uniquely determined by the family of probabilities $\{P_c^n(T_r \geq t) : t \in R^K\}$, and since C3 implies that $P_c^n(T_r \geq t) = P_c^1(T_1 \geq t)$ for all $r \leq n \in Z_{++}$ and $t \in R^K$, we may conclude that

$$P_c^n(T_r \in B) = P_c^1(T_1 \in B), \tag{3.95}$$

for all $r \leq n \in Z_{++}$ and $B \in \boldsymbol{B}(R^K)$, and hence by substituting (3.95) into (3.94) that (3.93) must hold. □

Given these general properties, our first result is to show that for any given TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, the corresponding potential interaction frequencies, $(N_{ij} : ij \in I \times J)$, are Poisson distributed. In particular, if the associated *total threshold interaction frequency* is denoted by $N = \sum_{ij} N_{ij}$, and if $\mathrm{E}_c(N_{ij})$ and $\mathrm{E}_c(N)$ denote the respective means of N_{ij} and N under distribution, P_c, then:

Theorem 3.3 *For any independent TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, and separation configuration, $c \in C$, each frequency function, $N_{ij} : \Omega \to Z_+, ij \in I \times J$, defines a Poisson random variable on $\langle \Omega, \overline{M}, P_c \rangle$ with positive mean frequency*

$$\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(N) P_c^1(I_1 = i, J_1 = j). \tag{3.96}$$

PROOF: To establish that the frequency functions, $(N_{ij} : ij \in I \times J)$, are Poisson distributed, observe first that since $M \subseteq \overline{M}$, it follows at once from a comparison of conditions (R1,R2,R3,A1,A2) and (P1,P3,P4,C1,C2) that the M-restriction of each probability measure, $P_c : \overline{M} \to [0, 1]$, in $\boldsymbol{P}$ yields a family of probability measures, $\{P_c : c \in C\}$, on $\langle \Omega, M \rangle$ which satisfy all the conditions of independent interaction processes in Definitions 3.1 and 3.2 above. Hence, by Theorem 3.2, the random variables, $(N_{ij} : ij \in I \times J)$, must necessarily be Poisson distributed with positive finite means. Finally, to establish (3.96), it suffices to observe from (3.49), (3.53), and (3.54) together with the properties of multinomial distributions that $\mathrm{E}_c(N_{ij} | N = n) = np_c^n(ij) = np_c^1(ij) = nP_c^1(I_1 = i, J_1 = j)$ and hence that for all $c \in C$

and $ij \in I \times J$,

$$\begin{aligned}
\mathrm{E}_c(N_{ij}) &= \sum_n \mathrm{E}_c(N_{ij} \mid N = n) P_c(N = n) \\
&= \sum_n n P_c^1(I_1 = i, J_1 = j) P_c(N = n) \\
&= P_c^1(I_1 = i, J_1 = j) \sum_n n P_c(N = n) \\
&= P_c^1(I_1 = i, J_1 = j) \mathrm{E}_c(N).
\end{aligned} \tag{3.97}$$

□

3.7.3 Threshold Frequency Processes

In a manner paralleling (3.10) above, for each $c \in C, r \in Z_+$, and $ij \in I \times J$, let the *indicator function*, $\delta_{ij}^{cr} : \Omega \to \{0, 1\}$, be defined for all $\omega \in \Omega$ by

$$\delta_{ij}^{cr}(\omega) = \begin{cases} 1, & I_r(\omega) = i, J_r(\omega) = j, \text{ and } T_r(\omega) \geq c_{ij} \\ 0, & \text{otherwise.} \end{cases} \tag{3.98}$$

The $\overline{M}$-measurability of all functions in $\boldsymbol{I} \cup \boldsymbol{J} \cup \boldsymbol{T}$ implies at once that all indicator functions in (3.98) are $\overline{M}$-measurable. Hence, in manner paralleling (3.11), the associated Z_+-valued functions, $N_{ij} : \Omega \to Z_+$, defined for all $\omega \in \Omega$ by

$$N_{ij}^c(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^n \delta_{ij}^{cr}(\omega), & \omega \in \Omega_n, n \in Z_{++}, \end{cases} \tag{3.99}$$

are also seen to be $\overline{M}$-measurable. Thus, for any given TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, and each $c \in C$, and $ij \in I \times J$, the function, N_{ij}^c, defines a Z_+-valued random variable on $\langle \Omega, \overline{M}, P_c \rangle$, which we now designate as the *threshold frequency* of realized interactions between actors at location, i, and opportunities at locations, j, under configuration, c. Our main result is then to show that these observable interaction frequencies are also Poisson distributed. To state this result in manner paralleling Theorem 3.2, let us now designate the Z_+-valued random vector, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, defined by (3.99) as the *threshold frequency profile* under configuration, c, and designate the family, $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\}$, of such profiles as the *threshold frequency process* generated by TI-process, $\boldsymbol{P}$. Finally, if N again denotes the total potential interaction frequency, as in Theorem 3.3, then we now have:

Theorem 3.4 (Threshold Frequency Theorem) *For each independent TI-process, $\boldsymbol{P}$, the threshold frequency process, $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\}$, generated by $\boldsymbol{P}$ is a Poisson process on $I \times J$, with positive finite mean frequencies* *given for all $c \in C$ and $ij \in I \times J$ by*

$$\mathrm{E}(N_{ij}^c) = \mathrm{E}_c(N) P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \geq c_{ij}). \tag{3.100}$$

PROOF: To establish that $\boldsymbol{N_P} \in \langle \text{POISSON} \rangle$, we begin by letting the random variable, $\overline{N}_{ij}^c = N_{ij} - N_{ij}^c$, denote the difference between the potential and realized (ij)-interactions under configuration, c, so that by definition, $N = \sum_{ij}(N_{ij}^c + \overline{N}_{ij}^c)$. Next, if we define the Borel sets, $B_{ij}^c = \{t \in R^K : t \geq c_{ij}\}$ and $\overline{B}_{ij}^c = R^K - B_{ij}^c$, then for any realized frequencies, $(n_{ij} : ij \in I \times J), (m_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ with $n = \sum_{ij}(n_{ij} + m_{ij})$, the event, $(N_{ij}^c = n_{ij}, \overline{N}_{ij}^c = m_{ij} : ij \in I \times J) \in \overline{M}$, is precisely the union of all (disjoint) events, $(I_r = i_r, J_r = j_r, T_r \in B_r : r = 1, \ldots, n) \in \overline{M}_n$, satisfying the condition that for each $ij \in I \times J, (i_r, j_r, B_r) = (i, j, B_{ij}^c)$ and $(i_r, j_r, B_r) = (i, j, \overline{B}_{ij}^c)$ hold for exactly n_{ij} and m_{ij} values of r, respectively. But since Theorem 3.3 together with (3.79) implies that

$$P_c^n(I_r = i, J_r = j) = P_c^1(I_1 = i, J_1 = j), \tag{3.101}$$

for all $r \leq n$ and $ij \in I \times J$ it follows from (3.93) that each of these disjoint events has the same conditional probability given by

$$\begin{aligned} P_c^n(I_r &= i_r, J_r = j_r, T_r \in B_r : r = 1, \ldots, n) \\ &= \prod_{r=1}^{n} P_c^1(I_1 = i_r, J_1 = j_r) P_c^1(T_1 \in B_r) \\ &= \prod_{ij} p_c(ij)^{n_{ij}} q_c(ij)^{m_{ij}}, \end{aligned} \tag{3.102}$$

where $p_c(ij)$ and $q_c(ij)$ are defined for all $ij \in I \times J$ by

$$p_c(ij) = P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \in B_{ij}^c), \tag{3.103}$$

$$q_c(ij) = P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \in \overline{B}_{ij}^c). \tag{3.104}$$

Moreover, since there are exactly $n!/\prod_{ij} n_{ij}! m_{ij}!$ such events, it may be concluded [in a manner paralleling (3.59) above] that for all $n \in Z_{++}$ and all $(n_{ij} : ij \in I \times J), (m_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ with $n = \sum_{ij}(n_{ij} + m_{ij})$,

$$P_c^n(N_{ij}^c = n_{ij}, \overline{N}_{ij}^c = m_{ij} : ij \in I \times J) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!} \frac{q_c(ij)^{m_{ij}}}{m_{ij}!}. \tag{3.105}$$

Finally, since the total potential interaction frequency, $N = \sum_{ij} N_{ij}$, is Poisson distributed by Theorem 3.3 together with (3.52), and since (3.103) and (3.104) imply that $\sum_{ij}[p_c(ij) + q_c(ij)] = 1$ for each $c \in C$, it follows from (3.105) that

$$\begin{aligned}
P_c(N_{ij}^c &= n_{ij}, \overline{N}_{ij}^c = m_{ij} : ij \in I \times J) \\
&= P_c^n(N_{ij}^c = n_{ij}, \overline{N}_{ij}^c = m_{ij} : ij \in I \times J) \cdot P_c(N = n) \\
&= n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!} \cdot \frac{q_c(ij)^{m_{ij}}}{m_{ij}!} \left(\frac{\mathrm{E}_c(N)^n}{n!} \exp[-\mathrm{E}_c(N)] \right) \\
&= \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!} \cdot \frac{q_c(ij)^{m_{ij}}}{m_{ij}!} \cdot \mathrm{E}_c(N)^{\Sigma_{ij}(n_{ij}+m_{ij})} \\
&\qquad \cdot \exp[-\mathrm{E}_c(N) \sum_{ij} (p_c(ij) + q_c(ij))] \\
&= \prod_{ij} \frac{[p_c(ij)\mathrm{E}_c(N)]^{n_{ij}}}{n_{ij}!} \cdot \frac{[q_c(ij)\mathrm{E}_c(N)]^{m_{ij}}}{m_{ij}!} \\
&\qquad \cdot \prod_{ij} \exp[-p_c(ij)\mathrm{E}_c(N)] \cdot \exp[-q_c(ij)\mathrm{E}_c(N)] \\
&= \prod_{ij} \left\{ \frac{[p_c(ij)\mathrm{E}_c(N)]^{n_{ij}}}{n_{ij}!} \exp[-p_c(ij)\mathrm{E}_c(N)] \right\} \\
&\qquad \cdot \left\{ \frac{[q_c(ij)\mathrm{E}_c(N)]^{m_{ij}}}{m_{ij}!} \exp[-q_c(ij)\mathrm{E}_c(N)] \right\}.
\end{aligned} \tag{3.106}$$

Hence by taking marginals of this joint distribution, it follows immediately from (3.103) that each N_{ij}^c is Poisson distributed with mean

$$\begin{aligned}
\mathrm{E}_c(N_{ij}^c) &= p_c(ij)\mathrm{E}_c(N) \\
&= \mathrm{E}_c(N) P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \in B_{ij}^c)
\end{aligned} \tag{3.107}$$

and similarly from (3.104) that each $\bar{N}_{ij}^c$ is Poisson distributed with mean

$$\begin{aligned}
\mathrm{E}_c(\overline{N}_{ij}^c) &= q_c(ij)\mathrm{E}_c(N) \\
&= \mathrm{E}_c(N) P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \in \overline{B}_{ij}^c).
\end{aligned} \tag{3.108}$$

Thus we may conclude in particular that $\boldsymbol{N_P} \in \langle \text{POISSON} \rangle$, and from (3.107) together with the definition of B_{ij}^c that (3.100) must hold. □

Note that the proof of Theorem 3.4 serves to clarify the formal role of threshold behavior itself. In particular, such behavior is seen to act as *filter operator* on the Poisson family $(N_{ij} : ij \in I \times J)$ of potential interaction frequencies defined for each $c \in C$ in Theorem 3.3. This filtering operation is made clear by observing from (3.100) together with (3.96) that

$$\mathrm{E}_c(N_{ij}^c) = \mathrm{E}_c(N_{ij}) P_c^1(T_1 \geq c_{ij}), \tag{3.109}$$

so that threshold behavior is seen to reduce the mean frequency of potential interactions to a smaller mean frequency of realized interactions. In this context, the formal role of the threshold-independence condition, C3, is to guarantee that this filter operator is *homogeneous* over the origin-destination space, $I \times J$. Hence, Theorem 3.4 can in fact be regarded as a special instance of the more general result that every homogeneous filtering of a Poisson process yields a new Poisson process with reduced mean intensity measure [as developed, for example, in Matthes, *et al.* (1978, Proposition 1.13.7)].

3.8 Search Processes

As a final application of the general probabilistic framework developed above, we now develop a class of *search processes* which extend those of Example 4 in Chapter 2. [The present results will be employed in Section 4.3.4 below to establish gravity model representations of the class of *simple search processes* in Chapter 2.] To do so, we begin by formalizing the basic notion of a *search event*.

3.8.1 Search Events

Consider a finite set of *actors*, $\alpha \in A$, distributed over *origin locations*, $i \in I$, and a finite set of *opportunities*, $\beta \in B$, distributed over *destination locations*, $j \in J$, where A_i and B_j denote the subsets of actors at origin i and opportunities at destination j, respectively. Let Ω_1 denote a universe of *individual search events*, $\omega \in \Omega_1$, each involving some actor, $\alpha(\omega) \in A$, who searches among opportunities in B until he either finds one which satisifies his current needs or exhausts all opportunities in B. The actor's location is representable by an *origin attribute function*, $\boldsymbol{i}:\Omega_1 \to I$, where $\boldsymbol{i}(\omega) = i$ iff $\alpha(\omega) \in A_i$. The ability of any given opportunity, $\beta \in B$, to satisfy the needs of actor $\alpha(\omega)$ is representable by a *satisfaction attribute function*, $\boldsymbol{\delta}_\beta:\Omega_1 \to \{0,1\}$, where $\boldsymbol{\delta}_\beta(\omega) = 1$ if opportunity β satisfies $\alpha(\omega)$'s current needs, and $\boldsymbol{\delta}_\beta(\omega) = 0$ otherwise. If each joint occurrence of n individual search events, $\omega = (\omega_r : r = 1,\ldots,n) \in \Omega_n = (\Omega_1)^n$, is designated as a *population search event of size*, n, and if the *null event* is denoted by $\Omega_0 = \{o\}$, then the relevant outcome space in the present setting is given by the class of all finite population search events, $\Omega = \cup_{n\geq 0}\Omega_n$. The above attributes of individual search events can then be extended to population search events in the usual way. In particular, the origin attribute function, $\boldsymbol{i}:\Omega_1 \to I$, generates the same family of *origin functions*, $\boldsymbol{I} = \{I_r:\Omega \to I_o | r \in Z_{++}\}$, as in expression (3.2) above. Similarly, the satisfaction attribute functions, $\boldsymbol{\delta}_\beta:\Omega_1 \to \{0,1\}, \beta \in B$, generate a family of *satisfaction functions*, $\boldsymbol{\Delta} = \{\Delta_{r\beta}:\Omega \to \{0,1\} \mid \beta \in B, r \in Z_{++}\}$ defined for all $n \in Z_+, \omega = (\omega_r : r \leq$

$n) \in \Omega$, and $\beta \in B$ by

$$\Delta_{r\beta}(\omega) = \begin{cases} \delta_\beta(\omega_r), & r \leq n \\ 0, & r > n. \end{cases} \tag{3.110}$$

Each vector of functions, $\Delta_r = (\Delta_{r\beta} : \beta \in B)$, thus describes the profile of relevant opportunity satisfaction attributes in each individual search event, ω_r, and is designated as the *satisfaction profile* for the r-th population search event. The combined family of functions, $\boldsymbol{I} \cup \boldsymbol{\Delta}$, represents the basic set of attributes for our present purposes, and hence generates the appropriate σ-field of events in Ω, namely, $\overline{M} = \sigma(\boldsymbol{I} \cup \boldsymbol{\Delta})$. If for each $n \in Z_{++}, i^n = (i_1, \ldots, i_n) \in I^n$, and $\delta^n = (\delta_1, \ldots, \delta_n) \in (\{0,1\}^B)^n$ we now let $\Omega(i^n, \delta^n) = \cap_{r=1}^n \left[I_r^{-1}(i_r) \cap \{\cap_{\beta \in B} \Delta_{r\beta}^{-1}(\delta_{r\beta})\}\right]$, and let

$$\overline{M}_n = \sigma\{\Omega(i^n, \delta^n) : i^n \in I^n, \delta^n \in (\{0,1\}^B)^n\}, \tag{3.111}$$

then essentially the same argument as in Proposition 3.4 again shows that the field σ-field, $\overline{M}$, can be equivalently written as $\overline{M} = \sigma(\cup_{n \geq 0} \overline{M}_n)$.

The relevant notion of spatial separation in this search context is implicitly generated by an underlying search scheme, which defines the order in which opportunities are searched from each origin. More specifically, it is assumed that the order in which opportunities are reached by all searchers from a common origin is identical, and in particular, that opportunities at each given destination are exhausted before proceeding to a new destination. Hence, if we let $b = |B|$, then a search scheme is taken to be formally representable by an integer-valued function, $c : I \times B \to \{1, \ldots, b\}$, satisfying the condition that for all distinct destinations, $j, h \in J$, and opportunities, $\beta_j, \beta_j' \in B_j$ and $\beta_h, \beta_h' \in B_h$,

$$c_i(\beta_j) < c_i(\beta_h) \Leftrightarrow c_i(\beta_j') < c_i(\beta_h') \tag{3.112}$$

Each value $c_i(\beta)$ denotes the order in which opportunity β is searched by actors at origin i [so that by definition the map, $c_i : B \to \{1, \ldots, b\}$, is one-to-one for each $i \in I$]. If the class of all search schemes, c, is denoted by C, then we may now formalize the desired class of *search processes* on $\langle \Omega, \overline{M} \rangle$ as follows:

Definition 3.8 A family of probability measures, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is designated as a *search process* iff $\boldsymbol{P}$ satisfies the following two conditions for all $c \in C, i \in I, \delta \in \{0,1\}^B, n \in Z_{++}, A \in \overline{M}_n$, and $\pi \in \Pi_n$,

S1. (Event Positivity) $P_c(I_1 = i, \Delta_1 = \delta) > 0$.

S2. (Population Symmetry) $P_c(A) = P_c(A_\pi)$.

[Note that the argument of Proposition 3.5 above applied to the present sets $\overline{M}_n$ shows that $A_\pi \in \overline{M}_n$ for all $A \in \overline{M}_n$, and hence that the probabilities, $P_c(A_\pi)$, in condition S2 are always well defined.]

3.8.2 REALIZED-INTERACTION FREQUENCIES

Of primary interest in each individual search event, ω, is the stopping point of the search, namely the first destination reached at which some opportunity satisfies the current need of actor, $\alpha(\omega)$. If we let $J_o = J \cup \{o\}$ (as in Section 3.3.2 above) then this property of individual search events can be formalized as a family of attribute functions, $\boldsymbol{J} = \{J_{cr} : \Omega \rightarrow J_o | r \in Z_{++}, c \in C\}$, as follows. First let $B_o = B \cup \{o\}$, and observe that for each $r \in Z_{++}$ and $c \in C$, the function, $\beta_{cr} : \Omega \rightarrow B_o$, defined for all $\omega \in \Omega$ by

$$\beta_{cr}(\omega) = \begin{cases} \beta, & \Delta_{r\beta}(\omega) = 1 \text{ and } \Delta_{r\beta'}(\omega) = 0 \text{ for all } \beta' \in B \\ & \quad \text{with } c_{I_r(\omega)}(\beta') < c_{I_r(\omega)}(\beta) \\ o, & \text{otherwise,} \end{cases} \tag{3.113}$$

identifies the first opportunity found (if any) which satisfies the actor's needs in the r-th individual search event of each population search event in Ω. In terms of these functions, the desired *destination functions*, $J_r : \Omega \rightarrow J_o$, can be defined for all $r \in Z_{++}$ and $\omega \in \Omega$ by

$$J_{cr}(\omega) = \begin{cases} j, & r \leq n, \ \beta_{cr}(\omega) \in B_j \\ o, & \text{otherwise.} \end{cases} \tag{3.114}$$

Given this definition, our objective is to show that each search process, $\boldsymbol{P}$, is an *interaction process* with respect to the above class of origin functions, $\boldsymbol{I}$, and destination functions, $\boldsymbol{J}$, in the sense of Definition 3.1. To do so, we begin by showing that:

Proposition 3.6 *The family of functions, $\boldsymbol{J}$, is $\overline{M}$-measurable.*

PROOF: If for each origin-destination pair, $ij \in I \times J$, we recall from (2.166) that the set of destinations searched prior to j by actors from i under search scheme, $c \in C$, is given by

$$J_c[j|i] = \{h \in J : (\beta' \in B_h, \beta \in B_j) \Rightarrow c_i(\beta') < c_i(\beta)\}, \tag{3.115}$$

then it follows from (3.112) through (3.115) that for each function, $J_{cr} \in \boldsymbol{J}$, and destination, $j \in J, J_{cr}(\omega) = j$ iff $\Delta_{r\beta}(\omega) = 0$ for all opportunities $\beta \in B_h$ with $h \in J_c[j|I_r(\omega)]$, and $\Delta_{r\beta}(\omega) = 1$ for some $\beta \in B_j$. Hence it follows by definition that for all $j \in J$,

$$\begin{aligned} J_{cr}^{-1}(j) = & \left[\cap\{\Delta_{r\beta}^{-1}(0) : \beta \in B_h, h \in J_c[j|I_r(\omega)]\}\right] \\ & \cap \left[\cup_{\beta \in B_j} \Delta_{r\beta}^{-1}(1)\right] \end{aligned} \tag{3.116}$$

and similarly, that $J_{cr}^{-1}(o) = \cap_{\beta \in B} \Delta_{r\beta}^{-1}(0)$. But since all functions, $I_r \in \boldsymbol{I}$, and $\Delta_{r\beta} \in \boldsymbol{\Delta}$, are $\overline{M}$-measurable by definition, and since every σ-field is closed under countable unions and intersections, we may conclude (from the finiteness of J_o) that each function, $J_{cr} : \Omega \rightarrow J_o$, is $\overline{M}$-measurable. □

Next observe that if the *indicator functions*, $\delta_{ij}^{cr}:\Omega \to \{0,1\}$ are defined for all $\omega \in \Omega$ and $ij \in I \times J$ by

$$\delta_{ij}^{cr}(\omega) = \begin{cases} 1, & I_r(\omega) = i, \text{ and } J_{cr}(\omega) = j \\ 0, & \text{otherwise}, \end{cases} \tag{3.117}$$

then, in a manner paralleling expression (3.11) above, the associated functions, $N_{ij}^c:\Omega \to Z_+$, defined for all $\omega \in \Omega$ and $ij \in I \times J$ by

$$N_{ij}^c(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^n \delta_{ij}^{cr}(\omega), & \omega \in \Omega_n, n \in Z_{++} \end{cases} \tag{3.118}$$

yield random frequency variables, N_{ij}^c, for each search process, $\boldsymbol{P}$, which shall be designated as the *realized-interaction frequencies* associated with search scheme, $c \in C$. Given these definitions, we are now ready to show that each search process, $\boldsymbol{P}$, constitutes an *interaction process* in the sense of Definition 3.1:

Proposition 3.7 *Each search process, $\boldsymbol{P} = \{P_c : c \in C\}$, is an interaction process on $\langle \Omega, \overline{M} \rangle$.*

PROOF: First observe from $\overline{M}$-measurability of all functions in $\boldsymbol{J}$ that the σ-field of interaction events, $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J})$, in expression (3.4) satisfies the condition of Definition 3.1 that $M \subseteq \overline{M}$. Thus it remains only to show that $\boldsymbol{P}$ satisfies the regularity conditions R1, R2, and R3. To establish R1, it must be shown that with respect to the realized-interaction frequencies defined by (3.117) and (3.118) above,

$$P_c(N_{ij}^c \neq 0) > 0, \quad ij \in I \times J, c \in C. \tag{3.119}$$

To establish (3.119) choose any vector, $\delta_j = (\delta_{j\beta} : \beta \in B) \in \{0,1\}^B$, with $\delta_{j\beta} = 0$ for all $\beta \in \cup\{B_h : h \in J_c[j|i]\}$ and $\delta_{j\beta} = 1$ for some $\beta \in B_j$, and observe from the event positivity condition (S1) that $0 < P_c(I_1 = i, \Delta_1 = \delta_j) = P_c\{I_1^{-1}(i) \cap [\cap_{\beta \in B} \Delta_{1\beta}^{-1}(\delta_{j\beta})]\}$. But since $I_1^{-1}(i) \cap [\cap_{\beta \in B} \Delta_{1\beta}^{-1}(\delta_{j\beta})] \subseteq \{\omega \in \Omega : N_{ij}^c(\omega) \neq 0\}$ by definition, it then follows that $0 < P_c\{\omega \in \Omega : N_{ij}^c(\omega) \neq 0\} = P_c(N_{ij}^c \neq 0)$, and hence that $\boldsymbol{P}$ satisfies R1. To establish R2, simply observe from (3.6) and (3.111) together with (3.116) that $M_n \subseteq \overline{M}_n$ for all $n \in Z_+$, and hence that R2 follows directly from S2. Finally, since the class, C, of search schemes is finite, it follows that the continuity condition R3 holds trivially for this case. In particular, for any choice of metric on C and sequence $\{c(m)\}$ in C, $c(m) \to c$ must imply that $c(m) = c$ for all large m, and hence that $\mathrm{E}_{c(m)}[N_{ij}^{c(m)}] = \mathrm{E}_c(N_{ij}^c)$ for all large m. Thus $\boldsymbol{P}$ satisfies (R1, R2, R3) and must be an interaction process. □

3.8.3 Independent Search Processes

Having established that the above search processes are interaction processes, we next introduce independence axioms on search behavior which will ensure that these interaction processes are *independent* in the sense of Definition 3.2. To do so, let the family of indicator functions, $\delta_{ri} : \Omega \to \{0,1\}, i \in I, r \in Z_{++}$, be defined for all $\omega \in \Omega$ by

$$\delta_{ri}(\omega) = \begin{cases} 1, & I_r(\omega) = i \\ 0, & \text{otherwise,} \end{cases} \tag{3.120}$$

so that [as in expression (3.118) above] the family of origin frequency functions, $N_i : \Omega \to Z_+, i \in I$, defined for all $\omega \in \Omega$ by

$$N_i(\omega) = \begin{cases} 0, & \omega \in \Omega_0 \\ \sum_{r=1}^{n} \delta_{ri}(\omega), & \omega \in \Omega_n, n \in Z_{++} \end{cases} \tag{3.121}$$

yields random variables representing the frequency of search activity at each origin. The associated random variable, $N = \sum_{i \in I} N_i$, then denotes the *total search frequency* for the process. In terms of these random variables [together with those in (3.2) and (3.110) above], we now say that

Definition 3.9 A search process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is designated as an *independent search process* iff $\boldsymbol{P}$ satisfies the following additional conditions for all search schemes, $c \in C$, and all $n \in Z_{++}$,

S3. (Origin location independence) *For all* $(i_r : r = 1, \ldots, n) \in I^n$,

$$P_c^n(I_r = i_r : r = 1, \ldots, n) = \prod_{r=1}^{n} P_c^n(I_r = i_r). \tag{3.122}$$

S4. (Origin-Frequency Independence) *For all* $(n_i : i \in I) \in Z_+^n$,

$$P_c(N_i = n_i : i \in I) = \prod_{i \in I} P_c(N_i = n_i). \tag{3.123}$$

S5. (Satisfaction Independence) *For all* $i_r \in I$ *and* $\delta_r \in \{0,1\}^B, r = 1, \ldots, n$, *with* $P_c^n(I_r = i_r : r = 1, \ldots, n) > 0$,

$$\begin{aligned} P_c^n(\Delta_r = \delta_r : r = 1, \ldots, n | I_r = i_r : r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^1(\Delta_1 = \delta_r | I_1 = i_r). \end{aligned} \tag{3.124}$$

[Condition S4 is identical with that in Definition 2.22, and conditions (S3, S5) together yield a weaker version of condition S3 in Definition 2.22]. Given this class of independent search processes, we now show that realized interaction frequencies for each such process define a *Poisson process*. More formally, if for each search process, $\boldsymbol{P} = \{P_c : c \in C\}$, we designate the corresponding family of random variables, $\boldsymbol{N_P} = \{N_{ij}^c : ij \in I \times J, c \in C\}$, in (3.118) above as the *realized-interaction frequency process* generated by $\boldsymbol{P}$, then we now have:

Theorem 3.5 *For each independent search process, $\boldsymbol{P} = \{P_c : c \in C\}$, the realized-interaction frequency process, $\boldsymbol{N_P} = \{N_{ij}^c : ij \in I \times J, c \in C\}$, generated by $\boldsymbol{P}$ is a Poisson process on $I \times J$ with positive finite mean interaction frequencies given for all $ij \in I \times J$ and $c \in C$ by*

$$\mathrm{E}_c(N_{ij}^c) = P_c^1(I_1 = i, J_{c1} = j)\mathrm{E}_c(N). \tag{3.125}$$

PROOF: First observe that independence conditions S3 and S4 are precisely conditions A1 and A2 for the index set, I. Hence if the index set, $I \times J$, in Theorem 3.2 is replaced by I, then precisely the same argument now shows that for each search scheme, $c \in C$, the family of origin frequencies, $\{N_i : i \in I\}$, must be a Poisson family [as in (3.35)] with joint distribution given for all $(n_i : i \in I) \in Z_+^n$ by

$$P_c(N_i = n_i : i \in I) = \prod_{i \in I} \frac{\mathrm{E}_c(N_i)^{n_i}}{n_i} \exp[-\mathrm{E}_c(N_i)]. \tag{3.126}$$

In particular this implies [as in expression (3.79)] that

$$P_c^n(I_r = i) = P_c^1(I_1 = i), \tag{3.127}$$

for all $i \in I$ and $r = 1, \ldots, n \in Z_{++}$ and in addition [as in expression (3.74)] that

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) > 0, \tag{3.128}$$

for all $(i_1, \ldots, i_n) \in I^n$ and $(j_1, \ldots, j_n) \in J^n$. Moreover, it also follows [as in expression (3.52)] that the corresponding total frequency variate, $N = \sum_{i \in I} N_i$, is Poisson distributed as

$$P_c(N = n) = \frac{\mathrm{E}_c(N)^n}{n} \exp[-\mathrm{E}_c(N)], \quad n \in Z_+. \tag{3.129}$$

Next observe from (3.113) and (3.114) above that for all $\omega, \omega' \in \Omega, c \in C$, and $r \in Z_{++}, [I_r(\omega), \Delta_r(\omega)] = [I_r(\omega'), \Delta_r(\omega')] \Rightarrow J_{cr}(\omega) = J_{cr}(\omega')$. Hence for each $c \in C, r \in Z_{++}$, and given value of I_r, the random variable, J_{cr}, is solely a function of the corresponding satisfaction profile, Δ_r, and it follows from the satisfaction independence condition (S5) and (3.128) above that for all $c \in C, n \in Z_{++}$, and $(i_r j_r : r = 1, \ldots, n) \in (I \times J_o)^n$

$$\begin{aligned} &P_c^n(J_{cr} = j_r : r = 1, \ldots, n | I_r = i_r \ : \ r = 1, \ldots, n) \\ &\qquad = \prod_{r=1}^n P_c^1(J_{c1} = j_r \mid I_1 = i_r), \end{aligned} \tag{3.130}$$

which together with the origin-location independence condition (S3) and (3.127) implies that

$$
\begin{aligned}
P_c^n(I_r = i_r, J_{cr} = j_r : r = 1, \ldots, n) \\
&= P_c^n(J_{cr} = j_r : r = 1, \ldots, n | I_r = i_r : r = 1, \ldots, n) \cdot \prod_{r=1}^{n} P_c^n(I_r = i_r) \\
&= \prod_{r=1}^{n} P_c^1(J_{c1} = j_r | I_1 = i_r) \cdot \prod_{r=1}^{n} P_c^1(I_1 = i_r) \\
&= \prod_{r=1}^{n} P_c^1(J_{c1} = j_r | I_1 = i_r) P_c^1(I_1 = i_r) \\
&= \prod_{r=1}^{n} P_c^1(I_1 = i_r, J_{c1} = j_r),
\end{aligned}
\tag{3.131}
$$

for all $c \in C, n \in Z_{++}$, and $(i_r j_r : r = 1, \ldots, n) \in (I \times J_o)^n$. But since this in turn implies [by summing i_r and j_r over all $r \neq 1$] that

$$P_c^n(I_1 = i, J_{c1} = j) = P_c^1(I_1 = i, J_{c1} = j), \tag{3.132}$$

for all $ij \in I \times J_o$ and $n \in Z_{++}$, it follows from (3.131) that $\boldsymbol{P}$ satisfies condition A1 for independent interaction processes (with J replaced by J_o). Hence, if we now let the *noninteraction frequencies*, $N_{io}^c : \Omega \to Z_+$, be defined for all $c \in C, i \in I$, and $\omega \in \Omega$ by $N_{io}^c(\omega) = N(\omega) - \sum_{ij \in I \times J} N_{ij}^c(\omega)$, then it follows from Lemma 3.1 [together with (3.132) above] that the conditional distributions of the frequency variates $(N_{ij}^c : ij \in I \times J_o)$ are given for all $n \in Z_{++}, c \in C$, and $(n_{ij} \in Z_+ : ij \in I \times J_o)$ with $n = \sum_{ij} n_{ij}$ by

$$P_c^n(N_{ij}^c = n_{ij} : ij \in I \times J_o) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}, \tag{3.133}$$

where $p_c(ij)$ is defined for all $ij \in I \times J_o$ and $c \in C$ by

$$p_c(ij) = P_c^1(I_1 = i, J_{c1} = j) > 0. \tag{3.134}$$

Thus, by combining (3.129) and (3.133) [and observing that by definition $\sum_{ij} p_c(ij) = 1$], we see that for $c \in C$, and $(n_{ij} \in Z_+ : ij \in I \times J_o)$ with

$\sum_{ij} n_{ij} = n \in Z_{++}$,

$$
\begin{aligned}
&P_c(N_{ij}^c = n_{ij} : ij \in I \times J_o) \\
&\quad = P_c^n(N_{ij}^c = n_{ij} : ij \in I \times J_o) \cdot P_c(N = n) \\
&\quad = \left[n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!} \right] \frac{\mathrm{E}_c(N)^n}{n!} \exp[-\mathrm{E}_c(N)] \\
&\quad = \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!} \mathrm{E}_c(N)^{\Sigma_{ij} n_{ij}} \exp[-\sum_{ij} p_c(ij)\mathrm{E}_c(N)] \\
&\quad = \prod_{ij} \left[\frac{[p_c(ij)\mathrm{E}_c(N)]^{n_{ij}}}{n_{ij}!} \exp[-p_c(ij)\mathrm{E}_c(N)] \right]
\end{aligned}
\tag{3.135}
$$

and hence that for each $c \in C$, the frequency variables $(N_{ij}^c : ij \in I \times J_o)$ are independently Poisson distributed with mean frequencies of the form

$$\mathrm{E}_c(N_{ij}^c) = p_c(ij)\mathrm{E}_c(N) = P_c^1(I_1 = i, J_{c1} = j)\mathrm{E}_c(N). \tag{3.136}$$

Finally, since this implies in particular that the subset of realized frequencies are distributed as independent Poisson variates, we may conclude from (3.136) that the realized-frequency process, $\boldsymbol{N_P} = \{N_{ij}^c : ij \in I \times J\}$, generated by $\boldsymbol{P}$ is a Poisson process on $I \times J$ satisfying (3.125). □

Note from (3.131) that an alternative proof of Theorem 3.5 could be constructed by establishing that $\boldsymbol{P}$ also satisfies the frequency independence condition (A2) for interaction processes, and then applying Theorem 3.2. However, the present proof is considerably more instructive in that it yields a direct construction of the corresponding Poisson process of realized interaction frequencies generated by $\boldsymbol{P}$. In particular, note that [as in the remarks following Theorem 3.4 above] this construction shows that the resulting Poisson process of realized-interaction frequencies is obtained as a *filtering* of the underlying Poisson process of search frequencies defined by (3.126) above.

3.9 Relaxations of Independence

Recall from Section 1.5 above that it is possible to relax each of the basic independence axioms A1 and A2 in ways which preserve many features of the analysis above. Hence the purpose of this section is to establish the results alluded to in Section 1.5. The present format will follow that of Section 1.5.

3.9.1 Relaxations of Frequency Independence

Our first result is to show that essential features of the Multinomial Sampling Corollary (Corollary 3.3) are preserved if frequency independence is relaxed to scale independence. To be more precise, if [as in expression (1.33)] we now designate an interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, as *scale independent* iff for all $n \in Z_{++}, c \in C$, and $ij \in I \times J$,

$$P_c^n(I_1 = i, J_1 = j) = P_c^1(I_1 = i, J_1 = j), \tag{3.137}$$

then we have the following extension of Corollary 3.3 to scale independent interaction processes:

Proposition 3.8 (Multinomial Sampling) *If* $\boldsymbol{P}$ *is a scale independent interaction process satisfying* A1, *then for each* $c \in C, n \in Z_{++}$, *and* $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$ *with* $\sum_{ij} n_{ij} = n$,

$$P_c^n(N_{ij} = n_{ij} : ij \in I \times J) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}, \tag{3.138}$$

where

$$p_c(ij) = \mathrm{E}_c(N_{ij}/N | N > 0) = \mathrm{E}_c(N_{ij})/\mathrm{E}_c(N). \tag{3.139}$$

Proof: Observe first that since Lemma 3.1 depends only on A1, it follows at once from Lemma 3.1 together with the definition of scale independence in (3.137) that (3.138) must hold with $p_c(ij) \equiv P_c^1(I_1 = i, J_1 = j)$. Moreover, this in turn implies from the same argument as in part (ii) of Corollary 3.3 that $p_c(ij) = \mathrm{E}_c(N_{ij}/N|N > 0)$, and hence that the first equality in (3.139) holds. Finally the identity $p_c(ij) = P_c^1(I_1 = i, J_1 = j)$ together with A1 and (3.137) also implies from the same argument as in (3.97) that

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \sum_n \mathrm{E}_c(N_{ij}|N = n)P_c(N = n) \\ &= \sum_n nP_c^n(I_1 = i, J_1 = j)P_c(N = n) \\ &= \sum_n np_c(ij)P_c(N = n) \\ &= p_c(ij)\mathrm{E}_c(N), \end{aligned} \tag{3.140}$$

and hence from the positivity of $\mathrm{E}_c(N)$ for all $c \in C$ that the second equality in (3.139) must hold as well. □

Weakly Independent Threshold-Interaction Processes

The above argument also leads to a similar extension of the class of independent threshold-interaction processes in Section 3.7 above. In particular, if the frequency independence condition (C2) is again replaced by the scale independence condition, then realized interactions will continue to be multinomially distributed with precisely the same interaction probabilities as in the case of independent threshold-interaction processes. To be more precise, we now say that:

Definition 3.10
(i) A threshold-interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is designated as a *weakly independent threshold-interaction process* iff $\boldsymbol{P}$ satisfies the following three independence conditions for all $c \in C$ and $n \in Z_{++}$ with $P_c(\Omega_n) > 0$,

C1. (Locational Independence) *For all* $i_i, \ldots . i_n \in I$ *and* $j_1, \ldots, j_n \in J$,

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r). \tag{3.141}$$

C2. (Scale Independence) *For all* $c \in C, ij \in I \times J$,

$$P_c^n(I_1 = i, J_1 = j) = P_c^1(I_1 = i, J_1 = j). \tag{3.142}$$

C3. (Threshold Independence) *For all* $t_1, \ldots, t_n \in R^K, i_1, \ldots, i_n \in I$, *and* $j_1, \ldots, j_n \in J$ *with* $P_c^n(I_r = i_r, J_r = j_r :\ r = 1, \ldots, n) > 0$,

$$P_c^n(T_1 \geq t_1, \ldots, T_n \geq t_n | I_r = i_r, J_r = j_r, r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^n(T_r \geq t_r). \tag{3.143}$$

(ii) In addition, $\boldsymbol{P}$ is said to satisfy *origin-destination independence* iff condition C1 above is replaced by the following stronger condition:

C1′. (Origin-Destination Independence) *For all* $i_1, \ldots, i_n \in I$ *and* $j_1, \ldots, j_n \in J$,

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ = \prod_{r=1}^{n} P_c^n(I_r = i_r) P_c^n(J_r = j_r). \tag{3.144}$$

To establish the desired multinomial property of realized interactions, it is necessary to distinguish between the number of potential interactions which are realized and those which are not. Hence, employing notation paralleling the proof of Theorem 3.4 above, we now consider events of the form $(N_{ij} = n_{ij}, \overline{N}_{ij} = \overline{n}_{ij} : ij \in I \times J)$ and denote the total number of *realized potential interactions* by $m = \sum_{ij} n_{ij}$ and denote the total number of *unrealized potential interactions* by $\overline{m} = \sum_{ij} \overline{n}_{ij}$. Hence, each possible total number of potential interactions is now given by the sum, $n = m + \overline{m}$ [so that the conditional distributions in Definition 3.10 above are of the form, $P_c^{m+\overline{m}}$]. In this context, we now let $P_c^{(m)}(N_{ij} = n_{ij} : ij \in I \times J)$ denote the conditional probability of realized interaction profile $(n_{ij} : ij \in I \times J)$ given that exactly m potential interactions are realized, and similarly, let $P_c^{(m)}(\overline{m})$ denote the conditional probability of $\overline{m}$ unrealized potential interactions given m, so that by definition

$$P_c^{(m)}(n_{ij} : ij \in I \times J) =$$
$$\sum_{\overline{m} \in Z_+} \sum_{(\overline{n}_{ij}) \in Z(\overline{m})} P_c^{m+\overline{m}}(N_{ij} = n_{ij}, \overline{N}_{ij} = \overline{n}_{ij} : ij \in I \times J) P_c^{(m)}(\overline{m}), \tag{3.145}$$

where $Z(\overline{m}) = \{(\overline{n}_{ij} : ij \in I \times J) \in Z_+^{I \times J} : \sum_{ij} \overline{n}_{ij} = \overline{m}\}$. With these conventions, we now have the following multinomial property of weakly independent threshold-interaction processes:

Proposition 3.9 (Multinomial Sampling)
(i) *If* $\boldsymbol{P}$ *is a weakly independent threshold-interaction process, then for each* $c \in C$ *and each realized interaction profile,* $(n_{ij} : ij \in I \times J)$*, with* $\sum_{ij} n_{ij} = m$*,*

$$P_c^{(m)}(N_{ij} = n_{ij} : ij \in I \times J) = m! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}, \tag{3.146}$$

where

$$p_c(ij) = \frac{P_c^1(I_1 = i, J_1 = j) P_c^1(T_1 \geq c_{ij})}{\sum_{gh} P_c^1(I_1 = g, J_1 = h) P_c^1(T_1 \geq c_{gh})}. \tag{3.147}$$

(ii) *In addition, if* $\boldsymbol{P}$ *is origin-destination independent and configuration free, with* $P_c = P$ *for all* $c \in C$*, then* (3.147) *takes the form*

$$p_c(ij) = \frac{P^1(I_1 = i) P^1(J_1 = j) P^1(T_1 \geq c_{ij})}{\sum_{gh} P^1(I_1 = g) P^1(J_1 = h) P^1(T_1 \geq c_{gh})}. \tag{3.148}$$

PROOF: **(i)** Observe first that since Lemma 3.3 depends only on properties C1 and C3, this result continues to hold for all weakly independent threshold-interaction processes. Moreover, since (3.101) is now obtained directly by applying (3.142) to the marginal distributions in (3.92), the argument in the proof of Theorem 3.4 shows that (3.105) continues to hold. Hence, in terms of the present notation, it must be true that for all $c \in C, n \in Z_{++}$ and $n_{ij}, \overline{n}_{ij} \in Z_+, ij \in I \times J$, with $n = \sum_{ij}(n_{ij} + \overline{n}_{ij})$,

$$P_c^n(N_{ij} = n_{ij}, \overline{N}_{ij} = \overline{n}_{ij} : ij \in I \times J) = n! \prod_{ij} \frac{q_c(ij)^{n_{ij}}}{n_{ij}!} \frac{\bar{q}_c(ij)^{\overline{n}_{ij}}}{\overline{n}_{ij}!}, \tag{3.149}$$

where

$$q_c(ij) = P_c^1(I_1 = i, J_1 = j)P_c^1(T_1 \in B_{ij}^c), \tag{3.150}$$

$$\bar{q}_c(ij) = P_c^1(I_1 = i, J_1 = j)P_c^1(T_1 \in \bar{B}_{ij}^c). \tag{3.151}$$

But since (3.150) together with the definition of B_{ij}^c in Theorem 3.4 implies that $p_c(ij)$ in (3.147) is of the form $p_c(ij) = q_c(ij)/\sum_{gh} q_c(gh)$, it then follows from the well-known conditional properties of multinomial distributions [Johnson and Kotz (1969),p.282] that

$$\begin{aligned} P_c^{(m)}(N_{ij} = n_{ij} : ij \in I \times J) &= m! \prod_{ij} \frac{[q_c(ij)/\sum_{gh} q_c(gh)]^{n_{ij}}}{n_{ij}!} \\ &= m! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}. \end{aligned} \tag{3.152}$$

Hence, we may conclude from (3.152) that (3.146) holds with $p_c(ij)$ given by (3.147).

(ii) Finally, if $\boldsymbol{P}$ is configuration free and satisfies C1′, then (3.148) follows at once by substituting (3.144) [with $n = 1$] into (3.147) and setting $P_c^1 = P^1$. □

Thus we see that for any given number of m realized interactions in a weakly independent threshold-interaction process, $\boldsymbol{P}$, these realizations can be treated as a random sample of size m from an interaction probability distribution of the form (3.147). In addition, if $\boldsymbol{P}$ is configuration free and origin-destination independent, then the resulting interaction probabilities (3.148) must have the separable gravity-type form in (2.20) of Chapter 2.

3.9.2 Relaxation of Locational Independence

Next we consider possible relaxation of locational independence (A1). To do so, observe first that if (as in Chapter 1) we let $P_c(s) \equiv P_c[\Omega(s)]$ and recall (from the proof of Theorem 3.2) that for each $s \in S$ and $ij \in I \times J$ we may write $N_{ij}(s) = N_{ij}(\omega)$ for all $\omega \in \Omega(s)$, then in terms of this notation, we have the following representation result for all interaction processes satisfying frequency independence:

Proposition 3.10 *If $\boldsymbol{P}$ is any interaction process satisfying frequency independence* (A2), *then for each $c \in C$ and $s \in S$ with $N_{ij}(s) = n_{ij}$, $ij \in I \times J$,*

$$P_c(s) = (n!)^{-1} \prod_{ij} n_{ij}! P_c(N_{ij} = n_{ij}). \tag{3.153}$$

PROOF: Simply observe from (3.67), (3.72) and the first line at (3.68) together with A2 that, in terms of the present notation, we must have

$$\begin{aligned} P_c(s) &= \left[\frac{n!}{\prod_{ij} n_{ij}!}\right]^{-1} P_c(N_{ij} = n_{ij} : ij \in I \times J) \\ &= \frac{\prod_{ij} n_{ij}!}{n!} \prod_{ij} P_c(N_{ij} = n_{ij}), \end{aligned} \tag{3.154}$$

and hence that (3.153) must hold identically for all $c \in C$ and $s \in S$. □

Thus, as mentioned in Section 1.5.2 above, all such processes are completely specified in terms of the frequency probabilities, $P_c(N_{ij} = n_{ij})$. To develop examples which are not Poisson, we begin by recalling that the variance of a Poisson variate is precisely equal to its mean. Hence, if we now say that a random variable exhibits *restricted variation (extra variation)* iff its variance is less than (greater than) its mean, then we may consider non-Poisson processes characterized by each of these properties.

(A) PROCESSES WITH RESTRICTED VARIATION

As a simple example of a process with restricted variation, one may consider a population of trip makers including both commuters and shoppers. If there are M_{ij} commuters living in zone i and working in zone j, and if work-trip decisions for each (ij)-commuter are treated as independently and identically distributed binary variates, then the probable number, N_{ij}^1, of work trips between i and j on any given day can be treated as a *binomial random variable* with distribution given for all $n_{ij} = 0, 1, \ldots, M_{ij}$ by

$$P_c(N_{ij}^1 = n_{ij}) = \frac{M_{ij}!}{n_{ij}!(M_{ij} - n_{ij})!} p_c(ij)^{n_{ij}} [1 - p_c(ij)]^{M_{ij} - n_{ij}} \tag{3.155}$$

where c denotes the prevailing configuration of travel times (or distances) between all locations, and where $p_c(ij)$ denotes the probability that any given work trip (ij) occurs. In most situations it may be assumed that $p_c(ij)$ is very close to one, so that day-to-day variations in N_{ij}^1 are very small. In contrast to commuting trips, shopping trips are much more discretionary in nature. Hence the assumptions of both locational independence and frequency independence may well be appropriate for such travel behavior. If so, then (by the Poisson Characterization Theorem) the numbers of shopping trips, N_{ij}^2, between each pair of zones, i and j, must be independent Poisson variates with distributions given for all $n_{ij} \in Z_+$ by

$$P_c(N_{ij}^2 = n_{ij}) = \frac{\lambda_c(ij)^{n_{ij}}}{n_{ij}!} \exp[-\lambda_c(ij)] \tag{3.156}$$

where $\lambda_c(ij) = \mathrm{E}_c(N_{ij}^2)$, $ij \in I \times J$. If these two processes are statistically independent, then distribution of the total trip frequencies

$$N_{ij} = N_{ij}^1 + N_{ij}^2 \tag{3.157}$$

between i and j is given by the *convolution* of the two distributions in (3.155) and (3.156). This distribution is easily specified in terms of probability generating functions. In particular, if the *probability generating function*, $G_X : R \rightarrow R$, of any Z_+-valued random variate, X, is defined for all $t \in [-1, 1]$ by

$$G_X(t) = \sum_{n \geq 0} t^n P(X = n) = \mathrm{E}(t^X) \tag{3.158}$$

and if the probability generating functions of N_{ij}, N_{ij}^1, and N_{ij}^2 are denoted respectively by G_{ij}, G_{ij}^1 and G_{ij}^2, then it is well known that G_{ij} is given by the product of the functions G_{ij}^1 and G_{ij}^2 [Feller, 1957, Section XI.2]. Hence, recalling that the probability generating functions for the binomial random variate in (3.155) and the Poisson random variate in (3.156) are given respectively by [Feller, 1957, Examples XI.2(b) and XI.2(c)],

$$G_{ij}^1(t) = [1 + (t-1)p_c(ij)]^{M_{ij}}, \tag{3.159}$$

$$G_{ij}^2(t) = \exp[(t-1)\lambda_c(ij)]. \tag{3.160}$$

it follows that G_{ij} is given for all $t \in R$ by

$$G_{ij}(t) = [1 + (t-1)p_c(ij)]^{M_{ij}} \exp[(t-1)\lambda_c(ij)]. \tag{3.161}$$

By the general definition of probability generating functions in (3.158), it also follows that the probabilities, $P_c(N_{ij} = n)$, are obtainable from the n-th derivatives of (3.161) evaluated at $t = 0$. To see that the resulting frequency process exhibits restricted variation, observe simply that if the variance of any frequency attribute, X, of interactions under configuration c is denoted by $\mathrm{var}_c(X)$, then by (3.157) together with independence, it follows that both $\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(N_{ij}^1) + \mathrm{E}_c(N_{ij}^2)$ and $\mathrm{var}_c(N_{ij}) = \mathrm{var}_c(N_{ij}^1) + \mathrm{var}_c(N_{ij}^2)$. But since $\mathrm{var}_c(N_{ij}^1) = [1 - p_c(ij)]\mathrm{E}_c(N_{ij}^1)$ for the binomial variate, N_{ij}^1, and $\mathrm{var}_c(N_{ij}^2) = \mathrm{E}_c(N_{ij}^2)$ for the Poisson variate, N_{ij}^2, we may conclude that

$$\mathrm{var}_c(N_{ij}) = [1 - p_c(ij)]\mathrm{E}_c(N_{ij}^1) + \mathrm{E}_c(N_{ij}^2) < \mathrm{E}_c(N_{ij}). \tag{3.162}$$

and hence that each frequency variate, N_{ij}, exhibits less variation than the Poisson.

(B) Processes with Extra Variation

While the above example illustrates a case in which population heterogeneity results in a process with less variation than the Poisson, most examples of population heterogeneity tend to increase variation. As mentioned in Section 1.5.2, a classic illustration is provided by the heterogeneity of family sizes among migrating households. This can be formalized probabilistically in terms of *compound distributions* as follows. If M_{ij} denotes the number of households in region i deciding to migrate to region j (during some relevant time period), and if $S_{ij}^r (r = 1, \ldots, \in M)$ denotes the size of the r^{th} migrating household, then the total number of migrants from i to j is given by

$$N_{ij} = \sum_{r=1}^{M_{ij}} S_{ij}^r. \tag{3.163}$$

Hence if household sizes are assumed to be independently and identically distributed Z_{++}-valued random variates, represented by a common random variable, S_{ij}, with probability generating function, H_{ij}, and if F_{ij} denotes the probability generating function of M_{ij}, then it is well known [Feller, 1957, Section XII.1] that the probability generating function of N_{ij} is given by the composition

$$G_{ij}(t) = F_{ij}[H_{ij}(t)], \quad t \in R. \tag{3.164}$$

Moreover, for any configuration, c, of relevant migration distances between regions, the corresponding mean and variance of N_{ij} are then given in terms of the means and variances of M_{ij} and S_{ij}, respectively, by [Feller, 1957, Problem XII.6(1)]:

$$\mathrm{E}_c(N_{ij}) = \mathrm{E}_c(M_{ij})\mathrm{E}_c(S_{ij}), \tag{3.165}$$

$$\mathrm{var}_c(N_{ij}) = \mathrm{E}_c(M_{ij}) \,\mathrm{var}_c(S_{ij}) + \mathrm{var}_c(M_{ij})[\mathrm{E}_c(S_{ij})]^2. \tag{3.166}$$

In particular, if M_{ij} is assumed to be Poisson distributed as in the illustration of Section 1.5.2, then N_{ij} is said to have a *compound Poisson distribution*, with probability generating function given from (3.160) and (3.164) by

$$G_{ij}(t) = \exp\{[1 - H_{ij}(t)]\mathrm{E}_c(M_{ij})\}. \tag{3.167}$$

To compare the mean and variance of such distributions, observe first that $\mathrm{E}_c(M_{ij}) = \mathrm{var}_c(M_{ij})$ by hypothesis. Moreover, since $\mathrm{var}_c(S_{ij}) = \mathrm{E}_c(S_{ij}^2) - [\mathrm{E}_c(S_{ij})]^2$ by definition, it follows from (3.166) that

$$\mathrm{var}_c(N_{ij}) = \mathrm{E}_c(M_{ij})\mathrm{E}_c(S_{ij}^2). \tag{3.168}$$

Hence if S_{ij} is any Z_{++}-valued random variable with $P_c(S_{ij} \geq 2) > 0$ [so that $\mathrm{E}_c(S_{ij}^2) > \mathrm{E}_c(S_{ij})$], then it follows from (3.165) and (3.168) that $\mathrm{var}_c(N_{ij}) > \mathrm{E}_c(N_{ij})$. Thus, whenever migration involves at least some *family* movement (i.e., with two or more members), the resulting compound Poisson distribution must exhibit greater variation than the Poisson.

The most elementary example of an explicit compound Poisson is one in which S_{ij} is also postulated to be Poisson distributed. The resulting distribution has probability generating function

$$G_{ij}(t) = \exp\{(1 - \exp[(1-t)\mathrm{E}_c(S_{ij})])\mathrm{E}_c(M_{ij})\} \tag{3.169}$$

and is often designated as *Neyman Type A* (or *double Poisson*) distribution. A second well-known example is given by assuming that S_{ij} is *logarithmically* distributed, i.e., that for some parameter, $0 < \alpha_{ij} < 1$,

$$P_c(S_{ij} = m) = \alpha_{ij}^m[-m\log(1-\alpha_{ij})]^{-1}. \tag{3.170}$$

The resulting distribution of N_{ij} in this case can be shown [as for example in Rogers (1974, Section 3.4.2)] to be the *negative binomial distribution* with probability generating function

$$G_{ij}(t) = [1 + p_{ij}(1-t)]^{-\beta_c(ij)} \tag{3.171}$$

where $p_{ij} = \alpha_{ij}/(1-\alpha_{ij})$ and $\beta_c(ij) = -[\log(1-\alpha_{ij})]^{-1}\mathrm{E}_c(M_{ij})$. A number of additional examples of compound Poisson distributions can be found in Rogers (1974, Section 3.5).

Finally, processes with extra variation can also be constructed in terms mixtures of Poisson populations. The most well known example involves gamma mixtures, which again leads to a family of negative binomial distributions for N_{ij}, as shown for example Rogers (1974, Section 3.3.2). [An interesting class of parametric specifications for these negative binomial models in the present spatial context be found in Davies and Guy (1987).]

3.9.3 More Complex Types of Interdependencies

As in Section 1.5.3 above, an additional appealing feature of the locational and frequency independence axioms is that they yield reasonable approximations to a wide range of weakly interacting populations. In particular, it is well known that a wide class of interdependent-interaction processes converge to the Poisson process as the relevant types of behavioral interdependencies diminish. Such results are well illustrated by the classes of *superimposed processes* and *congestive processes*, which we now consider in more detail.

(A) Superimposed Processes

As an extension of the 'coin flip' example in Section 1.5.3, it is of interest to consider interaction processes composed of a large number of small independent subprocesses. In the context of spatial interaction processes, each realized pattern of spatial interactions (shopping trips, commodity shipments) is typically the result of many independent decisions by individual households or firms, each constituting a relevant subsystem within the economy as a whole. Hence if the travel behavior of each household, α (or shipping firm, α) is modeled as a very 'small' interaction process, $\boldsymbol{P}_\alpha = \{P_c^\alpha : c \in C\}$, then the resultant interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, can be taken to be the *superposition* of all these processes. In particular, if the frequency process generated by $\boldsymbol{P}^\alpha$ is denoted by $\boldsymbol{N}^\alpha = \{N_c^\alpha : c \in C\}$, then the overall frequency process, $\boldsymbol{N} = \{N_c : c \in C\}$, for $\boldsymbol{P}$ can be taken to be the sum of these individual processes. To formalize these ideas, it is convenient to introduce the following more general representation of frequency processes on $I \times J$. If for any frequency process, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, we denote for each $c \in C$ and nonempty set, $A \subseteq I \times J$, the *A-frequency* of $\boldsymbol{N}_c$ by $N_A^c = \sum_{ij \in A} N_{ij}^c$, then each frequency profile $\boldsymbol{N}_c$ in $\boldsymbol{N}$ can be represented as $\boldsymbol{N}_c = \{N_A^c : \emptyset \neq A \subseteq I \times J\}$. In particular, the (ij)-frequency, N_{ij}^c, of $\boldsymbol{N}_c$ is now represented as $N_{\{ij\}}^c$, and the total interaction frequency, $N^c = \sum_{ij} N_{ij}^c$, is represented by $N_{I\times J}^c$. Hence if a collection of frequency processes, $\boldsymbol{N}^1, \ldots, \boldsymbol{N}^k$ is now said to be *independent* iff for each $c \in C$ and nonempty set $A \subseteq I \times J$ the random variables, $N_A^{1c}, \ldots, N_A^{kc}$, are independent, then our objective is to characterize the asymptotic behavior of sums of many small independent frequency processes on $I \times J$.

To do so, let us fix $c \in C$ and $A \subseteq I \times J$ and consider the following triangular array of random variables, where each row n corresponds to the A-frequencies of a collection of k_n independent frequency processes, $\boldsymbol{N}^{n1c}, \ldots, \boldsymbol{N}^{nk_nc}$, under configuration c,

$$\begin{array}{lllll} N_A^{11c}, & \ldots, & N_A^{1k_1c} & & \\ \vdots & & & \ddots & \\ N_A^{n1c}, & \ldots, & N_A^{n\alpha c} & \ldots, & N_A^{nk_nc} \\ \vdots & & & & \ddots \end{array} \tag{3.172}$$

If it is assumed that $k_n \to \infty$ as $n \to \infty$, then one may interpret this sequence of rows (for all A and c) as approaching an infinite collection $\{\boldsymbol{N}^\alpha\}$ of independent (uniformly) *small processes* [also designated as *sparse processes* in Cox and Isham (1980)] by requiring that the following two conditions hold for each $c \in C$ and $A \subseteq I \times J$:

$$\max\{P(N_A^{n\alpha c} > 0) : \alpha = 1, \ldots, k_n\} \to 0, \tag{3.173}$$

$$\sum_{\alpha=1}^{k_n} P(N_A^{n\alpha c} > 1) \to 0, \tag{3.174}$$

as $n \to \infty$. Condition (3.173) requires that the A-frequencies of processes in each row become uniformly small as $n \to \infty$, and (3.174) essentially requires that the possibility of any of these frequencies being greater than one can be disregarded as $n \to \infty$, i.e., that the individual frequencies, $N_A^{n\alpha c}$, in each set $A \subseteq I \times J$ can be treated as zero-one random variables when n becomes large. In particular, if for each row n we denote by, $\boldsymbol{N}^n = \{\boldsymbol{N}_c^n : c \in C\}$, the *cumulative frequency process* with frequency profiles, $\boldsymbol{N}_c^n = \{N_A^{nc} : \emptyset \neq A \subseteq I \times J\}$, defined for each $A \subseteq I \times J$ by $N_A^{nc} = \sum_{\alpha=1}^{k_n} N_A^{n\alpha c}$, then (3.174) implies that the associated mean frequencies, $\mathrm{E}(N_A^{nc})$, behave like sums of zero-one random variables in the sense that for large n,

$$\mathrm{E}(N_A^{nc}) = \sum_{\alpha=1}^{k_n} \mathrm{E}(N_A^{n\alpha c}) \approx \sum_{\alpha=1}^{k_n} P(N_A^{n\alpha c} = 1). \tag{3.175}$$

In this context, the main property of such infinite superpositions of processes [see for example Cinlar (1972, Theorem 3.10)] is that if their associated mean frequencies converge at all, i.e., if the limits

$$\mu_c(A) = \lim_{n \to \infty} \sum_{\alpha=1}^{k_n} P(N_A^{nc} = 1) \tag{3.176}$$

exist for all $A \subseteq I \times J$, then the sequence of frequency processes, $\boldsymbol{N}^n = \{\boldsymbol{N}_c^n : c \in C\}$, converges to the unique Poisson frequency process, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, with mean frequencies in each profile $\boldsymbol{N}_c = \{N_A^c : \emptyset \neq A \subseteq I \times J\}$, given by $\mathrm{E}(N_A^c) = \mu_c(A), A \subseteq I \times J$. Hence for large collections of small frequency processes, it may be argued that the resulting superimposed process is approximately Poisson. In fact, this result is seen to be essentially a generalization of the classical Poisson approximation of the binomial distribution (as for example in Feller, 1957, Section VI.5). [For more rigorous formulations of this general Poisson convergence property see Kallenberg (1976, Corollary 7.5) and Matthes, *et al.* (1978, Theorem 3.4.4)].

To relate this result to the present notion of an independent interaction process, observe that since the frequency variables, $N_{I \times J}^{n\alpha c}$, are almost surely zero-one for large n, it may be seen intuitively that each (arbitrarily small) component, $\boldsymbol{N}_c^\alpha$, of the limiting frequency profile, $\boldsymbol{N}_c = \sum_\alpha \boldsymbol{N}_c^\alpha$, contributes at most one interaction. But since these components are mutually independent by construction, it is clear that for each $c \in C$ the origin-destination attributes of individual interactions will be independent, and hence will satisfy locational independence (A1) for any given total number of interactions, $N_{I \times J}^c = n$. Moreover, since each component $\boldsymbol{N}_c^\alpha$, can influence the interaction frequency of at most one origin-destination pair, it is also clear from the independence of these components that frequency independence (A2) must hold as well, and hence that the limiting process, $\boldsymbol{N} = \sum_\alpha \boldsymbol{N}^\alpha$, can only be generated by an *independent* interaction process. Thus the Poisson Characterization Theorem (Theorem 3.2) helps to clarify the intuitive meaning of this convergence to Poisson frequency processes. Conversely, this convergence result shows that whenever population interactions occur only within small independent subpopulations (and are 'weak' in this sense), then overall population should correspond closely to an independent population process.

(B) Congestive Processes

Recall from the discussion of spatial Markov processes in Section 1.5.3 that an important class of spatial interdependencies involves 'neighborhood' effects, such as local congestion effects on road systems or local crowding effects in residential areas. To develop the travel example sketched in Section 1.5.3(B) above, it is convenient to adapt the approach of Ripley and Kelley (1976) to our present framework. Here the key spatial parameters relate to types of congestion effects which are present in the given transportation network $I \times J$. Hence we now take all relevant travel costs, other than those imposed by congestion effects, to be fixed and (with a slight abuse of notation) let c denote 'congestion costs' imposed on trips by the presence of other trips. In particular, if a congestion interdependency exists between any pair of trips ij and gh, this will be denoted by the congestion cost of $c(ij, gh) = 1$, and the absence of such an interdependency will be denoted by $c(ij, gh) = 0$ [examples involving a richer range of congestion costs are easily constructed]. The only prior conditions we impose on such functions are that trips between the same origin-destination pair always exhibit congestion interdependencies, and that congestion interdependencies between any trip pair are symmetric. Hence we now say that:

Definition 3.11 A function, $c:(I \times J)^2 \rightarrow \{0, 1\}$, is designated as a *congestion function* on $I \times J$ iff c satisfies the following two conditions for all $ij, gh \in I \times J$:

$$c(ij, ij) = 1, \tag{3.177}$$

$$c(ij, gh) = c(gh, ij). \tag{3.178}$$

By way of illustration, observe that the congestion function, $c_1:(I \times J)^2 \rightarrow \{0, 1\}$, defined for all $ij, gh \in I \times J$ by

$$c_1(ij, gh) = 1 \Leftrightarrow (i = g) \text{ or } (j = h) \tag{3.179}$$

implies that congestion interdependencies exist between all trips which share the same origin and/or destination. In addition, if 'head-on' effects also exist between trips [as discussed in Section 1.5.3(B)], then the appropriate congestion function, $c_2:(I \times J)^2 \rightarrow \{0, 1\}$, would be of the form

$$c_2(ij, gh) = 1 \Leftrightarrow (i = g) \text{ or } (j = h) \text{ or } (ij = hg). \tag{3.180}$$

Next observe that most realistic congestion functions tend to be local in nature. In particular, c_1 and c_2 above illustrate cases where direct congestion exists only between trips which are 'close together' in some appropriate sense. With this in mind, we now focus [following Ripley and Kelley (1976)] on those congestion functions for which there exists at least one pair of trips ab and de in $I \times J$ which are not directly congestion-interdependent but are indirectly congestion-interdependent in terms of a third trip fg:

Definition 3.12
(i) A congestion function c on $I \times J$ is said to be *proper* iff there exist origin-destination pairs, ab, de, $fg \in I \times J$ with

$$c(ab, fg) = c(de, fg) > c(ab, de). \tag{3.181}$$

(ii) The class of all *proper congestion functions* on $I \times J$ is denoted by C.

Observe from conditions (3.177) and (3.178) that ab, de and fg in (3.181) must all be distinct origin-destination pairs. Also note from the finiteness of $I \times J$ and $\{0,1\}$ that C is always finite. Finally, observe that C is always nonempty, even for the minimal case in which $|I| = |J| = 2$. In particular, if $I = \{i, g\}$ and $J = \{j, h\}$, then it follows from (3.179) and (3.180) that both c_1 and c_2 are proper congestion functions on $I \times J$ [since $c_\alpha(ij, gj) = c_\alpha(gh, gj) > c_\alpha(ij, gh)$ for both $\alpha = 1$ and $\alpha = 2$].

Assuming that congestion effects are additive, it follows by definition that for each trip $i_r j_r$ in a given pattern, $s = (i_r j_r : r = 1, \ldots, n) \in S$, the total cost imposed on $i_r j_r$ is $\Sigma_{s \neq r} c(i_r j_r, i_s j_s)$. Hence the *total congestion cost* for all trips in s is given by

$$c(s) = \begin{cases} 0, & \text{if} \quad n \leq 1 \\ \sum_{r=1}^{n} \sum_{s \neq r} c(i_r j_r, i_s j_s), & \text{if} \quad n > 1. \end{cases} \tag{3.182}$$

If a new trip ij is added to pattern s, then the total additional cost, $c(ij \mid s)$, created by ij must include both the total costs, $\Sigma_{r=1}^{n} c(ij, i_r j_r)$, imposed on ij by s plus the total costs, $\Sigma_{r=1}^{n} c(i_r j_r, ij)$ imposed on s by ij, which by symmetry must be of the form

$$c(ij \mid s) = \begin{cases} 0, & \text{if} \quad n = 0 \\ 2\sum_{r=1}^{n} c(ij, i_r j_r), & \text{if} \quad n > 0. \end{cases} \tag{3.183}$$

Hence, if for each n-trip pattern, $s \in S_n$, we let (s, ij) denote the $(n+1)$-trip pattern formed by adding ij to s, then it follows at once from (3.182) and (3.183) that

$$c(s, ij) = c(s) + c(ij \mid s). \tag{3.184}$$

Given this notion of proper congestion functions, our objective is to characterize those spatial interaction processes which might represent typical travel behavior under conditions of congestion. To develop an appropriate probabilistic framework for this task, we now consider *congestive spatial interaction processes*, $\boldsymbol{P} = (P_c : c \in C)$, for which the relevant parameter set is given by the set of proper congestion functions on $I \times J$, and for which all pattern probabilities, $P_c(s)$, are assumed to be positive. [Note that the positivity condition on $\boldsymbol{P}$ implies regularity condition R1, and that the finiteness of C implies (from the argument in Proposition 3.7 for search processes) that condition R3 also holds, so that regularity for $\boldsymbol{P}$ is reduced to the symmetry condition R2. Note also that the symmetry of total congestion costs for each pattern as implied by (3.178), (3.182) and (3.183) is consistent with R2].

Within this framework we seek to characterize those processes $\boldsymbol{P}$ which might be said to represent congestion-sensitive behavior. To do so, we begin [in a manner paralleling the development of Carroll-Bevis processes in Section 2.2.2(A) and Kullback-Leibler processes in Section 2.4.2(A) above] by asking how individuals might behave if they were completely *insensitive* to congestion effects. In particular, if it is postulated that the only interdependencies between trip makers are those due to congestion effects, then the behavior of trip makers insensitive to such effects should be indistinguishable from *independent* spatial interaction behavior, i.e., from behavior consistent with the locational and frequency independence axioms A1 and A2. It then follows from the Poisson Characterization Theorem that congestion-insensitive spatial interaction behavior must be representable by some Poisson frequency process, $\boldsymbol{N}$. Hence, recalling (from Definition 3.5) that $\boldsymbol{N}$ is in turn representable by its unique generator, $\boldsymbol{P_N}$, we now adopt spatial interaction processes of the form (3.76) as the appropriate model of 'congestion insensitivity'. More precisely, we now say that:

Definition 3.13 A congestive spatial interaction process, $\boldsymbol{P}^o = (P_c^o : c \in C)$, on $\langle \Omega, M \rangle$ with mean frequencies $[\mathrm{E}_c^o(N_{ij}) : ij \in I \times J]$ and $\mathrm{E}_c^o(N)$ $[\equiv \Sigma_{ij} \mathrm{E}_c^o(N_{ij})]$ for each $c \in C$ is designated as a *congestive-insensitive process* iff the pattern probabilities $P_c^o(s)$ are given for all $s \in S$ and $c \in C$ by

$$P_c^o(s) = \frac{\mathrm{E}_c^o(N)^{N(s)}}{N(s)!} \exp[-\mathrm{E}_c^o(N)] \prod_{ij} \left[\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N)} \right]^{N_{ij}(s)} . \tag{3.185}$$

In this framework, one can analyze 'congestive-sensitive' processes in terms of their deviations from congestive-insensitive processes. To do so, it is convenient to construct for each process $\boldsymbol{P}$ a *unique* 'congestion-insensitive' version, $\boldsymbol{P}^o$, which is comparable to $\boldsymbol{P}$ in some appropriate sense. Recall that in the case of the Kullback-Leibler processes in Section 2.4.2(A), a mean-activity-equivalence criteria was proposed for constructing a unique OD independent version of each KL-process [as established in Proposition 4.5 of Chapter 4 below]. However, if sensitivity to congestion *deters* travel behavior, then actual mean activity levels will generally be *lower* than would be expected under congestion-insensitive behavior. Hence we now develop an alternative criteria of comparability which is more appropriate to the present setting. To do so, observe that congestion sensitivity should have no effect on travel behavior in those spatial interaction patterns, s, which are *free* of congestion, i.e., which satisfy $c(s) = 0$. In particular, since both the null spatial interaction pattern, $s_o \in S_0$, and each singleton interaction pattern, $ij \in S_1$, are by definition free of congestion, the relative likelihood of ij versus s_o (i.e., of a single ij trip versus no trips at all) should be the same under congestion-sensitive and congestion-insensitive behavior. Hence any congestion-insensitive process, $\boldsymbol{P}^o$, which is comparable to $\boldsymbol{P}$ should yield the same relative likelihoods of ij versus s_o under each possible congestion function, $c \in C$. More formally, we now say that:

Definition 3.14 For each congestive spatial interaction process, $\boldsymbol{P} = (P_c : c \in C)$, and a congestion-insensitive process, $\boldsymbol{P}^o = (P_c^o : c \in C)$, is designated as an admissible *congestion-insensitive version* of $\boldsymbol{P}$ iff for all $c \in C$ and $ij \in I \times J$,

$$\frac{P_c(ij)}{P_c(s_o)} = \frac{P_c^o(ij)}{P_c^o(s_o)}. \tag{3.186}$$

This compatibility condition identifies a unique congestion-insensitive version of each congestive spatial interaction process, as we now show:

Proposition 3.11 *For each congestive spatial interaction process, $\boldsymbol{P} = (P_c : c \in C)$, there exists a unique congestion-insensitive version of $\boldsymbol{P}$.*

PROOF: By definition $N(s_o) = 0$ and $N(ij) = 1$ for all $ij \in I \times J$, so that (3.185) implies $P_c^o(ij)/P_c^o(s_o) = \mathrm{E}_c^o(N_{ij})$ for all $ij \in I \times J$ and $c \in C$. But since each Poisson process is uniquely defined by its mean frequencies, it then follows from (3.186) that the Poisson process, $\boldsymbol{P}^o$, with mean frequencies given for all $ij \in I \times J$ and $c \in C$ by $\mathrm{E}_c^o(N_{ij}) = P_c(ij)/P_c(s_o)$ must be the unique congestion-insensitive version of $\boldsymbol{P}$. □

Hence we now take this unique congestion-insensitive version of $\boldsymbol{P}$ to represent the 'null hypothesis' of congestion-insensitive behavior, and seek to compare behavior under $\boldsymbol{P}$ with behavior under the *congestion-insensitivity hypothesis*, $\boldsymbol{P}^o$. In particular, observe that for any $c \in C$ and spatial interaction patterns, $s, v \in S$, the relative likelihood of s versus v under P_c is by definition not greater than would be expected under the congestion-insensitivity hypothesis if and only if $P_c(s)/P_c(v) \leq P_c^o(s)/P_c^o(v)$. Equivalently, if the function, $L : C \times S \to R_{++}$, defined for all $c \in C$ and $s \in S$ by

$$L_c(s) = \frac{P_c(s)}{P_c^o(s)} \tag{3.187}$$

is now designated as the *standardized likelihood function* for $\boldsymbol{P}$, then the relative likelihood of s versus v under P_c is not greater than would be expected under the congestion-insensitivity hypothesis if and only if $L_c(s) \leq L_c(v)$, i.e., if and only if the *standardized relative likelihood*, $L_c(s)/L_c(v)$, of s versus v under P_c is less than or equal to one.

These general observations may now be applied to congestion-sensitive behavior in the following way. If interactions are in fact *deterred* by the presence of congestion, then compared to congestion-insensitive behavior under any prevailing conditions, those trips which add higher levels of congestion should be relatively less likely to occur than those which add lower congestion levels. In particular, if the addition of trip ij to pattern s creates at least as much new congestion as does the addition of trip gh to v, then it may be hypothesized that the *standardized* relative likelihood of (s, ij) versus (v, gh) does not exceed that of s versus v. [Note in particular that the standardization in (3.187) helps to remove 'size' effects, so that relative likelihoods of larger pattern, (s, ij) and (v, gh) can be compared with those of smaller patterns, (s) and (v).] Interaction processes satisfying this basic 'congestion-deterrence' postulate may now be formalized as follows:

Definition 3.15 A congestive spatial interaction process, $\boldsymbol{P} = (P_c : c \in C)$, with standardized likelihood function, L, is said to exhibit *congestion-deterrence behavior* iff for all $s, v \in S$, $ij, gh \in I \times J$ and $c \in C$,

$$c(ij \mid s) \geq c(gh \mid v) \Rightarrow \frac{L_c(s, ij)}{L_c(v, gh)} \leq \frac{L_c(s)}{L_c(v)}. \tag{3.188}$$

Equivalently, $\boldsymbol{P}$ is designated as *congestion-deterrence process.*

For this class of congestive spatial interaction processes we now have the following representational consequence:

Proposition 3.12 *A congestive spatial interaction process, $\boldsymbol{P} = (P_c : c \in C)$, exhibits congestion-deterrence behavior iff for each $c \in C$ there exists constants, $\alpha_c \in R_{++}$ and $\theta_c \in R_+$ such that for all $s \in S$,*

$$P_c(s) = P_c^o(s)\alpha_c \exp[-\theta_c c(s)]. \tag{3.189}$$

PROOF: (a) First observe that if (3.189) holds, then it follows that,

$$\frac{P_c(s)}{P_c^o(s)} = \alpha_c \exp[-\theta_c c(s)], \tag{3.190}$$

which together with (3.184) implies that for any patterns, $s, v \in S$, and $ij, gh \in I \times J$, with $c(ij \mid s) \geq c(gh \mid v)$ we must have

$$\begin{aligned} \frac{P_c(s, ij)/P_c^o(s, ij)}{P_c(v, gh)/P_c^o(v, gh)} &= \frac{\exp[-\theta_c c(s, ij)]}{\exp[-\theta_c c(v, gh)]} \\ &= \frac{\exp[-\theta_c\{c(s) + c(ij \mid s)\}]}{\exp[-\theta_c\{c(v) + c(gh \mid v)\}]} \\ &= \frac{\exp[-\theta_c c(s)] \exp[-\theta_c c(ij \mid s)]}{\exp[-\theta_c c(v)] \exp[-\theta_c c(gh \mid v)]} \\ &\leq \frac{\exp[-\theta_c c(s)]}{\exp[-\theta_c c(v)]} = \frac{P_c(s)/P_c^o(s)}{P_c(v)/P_c^o(v)}. \end{aligned} \tag{3.191}$$

Hence it follows from (3.191) together with (3.187) that $\boldsymbol{P}$ satisfies (3.188), and must exhibit congestion-deterrence behavior.

(b) To establish the converse, suppose that $\boldsymbol{P}$ satisfies (3.188). Then it follows at once that for all $s, v \in S$, $ij, gh \in I \times J$ and $c \in C$,

$$\begin{aligned} c(ij \mid s) = c(gh \mid v) &\Rightarrow c(ij \mid s) \begin{bmatrix} \geq \\ \leq \end{bmatrix} c(gh \mid v) \\ &\Rightarrow \frac{L_c(s, ij)}{L_c(v, gh)} \begin{bmatrix} \geq \\ \leq \end{bmatrix} \frac{L_c(s)}{L_c(v)} \Rightarrow \frac{L_c(s, ij)}{L_c(s)} \begin{bmatrix} \geq \\ \leq \end{bmatrix} \frac{L_c(v, gh)}{L_c(v)} \\ &\Rightarrow \frac{L_c(s, ij)}{L_c(s)} = \frac{L_c(v, gh)}{L_c(v)}. \end{aligned} \tag{3.192}$$

Next, recall from Definition 3.12 that for each $c \in C$ there must exist distinct elements, $ab, de, fg \in I \times J$ satisfying (3.181), i.e., with $c(ab, fg) = c(de, fg) = 1$ and $c(ab, de) = 0$. Hence for each $n > 0$ let $s_n \in S_n$ consist of n identical (ab)-interactions and let the function, $h_c : Z_+ \to R_{++}$, be defined for all $n > 0$ by

$$h_c(n) = \frac{L_c(s_n, ab)}{L_c(s_n)}. \tag{3.193}$$

For $n = 0$ it follows from (3.186), together with the identity $P_c(s_o, ab) = P_c(ab)$, that the appropriate value of h_c is given by

$$h_c(0) = \frac{L_c(s_o, ab)}{L_c(s_o)} = \frac{P_c(s_o, ab)/P_c^o(s_o, ab)}{P_c(s_o)/P_c^o(s_o)} = \frac{P_c(ab)/P_c^o(ab)}{P_c(s_o)/P_c^o(s_o)} = 1. \tag{3.194}$$

Then [recalling from (3.183) that $c(ij \mid s)$ is always an even integer] it follows from (3.192) that for any $s \in S$ and ij with $c(ij \mid s) = 2n$,

$$\begin{aligned} c(ij \mid s) = c(ab \mid s_n) \Rightarrow \frac{L_c(s, ij)}{L_c(s)} &= \frac{L_c(s_n, ab)}{L_c(s_n)} \\ &= h_c(n) = h_c\left[\frac{1}{2}c(ij \mid s)\right] \end{aligned} \tag{3.195}$$

so that we obtain the basic identity

$$\frac{L_c(s, ij)}{L_c(s)} = h_c\left[\frac{1}{2}c(ij \mid s)\right], \tag{3.196}$$

for all $c \in C$, $s \in S$ and $ij \in I \times J$. To employ this identity, observe that by construction $c[fg \mid (s_n, de)] = 2(n+1)$, $c[de \mid (s_n, fg)] = 2$, $c(fg \mid s_n) = 2n$, and $c(de \mid s_n) = 0$, so that by (3.196) [together with (3.194)] we must have both

$$\begin{aligned} L_c[(s_n, fg), de] &= \frac{L_c[(s_n, fg), de]}{L_c(s_n, fg)} \cdot \frac{L_c(s_n, fg)}{L_c(s_n)} \cdot L_c(s_n) \\ &= h_c(1)h_c(n)L_c(s_n) \end{aligned} \tag{3.197}$$

and

$$\begin{aligned}L_c[(s_n, de), fg] &= \frac{L_c[(s_n, de), fg]}{L_c(s_n, de)} \cdot \frac{L_c(s_n, de)}{L_c(s_n)} \cdot L_c(s_n) \\ &= h_c(n+1)h_c(0)L_c(s_n) = h_c(n+1)L_c(s_n),\end{aligned} \tag{3.198}$$

for each $n > 0$. But since the symmetry condition R2 for $\boldsymbol{P}$ together with (3.185) and (3.187) implies that L_c must be symmetric on S [i.e., that $L_c(s) = L_c(s_\pi)$ for each $s \in S_n$ and $\pi \in \Pi_n, n > 0$], it follows from (3.197) and (3.198) that,

$$\begin{aligned}L_c[(s_n, de), fg] &= L_c(s_n, de, fg) = L_c[(s_n, fg), de] \\ &\Rightarrow h_c(n+1)L_c(s_n) = h_c(1)h_c(n)L_c(s_n) \\ &\Rightarrow h_c(n+1) = h_c(1)h_c(n)\end{aligned} \tag{3.199}$$

and, hence by induction, that for all $n > 0$,

$$h_c(n) = h_c(1)^n = \exp[-\theta_c(2n)], \tag{3.200}$$

where $\theta_c = -\frac{1}{2}\log[h_c(1)]$. Note also from (3.196) and (3.200) that by setting $s = (ab, ab) = s_2$, $v = (ab) = s_1$ and $ij = ab = gh$ in (3.188), we must have

$$\begin{aligned}c(ab \mid s_2) = 4 > 2 = c(ab \mid s_1) &\Rightarrow \frac{L_c(s_2, ab)}{L_c(s_2)} \leq \frac{L_c(s_1, ab)}{L_c(s_1)} \\ &\Rightarrow h_c(2) \leq h_c(1) \Rightarrow \exp[-\theta_c 4] \leq \exp[-\theta_c 2] \Rightarrow \theta_c \geq 0.\end{aligned} \tag{3.201}$$

For this choice of θ_c we next show that if $\alpha_c = L_c(s_o) > 0$ then the standardized likelihood function, L, for $\boldsymbol{P}$ has the following explicit form for all $s \in S$ and $c \in C$,

$$L_c(s) = \alpha_c \exp[-\theta_c c(s)]. \tag{3.202}$$

First observe that if $N(s) = 0$, then $c(s) = 0$ by (3.182) and $s = s_o$ by definition, so that (3.202) holds trivially. Similarly if $N(s) = 1$, with say $s = (ij)$, then it again follows from (3.182) that $c(s) = 0$, which together with (3.194) now shows that (3.202) is given by the identity,

$$L_c(s) = L_c(ij) = L_c(s_o, ij) = L_c(s_o) \cdot \frac{L_c(s_o, ij)}{L_c(s_o)} = \alpha_c h_c(0) = \alpha_c. \tag{3.203}$$

Finally, if (3.202) holds for any $n \geq 1$, then for each $(s, ij) \in S_{n+1}$ [i.e., each $s \in S_n$ and $ij \in I \times J$] it follows from (3.184), (3.196) and (3.200) that

$$\begin{aligned}L_c(s, ij) &= \frac{L_c(s, ij)}{L_c(s)} \cdot L_c(s) \\ &= h_c\left[\frac{1}{2}c(ij \mid s)\right] \cdot \alpha_c \exp[-\theta_c c(s)] \\ &= \exp[-\theta_c c(ij \mid s)] \cdot \alpha_c \exp[-\theta_c c(s)] \\ &= \alpha_c \exp[-\theta_c\{c(ij \mid s) + c(s)\}] = \alpha_c \exp[-\theta_c c(s, ij)],\end{aligned} \tag{3.204}$$

so that (3.202) holds for $n+1$, and by induction, must hold for all $n \in Z_+$. Thus (3.189) follows at once from (3.204) together with (3.187), and the result is established. □

Remark 3.1 It is of interest to observe from the proof of Proposition 3.11 that the inequalities in (3.188) are not strictly necessary. In particular, this proof shows that the representation of L_c in (3.202) depends only on the 'equality' version of (3.188) and (3.192). Hence the only formal role of these inequalities is to ensure that $\theta_c \geq 0$. But [as observed by Kelly and Ripley(1976)] this sign condition is already implied by the requirement that $\sum_{s \in S} P_c(s) = 1$. For since the patterns $s_n \in S_n$ in (3.193) must have total congestion costs, $c(s_n) = n(n-1)$, by (3.177) and (3.182), it follows by setting $m_c = \min\{\mathrm{E}_c^o(N_{ij})/\mathrm{E}_c^o(N) : ij \in I \times J\} > 0$, $b_c = \mathrm{E}_c^o(N) m_c$, $a_c = \alpha_c \exp[-\mathrm{E}_c^o(N)]$ and $\sigma_c = \exp(-\theta_c) > 0$ that one must always have $\sum_{s \in S} P_c(s) \geq \sum_n P_c(s_n) = \sum_n [P_c^o(s_n)\alpha_c \sigma_c^{c(s_n)}] \geq \sum_n a_c(b_c^n/n!)\sigma_c^{n(n-1)} = a_c \sum_n [b_c^n \sigma_c^{n(n-1)}/n!$. But since $\theta_c < 0$ implies that $\sigma_c > 1$, and since $\lim_{n\to\infty} x^n y^{n(n-1)}/n! = \infty$ for all $x > 0$ and $y > 1$, it would then follow that $\sum_{s \in S} P_c(s) = \infty$. Hence every probability distribution of the form (3.189) must satisfy $\theta_c \geq 0$. However, in view of the *behavioral* significance of the inequalities in (3.188), we have chosen to employ this somewhat stronger form. •

As a first consequence of this result, it is of interest to observe that while the minimal compatibility condition in (3.186) served to guarantee the existence of a unique congestion-insensitive version for every *congestive* process, we now see [directly from (3.189)] that for congestion-deterrent processes, these congestion-insensitive versions exhibit the following stronger form of condition (3.186):

Proposition 3.13 *If $\boldsymbol{P}$ is a congestion-deterrent process with congestion-insensitive version, $\boldsymbol{P}^o$, then for all $c \in C$ and $s, v \in S$,*

$$c(s) = 0 = c(v) \Rightarrow \frac{P_c(s)}{P_c(v)} = \frac{P_c^o(s)}{P_c^o(v)}. \tag{3.205}$$

Equivalently, if the *zero-congestion event* in S under c is denoted by $S_c(0) = \{s \in S : c(s) = 0\}$, then the conditional distributions $P_c[\cdot \mid S_c(0)]$ and $P_c^o[\cdot \mid S_c(0)]$ are identical for each $c \in C$. Hence this class of congestion-deterrent processes exhibits the reasonable property that strict congestion-deterrence is only distinguishable from congestion-insensitive behavior when some congestion is present.

As a second behavioral consequence of Proposition 3.11, observe that if $\theta_c > 0$, then for all sufficiently large spatial interaction patterns, the negative exponential factor in (3.189) eventually drives the standardized likelihood function L_c to zero, and in particular implies that $L_c(s) < 1$

for all large s. Processes with this type of standardized likelihood function [first proposed by Strauss (1975)] are thus said to be 'self-inhibiting' in the sense that high levels of interaction activity are always less likely to occur than would be expected under independence. At the opposite extreme, when $\theta_c = 0$, we have the following simple characterization of *congestion-insensitive* processes within the present setting:

Proposition 3.14 *A congestion-deterrence process,* $\boldsymbol{P} = (P_c : c \in C)$*, is congestion-insensitive iff* $\theta_c = 0$ *for all* $c \in C$.

PROOF: If $\theta_c = 0$, then (3.189) becomes $P_c(s) \equiv P_c^o(s)\alpha_c$, which together with the normalization condition, $1 = \sum_s P_c(s) = \alpha_c \sum_s P_c^o(s) = \alpha_c$, implies that $P_c(s) \equiv P_c^o(s)$, and thus that $\boldsymbol{P}$ is congestion insensitive. Conversely, if $\boldsymbol{P}$ is congestion-insensitive, then we must have $P_c(s) \equiv P_c^o(s)$, so that $\alpha_c \exp[-\theta_c c(s)] = 1$ for all $s \in S$ and $c \in C$. But since $c(s_o) = 0$ then implies that $\alpha_c = \alpha_c \exp(0) = 1$, it may be concluded from (3.177) that $1 = \exp[-\theta_c c(ij, ij)] = \exp(-\theta_c)$, and, consequently, that $\theta_c = 0$ for all $c \in C$. □

As a consequence of this result, the parameter θ_c can thus be interpreted as a *congestion sensitivity* parameter for the population of spatial actors. When $\theta_c = 0$, spatial actors exhibit complete insensitivity to congestion effects in the sense that individual spatial interactions are mutually independent. Moreover, as θ_c increases, the likelihood of patterns with higher congestion costs decreases relative to what would be expected under independence.

Note finally from the continuity of (3.189) in θ_c that if spatial actors are not too sensitive to congestion effects (i.e., if θ_c is 'small'), then the corresponding frequency process must be well approximated by a Poisson process. Hence [as a parallel to Section (A) above] it may be concluded that Poisson processes provide a natural probabilistic model for this type of 'weak interaction' behavior.

Within the present framework, it is also important to ask under what conditions Poisson approximations are still possible when actors are *sensitive* to congestion effects. It should be clear that if θ_c is large then the pattern probabilities (3.189) will not be approximated by (3.185). However, if one focuses on overall levels of interaction activity, then the corresponding probabilities, $P_c(N = n)$, will still be approximately Poisson provided that congestion effects are 'sufficiently small' in an appropriate sense. To see that this is possible, we begin by denoting the (possibly empty) subclass of spatial interaction n-patterns in S_n which involve no congestion by $S_n(0) = \{s \in S_n : c(s) = 0\}$, and let the subclass involving some degree of congestion be denoted by $S_n(+) = S_n - S_n(0)$. In terms of this notation, each frequency probability, $P_c(N = n)$, can be written as

$$P_c(N = n) = \sum_{s \in S_n(0)} P_c(s) + \sum_{s \in S_n(+)} P_c(s). \tag{3.206}$$

Note also from (3.189) that

$$\sum_{s \in S_n(0)} P_c(s) = \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] P_c^n(0), \tag{3.207}$$

where

$$P_c^n(0) = \sum_{s \in S_n(0)} \prod_{ij} \left[\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N)}\right]^{N_{ij}(s)} \tag{3.208}$$

denotes the *probability of uncongested interaction behavior* in patterns of size n under congestion function $c \in C$. In addition, observe that since there are only $n(n-1)$ distinct ordered pairs in any interaction pattern, $s \in S$, it follows by definition that $c(s) \leq n(n-1)$ and hence that

$$\begin{aligned}
&\sum_{s \in S_n(+)} P_c(s) \\
&\geq \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] \sum_{s \in S_n(+)} \prod_{ij} \left[\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N)}\right]^{N_{ij}(s)} \\
&\qquad \cdot \exp[-\theta_c n(n-1)] \\
&= \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] P_c^n(+) \exp[-\theta_c n(n-1)],
\end{aligned} \tag{3.209}$$

where $P_c^n(+)[= 1 - P_c^n(0)]$ denotes the *probability of congested interaction behavior* in patterns of size n under congestion function $c \in C$. Finally, since $\exp[-\theta_c c(s)] \leq 1$ implies by definition that

$$\begin{aligned}
P_c(N = n) &\leq \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] \sum_{s \in S_n} \prod_{ij} \left[\frac{\mathrm{E}_c^o(N_{ij})}{\mathrm{E}_c^o(N)}\right]^{N_{ij}(s)} \\
&= \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] \frac{[\sum_{ij} \mathrm{E}_c^o(N_{ij})]^n}{\mathrm{E}_c^o(N)^n} \\
&= \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)],
\end{aligned} \tag{3.210}$$

it follows by combining (3.206) through (3.210) that

$$\begin{aligned}
\alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] &\geq P_c(N = n) \\
&\geq \alpha_c \frac{\mathrm{E}_c^o(N)^n}{n!} \exp[-\mathrm{E}_c^o(N)] \{[1 - P_c^n(+)] \\
&\qquad + P_c^n(+) \exp[-\theta_c n(n-1)]\}.
\end{aligned} \tag{3.211}$$

From this bounding relation, we see that if $P_c^n(+)$ is close to zero (so that the last bracketed expression is close to one) then these frequency probabilities for $\boldsymbol{P}$ are approximately proportional to those of the congestion-insensitive version, $\boldsymbol{P}^o$, of $\boldsymbol{P}$ [i.e., to Poisson distributions with mean frequencies $\mathrm{E}_c^o(N)$ for each $c \in C$]. Hence it may be concluded that if congested interaction behavior is sufficiently unlikely [i.e., if $P_c^n(+)$ is sufficiently close to zero for all n except for very large (and hence highly unlikely) pattern sizes], then even congestion-sensitive behavior can be well approximated by Poisson frequencies.

3.10 Notes and References

In this final section, it is appropriate to mention some of the more important background material upon which the present concepts and results are based. In addition, a number of related results in the literature are pointed out.

Section 3.1

The basic analytical concepts and results developed in this section are completely standard, and can be found in a variety of references, such as Halmos (1950).

Section 3.2

Our definition of an *interaction probability space* is closely related to the more general notion of a *population process*, first studied by Moyal (1962) [and extended by Carter and Prenter (1972), among others]. Such probability spaces have also been designated as *finite point processes*, as for example in Daley and Vere-Jones (1972) and Kallenberg (1976).

Section 3.3

Our definition of an *interaction process* is closely related to the specialized class of population processes studied in Smith (1986). The *positivity condition*, R1, for interaction processes is a slightly stronger form of the *positive divisibility condition* employed in Smith (1986). [This strengthening guarantees that all Poisson frequency variates for independent interaction processes will be nondegenerate, and in addition, that the probability ratios employed axioms A3 and A4 of Chapter 4 are well defined]. The *symmetry condition*, R2, is identical with the *exchangeability condition* in Smith (1986) [and formally guarantees that the locational attribute pairs, (I_r, J_r), $r = 1, \ldots, n$, in axiom, A1, of Chapter 4 are (conditionally) identically distributed for each given value of $n > 0$]. Finally, the *continuity condition*, R3, is closely related to the notion of C-continuity employed in

Smith (1978a) [and formally guarantees that certain functional representations of the mean frequencies generated by such processes are determined entirely by their restrictions to rational-valued separation configurations, as demonstrated in the proof of Theorem 4.15 in Chapter 4].

Our definition of an *independent interaction process* specializes the notion of an *independent population process* developed in Smith (1986) [and represents a slight generalization of the *independent flow processes* in Smith (1987)]. The *frequency independence axiom*, A2, for such processes is standard in the literature on point processes. General point processes satisfying this independence condition are often said to exhibit *independent increments* [as in Kallenberg (1976, p.6)], or equivalently, to be *free from after effects* [as in Matthes, *et al.* (1978, p.16)]. The *locational independence axiom*, A1, has been studied to a much lesser degree for general point processes [see for example Isham (1975) and Cox and Isham (1980, Sections 3.2 and 5.3)]. However, for measure spaces equipped with an atomless 'volume measure', locational independence is embodied in the well known condition that the locations of any finite number of points in a region of positive finite volume be independently and identically distributed according to the uniform distribution induced by the volume measure on the space [as for example in Karlin and Taylor (1981, Section 16.1), Cox and Isham (1980, Section 6.2), and Smith (1986, Section 3.5)].

Section 3.4

The present notion of *generators* for Poisson processes in Definition 3.5 [first introduced in Smith (1987)], is closely related to certain more general types of generators for Poisson random measures, as studied by Carter and Prenter (1972), Preston (1976), Kallenberg (1976), and Matthes, *et al.* (1978, Section 1.6). Theorem 3.1 for Poisson families is essentially based on the results of Moran (1952) and Chatterji (1963) [and, in its present form, combines the results of Lemmas 3.1 and 3.3 in Smith (1986) with Lemma 3.1 in Smith (1987)]. Related results can be found in Patil and Seshadri (1964), Volodin (1965), Govindarajulu and Leslie (1970), and Kagan, Linnik, and Rao (1973, Theorem 13.4.4). Finally, the fundamental Poisson Characterization Theorem (Theorem 3.2) essentially combines the results in Smith (1986, Theorem 3.1) and Smith (1987, Theorem 5.1).

Section 3.5

The classes of *threshold-interaction processes* and *threshold-frequency processes* developed here extend the formulation in Smith (1987b) to the broader class of interaction processes. Related models can also be found in Tellier and Sankoff (1975) and Luoma and Palomaki (1983).

SECTION 3.6

The class of *search processes* developed here is inspired by the original work of Schneider for the Chicago Area Transportation Study (1960), and also draws on the work of Okabe (1976). However, the present general formulation of such processes in terms of populations of searchers (and in particular, the notion of *independent search processes*) appears to be new.

SECTION 3.7

The notion of *scale independence* and its application to *weakly-independent threshold interaction processes* is new. In addition, the analysis of Poisson approximations to spatial Markov processes with weak neighborhood interaction effects also appears to be new. All other results in this section are well known, and appear in the literature cited.

CHAPTER 4

Gravity Models: Formal Development

4.1 Introduction

In this chapter, the classes of gravity models developed in Chapter 2 are formalized, and are shown to be characterized by the axioms in Chapter 2. We begin in Section 4.2 below by developing a formal specification of the various gravity model classes within the probabilistic framework of Chapter 3. In Section 4.3, we examine in more detail certain of the illustrations of these model classes developed in Chapter 2. In Section 4.4, the general behavioral axioms of Chapter 2 are then formalized within the framework of Chapter 3. The central results of the chapter are developed in Section 4.5. First, the aggregate characterizations of general gravity representations presented in Section 2.2.3(A) of Chapter 2 are formally established in Theorem 4.1. Next, the local characterizations of general gravity representations in Section 2.2.3(B) of Chapter 2 are established in Theorem 4.2. The characterizations of exponential gravity representations in Section 2.4.3 of Chapter 2 are then established in Theorem 4.3. Finally in Section 4.6, the generalizations of gravity models discussed in Sections 2.5.2 and 2.5.3 of Chapter 2 are formally developed within the framework of Chapter 3.

4.2 Definition of Gravity Model Classes

In this section, the gravity models developed in Chapter 2 are restated in a manner which is more suitable for analysis. Throughout this development, the basic notation of Chapter 3 will be employed. In particular, I and J denote the finite state spaces of possible *origins*, i, and *destinations*, j, associated with each possible individual interaction event. All aspects of spatial separation between i and j are then assumed to be summarized a *separation profile*, $c_{ij} \in V$, so that the relevant spatial structure of the origin-destination space, $I \times J$, is summarized by a *separation configuration*, $c = (c_{ij} : ij \in I \times J) \in C = V^{I \times J}$. For each configuration, $c \in C$, the set, $V_c = \{c_{ij} : ij \in I \times J\} \subseteq V \subseteq R^K$, contains all separation profiles which appear at least once in c. Also, each random vector, $\boldsymbol{N}_c = (N_{ij}^c : ij \in I \times J)$, denotes a possible *frequency profile* under configuration, c, and each family of jointly distributed frequency profiles, $\boldsymbol{N} = \{\boldsymbol{N}_c : c \in C\}$, denotes a possible *frequency process* on $I \times J$. The class of frequency processes on

$I \times J$ is again denoted by $\langle I \times J \rangle$, and the subclass of *Poisson processes* by $\langle \text{POISSON} \rangle \subseteq \langle I \times J \rangle$. Finally, recall that for any subset, $A \subseteq R^K$, a function $F : A \to R$, is *nonincreasing* iff for all $x, y \in A$, $x \geq y \Rightarrow F(x) \leq F(y)$. With this notation, the six general gravity model classes for representations of mean interaction frequencies stated in Chapter 2 can now be formalized within the setting of frequency processes, $\boldsymbol{N}$, developed in Chapter 3.

4.2.1 General Gravity Models

We begin by observing that while each of the model types in Chapter 2 is perfectly meaningful for general frequency processes, our statistical analyses in subsequent chapters shall be concerned exclusively with *Poisson frequency processes.* Hence, to avoid repetition of this assumption in later references to each model, we choose to incorporate this Poisson condition as part of the definition of each model. With this in mind, we now have the following formalizations of the six general gravity model types developed in Section 2.2.1 of Chapter 2:

MODEL G1. $\boldsymbol{N} \in \langle \text{POISSON} \rangle$ *and for each* $c \in C$, *there exist functions,* $A_c : I \to R_{++}, B_c : J \to R_{++}$, *and* $F_c : V_c \to R_{++}$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)F_c(c_{ij}). \tag{4.1}$$

MODEL G1*. $\boldsymbol{N}$ *satisfies* MODEL G1 *and in addition, the function,* F_c, *is nonincreasing for each* $c \in C$.

MODEL G2. $\boldsymbol{N} \in \langle \text{POISSON} \rangle$ *and there exists a function,* $F : V \to R_{++}$, *together with functions* $A_c : I \to R_{++}$, $B_c : J \to R_{++}$, *for each* $c \in C$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)F(c_{ij}). \tag{4.2}$$

MODEL G2*. $\boldsymbol{N}$ *satisfies* MODEL G2 *and in addition, the function,* F, *is nonincreasing for each* $c \in C$.

MODEL G3. $\boldsymbol{N} \in \langle \text{POISSON} \rangle$ *and there exist functions,* $B : J \to R_{++}$, *and* $F : V \to R_{++}$ *together with functions,* $A_c : I \to R_{++}$, *for each* $c \in C$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B(j)F(c_{ij}). \tag{4.3}$$

MODEL G3*. *N satisfies* MODEL G6 *and in addition, the function, F, is nonincreasing for each $c \in C$.*

MODEL G4. *$N \in$ ⟨POISSON⟩ and there exist functions, $A: I \to R_{++}$, and $F: V \to R_{++}$ together with functions, $B_c: J \to R_{++}$, for each $c \in C$ such that for all $ij \in I \times J$,*

$$\mathrm{E}(N_{ij}^c) = A(i)B_c(j)F(c_{ij}). \tag{4.4}$$

MODEL G4*. *N satisfies* MODEL G4 *and in addition, the function, F, is nonincreasing for each $c \in C$.*

MODEL G5. *$N \in$ ⟨POISSON⟩ and there exist functions,*

$$\lambda: C \to R_{++}, A: I \to R_{++}, B: J \to R_{++} \text{ and } F: V \to R_{++},$$

such that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = \lambda(c)A(i)B(j)F(c_{ij}). \tag{4.5}$$

MODEL G5*. *N satisfies* MODEL G5 *and in addition, the function, F, is nonincreasing.*

MODEL G6. *$N \in$ ⟨POISSON⟩ and there exist functions, $A: I \to R_{++}$, $B: J \to R_{++}$, and $F: V \to R_{++}$, such that for all $c \in C$ and $ij \in I \times J$,*

$$\mathrm{E}(N_{ij}^c) = A(i)B(j)F(c_{ij}). \tag{4.6}$$

MODEL G6*. *N satisfies* MODEL G6, *and in addition, the function, F, is nonincreasing.*

4.2.2 Exponential Gravity Models

Finally, the six exponential gravity models developed in Section 2.4.1 of Chapter 2 can be formalized in a similar manner. As in Section 2.4 of Chapter 2, it is implicitly assumed here that each component of separation profiles is extensively measurable, so that the following exponential specifications are meaningfully parameterized functions of cost configurations, $c \in C$:

MODEL E1. *$N \in$ ⟨POISSON⟩ and for each $c \in C$, there exists some vector, $\theta_c \in R^K$, and functions, $A_c: I \to R_{++}$ and $B_c: J \to R_{++}$, such that for all $ij \in I \times J$,*

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)\exp[-\theta_c^t c_{ij}]. \tag{4.7}$$

MODEL E1*. $\boldsymbol{N}$ *satisfies* MODEL E1, *and in addition,* $\theta_c \in R_+^K$ *for all* $c \in C$.

MODEL E2. $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ *and there exists a vector,* $\theta \in R^K$, *and functions,* $A_c : I \to R_{++}$, *and* $B_c : J \to R_{++}$, *for each* $c \in C$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)\exp[-\theta^t c_{ij}]. \tag{4.8}$$

MODEL E2*. $\boldsymbol{N}$ *satisfies* MODEL E5, *and in addition,* $\theta \in R_+^K$.

MODEL E3. $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ *and there exists a vector,* $\theta \in R^K$, *and function,* $B : J \to R_{++}$, *together with functions* $A_c : I \to R_{++}$, *for each* $c \in C$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B(j)\exp[-\theta^t c_{ij}]. \tag{4.9}$$

MODEL E3*. $\boldsymbol{N}$ *satisfies* MODEL E6, *and in addition,* $\theta \in R_+^K$.

MODEL E4. $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ *and there exists a vector,* $\theta \in R^K$, *and function,* $A : I \to R_{++}$, *together with functions* $B_c : J \to R_{++}$, *for each* $c \in C$ *such that for all* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A(i)B_c(j)\exp[-\theta^t c_{ij}]. \tag{4.10}$$

MODEL E4*. $\boldsymbol{N}$ *satisfies* MODEL E4, *and in addition,* $\theta \in R_+^K$.

MODEL E5. $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ *and there exists a vector,* $\theta \in R^K$, *and functions,* $\lambda : C \to R_{++}, A : I \to R_{++}$, *and* $B : J \to R_{++}$, *such that for all* $c \in C$ *and* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = \lambda(c)A(i)B(j)\exp[-\theta^t c_{ij}]. \tag{4.11}$$

MODEL E5*. $\boldsymbol{N}$ *satisfies* MODEL E5, *and in addition,* $\theta \in R_+^K$.

MODEL E6. $\boldsymbol{N} \in \langle\text{POISSON}\rangle$ *and there exists a vector,* $\theta \in R^K$, *and functions* $A : I \to R_{++}$ *and* $B : J \to R_{++}$ *such that for all* $c \in C$ *and* $ij \in I \times J$,

$$\mathrm{E}(N_{ij}^c) = A(i)B(j)\exp[-\theta^t c_{ij}]. \tag{4.12}$$

MODEL E6*. $\boldsymbol{N}$ *satisfies* MODEL E6, *and in addition,* $\theta \in R_+^K$.

4.2.3 Relationships Among Model Types

If we now let $\langle \text{MODEL } k\rangle \subseteq \langle I \times J\rangle$ denote the subclass of frequency processes representable by MODEL k for each model type k in Section 4.2.2 above, then the following relations among these subclasses are immediate consequences of the above definitions:

$$\langle\text{MODEL G3}\rangle \cap \langle\text{MODEL G4}\rangle \subseteq \langle\text{MODEL G2}\rangle \subseteq \langle\text{MODEL G1}\rangle, \quad (4.13)$$

$$\langle\text{MODEL G3}^*\rangle \cap \langle\text{MODEL G4}^*\rangle \subseteq \langle\text{MODEL G2}^*\rangle \subseteq \langle\text{MODEL G1}^*\rangle, \quad (4.14)$$

$$\langle\text{MODEL G6}\rangle \subseteq \langle\text{MODEL G5}\rangle \subseteq \langle\text{MODEL G2}\rangle, \quad (4.15)$$

$$\langle\text{MODEL G6}^*\rangle \subseteq \langle\text{MODEL G5}^*\rangle \subseteq \langle\text{MODEL G1}^*\rangle, \quad (4.16)$$

$$\langle\text{MODEL E3}\rangle \cap \langle\text{MODEL E4}\rangle \subseteq \langle\text{MODEL E2}\rangle \subseteq \langle\text{MODEL E1}\rangle, \quad (4.17)$$

$$\langle\text{MODEL E3}^*\rangle \cap \langle\text{MODEL E4}^*\rangle \subseteq \langle\text{MODEL E2}^*\rangle \subseteq \langle\text{MODEL E1}^*\rangle, \quad (4.18)$$

$$\langle\text{MODEL E6}\rangle \subseteq \langle\text{MODEL E5}\rangle \subseteq \langle\text{MODEL E1}\rangle, \quad (4.19)$$

$$\langle\text{MODEL E6}^*\rangle \subseteq \langle\text{MODEL E5}^*\rangle \subseteq \langle\text{MODEL E1}^*\rangle. \quad (4.20)$$

In addition, the following two relations hold for $k = 1, \ldots, 6$:

$$\langle\text{MODEL E}k^*\rangle \subseteq \langle\text{MODEL G}k^*\rangle \subseteq \langle\text{MODEL G}k\rangle, \quad (4.21)$$

$$\langle\text{MODEL E}k^*\rangle \subseteq \langle\text{MODEL E}k\rangle \subseteq \langle\text{MODEL G}k\rangle. \quad (4.22)$$

The above relationships among models can be depicted graphically as in Figure 4.1, where the arrows indicate directions of implication. For example, the relationships in expression (4.15) above are depicted graphically by the arrow sequence, G6$\rightarrow$ G5$\rightarrow$ G2.

4.3 Examples of Gravity Model Classes

In this section we establish the results on gravity representations of mean interaction frequencies discussed in Examples 1 through 4 of Chapter 2. To do so, we begin in Section 4.3.1 below with a formal development of the class of *Carroll-Bevis processes* introduced in Example 1 of Chapter 2. This is followed in Section 4.3.2 with an extension of the formal analysis of *threshold interaction processes* begun in Section 3.7 above. In particular, we formalize the additional conditions on such processes which yield the various gravity representations discussed in Example 2 of Chapter 2. Next, in Section 4.3.3, we present a formal development of the representational properties of *Kullback-Leibler processes* introduced in Example 3 of Chapter 2. Finally, we present a formal development in Section 4.3.3 of the class of *simple search processes* begun in Section 3.8 above.

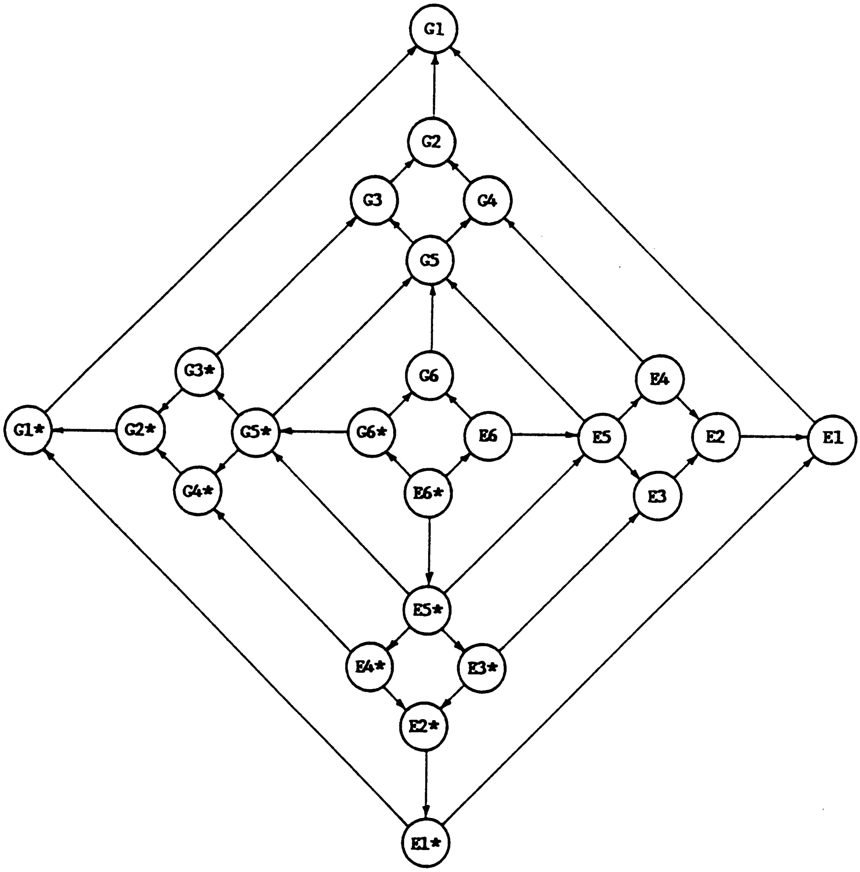

Exhibit 4.1: Relationships among Models

4.3.1 Carroll-Bevis Processes

To analyze the class of interaction processes introduced in Example 1 of Chapter 2, we begin formalizing the various notions of spatially insensitive interaction behavior upon which these processes were based:

Definition 4.1
(i) An interaction process, $\boldsymbol{P}^{\circ} = \{P_c^{\circ} : c \in C\}$, with mean frequencies, $\mathrm{E}_c^{\circ}(N_{ij}) > 0$, for each $c \in C$ and $ij \in I \times J$ is said to be *location-insensitive* iff for all $c \in C$, $i, g \in I$, and $j, h \in J$,

$$\frac{\mathrm{E}_c^{\circ}(N_{ij})}{\mathrm{E}_c^{\circ}(N_{ih})} = \frac{\mathrm{E}_c^{\circ}(N_{gj})}{\mathrm{E}_c^{\circ}(N_{gh})}. \tag{4.23}$$

(ii) In addition, $\boldsymbol{P}^{\circ}$ is said to be *configuration-insensitive* iff the following stronger conditions are satisfied for all $c, c' \in C$, $i, g \in I$, and $j, h \in J$,

$$\frac{\mathrm{E}_c^{\circ}(N_{ij})}{\mathrm{E}_c^{\circ}(N_{ih})} = \frac{\mathrm{E}_{c'}^{\circ}(N_{gj})}{\mathrm{E}_{c'}^{\circ}(N_{gh})}, \tag{4.24}$$

$$\frac{\mathrm{E}_c^{\circ}(N_i)}{\mathrm{E}_c^{\circ}(N_g)} = \frac{\mathrm{E}_{c'}^{\circ}(N_i)}{\mathrm{E}_{c'}^{\circ}(N_g)}. \tag{4.25}$$

(iii) Finally, a configuration-insensitive process, $\boldsymbol{P}^{\circ}$, is said to be *strongly configuration-insensitive* iff for all $c, c' \in C$ it is also true that

$$\mathrm{E}_c^{\circ}(N) = \mathrm{E}_{c'}^{\circ}(N). \tag{4.26}$$

With respect to these null hypotheses, the corresponding classes of Carroll-Bevis processes can now be defined as follows:

Definition 4.2
(i) An interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is designated as a *Carroll-Bevis process* (*CB-process*) iff there exists a location-insensitive process, $\boldsymbol{P}^{\circ}$, such that for all $c \in C$ and $ij, gh \in I \times J$,

$$c_{ij} = c_{gh} \Rightarrow \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c^{\circ}(N_{ij})} = \frac{\mathrm{E}_c(N_{gh})}{\mathrm{E}_c^{\circ}(N_{gh})}. \tag{4.27}$$

(ii) In addition, $\boldsymbol{P}$ is said to be a *uniform CB-process* iff there exists a configuration-insensitive process, $\boldsymbol{P}^{\circ}$, such that for all $c, c' \in C$ and $ij, gh \in I \times J$,

$$c_{ij} = c'_{gh} \Rightarrow \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c^{\circ}(N_{ij})} = \frac{\mathrm{E}_{c'}(N_{gh})}{\mathrm{E}_{c'}^{\circ}(N_{gh})}. \tag{4.28}$$

(iii) Finally, $\boldsymbol{P}$ is said to be a *simple CB-process* iff $\boldsymbol{P}$ satisfies (4.28) with respect to a strongly configuration-insensitive process, $\boldsymbol{P}^{\circ}$.

With these definitions, we now have the following representation properties for mean interaction frequencies of such processes:

Proposition 4.1
(i) *For every CB-process, $\boldsymbol{P} = \{P_c : c \in C\}$, there exist positive functions, $A_c : I \to R_{++}, B_c : J \to R_{++}$, and $F_c : V_c \to R_{++}$ for each $c \in C$, such that for all $ij \in I \times J$,*

$$\mathrm{E}_c(N_{ij}) = A_c(i)B_c(j)F_c(c_{ij}). \tag{4.29}$$

(ii) *If $\boldsymbol{P}$ is a uniform CB-process, then there exist positive functions, $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : V \to R_{++}$, such that for all $ij \in I \times J$,*

$$\mathrm{E}_c(N_{ij}) = \lambda(c)A(i)B(j)F(c_{ij}). \tag{4.30}$$

(iii) *Finally, if $\boldsymbol{P}$ is a simple CB-process, then the functions in* (4.30) *may be chosen so that $\lambda(c) = 1$ for all $c \in C$.*

PROOF: **(i)** To establish (4.29), let $\boldsymbol{P}^{\circ}$ denote any location-insensitive process satisfying (4.27) for $\boldsymbol{P}$. Next, for each $c \in C$ and $v \in V_c$, choose any origin-destination pair, $i_v j_v \in I \times J$ with $c_{i_v j_v} = v$, and define the function, $F_c : V_c \to R_{++}$ for all $v \in V_c$ by

$$F_c(v) = \frac{\mathrm{E}_c(N_{i_v j_v})}{\mathrm{E}_c^{\circ}(N_{i_v j_v})}. \tag{4.31}$$

Then for each $v \in V_c$ and all $ij \in I \times J$ with $c_{ij} = v$, it follows from (4.27) that $c_{ij} = v = c_{i_v j_v}$ implies

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \mathrm{E}_c^{\circ}(N_{ij}) \left[\frac{\mathrm{E}_c(N_{i_v j_v})}{\mathrm{E}_c^{\circ}(N_{i_v j_v})}\right] = \mathrm{E}_c^{\circ}(N_{ij})F_c(v) \\ &= \mathrm{E}_c^{\circ}(N_{ij})F_c(c_{ij}). \end{aligned} \tag{4.32}$$

Moreover, since $\boldsymbol{P}^{\circ}$ satisfies (4.23), it also follows [by rearranging (4.23)] that for any fixed origin, $a \in I$, the quantities

$$A_c(i) = \frac{\mathrm{E}_c^{\circ}(N_{ij})}{\mathrm{E}_c^{\circ}(N_{aj})}, \quad i \in I \tag{4.33}$$

are independent of the choice of $j \in J$, and hence yield a well-defined function, $A_c : I \to R_{++}$. Finally, letting the function, $B_c : J \to R_{++}$, be defined for each $j \in J$ by

$$B_c(j) = \mathrm{E}_c^{\circ}(N_{aj}) \tag{4.34}$$

it follows by combining (4.32), (4.33), and (4.34) that (4.29) must hold for this choice of functions.

(ii) Next, to establish (4.30), let $\boldsymbol{P}^{\circ}$ now denote any configuration-insensitive process satisfying (4.28) for $\boldsymbol{P}$, and observe that for any fixed origin-destination pair, $ab \in I \times J$, there exists for each $v \in V$ some configuration,

$c_v = (c_{vij} : ij \in I \times J) \in C$ $[= V^{I\times J}]$, with $c_{vab} = v$. Hence, letting the function, $F : V \to R_{++}$, be defined for all $v \in V$ by

$$F(v) = \frac{\mathrm{E}_{c_v}(N_{ab})}{\mathrm{E}^{\circ}_{c_v}(N_{ab})}, \tag{4.35}$$

it now follows from (4.28) that for any $v \in V$, $c \in C$, and $ij \in I \times J$ with $c_{ij} = v = c_{vab}$,

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \mathrm{E}^{\circ}_c(N_{ij}) \left[\frac{\mathrm{E}_{c_v}(N_{ab})}{\mathrm{E}^{\circ}_{c_v}(N_{ab})} \right] = \mathrm{E}^{\circ}_c(N_{ij}) F(v) \\ &= \mathrm{E}^{\circ}_c(N_{ij}) F(c_{ij}). \end{aligned} \tag{4.36}$$

Moreover, since $\boldsymbol{P}^{\circ}$ satisfies (4.24), it follows that if for any fixed $\sigma \in C$, $a \in I$, and $b \in J$ we let the function, $B : J \to R_{++}$, be defined for all $j \in J$ by

$$B(j) = \frac{\mathrm{E}^{\circ}_{\sigma}(N_{aj})}{\mathrm{E}^{\circ}_{\sigma}(N_{ab})}, \tag{4.37}$$

then we see from (4.24) that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}^{\circ}_c(N_{ij}) = \mathrm{E}^{\circ}_c(N_{ib}) \left[\frac{\mathrm{E}^{\circ}_{\sigma}(N_{aj})}{\mathrm{E}^{\circ}_{\sigma}(N_{ab})} \right] = \mathrm{E}^{\circ}_c(N_{ib}) B(j). \tag{4.38}$$

Similarly, if we define the function, $A : I \to R_{++}$, for all $i \in I$ by

$$A(i) = \frac{\mathrm{E}^{\circ}_{\sigma}(N_i)}{\mathrm{E}^{\circ}_{\sigma}(N_a)} \tag{4.39}$$

and observe from (4.38) that

$$\frac{\mathrm{E}^{\circ}_c(N_i)}{\mathrm{E}^{\circ}_c(N_a)} = \frac{\sum_j \mathrm{E}^{\circ}_c(N_{ij})}{\sum_j \mathrm{E}^{\circ}_c(N_{aj})} = \frac{\mathrm{E}^{\circ}_c(N_{ib}) \sum_j B(j)}{\mathrm{E}^{\circ}_c(N_{ab}) \sum_j B(j)} = \frac{\mathrm{E}^{\circ}_c(N_{ib})}{\mathrm{E}^{\circ}_c(N_{ab})}, \tag{4.40}$$

then we may conclude from (4.25) together with (4.39) and (4.40) that for all $c \in C$ and $i \in I$,

$$\begin{aligned} \mathrm{E}^{\circ}_c(N_{ib}) &= \mathrm{E}^{\circ}_c(N_{ab}) \left[\frac{\mathrm{E}^{\circ}_c(N_i)}{\mathrm{E}^{\circ}_c(N_a)} \right] = \mathrm{E}^{\circ}_c(N_{ab}) \left[\frac{\mathrm{E}^{\circ}_{\sigma}(N_i)}{\mathrm{E}^{\circ}_{\sigma}(N_a)} \right] \\ &= \mathrm{E}^{\circ}_c(N_{ab}) A(i). \end{aligned} \tag{4.41}$$

Finally, letting the function, $\lambda : C \to R_{++}$, be defined for all $c \in C$ by

$$\lambda(c) = \mathrm{E}^{\circ}_c(N_{ab}), \tag{4.42}$$

it follows by combining (4.36), (4.38), (4.41) and (4.42) that (4.30) must hold for this choice of functions.

(iii) To establish that (4.26) implies that λ may be chosen to be identically one, observe first from (4.38), (4.41) and (4.42) that

$$\begin{aligned} &\mathrm{E}_c^{\circ}(N_{ij}) = \lambda(c)A(i)B(j) \\ &\quad \Rightarrow \mathrm{E}_c^{\circ}(N) = \sum_{ij} \mathrm{E}_c^{\circ}(N_{ij}) = \lambda(c) \sum_{ij} A(i)B(j), \end{aligned} \tag{4.43}$$

and hence that for all $c, c' \in C$,

$$\frac{\mathrm{E}_c^{\circ}(N)}{\mathrm{E}_{c'}^{\circ}(N)} = \frac{\lambda(c)}{\lambda(c')}. \tag{4.44}$$

Thus (4.26) implies that λ must be constant, and the desired representation is obtained by absorbing this constant into the definition of A, B, or F. □

As an immediate consequence of this result, we obtain the following instances of models G1, G5, and G6, where $\boldsymbol{N_P}$ denotes the frequency process generated by interaction process, $\boldsymbol{P}$, as in Chapter 3:

Corollary 4.1 *For any given independent interaction process,* $\boldsymbol{P}$,

(i) *If* $\boldsymbol{P}$ *is a CB-process, then* $\boldsymbol{N_P} \in \langle \text{MODEL G1} \rangle$.

(ii) *If* $\boldsymbol{P}$ *is a uniform CB-process, then* $\boldsymbol{N_P} \in \langle \text{MODEL G5} \rangle$.

(iii) *If* $\boldsymbol{P}$ *is a simple CB-process, then* $\boldsymbol{N_P} \in \langle \text{MODEL G6} \rangle$.

PROOF: If $\boldsymbol{P}$ is an independent interaction process, then

$$\boldsymbol{N_P} \in \langle \text{POISSON} \rangle$$

by Theorem 3.2. Hence assertions (i), (ii), and (iii) follow at once from expression (3.69) together with parts (i), (ii), and (iii) of Proposition 4.1, respectively. □

4.3.2 Threshold Interaction Processes

To establish the representation properties of mean threshold interaction frequencies, we begin by formalizing the notion of fully independent threshold interaction processes employed in Example 2 of Chapter 2:

Definition 4.3 A TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, is designated as a *fully independent TI-process* iff $\boldsymbol{P}$ satisfies conditions C2 and C3 of Definition 3.7 together with the following stronger version of condition C1:

C1′. (Origin-Destination Independence)

$$P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) = \prod_{r=1}^{n} P_c^n(I_r = i_r) P_c^n(J_r = j_r), \tag{4.45}$$

for all $i_1, \ldots, i_n \in I$ *and all* $j_1, \ldots, j_n \in J$.

With this definition, we now have the following instance of MODEL G1* for threshold interaction processes:

Proposition 4.2 *If* $\boldsymbol{P}$ *is a fully independent TI-process, then*

$$\boldsymbol{N_P} \in \langle \text{MODEL G1}^* \rangle.$$

PROOF: First observe from expression (3.100) in Theorem 3.4 of Chapter 3 together with condition C1′ above (evaluated at $n = 1$) that mean interaction frequencies for such processes have the following representation for all $c \in C$ and $ij \in I \times J$:

$$\mathrm{E}(N_{ij}^c) = \mathrm{E}_c(N) P_c^1(I_1 = i) P_c^1(J_1 = j) P_c^1(T_1 \geq c_{ij}). \tag{4.46}$$

Hence letting the functions, $A_c : I \to R_{++}$, $B_c : J \to R_{++}$, and $F_c : V_c \to R_{++}$ be defined, respectively, by $A_c(i) = \mathrm{E}_c(N) P_c^1(I_1 = i), B_c(j) = P_c^1(J_1 = j)$, and $F_c(c_{ij}) = P_c^1(T_1 \geq c_{ij})$, for all $ij \in I \times J$, it follows at once from (4.46) that these mean frequencies are the form (4.1). Moreover, it must be true for any profiles, $v, v' \in R^K$, that $v \geq v' \Rightarrow \{\omega \in \Omega_1 : T_1(\omega) \geq v\} \subseteq \{\omega \in \Omega_1 : T_1(\omega) \geq v'\} \Rightarrow P_c^1(T_1 \geq v) \leq P_c^1(T_1 \geq v') \Rightarrow F_c(v) \leq F_c(v')$. Thus F_c is nonincreasing, and it follows that $\boldsymbol{N_P} \in \langle \text{MODEL G1}^* \rangle$.
□

Next we formalize the configuration-free properties of Example 2 in Chapter 2 within the present framework. In particular, if we recall the σ-fields, $\overline{M}$ and $\overline{M}_n$, defined in expressions (3.81) and (3.86) above, then in terms of these concepts we now have:

Definition 4.4
(i) A TI-process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be *conditionally configuration-free* iff for all $c, c' \in C$, $n \in Z_{++}$ and conditional events, $A \in \overline{M}_n$,

$$P_c^n(A) = P_{c'}^n(A). \tag{4.47}$$

(ii) Similarly, $\boldsymbol{P}$ is said to be *configuration-free* iff for all $c, c' \in C$ and (unconditional) events, $A \in \overline{M}$,

$$P_c(A) = P_{c'}(A). \tag{4.48}$$

With these definitions, we now have the following additional representational properties of mean interaction frequencies for fully independent TI-processes:

Proposition 4.3 *For any fully independent TI-process,* $\boldsymbol{P} = \{P_c : c \in C\}$:

(i) *If* $\boldsymbol{P}$ *is conditionally configuration-free then* $\boldsymbol{N_P} \in \langle \text{MODEL G5}^* \rangle$.

(ii) *If* $\boldsymbol{P}$ *is configuration-free then* $\boldsymbol{N_P} \in \langle \text{MODEL G6}^* \rangle$.

PROOF: **(i)** If a fully independent TI-process, $\boldsymbol{P}$, is conditionally configuration-free, then by setting $n = 1$ in (4.47), it follows that the event probabilities, $P_c^1(I_1 = i), P_c^1(J_1 = j)$, and $P_c^1(T_1 \geq c_{ij})$ in (4.46) are independent of the choice of configuration, c. Hence, if for any fixed configuration, $\sigma \in C$, we now define the functions, $\lambda : C \to R_{++}$, $A : I \to R_{++}$, $B : J \to R_{++}$, and $F : V \to R_{++}$ respectively by $\lambda(c) = \mathrm{E}_c(N)$, $A(i) = P_\sigma^1(I_1 = i)$, $B(j) = P_\sigma^1(J_1 = j)$, and $F(v) = P_\sigma^1(T_1 \geq v)$ for all $c \in C, ij \in I \times J$ and $v \in V(\subseteq R^K)$, it follows at once from (4.46) that

$$\begin{aligned} \mathrm{E}(N_{ij}^c) &= \mathrm{E}_c(N) P_\sigma^1(I_1 = i) P_\sigma^1(J_1 = j) P_\sigma^1(T_1 \geq c_{ij}) \\ &= \lambda(c) A(i) B(j) F(c_{ij}), \end{aligned} \tag{4.49}$$

and hence that these mean interaction frequencies are the form (4.5). But since F is also nonincreasing by the argument in Proposition 4.2 above, we must have $\boldsymbol{N_P} \in \langle \text{MODEL G5}^* \rangle$.

(ii) If P satisfies the stronger configuration-free condition in (4.48) above, then the probability $P_c(A)$ of every $\overline{M}$-measurable event is independent of the choice of configuration, c. Hence, by choosing the fixed configuration, $\sigma \in C$, in part (i) above and considering the $\overline{M}$-measurable events, $(N = n) \equiv \{\omega \in \Omega : N(\omega) = n\}$, it follows now follows that $\mathrm{E}_c(N) = \sum_n n P_c(N = n) = \sum_n n P_\sigma(N = n) = \mathrm{E}_\sigma(N)$ for all $c \in C$. Moreover, since $\overline{M}_1 \subseteq \overline{M}$ implies that the event probabilities, $P_c^1(I_1 = i), P_c^1(J_1 = j)$, and $P_c^1(T_1 \geq c_{ij})$ in (4.46) are again independent of c, if we choose the same functions B and F in part (i) above, and let the new function, $A : I \to R_{++}$, be defined for all $i \in I$ by $A(i) = \mathrm{E}_\sigma(N) P_\sigma^1(I_1 = i)$, it now follows that the mean interaction frequencies in (4.49) are of the form (4.6), and hence that $\boldsymbol{N_P} \in \langle \text{MODEL G6}^* \rangle$. □

4.3.3 KULLBACK-LEIBLER PROCESSES

To analyze the class of interaction processes introduced in Example 3 of Chapter 2, we now focus on *independent* interaction processes, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, M \rangle$. Since each such process is uniquely defined by the pattern probabilities, $P_c[\Omega(s)]$, for each $c \in C$ and $s \in S$, and since the

Poisson Characterization Theorem (Theorem 3.2) implies [as in expression (3.76)] that

$$P_c[\Omega(s)] = \frac{\mathrm{E}_c(N)^{N(s)}}{N(s)!}\exp[-\mathrm{E}_c(N)]\prod_{ij}\left[\frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N)}\right]^{N_{ij}(s)}, \tag{4.50}$$

it follows that each independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle\Omega, M\rangle$ is uniquely defined by its associated family of *mean interaction frequencies*, $\{\mathrm{E}_c(N_{ij}) : ij \in I \times J, c \in C\}$. In this context, if the marginal interaction probabilities for each such process are defined [as in (2.137) and (2.138)] for all $ij \in I \times J$ and $c \in C$ by

$$p_c(i\cdot) = \sum_h p_c(ih) = \sum_h [\mathrm{E}_c(N_{ih})/\mathrm{E}_c(N)] = \mathrm{E}_c(N_i)/\mathrm{E}_c(N), \tag{4.51}$$

$$p_c(\cdot j) = \sum_g p_c(gj) = \sum_g [\mathrm{E}_c(N_{gj})/\mathrm{E}_c(N)] = \mathrm{E}_c(N_j)/\mathrm{E}_c(N), \tag{4.52}$$

we can say that:

Definition 4.5 An independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle\Omega, M\rangle$ is designated as an *origin-destination independent* (*OD-independent*) *process* iff for all $ij \in I \times J$ and $c \in C$,

$$p_c(ij) = p_c(i\cdot)p_c(\cdot j). \tag{4.53}$$

With this definition, we have the following equivalent characterization of location-insensitivity for such processes, as in Definition 4.1 above:

Proposition 4.4 *An independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is location-insensitive iff it is OD-independent.*

PROOF: **(i)** If $\boldsymbol{P}$ is location-insensitive, then by (3.69) and (3.77) together with (4.23) it follows that for all $i, g \in I, j, h \in J$, and $c \in C$,

$$\begin{aligned} \frac{p_c(ij)}{p_c(ih)} = \frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{ih})} &= \frac{\mathrm{E}_c(N_{gj})}{\mathrm{E}_c(N_{gh})} = \frac{p_c(gj)}{p_c(gh)} \\ \Rightarrow p_c(ij)p_c(gh) &= p_c(ih)p_c(gj). \end{aligned} \tag{4.54}$$

Hence, summing both sides of (4.54) with respect to g and h we obtain

$$\begin{aligned} p_c(ij) = p_c(ij)\sum_{gh} p_c(gh) &= \sum_{gh}[p_c(ih)p_c(gj)] \\ &= \sum_h p_c(ih)\cdot\sum_g p_c(gj) = p_c(i\cdot)p_c(\cdot j), \end{aligned} \tag{4.55}$$

and may conclude that $\boldsymbol{P}$ is OD-independent.

(ii) Conversely, if $\boldsymbol{P}$ is OD-independent, then it follows at once from (4.53) that for all $i, g \in I, j, h \in J$, and $c \in C$,

$$\begin{aligned}\frac{\mathrm{E}_c(N_{ij})}{\mathrm{E}_c(N_{ih})} &= \frac{p_c(ij)}{p_c(ih)} = \frac{p_c(i\cdot)p_c(\cdot j)}{p_c(i\cdot)p_c(\cdot h)} = \frac{p_c(\cdot j)}{p_c(\cdot h)} \\ &= \frac{p_c(g\cdot)p_c(\cdot j)}{p_c(g\cdot)p_c(\cdot h)} = \frac{p_c(gj)}{p_c(gh)} = \frac{\mathrm{E}_c(N_{gj})}{\mathrm{E}_c(N_{gh})},\end{aligned} \tag{4.56}$$

and hence that $\boldsymbol{P}$ is location-insensitive. □

Next, if [as in (2.139)] we designate a pair of independent interaction processes, $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$, on $\langle \Omega, M \rangle$ as *mean-activity-equivalent processes* iff for all $ij \in I \times J$ and $c \in C$ it is true that

$$[\mathrm{E}_c(N_i), \mathrm{E}_c(N_j)] = [\tilde{\mathrm{E}}_c(N_i), \tilde{\mathrm{E}}_c(N_j)], \tag{4.57}$$

then we have the following uniqueness result:

Proposition 4.5 *For any independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, M \rangle$, there exists a unique OD-independent process, $\boldsymbol{P}^{\circ} = \{P_c^{\circ} : c \in C\}$, which is mean-activity-equivalent to $\boldsymbol{P}$.*

PROOF: First recall from (4.50) that each independent interaction process on $\langle \Omega, M \rangle$ is uniquely defined by its associated mean interaction frequencies. Hence if the mean interaction frequencies for $\boldsymbol{P}$ are denoted by $\mathrm{E}_c(N_{ij})$ [with corresponding mean activity levels, $\mathrm{E}_c(N_i), \mathrm{E}_c(N_j)$, and $\mathrm{E}_c(N)$], then the independent interaction process, $\boldsymbol{P}^{\circ} = \{P_c : c \in C\}$, on $\langle \Omega, M \rangle$, defined by the mean interaction frequencies

$$\mathrm{E}_c^{\circ}(N_{ij}) = \frac{\mathrm{E}_c(N_i)\mathrm{E}_c(N_j)}{\mathrm{E}_c(N)}, \tag{4.58}$$

is clearly unique. Moreover, since (4.58) implies that

$$\mathrm{E}_c^{\circ}(N_i) = \sum_j \mathrm{E}_c^{\circ}(N_{ij}) = \sum_j [\mathrm{E}_c(N_i)\mathrm{E}_c(N_j)/\mathrm{E}_c(N)] = \mathrm{E}_c(N_i),$$

and similarly that $\mathrm{E}_c^{\circ}(N_j) = \mathrm{E}_c(N_j)$, it follows that $\boldsymbol{P}^{\circ}$ and $\boldsymbol{P}$ are mean-activity-equivalent. But this in turn implies from (4.51), (4.52) and (4.58) that

$$p_c^{\circ}(ij) = \frac{\mathrm{E}_c^{\circ}(N_{ij})}{\mathrm{E}_c^{\circ}(N)} = \frac{\mathrm{E}_c^{\circ}(N_i)}{\mathrm{E}_c^{\circ}(N)} \cdot \frac{\mathrm{E}_c^{\circ}(N_j)}{\mathrm{E}_c^{\circ}(N)} = p_c^{\circ}(i\cdot)p_c^{\circ}(\cdot j) \tag{4.59}$$

holds for all $ij \in I \times J$ and $c \in C$, and hence that $\boldsymbol{P}^{\circ}$ is OD-independent. Finally, since (4.59) also implies that any other OD-independent process which is mean-activity-equivalent to $\boldsymbol{P}$ must necessarily satisfy (4.58), we may conclude that there is exactly one such process. □

Hence [in a manner paralleling Example 3 of Chapter 2], we now designate this unique process, $\boldsymbol{P}^{\circ}$, as the *OD-independent version* of $\boldsymbol{P}$. To compare $\boldsymbol{P}$ and $\boldsymbol{P}^{\circ}$, recall from Corollary 3.3 on multinomial sampling that if the random vector, $\boldsymbol{N}:\Omega \to Z_{+}^{I\times J}$, is defined for all $\omega \in \Omega$ by $\boldsymbol{N}(\omega) = [N_{ij}(\omega): ij \in I \times J]$, then for each *frequency profile*, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, with $\sum_{ij} n_{ij} = n \in Z_{++}$, and each $c \in C$,

$$P_c(\boldsymbol{N} = \boldsymbol{n} \,|\, N = n) = n! \prod_{ij} \frac{p_c(ij)^{n_{ij}}}{n_{ij}!}, \tag{4.60}$$

under distribution P_c. Next observe that if for each n we define the $\boldsymbol{P}$*-modal profile*, $\boldsymbol{m}_n^c = (m_{nij}^c : ij \in I \times J)$, by

$$m_{nij}^c = p_c(ij)n, \quad ij \in I \times J, \tag{4.61}$$

then for large population sizes, n, the most probable conditional frequency profiles given $N = n$ [i.e., those which maximize (4.60)] are well approximated by these unique $\boldsymbol{P}$-modal profiles. In fact, it can be shown [as for example in Ellis (1985, Theorem I.4.3)] that for each $\varepsilon > 0$ there is a constant $I(\varepsilon) > 0$ such that for all sufficiently large population sizes, $n \in Z_{++}$,

$$P_c[\max\{|1 - [N_{ij}/m_{nij}^c]| : ij \in I \times J\} \geq \varepsilon] \leq \exp[-nI(\varepsilon)]. \tag{4.62}$$

In other words, the ratios of realized frequencies, N_{ij}, to modal frequencies, m_{nij}^c, *converge uniformly to one in probability* as n becomes large, and in addition, the rate of convergence is *exponentially* fast [Ellis (1985, p.48)]. Thus for large population sizes, n, the $\boldsymbol{P}$-modal profile, $\boldsymbol{m}_n^c$, is seen to represent the *overwhelmingly most probable* interaction frequency behavior under distribution P_c. With this in mind, observe next from (4.60) that for any frequency profile, $\boldsymbol{n} = (n_{ij} : ij \in I \times J)$, the *relative likelihood* of $\boldsymbol{n}$ under $\boldsymbol{P}$ versus $\boldsymbol{P}^{\circ}$ is given [as in (2.140)] by

$$\frac{P_c(\boldsymbol{N} = \boldsymbol{n} \,|\, N = n)}{P_c^{\circ}(\boldsymbol{N} = \boldsymbol{n} \,|\, N = n)} = \frac{\prod_{ij} p_c(ij)^{n_{ij}}}{\prod_{ij} p_c^{\circ}(ij)^{n_{ij}}}. \tag{4.63}$$

Hence for large population sizes, n, it follows in particular that with respect to the overwhelmingly most probable frequency behavior under $\boldsymbol{P}$, the relative likelihood of this behavior under $\boldsymbol{P}$ versus $\boldsymbol{P}^{\circ}$ is well approximated by the quantity [as in (2.143)]:

$$\frac{P_c(\boldsymbol{m}_n^c \,|\, n)}{P_c^{\circ}(\boldsymbol{m}_n^c \,|\, n)} = \frac{\prod_{ij} p_c(ij)^{p_c(ij)n}}{\prod_{ij} p_c^{\circ}(ij)^{p_c(ij)n}}. \tag{4.64}$$

These ratios exhibit the following well known property:

Proposition 4.6 *For all $n \in Z_{++}$ and $c \in C$,*

$$\frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^o(\boldsymbol{m}_n^c \mid n)} \geq 1 \tag{4.65}$$

with equality holding iff $P_c = P_c^o$.

PROOF: Recall [as in (2.153) and (2.154)] that if the *mean discrimination information* for p_c against p_c^o under hypothesis p_c is defined by

$$I_{p_c|p_c^o} = \sum_{ij} p_c(ij) \log[p_c(ij)/p_c^o(ij)], \tag{4.66}$$

then the ratio in (4.65) is given in terms of (4.66) by

$$P_c(\boldsymbol{m}_n^c \mid n)/P_c^o(\boldsymbol{m}_n^c \mid n) = \exp[nI_{p_c \mid p_c^o}]. \tag{4.67}$$

But since it is well known [as for example in Ellis (1985, Proposition I.4.1)] that $I_{p_c|p_c^o} \geq 0$, with equality holding iff $p_c = p_c^o$, the desired result follows at once from (4.67). □

Next observe that if we designate the associated cost vector

$$\mathrm{E}_c(c \mid n) = \sum_{ij} c_{ij} m_{nij}^c = n \sum_{ij} c_{ij} p_c(ij) \tag{4.68}$$

as the *conditional mean-cost profile* for population size n under distribution P_c, then the above observations are also seen to imply that for large n, the overwhelmingly most probable profile of total interaction costs exhibited by such populations must be well approximated by $\mathrm{E}_c(c \mid n)$. Given these concepts, we may now define the class of *Kullback-Leibler* processes [paralleling Definition 2.19] as follows:

Definition 4.6 An independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, M \rangle$ is designated as a *Kullback-Leibler process (KL-process)* iff for each $c \in C, n \in Z_{++}$, and each activity-equivalent independent process, $\tilde{\boldsymbol{P}} = \{\tilde{P}_c : c \in C\}$, on $\langle \Omega, M \rangle$ with associated conditional $\tilde{\boldsymbol{P}}$-modal profiles, $\tilde{\boldsymbol{m}}_n^c$, conditional mean-cost profiles, $\tilde{\mathrm{E}}_c(c \mid n)$, and OD-independent version, $\tilde{\boldsymbol{P}}^o$,

$$\tilde{\mathrm{E}}_c(c \mid n) = \mathrm{E}_c(c \mid n) \Rightarrow \frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^o(\boldsymbol{m}_n^c \mid n)} \leq \frac{\tilde{P}_c(\tilde{\boldsymbol{m}}_n^c \mid n)}{\tilde{P}_c^o(\boldsymbol{m}_n^c \mid n)}. \tag{4.69}$$

With this definition, we can now establish the following representation property of mean interaction frequencies for KL-processes:

Proposition 4.7 *If $\boldsymbol{P}$ is a KL-process on $\langle \Omega, M \rangle$, then*

$$\boldsymbol{N_P} \in \langle \text{MODEL E1} \rangle.$$

PROOF: First observe that if $\tilde{p}_c(ij)$ denotes the interaction probabilities for an independent interaction process, $\tilde{\boldsymbol{P}}$, which is mean-activity-equivalent to $\boldsymbol{P}$, then by (4.67), (4.68), and (4.69) it follows that any choice of $n \in Z_{++}$,

$$\begin{aligned}
&\sum_{ij} c_{ij}\tilde{p}_c(ij) = \sum_{ij} c_{ij}p_c(ij) \\
&\quad \Rightarrow n\sum_{ij} c_{ij}\tilde{p}_c(ij) = n\sum_{ij} c_{ij}p_c(ij) \\
&\quad \Rightarrow \tilde{\mathrm{E}}_c(c \mid n) = \mathrm{E}_c(c \mid n) \\
&\quad \Rightarrow \frac{P_c(\boldsymbol{m}_n^c \mid n)}{P_c^\circ(\boldsymbol{m}_n^c \mid n)} \le \frac{\tilde{P}_c(\boldsymbol{m}_n^c \mid n)}{\tilde{P}_c^\circ(\boldsymbol{m}_n^c \mid n)} \qquad (4.70) \\
&\quad \Rightarrow \exp[nI_{p_c \mid p_c^\circ}] \le \exp[nI_{\tilde{p}_c \mid \tilde{p}_c^\circ}] \\
&\quad \Rightarrow I_{p_c \mid p_c^\circ} \le I_{\tilde{p}_c \mid \tilde{p}_c^\circ} \\
&\quad \Rightarrow \sum_{ij} p_c(ij)\log[p_c(ij)/p_c^\circ(ij)] \le \sum_{ij} \tilde{p}_c(ij)\log[\tilde{p}_c(ij)/\tilde{p}_c^\circ(ij)].
\end{aligned}$$

Next observe from the mean-activity-equivalence of $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$, together with the OD-independence of both $\boldsymbol{P}^\circ$ and $\tilde{\boldsymbol{P}}^\circ$, that

$$\begin{aligned}
\tilde{p}_c^\circ(ij) &= \tilde{p}_c^\circ(i\cdot)\tilde{p}_c^\circ(\cdot j) \\
&= \frac{\tilde{\mathrm{E}}_c(N_i)}{\tilde{\mathrm{E}}_c(N)} \cdot \frac{\tilde{\mathrm{E}}_c(N_j)}{\tilde{\mathrm{E}}_c(N)} = \frac{\mathrm{E}_c(N_i)}{\mathrm{E}_c(N)} \cdot \frac{\mathrm{E}_c(N_j)}{\mathrm{E}_c(N)} \qquad (4.71) \\
&= p_c^\circ(i\cdot)p_c^\circ(\cdot j) = p_c^\circ(ij).
\end{aligned}$$

Moreover, since for each positive probability function, $q : I \times J \to R_{++}$, there exists some independent interaction process, $\tilde{\boldsymbol{P}}$, with $\tilde{p}_c(ij) = q(ij)$, $\tilde{\mathrm{E}}_c(N_i) = \mathrm{E}_c(N_i)$ and $\tilde{\mathrm{E}}_c(N_j) = E_c(N_j)$ for all $i, j \in I \times J$, it follows from (4.70) and (4.71) that for all probability functions, $q : I \times J \to R_{++}$,

$$\begin{aligned}
&\sum_{ij} c_{ij}q(ij) = \sum_{ij} c_{ij}p_c(ij) \\
&\quad \Rightarrow \sum_{ij} p_c(ij)\log[p_c(ij)/p_c^\circ(ij)] \le \sum_{ij} q(ij)\log[q(ij)/p_c^\circ(ij)].
\end{aligned} \qquad (4.72)$$

Finally, since all expressions are continuous functions of each probability value, $q(ij), i, j \in I \times J$, [with the standard convention that $0\log(0) = 0$], it follows that if for each nonnegative probability function, $q : I \times J \to R_+$, we choose any sequence of positive probability functions, $q_n : I \times J \to R_{++}$, with $\lim_{n\to\infty} q_n(ij) = q(ij), i, j \in I \times J$, then (4.72) must continue to hold for q. Hence, we see that for each KL-process, $\boldsymbol{P}$, and $c \in C$, the interaction probability distribution, $p_c : I \times J \rightarrow R_{++}$, must be a solution of the programming problem:

$$\text{minimize:} \sum_{ij} q(ij) \log[q(ij)/p_c^{\circ}(ij)] \tag{4.73}$$

subject to

$$\sum_{ij} c_{ij} q(ij) = \sum_{ij} c_{ij} p_c(ij), \tag{4.74}$$

$$\sum_{ij} q(ij) = 1, \tag{4.75}$$

$$q(ij) \geq 0 \quad \text{for all } i, j \in I \times J. \tag{4.76}$$

But since the positivity of p_c ensures the existence of a positive feasible solution, it follows from the positivity of p_c° in (4.73) that this (convex) programming problem has a unique positive solution [see for example Hoeffding (1965, Lemma 4.8)], which is characterized by the first order conditions of the Lagrangian function:

$$\begin{aligned} \mathcal{L}(q, \mu_c, \theta_c) = &\sum_{ij} q(ij) \log[q(ij)/p_c^{\circ}(ij)] + \mu_c[1 - \sum_{ij} q(ij)] \\ &+ \theta_c^t[\sum_{ij} c_{ij} q(ij) - \sum_{ij} c_{ij} p_c(ij)]. \end{aligned} \tag{4.77}$$

Hence $p_c(ij)$ satisfies the condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial q(i,j)} \mathcal{L}(p_c, \mu_c, \theta_c) \\ &= \{1 + \log[p_c(ij)] - \log[p_c^{\circ}(ij)\} + \mu_c - \theta_c^t c_{ij}, \end{aligned} \tag{4.78}$$

which together with the OD-independence of $\boldsymbol{P}^{\circ}$, implies that

$$\begin{aligned} p_c(ij) &= p_c^{\circ}(ij) \exp(\mu_c) \exp[-\theta_c^t c_{ij}] \\ &= p_c^{\circ}(i\cdot) p_c^{\circ}(\cdot j) \exp(\mu_c) \exp[-\theta_c^t c_{ij}]. \end{aligned} \tag{4.79}$$

Thus, letting the functions, $A_c : I \rightarrow R_{++}$ and $B_c : J \rightarrow R_{++}$, be defined respectively for all $i \in I, j \in J$, and $c \in C$, by

$$A_c(i) = \mathrm{E}_c(N) p_c^{\circ}(i\cdot) > 0, \tag{4.80}$$

$$B_c(j) = p_c^{\circ}(\cdot j) \exp(\mu_c) > 0, \tag{4.81}$$

we see from (4.79) that

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \mathrm{E}_c(N) p_c(ij) \\ &= [\mathrm{E}_c(N) p_c(i\cdot)] \cdot [p_c(\cdot j) \exp(\mu_c)] \exp[-\theta_c^t c_{ij}] \\ &= A_c(i) B_c(j) \exp[-\theta_c^t c_{ij}], \end{aligned} \tag{4.82}$$

for this choice of functions. Finally, since the independence of $\boldsymbol{P}$ together with the Poisson Characterization Theorem (Theorem 3.2) implies that $\boldsymbol{N_P} \in \langle\text{POISSON}\rangle$, we may conclude from (3.69) and (4.82) that $\boldsymbol{N_P} \in \langle\text{MODEL E1}\rangle$. □

4.3.4 Simple Search Processes

To establish the mean representation properties of simple search processes developed in Example 4 of Chapter 2, we begin by formalizing this class of processes within the framework of Section 3.8 above. In particular, we now strengthen the class of independent search processes in Definition 3.9 as follows:

Definition 4.7 An independent search process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is designated as a *simple search process* iff $\boldsymbol{P}$ satisfies the following two additional conditions

S6. (Opportunity Independence)*For all* $c \in C$, $\delta = (\delta_\beta : \beta \in B) \in \{0,1\}^B$, *and* $i \in I$,

$$P_c^1(\Delta_1 = \delta \mid I_1 = i) = \prod_{\beta \in B} P_c^1(\Delta_{1\beta} = \delta_\beta). \tag{4.83}$$

S7. (Opportunity Homogeneity)*For all* $\beta, \beta' \in B$ *and* $c, c' \in C$,

$$P_c^1(\Delta_{1\beta} = 1) = P_{c'}^1(\Delta_{1\beta'} = 1). \tag{4.84}$$

[Condition S6 together with conditions S3 and S5 yield a weaker version of condition S3 in Definition 2.20, and condition S7 is identical with condition S5 of Definition 2.20.]

Representations of Realized-Interaction Frequencies

To analyze the properties of simple search processes, we begin by observing from the opportunity homogeneity condition (S7) that the satisfaction probability, $P_c^1(\Delta_{1\beta} = 1)$, is independent of both β and c. Hence the corresponding *rejection probability* that an opportunity will fail to satisfy the searcher's needs can be denoted for any representative choice of c and β by

$$P(\Delta = 0) = P_c^1(\Delta_{1\beta} = 0) = 1 - P_c^1(\Delta_{1\beta} = 1) > 0. \tag{4.85}$$

Next observe that [as in expression (2.168)] if the cardinality of each opportunity set B_j is denoted by b_j, then for each origin-destination pair, $i, j \in I \times J$, the *intervening-opportunity distance* between i and j generated by search scheme, $c \in C$, is given in terms of (3.115) by

$$c_{ij} = \sum_{h \in J_c[j \mid i]} b_h \tag{4.86}$$

(where by definition, $c_{ij} = 0$ if $J[j \mid i] = \emptyset$). In terms of this notation, we now have the following representation of mean realized-interaction frequencies for simple search processes:

Proposition 4.8 *For each simple search process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *on* $\langle \Omega, M \rangle$ *the associated mean realized-interaction frequencies,* $\mathrm{E}_c(N_{ij}^c)$, *are given for* $c \in C$ *and* $i, j \in I \times J$ *by*

$$\mathrm{E}_c(N_{ij}^c) = P(\Delta = 0)^{c_{ij}} \left[1 - P(\Delta = 0)^{b_j}\right] P_c^1(I_1 = i)\mathrm{E}_c(N). \tag{4.87}$$

PROOF: First recall from Theorem 3.5 that for all $c \in C$ and $i, j \in I \times J$,

$$\begin{aligned} \mathrm{E}_c(N_{ij}^c) &= P_c^1(I_1 = i, J_{c1} = j)\mathrm{E}_c(N) \\ &= P_c^1(J_{c1} = j \mid I_1 = i)P_c^1(I_1 = i)\mathrm{E}_c(N). \end{aligned} \tag{4.88}$$

Next observe that if for each $j \in J$ we define the ($\overline{M}$-measurable) function, $\Delta_{1j} : \Omega \to \{0, 1\}$, for all $\omega \in \Omega$ by

$$\Delta_{1j}(\omega) = \max\{\Delta_{1\beta}(\omega) : \beta \in B_j\}, \tag{4.89}$$

then it follows from (3.113) and (3.114) that $\beta_{c1}(\omega) \in B_j$ iff $\Delta_{1j}(\omega) = 1$ and $\Delta_{1\beta}(\omega) = 0$ for all $\beta \in \cup\{B_h : h \in J_c[j \mid i]\}$. Moreover, conditions S6 and S7 also imply that the distribution of Δ_{1j} is independent of the underlying search scheme, and hence [in a manner paralleling (4.85)] that for all $c \in C$ we may write $P_c^1(\Delta_{1j} = 1) = P^1(\Delta_{1j} = 1)$. Thus it follows from conditions S6 and S7 together with (4.85) and (4.86) that

$$\begin{aligned} P_c^1(J_{c1} = j \mid I_1 = i) &= P_c^1(\beta_{c1} \in B_j \mid I_1 = i) \\ &= P_c^1\{(\Delta_{1\beta} = 0 : \beta \in B_h, h \in J_c[j \mid i]), \Delta_{1j} = 1 | I_1 = i\} \\ &= \prod_{h \in J_c[j \mid i]} \left[\prod_{\beta \in B_h} P_c^1(\Delta_{1\beta} = 0)\right] P_c^1(\Delta_{1j} = 1) \\ &= \left[P(\Delta = 0)^{\sum_{h \in J_c[j \mid i]} b_h}\right] P^1(\Delta_{1j} = 1) \\ &= P(\Delta = 0)^{c_{ij}} P^1(\Delta_{1j} = 1). \end{aligned} \tag{4.90}$$

Finally, since S6 and S7 also imply from (4.85) and (4.89) that

$$\begin{aligned} P^1(\Delta_{1j} = 1) &= 1 - P_c^1(\Delta_{1\beta} = 0 : \beta \in B_j) \\ &= 1 - P(\Delta = 0)^{b_j}, \end{aligned} \tag{4.91}$$

we may conclude from (4.88) together with (4.90) and (4.91) that (4.87) must hold. □

As an immediate consequence of this result, we obtain the following instance of MODEL E1* for the realized-interaction frequency processes generated by simple search processes:

Corollary 4.2 *If $\boldsymbol{N_P}$ is the realized-interaction frequency process generated by a simple search process $\boldsymbol{P}$, then $\boldsymbol{N_P} \in \langle$MODEL E1*$\rangle$.*

PROOF: Since $\boldsymbol{P}$ is an independent search process, it follows at once from Theorem 3.5 that $\boldsymbol{N_P} \in \langle$POISSON$\rangle$. Moreover, if the functions, $A_c : I \rightarrow R_{++}$ and $B : J \rightarrow R_{++}$, are defined respectively for all $c \in C$ and $i, j \in I \times J$ by

$$A_c(i) = P_c^1(I_1 = i)\mathrm{E}_c(N), \tag{4.92}$$

$$B(j) = 1 - P(\Delta = 0)^{b_j}, \tag{4.93}$$

and if the scalar $\theta \in R_{++}$ is defined by

$$\theta = -\log P(\Delta = 0) > 0, \tag{4.94}$$

then it follows at once from (4.87) that

$$\begin{aligned} \mathrm{E}_c(N_{ij}^c) &= A_c(i)B(j)\exp\{-[-\log P(\Delta = 0)]c_{ij}\} \\ &= A_c(i)B(j)\exp[-\theta c_{ij}], \end{aligned} \tag{4.95}$$

and hence from the positivity of θ in (4.94) that $\boldsymbol{N_P} \in \langle$MODEL E1*$\rangle$. □

Notice also from the definitions of $P(\Delta = 0)$, c_{ij}, and θ that if b denotes the cardinality of B and if the ($\overline{M}$-measurable) *length function*, $L_{c1} : \Omega \rightarrow Z_+$, is defined for all $\omega \in \Omega$ by

$$L_{c1}(\omega) = \begin{cases} c_i[\beta_{c1}(\omega)], & \beta_{c1}(\omega) \in B \\ b + 1, & \beta_{c1}(\omega) \in o, \end{cases} \tag{4.96}$$

then the argument in (4.90) shows that for all $c \in C$ and $i, j \in I \times J$,

$$\begin{aligned} P_c^1(L_{c1} > c_{ij} \mid I_1 = i) &= P_c^1\{(\Delta_{c1}(\beta) = 0 : \beta \in J_c[j \mid i])| I_1 = i\} \\ &= P(\Delta = 0)^{c_{ij}} \\ &= \exp[-\theta c_{ij}], \end{aligned} \tag{4.97}$$

and hence that the deterrence function in (4.95) defined by (4.94) is precisely the probability that destination j will be reached by a searcher starting at i. Notice also from the Poisson property of N_i in (3.126), together with the general argument in (3.99), that for all $c \in C$ and $i \in I$,

$$\mathrm{E}_c(N_i) = P_c^1(I_1 = i)\mathrm{E}_c(N) = A_c(i). \tag{4.98}$$

Hence, the origin weight function in (4.95) defined by (4.92) is given precisely by the mean search activity at each origin.

Search-Neutral Processes

To establish the stronger representations of mean realized-interaction frequencies under the conditions of search neutrality discussed in Example 4 of Chapter 2, we first restate these conditions as follows:

Definition 4.8
(i) A search process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is said to be *conditionally search-neutral* iff for all $n \in Z_{++}$, $A \in \overline{M}_n$, and $c, c' \in C$,

$$P_c^n(A) = P_{c'}^n(A). \tag{4.99}$$

(ii) In addition, $\boldsymbol{P}$ is simply said to be *search-neutral* iff for all $A \in \overline{M}$ and $c, c' \in C$,

$$P_c(A) = P_{c'}(A). \tag{4.100}$$

With these definitions, we can now establish the following stronger representation properties of mean realized-interaction frequencies under conditions of search neutrality [as parallel of Proposition 4.3 above]:

Proposition 4.9 *For any simple search process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *on* $\langle \omega, \overline{M} \rangle$,

(i) *If* $\boldsymbol{P}$ *is conditionally search-neutral then* $\boldsymbol{N_P} \in \langle$MODEL E5$^*\rangle$.

(ii) *If* $\boldsymbol{P}$ *is search-neutral then* $\boldsymbol{N_P} \in \langle$MODEL E6$^*\rangle$.

Proof: **(i)** If $\boldsymbol{P}$ is conditionally search-neutral, then it follows in particular that the origin probabilities, $P_c^1(I_1 = i)$, in (4.92) are independent of the search scheme, c, and hence can be written as $P^1(I_1 = i)$. Thus if we define the functions $A : I \to R_{++}$ and $\lambda : C \to R_{++}$ for all $i \in I$ and $c \in C$, respectively, by

$$A(i) = P^1(I_1 = i), \tag{4.101}$$

$$\lambda(c) = \mathrm{E}_c(N), \tag{4.102}$$

then it follows at once from (4.92) and (4.95) that each mean realized interaction frequency, $\mathrm{E}_c(N_{ij}^c)$, is now of the form

$$\mathrm{E}_c(N_{ij}^c) = \lambda(c)A(i)B(j)\exp[-\theta c_{ij}]. \tag{4.103}$$

Hence we may conclude from (4.11) and Corollary 4.2 that

$$\boldsymbol{N_P} \in \langle \text{MODEL E5}^* \rangle.$$

(ii) Finally, if it is true in addition that $\boldsymbol{P}$ is search-neutral, then the argument in part (ii) of Proposition 4.3 again shows that $\mathrm{E}_c(N)$ is independent of the search scheme, $c \in C$. Thus we may conclude that the function A_c in expression (4.92) is now independent of c, and hence from (4.95) that $N_{\boldsymbol{P}} \in \langle$MODEL E6$^*\rangle$. □

4.4 Axioms for Interaction Processes

Our objective is to establish characterizations of the above frequency models in terms of those behavioral properties of interaction processes developed in Chapter 2. To do so, we begin by recalling from Chapter 2 that all these properties were stated in terms of the *spatial interaction patterns* corresponding to interaction patterns in Ω (and hence are M-measurable in the sense of Chapter 3). Thus, to facilitate a comparison of the present development with Chapter 2, it is convenient to formalize the notion of *spatial interaction processes* within the setting of Chapter 3. To do so, recall first from the development preceding Lemma 3.2 in Chapter 3 that each possible realization, $s = (i_r, j_r : r = 1, \ldots, n)$, of origin-destination pairs for an n-interaction pattern in Ω_n was designated as a spatial n-interaction pattern, $s \in S_n = (I \times J)^n$, and that the (countable) set of all spatial interaction patterns was given by $S = \cup_{n \geq 0} S_n$, where the singleton set, $S_0 = \{s_o\}$, consists only of the null interaction pattern, s_o, corresponding to the noninteraction event, $\Omega_0 = \{o\}$. With these definitions, it again follows that for each interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, and configuration, $c \in C$, the probability measure, P_c, on $\langle \Omega, \overline{M} \rangle$ generates an associated pattern probability function, $P_c : S \rightarrow [0, 1]$, defined for the null pattern, $s_o \in S_0$, by $P_c(s_o) = P_c(\{o\})$, and for each non-null pattern, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$, by

$$P_c(s) = P_c[\Omega(s)] = P_c(I_r = i_r, J_r = j_r : r = 1, \ldots, n), \tag{4.104}$$

where the interaction event, $\Omega(s)$, is given by (3.60). In particular, for each *singleton pattern*, $s = (ij) \in S_1$, we write $P_c(s) = P_c(ij)$. With this notation, the resulting family of probability functions, $P_c : S \rightarrow [0, 1], c \in C$, was designated in Chapter 1 as *the spatial interaction process* generated by $\boldsymbol{P}$ (Definitions 1.1 and 1.2). This terminology is not essential for the present development.

4.4.1 Positive Interaction Processes

Within the general setting of interaction processes, $\boldsymbol{P} = \{P_c : c \in C\}$, the only additional condition we require in order to state the desired properties of such processes is that *relative likelihoods*, $P_c(s)/P_c(s')$, of each pair of spatial interaction patterns, $s, s' \in S$, be defined for all $c \in C$. Hence we now say that:

Definition 4.9 Each interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying the condition that $P_c(s) > 0$ for all $s \in S$ and $c \in C$ is designated as a *positive interaction process.*

While this terminology is convenient for purposes of the present development, it is important to note that for the important class of *independent interaction processes*, this positivity condition is automatically satisfied:

Lemma 4.1 *Each independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is positive.*

PROOF: In terms of the frequency process, $\boldsymbol{N_P}$, generated by $\boldsymbol{P}$, it follows from Theorem 3.2 together with Definition 3.4 that $P_c(s_o) = P(N^c = 0) = \exp[-\mathrm{E}(N^c)] > 0$, and similarly from part (ii) of Corollary 3.3 together with expressions (3.25) and (4.104) that for all spatial interaction patterns, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$,

$$\begin{aligned} P_c(s) &= P_c(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ &= P_c(N = n) P_c^n(I_r = i_r, J_r = j_r : r = 1, \ldots, n) \\ &= P(N^c = n) \prod_{r=1}^{n} P_c^n(I_r = i_r, J_r = j_r) \qquad (4.105) \\ &= \{(\mathrm{E}(N^c)^n / n!) \exp[-\mathrm{E}(N^c)]\} \prod_{r=1}^{n} [\mathrm{E}(N^c_{i_r j_r}) / \mathrm{E}(N^c)] > 0 \end{aligned}$$

holds identically for each $c \in C$. □

In view of the central role of independent interaction processes in the present analysis, we shall hereafter focus almost exclusively on positive processes. For each such process, $\boldsymbol{P} = \{P_c : c \in C\}$, and spatial interaction pattern, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n, n \in Z_{++}$, it follows that the *conditional probability*, $P_c^n(s)$, of s given an interaction pattern of size, n, under configuration, c, is well defined, and is given by

$$P_c^n(s) = P_c(I_r = i_r, J_r = i_r : r = 1, \ldots, n) / P_c(N = n) > 0. \qquad (4.106)$$

For the case of independent interaction processes, we have a number of additional identities which will prove useful in the analysis to follow. First, observe from (4.106) together with (3.77) and part (ii) of Corollary 3.3 that for each $c \in C$, the *interaction probability function*, $p_c : I \times J \to (0,1)$, associated with the pattern probability function, $P_c : S \to (0,1)$, satisfies the following identity

$$\mathrm{E}(N^c_{ij}) / \mathrm{E}(N^c) = p_c(ij) = P_c^1(ij) = P_c(ij) / P_c(S_1) \qquad (4.107)$$

for all $i, j \in I \times J$. Moreover, by the argument in (4.105), it also follows that for all $c \in C$ and $n \in Z_{++}$, the probability, $P_c(s)$, of each spatial interaction pattern, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n \subseteq S$, can be expressed in terms of p_c as:

$$\begin{aligned} P_c(s) &= P_c(N = n) \prod_{r=1}^{n} [\mathrm{E}(N^c_{i_r j_r}) / \mathrm{E}(N^c)] \\ &= P_c(N = n) \prod_{r=1}^{n} p_c(i_r j_r) P_c(N = n) \prod_{ij} p_c(ij)^{N_{ij}(s)}. \end{aligned} \qquad (4.108)$$

Finally, observe that for any spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n), t = (g_r h_r : r = 1, \ldots, n) \in S$, it follows from (4.107) and (4.108) together with the identity $n = \sum_{ij} N_{ij}(s) = \sum_{ij} N_{ij}(t)$ that

$$\begin{aligned}
\frac{P_c(s)}{P_c(t)} &= \frac{\prod_{r=1}^n [P_c(i_r j_r)/P_c(S_1)]}{\prod_{r=1}^n [P_c(g_r h_r)/P_c(S_1)]} = \frac{\prod_{r=1}^n P_c(i_r j_r)}{\prod_{r=1}^n P_c(g_r h_r)} \\
&= \frac{\prod_{ij} [\mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c)]^{N_{ij}(s)}}{\prod_{ij} [\mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c)]^{N_{ij}(t)}} \qquad (4.109) \\
&= \frac{\mathrm{E}(N^c)^{-n} \prod_{ij} \mathrm{E}(N_{ij}^c)^{N_{ij}(s)}}{\mathrm{E}(N^c)^{-n} \prod_{ij} \mathrm{E}(N_{ij}^c)^{N_{ij}(t)}} \frac{\prod_{ij} \mathrm{E}(N_{ij}^c)^{N_{ij}(s)}}{\prod_{ij} \mathrm{E}(N_{ij}^c)^{N_{ij}(t)}}.
\end{aligned}$$

4.4.2 Behavioral Axioms

Next, it follows that if the indicator functions, $\delta_{ij}^r : S \to \{0, 1\}$, are defined for all $r \in Z_{++}$, $i, j \in I \times J$, and $s = (i_r, j_r : r = 1, \ldots, n) \in S$ [as in (3.10)] by

$$\delta_{ij}^r(s) = \begin{cases} 1 & \text{when } (i_r, j_r) = (i, j) \\ 0 & \text{otherwise,} \end{cases} \qquad (4.110)$$

and if the frequency functions, $N_{ij} : S \to Z_+$, are defined for all $i, j \in I \times J$ and $s = (i_r, j_r : r = 1, \ldots, n) \in S$ [as in (3.11)] by

$$N_{ij}(s) = \sum_{r=1}^n \delta_{ij}^r(s), \qquad (4.111)$$

then [as in (2.49), (2.50), and (2.52) of Chapter 2] we may set $N(s) = \sum_{ij} N_{ij}(s)$ and set

$$N_i(s) = \sum_{j \in J} N_{ij}(s), \qquad (4.112)$$

$$N_j(s) = \sum_{i \in I} N_{ij}(s), \qquad (4.113)$$

$$A(s) = [(N_i(s) : i \in I), (N_j(s) : j \in J)], \qquad (4.114)$$

for each $s \in S$. Two spatial interaction patterns, $s, s' \in S$, are designated as *comparable* iff $N(s) = N(s')$, and are designated as *activity-equivalent* iff $A(s) = A(s')$. In addition, for each $c \in C$ and $s = (i_r, j_r : r = 1, \ldots, n) \in S$, we now let

$$c_s = (c_{i_r j_r} : r = 1, \ldots, n) \in V^n \qquad (4.115)$$

denote the *separation array* for s. For the case of *cost profiles*, in which each separation component is extensively measurable (and hence in which it is meaningful to add profiles), we denote the associated *total-cost profile* for pattern, s, under cost configuration, c, by

$$c(s) = \sum_{ij} c_{ij} N_{ij}(s). \qquad (4.116)$$

Finally, if for each separation configuration, $c \in C$, and origin, $i \in I$, we denote the *sub-configuration* at origin i by $c_i = (c_{ij} : j \in J)$, then the axioms developed in Chapter 2 can now be stated as formal properties of *positive interaction processes*, $\boldsymbol{P} = \{P_c : c \in C\}$. We begin with the axioms characterizing general gravity models [where axioms A1 and A2 in Definition 3.2 are here restated for convenience]:

A1. (Locational Independence)*For all configurations, $c \in C$, and spatial interaction patterns, $s = (i_r, j_r : r = 1, \ldots, n) \in S_n \subseteq S, n \in Z_{++}$,*

$$P_c^n(s) = \prod_{r=1}^{n} P_c^n(i_r j_r). \tag{4.117}$$

A2. (Frequency Independence)*For all configurations, $c \in C$, and frequency profiles, $(n_{ij} : ij \in I \times J) \in Z_+^{I \times J}$,*

$$P_c(N_{ij} = n_{ij} : ij \in I \times J) = \prod_{ij} P_c(N_{ij} = n_{ij}). \tag{4.118}$$

Next we develop number of aggregate behavioral axioms involving the spatial separation properties exhibited by realized interaction patterns:

A3. (Separation Dependence)*For all configurations, $c \in C$, and spatial interaction patterns, $s, s' \in S$, with $A(s) = A(s')$,*

$$c_s = c_{s'} \Rightarrow P_c(s) = P_c(s'). \tag{4.119}$$

A3′. (Relative Separation Dependence) *For all configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(t)$ and $A(s') = A(t')$,*

$$(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.120}$$

A3″. (Uniform Separation Dependence)*For all configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(s')$ and $A(t) = A(t')$,*

$$(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.121}$$

A4. (Separation Efficiency) *For all configurations, $c \in C$, and spatial interaction patterns, $s, s' \in S$, with $A(s) = A(s')$,*

$$c_s \leq c_{s'} \Rightarrow P_c(s) \geq P_c(s'). \tag{4.122}$$

A4′. (Relative Separation Efficiency) *For all configurations,* $c, c' \in C$*, and comparable spatial interaction patterns,* $s, t, s', t' \in S$ *with* $A(s) = A(t)$ *and* $A(s') = A(t')$,

$$(c_s \leq c'_{s'}, c_t \geq c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \quad (4.123)$$

A4″. (Uniform Separation Efficiency) *For all configurations,* $c, c' \in C$*, and comparable spatial interaction patterns,* $s, t, s', t' \in S$ *with* $A(s) = A(s')$ *and* $A(t) = A(t')$,

$$(c_s \leq c'_{s'}, c_t \geq c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \quad (4.124)$$

A5. (Sub-Configuration Dependence) *For all origins,* $i \in I$*, and configurations,* $c, c' \in C$,

$$c_i = c'_i \Rightarrow \mathrm{E}_c(N_i) = \mathrm{E}_{c'}(N_i). \quad (4.125)$$

In addition, we have the following six local axioms which can also be employed to characterize general gravity models:

A6. (Destination Proportionality) *For all origins,* $i, g \in I$*, destinations,* $j, h \in J$*, and configurations,* $c, c' \in C$,

$$(c_{ij} = c'_{gj}, c_{ih} = c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \quad (4.126)$$

A7. (Origin Proportionality) *For all origins,* $i, g \in I$*, destinations,* $j, h \in J$*, and configurations,* $c, c' \in C$,

$$(c_{ij} = c'_{ih}, c_{gj} = c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ih)}{P_{c'}(gh)}. \quad (4.127)$$

A8. (Destination Separability) *For any origin,* $i \in I$*, destinations,* $j, h \in J$*, and configurations,* $c, c' \in C$,

$$(c_{ij} = c_{ih}, c'_{ij} = c'_{ih}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(ij)}{P_{c'}(ih)}. \quad (4.128)$$

A9. (Origin Separability) *For any destination,* $j \in J$*, origins,* $i, g \in I$*, and configurations,* $c, c' \in C$,

$$(c_{ij} = c_{gj}, c'_{ij} = c'_{gj}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ij)}{P_{c'}(gj)}. \quad (4.129)$$

A10. (Destination Monotonicity) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and configurations, $c, c' \in C$,*

$$(c_{ij} \leq c'_{gj}, c_{ih} \geq c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(ih)} \geq \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{4.130}$$

A11. (Origin Monotonicity) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and configurations, $c, c' \in C$,*

$$(c_{ij} \leq c'_{ih}, c_{gj} \geq c'_{gh}) \Rightarrow \frac{P_c(ij)}{P_c(gj)} \geq \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{4.131}$$

Finally, we have the following six axioms for cost configurations which characterize the special class of exponential gravity models:

A12. (Cost Dependence) *For all configurations, $c \in C$, and spatial interaction patterns, $s, s' \in S$, with $A(s) = A(s')$,*

$$c(s) = c(s') \Rightarrow P_c(s) = P_c(s'). \tag{4.132}$$

A12′. (Relative Cost Dependence) *For all cost configurations, $c, c' \in C$ and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(t)$ and $A(s') = A(t')$,*

$$[c(s) = c'(s'), c(t) = c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.133}$$

A12″. (Uniform Cost Dependence) *For all cost configurations, $c, c' \in C$ and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(s')$ and $A(t) = A(t')$,*

$$[c(s) = c'(s'), c(t) = c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.134}$$

A13. (Cost Efficiency) *For all configurations, $c \in C$, and spatial interaction patterns, $s, s' \in S$, with $A(s) = A(s')$,*

$$c(s) \leq c(s') \Rightarrow P_c(s) \geq P_c(s'). \tag{4.135}$$

A13′. (Relative Cost Efficiency) *For all cost configurations, $c, c' \in C$ and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(t)$ and $A(s') = A(t')$,*

$$[c(s) \leq c'(s'), c(t) \geq c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.136}$$

A13″. (Uniform Cost Efficiency) *For all cost configurations, $c, c' \in C$ and comparable spatial interaction patterns, $s, t, s', t' \in S$ with $A(s) = A(s')$ and $A(t) = A(t')$,*

$$[c(s) \leq c'(s'), c(t) \geq c'(t')] \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.137}$$

4.4.3 Relations among Axioms

In a manner paralleling the model classes in Section 4.2 above, if for each axiom Ak, in Section 4.4.2 above we now let $\langle \mathrm{A}k \rangle \subseteq \langle I \times J \rangle$ denote the class of frequency processes, $\boldsymbol{N}$, on $I \times J$ which are generated by positive interaction processes, $\boldsymbol{P}$, satisfying axiom Ak, then we have the following relationships among the behavioral axioms above:

Proposition 4.10 *The following inclusion relations hold among the frequency process classes* $\langle \mathrm{A}k \rangle$:

$$\begin{array}{rclcl}
\langle \mathrm{A4}' \rangle & \subseteq & \langle \mathrm{A3}' \rangle & \subseteq & \langle \mathrm{A3} \rangle, \qquad (4.138)\\
\langle \mathrm{A4}'' \rangle & \subseteq & \langle \mathrm{A3}'' \rangle & \subseteq & \langle \mathrm{A3} \rangle, \qquad (4.139)\\
\langle \mathrm{A4}'' \rangle \cup \langle \mathrm{A4}' \rangle & \subseteq & \langle \mathrm{A4} \rangle & \subseteq & \langle \mathrm{A3} \rangle, \qquad (4.140)\\
\langle \mathrm{A10} \rangle \subseteq \langle \mathrm{A6} \rangle, & & \langle \mathrm{A11} \rangle & \subseteq & \langle \mathrm{A7} \rangle, \qquad (4.141)\\
\langle \mathrm{A12}' \rangle \cup \langle \mathrm{A12}'' \rangle & \subseteq & \langle \mathrm{A12} \rangle & \subseteq & \langle \mathrm{A3} \rangle, \qquad (4.142)\\
\langle \mathrm{A13}' \rangle \cup \langle \mathrm{A13}'' \rangle & \subseteq & \langle \mathrm{A13} \rangle & \subseteq & \langle \mathrm{A4} \rangle, \qquad (4.143)\\
\langle \mathrm{A13}' \rangle & \subseteq & \langle \mathrm{A12}' \rangle & \subseteq & \langle \mathrm{A3}' \rangle, \qquad (4.144)\\
\langle \mathrm{A13}'' \rangle & \subseteq & \langle \mathrm{A12}'' \rangle & \subseteq & \langle \mathrm{A3}'' \rangle, \qquad (4.145)\\
\langle \mathrm{A13}' \rangle \subseteq \langle \mathrm{A4}' \rangle, & & \langle \mathrm{A13}'' \rangle & \subseteq & \langle \mathrm{A4}'' \rangle. \qquad (4.146)
\end{array}$$

Proof: To see that $\langle \mathrm{A3}' \rangle \subseteq \langle \mathrm{A3} \rangle$, observe simply that by setting $c = c'$ and $t = t'$ in (4.120) it follows at once that (4.120) reduces to (4.119). The inclusion relations $\langle \mathrm{A3}'' \rangle \subseteq \langle \mathrm{A3} \rangle, \langle \mathrm{A4}' \rangle \subseteq \langle \mathrm{A4} \rangle, \langle \mathrm{A4}'' \rangle \subseteq \langle \mathrm{A4} \rangle, \langle \mathrm{A12}' \rangle \subseteq \langle \mathrm{A12} \rangle, \langle \mathrm{A12}'' \rangle \subseteq \langle \mathrm{A12} \rangle, \langle \mathrm{A13}' \rangle \subseteq \langle \mathrm{A13} \rangle$ and $\langle \mathrm{A13}'' \rangle \subseteq \langle \mathrm{A13} \rangle$ are each established in a similar manner. Next, to see that $\langle \mathrm{A4} \rangle \subseteq \langle \mathrm{A3} \rangle$, observe that for all configurations, $c \in C$, and spatial interaction patterns, $s, s' \in S$ with $A(s) = A(s'), c_s = c'_{s'} \Rightarrow [c_s \leq c'_{s'}$ and $c_c \geq c'_{s'}] \Rightarrow [P_c(s) \geq P_c(s')$ and $P_c(s) \leq P_c(s')] \Rightarrow P_c(s) = P_c(s')$. The inclusion relations $\langle \mathrm{A4}' \rangle \subseteq \langle \mathrm{A3}' \rangle, \langle \mathrm{A4}'' \rangle \subseteq \langle \mathrm{A3}'' \rangle, \langle \mathrm{A10} \rangle \subseteq \langle \mathrm{A6} \rangle, \langle \mathrm{A11} \rangle \subseteq \langle \mathrm{A7} \rangle, \langle \mathrm{A13}' \rangle \subseteq \langle \mathrm{A12}' \rangle$ and $\langle \mathrm{A13}'' \rangle \subseteq \langle \mathrm{A12}'' \rangle$ are each established in a similar manner. To see that $\langle \mathrm{A12} \rangle \subseteq \langle \mathrm{A3} \rangle$, observe that for any spatial interaction patterns $s = (i_r j_r : r = 1, \ldots, n)$ and $s' = (i'_r j'_r : r = 1, \ldots, n)$, if (4.132) holds then

$$\begin{aligned}
c_s = c'_{s'} &\Rightarrow (c_{i_r j_r} = c'_{i'_r j'_r} : r = 1, \ldots, n) \\
&\Rightarrow \sum_{r=1}^{n} c_{i_r j_r} = \sum_{r=1}^{n} c'_{i'_r j'_r} \\
&\Rightarrow c(s) = c'(s') \\
&\Rightarrow P_c(s) = P_{c'}(s'),
\end{aligned} \qquad (4.147)$$

so that (4.119) holds. Finally, the remaining inclusion relations, $\langle \mathrm{A12}' \rangle \subseteq \langle \mathrm{A3}' \rangle, \langle \mathrm{A12}'' \rangle \subseteq \langle \mathrm{A3}'' \rangle, \langle \mathrm{A13} \rangle \subseteq \langle \mathrm{A4} \rangle, \langle \mathrm{A13}' \rangle \subseteq \langle \mathrm{A4}' \rangle$ and $\langle \mathrm{A13}'' \rangle \subseteq \langle \mathrm{A4}'' \rangle$, all follow from exactly the same argument as in (4.147), and the result is established. □

Note that the basic independence axioms, A1 and A2, are formally independent of all other axioms, and hence are not included in the above relations. In addition, it is also of interest to note that the sub-configuration axiom (A5) is not included for the same reason. To see that A5 is formally independent of all other axioms, observe that each of the other axioms which assert any type of configuration-free property (namely axioms A3′, A3″, A4′, A4″, A6, A7, A8, A9, A10, A11, A12′, A12″, A13′ and A13″) all involve assertions about pattern probabilities for *comparable* spatial interaction patterns (where in the case of axioms A6 through A11 the patterns are implicitly of size one). However, A5 makes direct assertions about mean frequencies, which depend on *all* pattern probabilities. Hence the assertion in A5 is qualitatively different from all other axioms.

4.5 Characterizations of Gravity Models

Given this set of axioms for positive interaction processes, our objective in this section is to characterize the gravity model classes of Section 4.2 in terms of these axioms. To do so, we begin in Section 4.5.1 below with number of preliminary results which will prove useful in establishing the desired characterizations. Given these results, the classes of general gravity models in Section 4.2.1 are then characterized in Section 4.5.2. Finally, the classes of exponential gravity models in Section 4.2.2 are characterized in Section 4.5.3.

4.5.1 Analytical Preliminaries

To establish the desired characterizations of gravity models, we begin with a number of preliminary results leading to two extension lemmas (Lemmas 4.4 and 4.5 below) for systems of linear equalities and inequalities. The following notation will be employed throughout. Let $Q \subseteq R$ denote the *rational numbers*, and let $Q_+ = Q \cap R_+$ denote the *nonnegative* rational numbers. For any $p \times k$ matrix, $A \in R^{p \times k}$, let $span(A) = \{Ax \in R^p : x \in R^k\}$, and for any vector, $b \in R^p$, let $A^{-1}[b] = \{x \in R^k : Ax = b\}$ denote the *inverse image set* for b. In particular, the inverse image set for the *zero element*, $0 \in R^p$, is usually denoted by $null(A) = A^{-1}[0]$. Next, for any set, $X \subseteq R^p$, a subset, $Y \subseteq X$, is said to be *dense* in X iff for each $x \in X$ there exists some sequence $\{y_n\}$ in Y with $y_n \to x$. In particular, it is well known that for every open subset, $X \subseteq R^p$, the rational subset, $X \cap Q^p$, is dense in X [so that, in particular, Q^p_{++} is dense in R^p_{++}]. Moreover, if Y is dense in X then Y is necessarily dense in the closure of X [so that, in particular, Q^p_{++} is also dense in R^p_{+}]. With these definitions, we now have the following denseness properties of rational-valued matrices:

Lemma 4.2 *For any rational-valued matrix, $A \in Q^{p \times k}$, and rational vector, $b \in Q^p$*

(i) *$span(A) \cap Q^p$ is dense in $span(A)$.*

(ii) *$span(A) \cap Q^p_+$ is dense in $span(A) \cap R^p_+$.*

(iii) *$A^{-1}[b] \cap Q^k$ is dense in $A^{-1}[b]$.*

(iv) *$A^{-1}[b] \cap Q^k_+$ is dense in $A^{-1}[b] \cap R^k_+$.*

(v) *$A^{-1}[b] \cap Q^k_{++}$ is dense in $A^{-1}[b] \cap R^k_{++}$.*

PROOF: For analytical convenience, we prove these properties in the following order: (i), (iii), (ii), (v), and (iv):

(i) The first property follows at once from the fact that for any rational matrix, $A \in Q^{p \times k}$, the set, $span(A) \cap Q^p$, is a linear subspace of the rational vector space, Q^p, whose completion (with respect to the euclidean metric) is precisely the linear subspace, $span(A)$, of R^p.

(iii) To establish property (iii), observe simply that by choosing any rational basis, B, for the linear subspace, $null(A) \cap Q^k$, of the rational vector space, Q^k, it follows by construction that $null(A) = span(B)$, and hence that (iii) is a consequence of (i) for $b = 0$. To extend this result to all $b \in Q^p$, we begin by observing that for all $b \in R^p$ with $A^{-1}[b] \neq \emptyset$, and all $y \in A^{-1}[b]$,

$$A^{-1}[b] = y + null(A). \tag{4.148}$$

To see this, note simply that $\{y \in A^{-1}[b], x \in null(A)\} \Rightarrow A(x+y) = b \Rightarrow (x+y) \in A^{-1}[b] \Rightarrow y + null(A) \subseteq A^{-1}[b]$, and conversely, that $\{y \in A^{-1}[b], 0 \in null(A)\} \Rightarrow y = (y+0) \in y + null(A) \Rightarrow A^{-1}[b] \subseteq y + null(A)$. Next we observe that if $b \in Q^p$ then

$$A^{-1}[b] \neq \emptyset \Rightarrow A^{-1}[b] \cap Q^k \neq \emptyset \tag{4.149}$$

[i.e., that the equation, $Ax = b$, must have rational solutions whenever it has real solutions]. To see this, note first that if A is of full column rank, then A^tA is nonsingular, so that $x \in A^{-1}[b] \Rightarrow Ax = b \Rightarrow A^tAx = A^tb \Rightarrow x = (A^tA)^{-1}A^tb \in Q^k \Rightarrow A^{-1}[b] \subseteq Q^k \Rightarrow A^{-1}[b] \cap Q^k = A^{-1}[b]$. Hence we may assume that A is singular, in which case (by relabelling the columns of A if necessary) we may take the first $m(< k)$ columns, $M = (a_1, \ldots, a_m)$, to form a maximal linearly independent subset of the columns of A, so that by definition, $span(M) = span(A)$. Hence $A^{-1}[b] \neq \emptyset$ implies the existence of some $y \in R^m$ with $My = b$. But since M is of full column rank, the matrix M^tM is nonsingular, so that $M^tMy = M^tb \Rightarrow y = (M^tM)^{-1}M^tb \in Q^m$. Thus, letting $z^t = (y^t, 0^t) \in Q^k$ [where $0 \in Q^{k-m}$], it follows by construction that $Az = b$ and hence that $z \in A^{-1}[b] \cap Q^k \neq \emptyset$. Finally, choosing y in (4.148) to be any element of $A^{-1}[b] \cap Q^k$, and recalling that $null(A) \cap Q^k$ is dense in $null(A)$, it follows at one that $A^{-1}[b] \cap Q^k = y + [null(A) \cap Q^k]$ is dense in $A^{-1}[b]$, and hence that (iii) holds.

(ii) To establish (ii), choose any $x \in span(A) \cap R^p_+$ and observe by definition that $x = Az$ for some $z \in R^k$. Hence, letting $(a_i : i = 1, \ldots p)$ denote the columns of the matrix, A, and observing from the nonnegativity of $x = Az$ that there exists a partition of these columns into two matrices, $A_1 = (a_i : a_i^t z = 0)$ and $A_2 = (a_i : a_i^t z > 0)$, satisfying $z \in null(A_1)$ and $A_2 z > 0$ by construction, it follows from (iii) [with $b = 0$] that there is some sequence, $\{z_n\}$, in $null(A_1) \cap Q^k$ with $z_n \to z$. But if we set $x_n = Az_n \in Q^p$, so that $A_1 z_n = 0$ for all n by construction, and observe from the continuity of matrix operators that $A_2 z > 0 \Rightarrow A_2 z_n > 0$ for all large n, it follows that for sufficiently large n, $\{x_n\}$ must be a sequence in $span(A) \cap Q^p_+$ with $x_n \to x$. Hence $span(A) \cap Q^p_+$ is dense in $span(A) \cap R^p_+$, and (ii) must hold.

(v) To establish property (v) choose any $x \in A^{-1}[b] \cap R^k_{++}$ and observe from (iii) that there is some sequence (x_n) in $A^{-1}[b] \cap Q^k$ with $x_n \to x$. But since R^k_{++} is an open set, $x \in R^k_{++}$ implies that $x_n \in R^k_{++}$ for all large n, and hence that $x_n \in A^{-1}[b] \cap Q^k_{++}$ for large n. Thus $A^{-1}[b] \cap Q^k_{++}$ must be dense in $A^{-1}[b] \cap R^k_{++}$.

(iv) Finally to establish property (iv), again choose B with $null(A) = span(B)$ [as in the proof of (iii) above], and observe that for the special case, $b = 0$, (iv) now follows from (ii). To extend this result to all $b \in Q^k$, it again follows from the argument in (iii) that it is enough to find one $y \in A^{-1}[b] \cap Q^k_+$, [and combine this together with (4.148) and property (iv) for $b = 0$]. To produce an element of $A^{-1}[b] \cap Q^k_+$, choose any element, $x \in A^{-1}[b] \cap R^k_+$, and observe that we may assume x has at least one positive component [since $0 \in Q^k$ would yield the desired element of $A^{-1}[b] \cap Q^k_+$]. Hence (by relabelling x if necessary we may assume that $x^t = (z^t, 0^t)$ for some $z \in R^m_{++}, m < k$. Thus, by partitioning the columns of A accordingly, we have $b = Ax = A_1 z + A_2 0 = A_1 z$, which in turn implies that $A_1[b] \cap R^m_{++} \neq \emptyset$. But this, together with the rationality of A_1, implies from (v) that $A_1[b] \cap Q^m_{++} \neq \emptyset$, so that for any $w \in A_1[b] \cap Q^m_{++}$ we then have $b = A_1 y = A_1 w + A_2 0$. Hence the desired element of $A^{-1}[b] \cap Q^k_+$ is given by $y^t = (w^t, 0^t)$, and the result is established. □

Next, we require the following conditions for the existence of solutions to systems of linear equalities and inequalities:

Lemma 4.3
(i) *For any matrix, $A \in R^{p \times k}$, and vector, $b \in R^p$, there exists some $\lambda \in R^k$ such that*

$$A\lambda \geq b \tag{4.150}$$

iff for all nonnegative vectors, $x \in R^p_+$,

$$A^t x = 0 \Rightarrow b^t x \leq 0. \tag{4.151}$$

(ii) *If the inequality in* (4.151) *is replaced by an equality, and* R_+ *is replaced by* R, *then* (4.150) *becomes an equality.*

PROOF: **(i)** By Theorem 2.7 in Gale (1960, p. 46), it follows that (4.150) fails to hold iff there exists some $x \in R^p_+$ with $A^t x = 0$ and $b^t x > 0$, i.e., iff (4.151) fails for some $x \in R^p_+$. Hence there exists some $\lambda \in R^k$ satisfying (4.150) iff (4.151) holds for all $x \in R^p_+$.

(ii) To establish property (ii), observe first that since $null(A^t)$ is the *orthogonal complement* of $span(A)$ in R^p [see Halmos (1958, Section 49)], it follows by definition that for each $b \in R^p$ there exist vectors, $x \in null(A^t)$ and $y \in span(A)$ with $b = x + y$ and $x^t y = 0$. But by (the equality form of) expression (4.150), $x \in null(A^t) \Rightarrow A^t x = 0 \Rightarrow b^t x = 0 \Rightarrow (x+y)^t x = 0 \Rightarrow x^t x + x^t y = 0 \Rightarrow x^t x = 0 \Rightarrow x = 0$, so that $b = y \in span(A)$. Hence, by definition, $A\lambda = b$ must hold for some $\lambda \in R^k$. □

Using these results (and recalling the definitions of the special sets I, J, R, Q, Z, Z_+, and Z_{++}), we can now establish our first extension lemma:

Lemma 4.4
(i) *For any nonempty finite sets,* T *and* W, *and any functions,* $a: T \times W \rightarrow Q$ *and* $b: W \rightarrow R$, *if*

$$[\sum_{w \in W} a_t(w) z(w) = 0, \text{ for all } t \in T] \Rightarrow \sum_{w \in W} b(w) z(w) \leq 0, \tag{4.152}$$

for all nonnegative integer-valued functions, $z \in Z_+^{I \times J}$, *then there exists a function,* $\lambda: T \rightarrow R$, *such that for all* $w \in W$

$$b(w) \leq \sum_{t \in T} \lambda(t) a_t(w). \tag{4.153}$$

(ii) *If the inequality on the right hand side of* (4.152) *is replaced by an equality, and* Z_+ *is replaced by* Z, *then* (4.153) *becomes an equality.*

PROOF: **(i)** If we write $T = \{t_1, \ldots, t_k\}$ and $W = \{w_1, \ldots, w_p\}$ and let the rational matrix, $A \in Q^{p \times k}$, be defined for all $i = 1, \ldots, p$ and $j = 1, \ldots, k$ by $a_{ij} = a_{t_j}(w_i)$, and if for each pair of functions, $b: W \rightarrow R$ and $\lambda: T \rightarrow R$, we let the vectors, $b^t = (b_1, \ldots, b_p)$ and $\lambda^t = (\lambda_1, \ldots, \lambda_k)$ be defined by $b_i = b(w_i), i = 1, \ldots, p$, and $\lambda_j = \lambda(t_j), j = 1, \ldots, k$, then assertion (4.153) is seen to be equivalent to the existence of some $\lambda \in R^k$ with $A\lambda \geq b$. Hence by Lemma 4.3(i), (4.141) will be established if $b^t x \leq 0$ can be shown to hold for all $x \in R^p_+$ with $A^t x = 0$, i.e., for all $x \in null(A^t) \cap R^p_+$. But by Lemma 4.2(iv), there exists for each $x \in null(A^t) \cap R^p_+$ a sequence, $\{q_n\}$, in $null(A^t) \cap Q^p_+$ with $q_n \rightarrow x$. Hence, writing each component of $q_n = (q_{n1}, \ldots, q_{np}) \in Q^p_+$ in integer form as $q_{ni} = h_{ni}/m_{ni}$ (where $h_{ni} \in Z_+$ and $m_{ni} \in Z_{++}$), and letting $m_n = \prod_{i=1}^p m_{ni} > 0$ denote the common denominator of the q_{ni}'s, so that $m_n q_{ni} = h_{ni} \prod_{j \neq i} m_{nj} \in Z_+$ for all $i = 1, \ldots, p$, it follows that $A^t q_n = 0 \Leftrightarrow m_n(A^t q_n) = 0 \Leftrightarrow A^t z_n = 0$, where $z_n = m_n q_n \in Z_+^p$. But by (4.152), $A^t z_n = 0 \Rightarrow b^t z_n \leq 0 \Rightarrow b^t(m_n q_n) \leq 0 \Rightarrow m_n(b^t q_n) \leq 0 \Rightarrow b^t q_n \leq 0$ for all n. Hence, by the continuity of inner products, it follows that $q_n \to x \Rightarrow b^t q_n \to b^t x \leq 0$, and we may conclude that $b^t x \leq 0$ holds for all $x \in R_+^p$ with $A^t x = 0$.

(ii) To establish assertion (ii), observe that by replacing $null(A^t) \cap R_+^p$ with $null(A^t) \cap R^p$ and employing Lemma 4.2(iii) and Lemma 4.3(ii), exactly the same argument as in part (i) above now shows that (4.153) must hold as an equality. □

Our second extension lemma depends on certain separation properties of polyhedral convex sets which are beyond the scope of this book, and hence will be stated without proof. [A detailed development of this result is given in Smith (1986b)]. If the rational-valued separation configurations in C are denoted by $C(Q) = C \cap Q^K$, then:

Lemma 4.5 *For any nonempty finite set, Y, rational-valued function, $a: Y \to Q^{I \times J}$, and continuous function, $b: C \to R^{I \times J}$*

(i) *If there exist functions, $w: C \times Y \to R$ and $\theta: C \times K \to R$, such that*

$$b_c(ij) = \sum_{y \in Y} w_c(y) a_y(ij) - \sum_{k \in K} \theta_c(k) c_{ij}^k \tag{4.154}$$

for all $c \in C(Q)$, then w and θ can be chosen so that (4.154) *holds for all $c \in C$.*

(ii) *If in addition it is true that*

$$\theta_c(k) \geq 0, \quad k \in K \tag{4.155}$$

for all $c \in C(Q)$, then θ may be chosen so that (4.155) *holds for all $c \in C$.*

Next we record certain results on linear programming problems. First, we have the following version of the well known 'Complementary Slackness Theorem' [see for example Dantzig (1963, Theorem 6-4.4, p. 136)]:

Lemma 4.6 *For any matrices, $A \in R^{p \times m}$, $B \in R^{q \times m}$, and vectors, $b \in R^p$, $d \in R^q$, $f \in R^m$, if the linear programming problem,*

$$\min: f^t x \quad \textit{subject to:} \quad Ax = b, Bx \geq d, x \in R_+^m, \tag{4.156}$$

has a positive solution, $x \in R_{++}^m$, then there exist (dual) vectors, $w \in R^p$ and $\theta \in R_+^q$, such that

$$A^t w + B^t \theta = f. \tag{4.157}$$

For the special case of rational-valued coefficient matrices and vectors, we have the following additional result:

Lemma 4.7 *For any real vector, $f \in R^m$, rational vectors, $b \in Q^p, d \in Q^q$, and rational matrices, $A \in Q^{p\times m}, B \in Q^{q\times m}$, if the linear programming problem in* (4.156) *has at least one solution, then it must have a rational solution, $x \in Q_+^m$.*

PROOF: Since the *feasibility space*, $F = \{x \in R_+^m : Ax = b, Bx \geq d\}$ for (4.156) is a polyhedral convex set, it follows that the solution set for (4.156) can only be nonempty if it contains at least one vertex of the set F. Hence, it suffices to show that each vertex of F is rational valued. To do so, observe that if the $m \times m$ identity matrix is denoted by I_m, then each vertex, v, of F is by definition a solution to the system linear equalities and inequalities

$$Av = b, Bv \geq d, I_m v \geq 0, \tag{4.158}$$

which involves at least m equalities. Moreover, m of these equalities must be non-redundant, i.e., must correspond to a set of m linearly independent columns of the $m\times(p+q+m)$ matrix, $[A^t, B^t, I_m]$. Hence, letting M_v^t denote the $m \times m$ matrix defined by these columns, and letting g_v^t denote the vector of elements in the corresponding positions of the coefficient vector, $(b^t, d^t, 0^t)$, it follows that each vertex, v, satisfies

$$M_v v = g_v, \tag{4.159}$$

for this choice of M_v and g_v. But since the rational matrix, M_v, is nonsingular by construction, it follows that the inverse matrix, M_v^{-1}, exists and is also rational (as a linear operator on the m-dimensional rational vector space, Q^m). Thus (4.159) together with the rationality of $g_v \in Q^m$ imply that $v = M_v^{-1} g_v \in Q_+^m$, and hence that each vertex, v, of F is rational. □

Finally, we consider certain characteristic properties of linear functions. First we record the following characterization of all linear functions on the set, V, of separation profile values:

Lemma 4.8 *For any continuous function, $f : V \to R$, satisfying the (Cauchy) equation*

$$f(x+y) = f(x) + f(y) \tag{4.160}$$

for all $x, y \in V$, there exists a coefficient vector, $\theta \in R^K$, such that for all $x \in V \subseteq R^K$,

$$f(x) = \theta^t x. \tag{4.161}$$

PROOF: This result follows for example from Theorem II.3.1.5 in Eichhorn (1978), by applying Theorem I.1.3.2 in Eichhorn (1978) to the nonnegative components, $k \in K_+$, of profiles, and applying Corollary I.1.2.10 in Eichhorn (1978) to all other components, $k \in K - K_+$, of profiles in the set, $V \subseteq R^K$. □

Next we consider the properties of *positive linear* functions, $f: R^n \to R$ of the form, $f(x) = \theta^t x$, for some positive coefficient vector, $\theta \in R^n_{++}$. Each such function is of course continuous and increasing in all components. Moreover, each such function is linearly *homogeneous* on all of R^n, i.e., satisfies the condition

$$f(\lambda x) = \lambda f(x) \tag{4.162}$$

for all $\lambda \in R$ and $x \in R^n$. For the scalar case, $n = 1$, these conditions completely characterize the class of positive linear functions, since each increasing function $f: R \to R$, satisfying (4.162) must clearly be of the form $f(x) = \theta x$ with $\theta = f(1) > 0$. However, for $n \geq 2$, there exist continuous increasing functions on R^n satisfying (4.162) which fail to be linear, i.e., fail to be of the form (4.161). In particular, if for each vector $x = (x_1, \ldots, x_n) \in R^n$ we let $\min(x) = \min\{x_1, \ldots, x_n\}, \max(x) = \max\{x_1, \ldots, x_n\}$, and denote the *unit* vector in R^n by $1 = (1, \ldots, 1)$, then we have the following specific instance of such a function:

Lemma 4.9 *For each $n \geq 2$ the function, $\ell : R^n \to R$, defined for all $x \in R^n$ by*

$$\ell(x) = \begin{cases} 1^t x + \min(x), & x \in R^n_{++} \\ 1^t x + \max(x), & -x \in R^n_{++} \\ 1^t x, & \textit{otherwise}, \end{cases} \tag{4.163}$$

is continuous, increasing, and linearly homogeneous on R^n, but fails to be linear.

PROOF: First to establish continuity of ℓ, observe that for any sequence (x_m) in R^n_{++} and $x \notin R^n_{++}$ with $x_m \to x$, we must have $\min(x_m) \to 0 = \min(x)$, and hence from (4.163) [together with the continuity of the linear function $1^t x$] must have $\ell(x_m) \to \ell(x)$. Similarly, if the sequence $(-x_m)$ is in R^n_{++}, but $-x \notin R^n_{++}$, then $x_m \to x$ now implies that $\max(x_m) \to 0 = \max(x)$, and hence again that $\ell(x_m) \to \ell(x)$. Moreover, since $x \geq y \neq x$ implies that $1^t x > 1^t y, \min(x) \geq \min(y)$, and $\max(x) \geq \max(y)$, it also follows from (4.163) that ℓ must be increasing on all of R^n. To see that ℓ is linearly homogeneous on R^n, observe first that for any $x \in R^n$ and nonnegative scalar, $\lambda \geq 0$, we must have $1^t(\lambda x) = \lambda(1^t x), \min(\lambda x) = \lambda \min(x)$, and $\max(\lambda x) = \lambda \max(x)$, so that (4.162) follows from (4.163). Moreover, since $x \in R^n_{++} \Rightarrow \max(-x) = -\min(x), -x \in R^n_{++} \Rightarrow \min(-x) = -\max(x)$, and $1^t(-x) = -(1^t x)$ for all $x \in R^n$, it also follows from (4.163) that $f(-x) = -f(x)$ identically on R^n, and hence that (4.162) holds for all $\lambda \in R$ and $x \in R^n$. Finally, to see that f is not linear for $n \geq 2$, let $x = (2, \ldots, 2, -1), y = (-1, 2, \ldots, 2)$ and observe from (4.163) that $x, -x \notin R^n_{++} \Rightarrow f(x) = 1^t x$ and $y, -y \notin R^n_{++} \Rightarrow f(y) = 1^t y$. But $\min(x + y) \geq 1 \Rightarrow x + y \in R^n_{++} \Rightarrow f(x + y) = 1^t(x + y) + \min(x + y) > 1^t(x + y) = 1^t x + 1^t y$, so that $f(x + y) > f(x) + f(y)$, and we may conclude that f is not linear. □

4.5.2 Characterizations of General Gravity Models

Given the preliminary results above, we are now ready to establish the behavioral characterizations of general gravity models discussed in Section 2.2.3 of Chapter 2 above. For purposes of the characterization theorems to follow, it is convenient to introduce the following notational conventions. First recall that the origin set, I, and destination set, J, may in general overlap (i.e., that origins may also serve as destinations). In order to separate the dual roles of such locations in the proofs to follow, it will often be convenient to employ *disjoint copies*, $I \times \{1\}$ and $J \times \{2\}$, of the sets, I and J. Hence, for each element, $i \in I \cap J$, the pair of elements, $(i, 1)$ and $(i, 2)$, may be interpreted as representing location, i, in each of its distinct roles as an origin and destination. In terms of this notation, the set, $I + J = (I \times \{1\}) \cup (J \times \{2\})$, is usually designated as the *disjoint union* of I and J [or the *cartesian sum* of I and J (Berge, 1963, p.3)]. In addition, we shall employ the following extension of the notation in Proposition 4.10 above. For each subset, $\{k_1, \ldots, k_m\}$ of the axioms in Section 4.4 above, let $\langle k_1, \ldots, k_m \rangle$ denote the class of frequency processes generated by those interaction processes satisfying axioms, $k_1, \ldots, k_m$. In terms of this notation, the equality relation, "$\langle \text{MODEL } k \rangle = \langle \text{A1}, \text{A2}, k_1, \ldots, k_m \rangle$", asserts that the class of frequency processes satisfying the conditions of MODEL k is precisely the class of frequency process generated by interaction processes satisfying axioms $(\text{A1}, \text{A2}, k_1, \ldots, k_m)$. Hence, to establish each assertion of this form, it suffices from Theorem 3.2 to show that (a) every independent interaction processes, $\boldsymbol{P}$, satisfying axioms, $k_1, \ldots, k_m$, generates a $\boldsymbol{P}$-frequency process, $\boldsymbol{N_P} \in \langle \text{MODEL } k \rangle$ [so that $\langle \text{A1}, \text{A2}, k_1, \ldots, k_m \rangle \subseteq \langle \text{MODEL } k \rangle$], and conversely, that (b) for every frequency process, $\boldsymbol{N} \in \langle \text{MODEL } k \rangle$, the generator, $\boldsymbol{P_N}$, of $\boldsymbol{N}$ satisfies axioms, $k_1, \ldots, k_m$ [so that $\langle \text{MODEL } k \rangle \subseteq \langle \text{A1}, \text{A2}, k_1, \ldots, k_m \rangle$].

Given this general proof format, we begin below with a formal development of *aggregate* behavioral characterizations, and then proceed to *local* behavioral characterizations, in a manner paralleling Section 2.2.3

(A) Aggregate Behavioral Characterizations

To establish the characterization of general gravity models (G1) in terms of the separation dependence axiom (A3) stated in Section 2.2.3(A), we begin by observing that axiom A3 can be equivalently stated in terms of the separation profile frequencies associated with spatial interaction patterns. To do so, we employ the following notation. If for each spatial interaction pattern, $s = (i_r j_r : r = 1, \ldots, n) \in S$, separation configuration, $c \in C$, and separation profile, $v \in V$, we denote the number of (ij)-interactions in s with $c_{ij} = v$ by

$$f_{cv}(s) = |\{r : c_{i_r j_r} = v\}|, \tag{4.164}$$

then we now designate the full array of such frequencies,

$$f_c(s) = [f_{cv}(s) : v \in V] \in Z_+^V, \tag{4.165}$$

as the *separation frequency array* generated by pattern s under configuration c. [Note from the definition of V that the dimension of each frequency array $f_c(s)$ is *uncountably infinite*. However, since $N(s)$ is finite, it follows that $f_{cr}(s) = 0$ for all but finitely many values of $v \in V$. Hence the summation identity, $N(s) = \sum_{v \in V} f_{cr}(s)$, is always well defined.] With this definition, axiom A3 can be given the following equivalent form:

Lemma 4.10 *An interaction process,* $\boldsymbol{P} = \{P_c : c \in C\}$*, satisfies axiom* A3 *iff for all separation configurations,* $c \in C$*, and spatial interaction patterns,* $s, s' \in S$,

$$[A(s) = A(s'), f_c(s) = f_c(s')] \Rightarrow P_c(s) = P_c(s'). \tag{4.166}$$

PROOF: First observe that if $\boldsymbol{P}$ satisfies (4.166), then since by definition, $c_s = c_{s'} \Rightarrow f_c(s) = f_c(s')$ for all $c \in C$, it follows at once that

$$\begin{aligned} [A(s) = A(s'), c_s = c_{s'}] &\Rightarrow [A(s) = A(s'), f_c(s) = f_c(s')] \\ &\Rightarrow P_c(s) = P_c(s'), \end{aligned} \tag{4.167}$$

and hence that $\boldsymbol{P}$ satisfies A3. To establish the converse, observe first that if $f_c(s) = f_c(s')$ then by (4.164) it follows that $N(s) = \sum_v f_{cv}(s) = \sum_v f_{cv}(s') = N(s')$, so that if $N(s) = n = N(s')$ then we may write $s = (i_r j_r : r = 1, \ldots, n)$ and $s' = (i'_r j'_r : r = 1, \ldots, n)$. Hence, if the function, $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$, is constructed by setting $\pi(1)$ equal to the first $r \in \{1, \ldots, n\}$, with $c_{i_1 j_1} = c_{i'_r j'_r}$, and for each $m = 2, \ldots, n$ setting $\pi(m)$ equal to the first integer, $r \in \{1, \ldots, n\} - \{\pi(1), \ldots, \pi(m-1)\}$, with $c_{i_m j_m} = c_{i'_r j'_r}$, then it follows from equality, $f_c(s) = f_c(s')$, that this construction yields a well defined permutation, $\pi \in \Pi_n$. Thus if [as in expression (3.17)] we denote the corresponding reordering of pattern s by $s(\pi) = [i_{\pi(r)} j_{\pi(r)} : r = 1, \ldots, n]$ then it follows by construction that $c_{s(\pi)} = c_{s'}$. Moreover, since $A(s) = A[s(\pi)]$ also holds for every reordering of s, we may conclude from the symmetry property (R2) of interaction processes that if $\boldsymbol{P}$ satisfies A3, then for all $c \in C$ and $s, s' \in S$,

$$\begin{aligned} &[A(s) = A(s'), f_c(s) = f_c(s')] \Rightarrow [A(s(\pi)) = A(s'), c_{s(\pi)} = c_{s'}] \\ &\Rightarrow P_c(s(\pi)) = P_c(s') \Rightarrow P_c(s) = P_c(s'). \end{aligned} \tag{4.168}$$

Hence $\boldsymbol{P}$ must also satisfy (4.166), and the result is established. □

Given this equivalent form of axiom A3, we are now ready to establish the characterization of frequency processes representable by general gravity models (G1) stated in Section 2.2.3(A). Recalling the discussion at the beginning of this section, the desired result can be stated in terms of our present notation as follows:

Theorem 4.1 (G1 Characterization) *The class of frequency processes, ⟨MODEL G1⟩ is characterized by the following properties of its generators,*

$$\langle \text{MODEL G1} \rangle = \langle \text{A1}, \text{A2}, \text{A3} \rangle. \tag{4.169}$$

PROOF: (a) First consider any independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying axiom A3. To establish that $\boldsymbol{N_P} \in \langle \text{MODEL G1} \rangle$, let the *indicator functions* δ_g, σ_h, $\tau_{cv} \in \{0,1\}^{I \times J}$, be defined for all $ij \in I \times J$ by

$$\delta_g(ij) = 1 \quad \Leftrightarrow i = g, \qquad g \in I, \tag{4.170}$$
$$\sigma_h(ij) = 1 \quad \Leftrightarrow j = h, \qquad h \in J, \tag{4.171}$$
$$\tau_{cv}(ij) = 1 \quad \Leftrightarrow c_{ij} = v, \qquad v \in V, \tag{4.172}$$

so that for all $s \in S$

$$\sum_{ij} \delta_g(ij) N_{ij}(s) \quad = N_g(s), \qquad g \in I, \tag{4.173}$$
$$\sum_{ij} \sigma_h(ij) N_{ij}(s) \quad = N_h(s), \qquad h \in J, \tag{4.174}$$
$$\sum_{ij} \tau_{cv}(ij) N_{ij}(s) \quad = f_{cv}(s), \qquad v \in V_c. \tag{4.175}$$

The results of Lemma 4.4 can then be applied to this system of equalities in the following way. For any given integer-valued function, $z \in Z^{I \times J}$, choose any pair of spatial interaction patterns, $s_z, s'_z \in S$, satisfying $N_{ij}(s_z) \geq z(ij)$ and $N_{ij}(s'_z) = N_{ij}(s_z) - z(ij) \geq 0$ for all $ij \in I \times J$. Then by (4.173), (4.174), and (4.175) it follows that

$$N_g(s_z) = N_g(s'_z) \quad \Leftrightarrow \quad \textstyle\sum_{ij} \delta_g(ij) z(ij) \quad = \quad 0, \quad g \in I, \tag{4.176}$$
$$N_h(s_z) = N_h(s'_z) \quad \Leftrightarrow \quad \textstyle\sum_{ij} \sigma_h(ij) z(ij) \quad = \quad 0, \quad h \in J, \tag{4.177}$$
$$f_{cv}(s_z) = f_{cv}(s'_z) \quad \Leftrightarrow \quad \textstyle\sum_{ij} \tau_{cv}(ij) z(ij) \quad = \quad 0, \quad v \in V_c. \tag{4.178}$$

Next, recall from (4.108) [together with the independence axioms A1 and A2] that for all $s, s' \in S$ with $N(s) = N(s')$,

$$\begin{aligned} P_c(s) = P_c(s') &\Leftrightarrow \prod_{ij} p_c(ij)^{N_{ij}(s)} = \prod_{ij} p_c(ij)^{N_{ij}(s')} \\ &\Leftrightarrow \sum_{ij} N_{ij}(s) \ln[p_c(ij)] = \sum_{ij} N_{ij}(s') \ln[p_c(ij)] \\ &\Leftrightarrow \sum_{ij} \ln[p_c(ij)]\{N_{ij}(s) - N_{ij}(s')\} = 0. \end{aligned} \tag{4.179}$$

Hence, for s_z and s'_z in particular, one may infer from (4.168) that whenever the right hand sides of (4.176), (4.177), and (4.178) hold, the equality,

$P_c(s_z) = P_c(s'_z)$, must also hold, so that by (4.179),

$$\sum_{ij} \ln[p_c(ij)]z(ij) = 0. \tag{4.180}$$

Finally, since the function, $z \in Z^{I\times J}$, was chosen arbitrarily, it may be concluded from part (ii) of Lemma 4.4 that by setting $T = (I + J) \cup V_c$, $W = I \times J$, and defining the functions, $b_c : I \times J \to R$ and $a_c : T \times (I \times J) \to Q$, respectively, for all $c \in C$ and $ij \in I \times J$ by $b_c(ij) = \ln[p_c(ij)]$ and

$$a_{ct}(ij) = \begin{cases} \delta_g(ij), & t = (g,1) \in I \times \{1\} \\ \sigma_h(ij), & t = (h,2) \in J \times \{2\} \\ \tau_{ct}(ij), & t \in V_c, \end{cases} \tag{4.181}$$

there must exist for each configuration, $c \in C$ a function, $\lambda_c : T \to R$, such that for all $ij \in I \times J$,

$$\begin{aligned} \ln[p_c(ij)] &= \sum_{t\in T} \lambda_c(t) a_{ct}(ij) \\ &= \sum_{g\in I} \lambda_c(g,1)\delta_g(ij) + \sum_{h\in J} \lambda_c(h,2)\sigma_h(ij) + \sum_{v\in V_c} \lambda_c(v)\tau_{cv}(ij) \\ &= \lambda_c(i,1) + \lambda_c(j,2) + \lambda_c(c_{ij}), \end{aligned} \tag{4.182}$$

which in turn may be written as

$$p_c(ij) = \exp[\lambda_c(i,1)] \exp[\lambda_c(j,2)] \exp[\lambda_c(c_{ij})]. \tag{4.183}$$

The desired representation of $\mathrm{E}(N^c_{ij})$ now follows at once from (4.183). For if we define the functions, $A_c : I \to R_{++}$, $B_c : J \to R_{++}$ and $F_c : V_c \to R_{++}$, respectively, for all $c \in C$ and $ij \in I \times J$ by $A_c(i) = \mathrm{E}(N^c)\exp[\lambda_c(i,1)] > 0$, $B_c(j) = \exp[\lambda_c(j,2)] > 0$, and $F_c(c_{ij}) = \exp[\lambda_c(c_{ij})] > 0$, then it follows from (4.107) that

$$\mathrm{E}(N^c_{ij}) = \mathrm{E}(N^c)p_c(ij) = A_c(i)B_c(j)F_c(c_{ij}) \tag{4.184}$$

for all $ij \in I \times J$ and $c \in C$. Thus, for each $\boldsymbol{P}$ satisfying (A1,A2,A3), it may be concluded from (4.184) that $\boldsymbol{N_P} \in \langle\text{MODEL G1}\rangle$, and hence that $\langle \text{A1, A2, A3}\rangle \subseteq \langle\text{MODEL 1}\rangle$.

(b) To establish the converse, observe first from Corollary 3.2 that for any frequency process, $\boldsymbol{N} \in \langle\text{MODEL G1}\rangle$, the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ must satisfy A1, and hence must satisfy (4.108) for all $c \in C$ and $s \in S$. But if we set $A^*_c(i) = A_c(i)/\mathrm{E}(N^c)$ in (4.1), then by (4.108) it follows that for all $s \in S$,

$$\begin{aligned} P_c(s) &= P_c[N = N(s)] \prod_{ij} \{A^*_c(i)B_c(j)F_c(c_{ij})\}^{N_{ij}(s)} \\ &= P_c[N = N(s)] \left(\prod_i A^*_c(i)^{N_i(s)}\right) \left(\prod_j B_c(j)^{N_j(s)}\right) \prod_v F_c(v)^{f_{cv}(s)}. \end{aligned} \tag{4.185}$$

Moreover, since $A(s) = A(s')$ implies that both $N_i(s) = N_i(s')$ and $N_j(s) = N_j(s')$ for all $ij \in I \times J$ [and hence that $N(s) = N(s')$], and since $c_s = c_{s'}$ implies that $f_{cv}(s) = f_{cv}(s')$ for all $v \in V$, it follows from (4.185) that $A(s) = A(s')$ and $c_s = c_{s'}$ together imply $P_c(s) = P_c(s')$, and thus that $\boldsymbol{P_N}$ must satisfy A3. Hence we also have $\langle$MODEL G1$\rangle \subseteq \langle$A1, A2, A3$\rangle$, and may conclude that assertion (4.169) holds. □

Given this characterization of frequency processes representable by general gravity models with unrestricted deterrence functions, we turn next to a consideration of general gravity models with monotone deterrence functions:

Theorem 4.2 (G1* Characterization) *The class of frequency processes, $\langle$MODEL G1*$\rangle$, is characterized by the following properties of its generators:*

$$\langle \text{MODEL G1}^* \rangle = \langle \text{A1, A2, A4} \rangle. \tag{4.186}$$

PROOF: To establish assertion (4.186), we begin by observing that axiom A4 is closely related to the general efficiency principle studied in Smith (1978b), and thus that the proof of assertion (4.186) closely parallels the arguments developed there. In particular, the vector inequality relation, $v \leq w$, among separation profiles, $v, w \in V$, can be employed to define a weak form of 'proximity relation' on $I \times J$ as follows. For any given spatial configuration, $c \in C$, the binary relation, $\wp_c$, on $I \times J$ defined for all origin-destination pairs, $(ij, kh) \in (I \times J)^2$, by

$$(ij, kh) \in \wp_c \Leftrightarrow c_{ij} \leq c_{kh}, \tag{4.187}$$

is designated as the *proximity relation* generated by c. Each element, $\rho = (ij, kh) \in \wp_c \subseteq (I \times J)^2$ is designated as a proximity pair in $(I \times J)^2$. Since $(ij, ij) \in \wp_c$ for all $ij \in I \times J$, it follows that $\wp_c$ is nonempty.

(a) One may now employ these definitions to establish that $\langle$A1, A2, A4$\rangle \subseteq \langle$MODEL G1*$\rangle$ as follows. Consider any independent interaction process, $\boldsymbol{P}$, satisfying A4, and observe that for any $n \in Z_{++}$, $s = ((i_r j_r) : r = 1, \ldots, n)$, $s' = ((i'_r j'_r) : r = 1, \ldots, n) \in S_n$, and $c \in C$, it follows from (4.187) that $c_s \leq c_{s'}$ iff $(i_r j_r, i'_r j'_r) \in \wp_c$ for all $r = 1, \ldots, n$. Hence, if for each $c \in C$, we denote by, $S_c^2 = \{(s, s') \in S^2 : N(s) = N(s'), c_s \leq c_{s'}\}$, the relevant domain of this inequality relation, and for each pair of patterns, $(s, s') = ((i_r j_r, i'_r j'_r) : r = 1, \ldots, n) \in S_c^2$ and proximity pair, $\rho \in \wp_c$, denote the number of occurrences, of ρ in (s, s') by, $z_{(s,s')}(\rho) = |\{r : \rho = (i_r j_r, i'_r j'_r)\}|$, then the resulting function, $z_{(s,s')} : \wp_c \rightarrow Z_+$, defines the *frequency distribution* of proximity pairs appearing in (s, s'). Axiom A4 can then be reformulated in terms of these frequencies as follows. Let the indicator functions, $\delta_\rho^\alpha : I \rightarrow \{0, 1\}$ and $\sigma_\rho^\alpha : J \rightarrow \{0, 1\}$, be defined for all $\rho = (i_\rho^1 j_\rho^1, i_\rho^2 j_\rho^2) \in \wp_c$ and $\alpha = 1, 2$ by $\delta_\rho^\alpha(i) = 1 \Leftrightarrow i = i_\rho^\alpha$ and

$\sigma_\rho^\alpha(j) = 1 \Leftrightarrow j = j_\rho^\alpha$. Then for each $(s, s') \in S_c^2$, it follows that the origin and destination activity levels for s and s' can be written as

$$\begin{aligned} N_i(s) &= \textstyle\sum_{\rho \in \wp_c} \delta_\rho^1(i) z_{(s,s')}(\rho), & i \in I, \quad (4.188)\\ N_j(s) &= \textstyle\sum_{\rho \in \wp_c} \sigma_\rho^1(j) z_{(s,s')}(\rho), & j \in J, \quad (4.189)\\ N_i(s') &= \textstyle\sum_{\rho \in \wp_c} \delta_\rho^2(i) z_{(s,s')}(\rho), & i \in I, \quad (4.190)\\ N_j(s') &= \textstyle\sum_{\rho \in \wp_c} \sigma_\rho^2(j) z_{(s,s')}(\rho), & j \in J. \quad (4.191) \end{aligned}$$

Hence, for all $i \in I, j \in J$, and $(s, s') \in S_c^2$,

$$N_i(s) = N_i(s') \Leftrightarrow \sum_{\rho \in \wp_c} [\delta_\rho^2(i) - \delta_\rho^1(i)] z_{(s,s')}(\rho) = 0 \qquad (4.192)$$

$$N_j(s) = N_j(s') \Leftrightarrow \sum_{\rho \in \wp_c} [\sigma_\rho^2(j) - \sigma_\rho^1(j)] z_{(s,s')}(\rho) = 0. \qquad (4.193)$$

Similarly, since $(s, s') \in S_c^2 \Rightarrow N(s) = N(s')$, it follows from (4.108) that

$$\begin{aligned} P_c(s) \geq P_c(s') &\Leftrightarrow \prod_{\rho \in \wp_c} p_c(i_\rho^1 j_\rho^1)^{z_{(s,s')}(\rho)} \geq \prod_{\rho \in \wp_c} p_c(i_\rho^2 j_\rho^2)^{z_{(s,s')}(\rho)} \\ &\Leftrightarrow \sum_{\rho \in \wp_c} z_{(s,s')}(\rho)(\ln[p_c(i_\rho^2 j_\rho^2)] - \ln[p_c(i_\rho^1 j_\rho^1)]) \leq 0. \end{aligned} \qquad (4.194)$$

Hence, if the equalities on the right hand side of (4.192) and (4.193) hold for any given pattern pair, $(s, s') \in S_c^2$, then one has $A(s) = A(s')$, and it follows from A4 that the inequality on the right hand side of (4.194) must also hold. Moreover, the frequency function, $z_{(s,s')}$, may be chosen to be *any* function from $\wp_c$ to Z_+. To see this, let $M = |\wp_c|$ and let $\{\rho_1, \ldots, \rho_m, \ldots, \rho_M\}$ denote any fixed enumeration of the finite set, $\wp_c$. Then for any choice of nonnegative integer-valued function, $z : \wp_c \to Z_+$, if the interaction patterns, $s_z = (\ldots, (i_{zr} j_{zr}), \ldots)$ and $s_z' = (\ldots, (i_{zr}' j_{zr}'), \ldots)$ in S_c^2 are defined by setting $i_{zr} = i_{\rho_m}^1$, $j_{zr} = j_{\rho_m}^1$, $i_{zr}' = i_{\rho_m}^2$, and $j_{zr}' = j_{\rho_m}^2$ for all $m = 1, \ldots, M$ and all r with $\sum_{t<m} z(\rho_t) < r \leq \sum_{t \leq m} z(\rho_t)$, it follows by construction that $(s, s') \in S_c^2$ and that $z_{(s,s')}(\rho) = z(\rho)$ for all $\rho \in \wp_c$. Thus it may be seen that if the equalities on the right hand side of (4.192) and (4.193) hold for any function, $z : \wp_c \to Z_+$, then the inequality on the right hand side of (4.194) also holds for z. This means that if in Lemma 4.4 we now set $T = I + J$, $W = \wp_c$, and define the functions $b_c : \wp_c \to R$, and, $a_c : T \times \wp_c \to Z \subseteq Q$, for all proximity pairs, $\rho = (i_\rho^1 j_\rho^1, i_\rho^2 j_\rho^2) \in \wp_c$, by $b_c(\rho) = \ln[p_c(i_\rho^2 j_\rho^2)] - \ln[p_c(i_\rho^1 j_\rho^1)]$ and

$$a_{ct}(\rho) = \begin{cases} \delta_\rho^2(i) - \delta_\rho^1(i), & t = (i, 1) \in I \times \{1\} \\ \sigma_\rho^2(j) - \sigma_\rho^1(j), & t = (j, 2) \in J \times \{2\}, \end{cases} \qquad (4.195)$$

respectively, then it follows from part (i) of Lemma 4.4 that for each $c \in C$ there must exist a function, $\lambda_c : T \to R$, such that for all $\rho \in \wp_c$,

$$\begin{aligned}
\ln[p_c(i_\rho^2 j_\rho^2)] - \ln[p_c(i_\rho^1 j_\rho^1)] &= b_c(\rho) \leq \sum_{t \in T} \lambda_c(t) a_{ct}(\rho) \\
&= \sum_{i \in I} \lambda_c(i,1)[\delta_\rho^2(i) - \delta_\rho^1(i)] + \sum_{j \in J} \lambda_c(j,2)[\sigma_\rho^2(j) - \sigma_\rho^1(j)] \\
&= [\lambda_c(i_\rho^2, 1) - \lambda_c(i_\rho^1, 1)] + [\lambda_c(j_\rho^2, 2) - \lambda_c(j_\rho^1, 2)].
\end{aligned} \tag{4.196}$$

Using this result, we may now verify that $\boldsymbol{N_P} \in \langle \text{MODEL G1}^* \rangle$ as follows. Let the functions, $A_c : I \to R_{++}$ and $B_c : J \to R_{++}$, be defined, respectively, by $A_c(i) = \exp[\lambda_c(i,1)] > 0, i \in I$, and $B_c(j) = \exp[\lambda_c(j,2)] > 0, j \in J$, and for each separation profile, $c_{ij} \in V_c$, let

$$F_c(c_{ij}) = \frac{\mathrm{E}(N_{ij}^c)}{A_c(i)B_c(j)} > 0. \tag{4.197}$$

With these definitions, if (4.197) can be shown to yield a nonincreasing function, $F_c : V_c \to R_{++}$, for each $c \in C$, then it will follow at once from (4.1) and (4.197) that $\boldsymbol{N_P} \in \langle \text{MODEL G1}^* \rangle$. To so do, observe from (4.196) and (4.197), together with (4.107) and the definitions of A_c and B_c, that for any profiles, $v, w, \in V_c$, and origin-destination pairs, $(ij), (kh) \in I \times J$ with $c_{ij} = v$ and $c_{kh} = w$ (which exist by the definition of V_c), we must have:

$$\begin{aligned}
v \geq w \Rightarrow c_{ij} \geq c_{kh} &\Rightarrow (kh, ij) \in \wp_c \\
&\Rightarrow \ln[p_c(ij)] - \ln[p_c(kh)] \\
&\qquad \leq \lambda_c(i,1) - \lambda_c(k,1) + \lambda_c(j,2) - \lambda_c(h,2) \\
&\Rightarrow \frac{p_c(ij)}{p_c(kh)} \leq \frac{\exp[\lambda_c(i,1)] \exp[\lambda_c(j,2)]}{\exp[\lambda_c(k,1)] \exp[\lambda_c(h,2)]} \\
&\Rightarrow \frac{\mathrm{E}(N_{ij}^c)}{A_c(i)B_c(j)} \leq \frac{\mathrm{E}(N_{kh}^c)}{A_c(k)B_c(h)} \\
&\Rightarrow F_c(v) = F_c(c_{ij}) \leq F_c(c_{kh}) = F_c(w).
\end{aligned} \tag{4.198}$$

Thus, F_c is seen to be a nonincreasing function on V_c, and we may conclude that $\langle \text{A1, A2, A4} \rangle \subseteq \langle \text{MODEL G1}^* \rangle$.

(b) To establish the converse, observe first that if $\boldsymbol{N} \in \langle \text{MODEL G1}^* \rangle$, then it follows from (4.1), (4.107), and (4.108) that by setting $A_c^*(i) = A_c(i)/E(N^c)$ as in (4.185), the generator, $\boldsymbol{P_N} = (P_c : c \in C)$, of $\boldsymbol{N}$ can always be written for all $c \in C$ and $s = ((i_r j_r) : r = 1, \ldots, n) \in S$ as

$$\begin{aligned} P_c(s) &= P_c[N = N(s)] \prod_{r=1}^{n} p_c(i_r j_r) \\ &= P_c[N = N(s)] \prod_{r=1}^{n} [A_c^*(i_r) B_c(j_r) F_c(c_{i_r j_r})] \\ &= P_c[N = N(s)] \left(\prod_i A_c^*(i)^{N_i(s)} \right) \\ &\quad \cdot \left(\prod_j B_c(j)^{N_j(s)} \right) \prod_r F_c(c_{i_r j_r}). \end{aligned} \tag{4.199}$$

Hence, for any patterns $s, s' \in S$ with $A(s) = A(s')$, we must have

$$P_c(s) \geq P_c(s') \Leftrightarrow \prod_r F_c(c_{i_r j_r}) \geq \prod_r F_c(c_{i'_r j'_r}). \tag{4.200}$$

But since the function, F_c, in MODEL G1* is nonincreasing by definition, it follows that for all $s = ((i_r j_r) : r = 1, \ldots, n)$, $s' = ((i'_r j'_r) : r = 1, \ldots, n) \in S$ with $A(s) = A(s')$,

$$\begin{aligned} c_s \leq c_{s'} &\Rightarrow c_{i_r j_r} \leq c_{i'_r j'_r}, r = 1, \ldots, n \\ &\Rightarrow F_c(c_{i_r j_r}) \geq F_c(c_{i'_r j'_r}), r = 1, \ldots, n \\ &\Rightarrow \prod_r F_c(c_{i_r j_r}) \geq \prod_r F_c(c_{i'_r j'_r}) \\ &\Rightarrow P_c(s) \geq P_c(s') \end{aligned} \tag{4.201}$$

and hence that $\boldsymbol{P_N}$ satisfies A4 whenever $\boldsymbol{N} \in \langle$MODEL G1*$\rangle$. Thus we also have $\langle$MODEL G1*$\rangle \subseteq \langle$A1, A2, A4$\rangle$, and may conclude that assertion (4.186) must hold. □

Turning next to deterrence-invariant gravity models, we begin by recalling from Section 2.2.3(A) that the characterization of these models in terms of the relative separation dependence axiom (A3′) required that either $|I| \geq 3$ or $|J| \geq 3$. Hence it is of interest to develop a stronger axiom which characterizes these models under the more general conditions (assumed for all models) that $|I| \geq 2$ and $|J| \geq 2$. To do so, we begin by introducing the following vector generalizations of (4.104), (4.114), (4.164) and (4.165). For any interaction process, $\boldsymbol{P} = \{P_c : c \in C\}, n \in Z_{++}$, and corresponding n-tuple of spatial interaction patterns, $\boldsymbol{s} = (s_\alpha : \alpha = 1, \ldots, n) \in S^n$, and separation configurations, $\boldsymbol{c} = (c_\alpha : \alpha = 1, \ldots, n) \in C^n$, let

$$P_{\boldsymbol{c}}(\boldsymbol{s}) \;=\; \prod_{\alpha=1}^{n} P_{c_\alpha}(s_\alpha), \tag{4.202}$$

$$A(\boldsymbol{s}) = [A(s_\alpha) : \alpha = 1, \ldots, n], \tag{4.203}$$

$$f_{\boldsymbol{c}v}(\boldsymbol{s}) = \sum_{\alpha=1}^{n} f_{c_\alpha v}(s_\alpha), \quad v \in V, \tag{4.204}$$

$$f_{\boldsymbol{c}}(\boldsymbol{s}) = [f_{\boldsymbol{c}v}(\boldsymbol{s}) : v \in V] = \sum_{\alpha=1}^{n} f_{c_\alpha}(s_\alpha). \tag{4.205}$$

In terms of this notation, we then have the following strengthening of the frequency version of separation dependence given in Lemma 4.10 above:

A3°. (Strong Separation Dependence) *For all $n \in Z_{++}$, and all n-tuples of separation configurations, $\boldsymbol{c}, \boldsymbol{c}' \in C^n$, and spatial interaction patterns, $\boldsymbol{s}, \boldsymbol{s}' \in S^n$, with $A(\boldsymbol{s}) = A(\boldsymbol{s}')$,*

$$f_{\boldsymbol{c}}(\boldsymbol{s}) = f_{\boldsymbol{c}'}(\boldsymbol{s}') \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}'}(\boldsymbol{s}'). \tag{4.206}$$

This axiom clearly implies A3. Moreover, it in fact implies A3′, i.e.,

$$\langle \text{A3°} \rangle \subseteq \langle \text{A3}' \rangle. \tag{4.207}$$

To see this, observe that if for any comparable $s, t, s', t' \in S$ it is true that $A(s) = A(t)$ and $A(s') = A(t')$ then letting $\boldsymbol{s} = (s, t')$ and $\boldsymbol{s}' = (t, s')$ it follows from (4.202) that $A(\boldsymbol{s}) = A(\boldsymbol{s}')$. Hence if A3° holds, then by letting $\boldsymbol{c} = (c, c') = \boldsymbol{c}'$, and recalling by definition that $c_s = c_t \Rightarrow f_c(s) = f_c(t)$, it follows from (4.203) and (4.206) that

$$\begin{aligned}
&(c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow [f_c(s) = f_{c'}(s'), f_c(t) = f_{c'}(t')] \\
&\quad \Rightarrow f_{\boldsymbol{c}}(\boldsymbol{s}) = f_c(s) + f_{c'}(t') = f_c(t) + f_{c'}(s') = f_{\boldsymbol{c}'}(\boldsymbol{s}') \\
&\quad \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}'}(\boldsymbol{s}') \Rightarrow P_c(s)P_{c'}(t') = P_c(t)P_{c'}(s') \\
&\quad \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}.
\end{aligned} \tag{4.208}$$

With this observation, we now have the following characterizations of deterrence-invariant gravity models:

Theorem 4.3 (G2 Characterization)
(i) *The class of frequency processes,* ⟨MODEL G2⟩, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G2} \rangle = \langle \text{A1, A2, A3°} \rangle. \tag{4.209}$$

(ii) *If in addition it is true that either $|I| \geq 3$ or $|J| \geq 3$ then*

$$\langle \text{MODEL G2} \rangle = \langle \text{A1, A2, A3}' \rangle. \tag{4.210}$$

PROOF: It is convenient to begin by establishing the assertion in part (ii), and then employ this result to establish the more general assertion in part (i):

(ii). (a) To establish that $\langle A1, A2, A3' \rangle \subseteq \langle \text{MODEL G2} \rangle$ when either $|I| \geq 3$ or $|J| \geq 3$, observe from the symmetric roles of I and J in $\langle \text{MODEL G2} \rangle$ that it suffices to consider the case $|J| \geq 3$. For this case, we begin by showing that for each independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying A3′ on $I \times J$, there exists a function, $F : V \to R_{++}$, satisfying the condition that for all $c \in C, i, g \in I$ and $j, h \in J$

$$\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gj})}. \tag{4.211}$$

To do so, choose any separation profiles, $x, y, z, w \in V$, and let

$$\begin{aligned}\phi(x, y, z, w) = &\{[P_c(ij)P_c(gh)/P_c(ih)P_c(gj)] : c \in C, i, g \in I, \\ &j, h \in J, (c_{ij}, c_{gh}, c_{ih}, c_{gj}) = (x, y, z, w)\}.\end{aligned} \tag{4.212}$$

Then the relations, $C = V^{I \times J}, |I| \geq 2$ and $|J| \geq 2$, together imply that $\phi(x, y, z, w) \neq \emptyset$. Moreover, if for any $c, c' \in C, i, g, a, b \in I$, and $j, h, d, e \in J$, with $c_{ij} = x = c'_{ad}, c_{gh} = y = c'_{be}, c_{ih} = z = c'_{ae}$, and $c_{gj} = w = c'_{bd}$, we let $s = (ij, gh), t = (ih, gj), s' = (ad, be)$, and $t' = (ae, bd)$, then by construction, $A(s) = A(t), A(s') = A(t'), c_s = c'_{s'}$, and $c_t = c'_{t'}$, which together with (4.109) and A3′ implies that

$$\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} = \frac{P_{c'}(ad)P_{c'}(be)}{P_{c'}(ae)P_{c'}(bd)}. \tag{4.213}$$

Hence we see that $|\phi(x, y, z, w)| = 1$ for all $x, y, z, w \in V$, and it follows that (4.212) yields a well defined function, $\phi : V^4 \to R_{++}$, which satisfies the identity

$$\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \phi(c_{ij}, c_{gh}, c_{ih}, c_{gj}), \tag{4.214}$$

for all $c \in C, i, g \in I$, and $j, h \in J$. To establish the desired separability properties of ϕ, observe first that for any $x, y, z, w \in V$ and $c \in C$, $i, g \in I$, and $j, h \in J$ with $(c_{ij}, c_{gh}, c_{ih}, c_{gj}) = (x, y, z, w)$, it follows at once from (4.214) that

$$\phi(x, y, z, w)\phi(z, w, x, y) = \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} \cdot \frac{P_c(ih)P_c(gj)}{P_c(ij)P_c(gh)} = 1. \tag{4.215}$$

Similar rearrangements of the left hand side of (4.214) show that

$$\phi(x, y, z, w) = \phi(y, x, z, w) = \phi(x, y, w, z) \tag{4.216}$$

hold identically for all $x, y, z, w \in V$. Moreover, since $|J| \geq 3$ implies the existence of *distinct* destinations, $j, h, d \in J$, it follows that if for any

$x, y, z, w, \sigma \in V$ and distinct origins, $i, g \in I$, we choose $c \in C$ with $c_{ij} = x, c_{ih} = z, c_{id} = \sigma, c_{gj} = w, c_{gh} = y$, and $c_{gd} = \sigma$, then (4.214) also yields the identity

$$\begin{aligned}\phi(x,y,z,w) = \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} &= \frac{P_c(ij)P_c(gd)}{P_c(id)P_c(gj)} \cdot \frac{P_c(gh)P_c(id)}{P_c(gd)P_c(ih)} \\ &= \phi(x,\sigma,\sigma,w)\phi(y,\sigma,\sigma,z).\end{aligned} \tag{4.217}$$

In particular, by setting $y = w = \sigma$ in (4.217), we obtain

$$\phi(x,\sigma,z,\sigma) = \phi(x,\sigma,\sigma,\sigma)\phi(\sigma,\sigma,\sigma,z). \tag{4.218}$$

Hence, by combining (4.215), (4.216), and (4.218) we see that

$$\begin{aligned}\phi(x,\sigma,\sigma,w) &= \phi(x,\sigma,w,\sigma) = \phi(x,\sigma,\sigma,\sigma)\phi(\sigma,\sigma,\sigma,z) \\ &= \phi(x,\sigma,\sigma,\sigma)\phi(\sigma,\sigma,z,\sigma) = \phi(x,\sigma,\sigma,\sigma)/\phi(z,\sigma,\sigma,\sigma),\end{aligned} \tag{4.219}$$

for all $x, w, \sigma \in V$. Finally, letting the function, $F: V \to R_{++}$, be defined for all $v \in V$ and any fixed choice of $\sigma \in V$ by

$$F(x) = \phi(x,\sigma,\sigma,\sigma), \tag{4.220}$$

it follows from (4.217), (4.219), and (4.220) that

$$\phi(x,y,z,w) = \left[\frac{\phi(x,\sigma,\sigma,\sigma)}{\phi(z,\sigma,\sigma,\sigma)}\right] \cdot \left[\frac{\phi(y,\sigma,\sigma,\sigma)}{\phi(w,\sigma,\sigma,\sigma)}\right] = \frac{F(x)F(y)}{F(z)F(w)}, \tag{4.221}$$

and we see from (4.214) and (4.221) that (4.211) must hold identically. Given this relation, the desired representation follows from the observation that by (4.107) the frequency process, $\boldsymbol{N_P} = \{N^c : c \in C\}$ generated by $\boldsymbol{P}$ satisfies the following identity for all $c \in C, i, g \in I$ and $j, h \in J$,

$$\frac{\mathrm{E}(N_{ij}^c)\mathrm{E}(N_{gh}^c)}{\mathrm{E}(N_{ih}^c)\mathrm{E}(N_{gj}^c)} = \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)}. \tag{4.222}$$

Hence, for any fixed origin, $a \in I$, and destination, $d \in J$, we see from (4.211) and (4.222) that

$$\begin{aligned}&\frac{\mathrm{E}(N_{ij}^c)\mathrm{E}(N_{ad}^c)}{\mathrm{E}(N_{id}^c)\mathrm{E}(N_{aj}^c)} = \frac{F(c_{ij})F(c_{ad})}{F(c_{id})F(c_{aj})} \\ &\quad\Rightarrow \mathrm{E}(N_{ij}^c) = \left[\frac{\mathrm{E}(N_{id}^c)F(c_{ad})}{\mathrm{E}(N_{ad}^c)F(c_{id})}\right] \cdot [\mathrm{E}(N_{aj}^c)/F(c_{aj})] \cdot F(c_{ij}).\end{aligned} \tag{4.223}$$

Finally, letting the functions, $A_c : I \to R_{++}$, and $B_c : J \to R_{++}$, be defined for each separation configuration $c \in C$, by

$$A_c(i) = \frac{\mathrm{E}(N_{id}^c)F(c_{ad})}{\mathrm{E}(N_{ad}^c)F(c_{id})}, \quad i \in I, \tag{4.224}$$

$$B_c(j) = \mathrm{E}(N_{aj}^c)/F(c_{aj}), \quad j \in J \tag{4.225}$$

[so that in particular, $A_c(a) = 1$], it follows at once from (4.223), (4.224), and (4.225) that,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)F(c_{ij}) \tag{4.226}$$

holds identically for all $c \in C$ and $(ij) \in I \times J$, and thus that $\boldsymbol{N_P} \in \langle\text{MODEL G2}\rangle$.

(b) To establish the converse, observe from (4.207) that $\langle \text{A1, A2, A3}^\circ\rangle \subseteq \langle \text{A1, A2, A3}'\rangle$, and hence that it suffices to establish that $\langle\text{MODEL G2}\rangle \subseteq \langle \text{A1, A2, A3}^\circ\rangle$. With this observation, we turn now to the general characterization of $\langle\text{MODEL G2}\rangle$ in part (i).

(i). (a) To establish that $\langle \text{A1, A2, A3}^\circ\rangle \subseteq \langle\text{MODEL G2}\rangle$, observe first that since $\langle \text{A1, A2, A3}^\circ\rangle \subseteq \langle \text{A1, A2, A3}'\rangle$, it suffices from the proof of part ii(a) above to consider the case $|I| = |J| = 2$, say with $I = \{i, g\}$ and $J = \{j, h\}$. For this case, choose any fixed profile, $\sigma \in V$, and let the separation configurations, $c_v \in C$, be defined for each profile, $v \in V$, by $c_{vij} = v$ and $c_{vih} = c_{vgj} = c_{vgh} = \sigma$. Then by construction, each configuration c_v depends only on v, so that for any independent interaction process $\boldsymbol{P} = \{P_c : c \in C\}$ on $I \times J$, the quantities

$$F(v) = \frac{P_{c_v}(ij)P_{c_v}(gh)}{P_{c_v}(ih)P_{c_v}(gj)} \tag{4.227}$$

depend only on v, and yield a well defined function, $F : V \to R_{++}$. Thus, if it can be shown that (4.211) holds identically for this choice of F, then the desired result will follow from the argument in expressions (4.222) through (4.226) above (which was independent of the cardinality of I or J). To establish (4.211) for F in (4.227), let $n = 5$ and let the 5-tuples of separation arrays, $\boldsymbol{s} = (s_1, \ldots, s_5)$ and $\boldsymbol{s}' = (s_1', \ldots, s_5')$ be defined by $s_1 = s_2' = s_3' = s_4 = s_5 = (ij, gh)$ and $s_1' = s_2 = s_3 = s_4' = s_5' = (ih, gj)$. Then since $A(s_\alpha) = A(s_\alpha')$ for all $\alpha = 1, \ldots, 5$, it follows at once that $A(\boldsymbol{s}) = A(\boldsymbol{s}')$. Next if for any choice of separation configuration, $c \in C$, we let $c_{ij} = x, c_{gh} = y, c_{ih} = z, c_{gj} = w$, and consider the 5-tuple of configurations, $\boldsymbol{c} = (c, c_x, c_y, c_z, c_w)$, then it may be verified by inspection that if the profiles $\{x, y, z, w, \sigma\}$ are all distinct, the nonzero components of $f_{\boldsymbol{c}}(\boldsymbol{s})$ are given by

$$\begin{aligned}
f_{\boldsymbol{c}x}(\boldsymbol{s}) &= f_{cx}(s_1) + f_{c_x x}(s_2) + f_{c_y x}(s_3) + f_{c_z x}(s_4) + f_{c_w x}(s_5) = 1, \\
f_{\boldsymbol{c}y}(\boldsymbol{s}) &= f_{cy}(s_1) + f_{c_x y}(s_2) + f_{c_y y}(s_3) + f_{c_z y}(s_4) + f_{c_w y}(s_5) = 1, \\
f_{\boldsymbol{c}z}(\boldsymbol{s}) &= f_{cz}(s_1) + f_{c_x z}(s_2) + f_{c_y z}(s_3) + f_{c_z z}(s_4) + f_{c_w z}(s_5) = 1, \\
f_{\boldsymbol{c}w}(\boldsymbol{s}) &= f_{cw}(s_1) + f_{c_x w}(s_2) + f_{c_y w}(s_3) + f_{c_z w}(s_4) + f_{c_w w}(s_5) = 1, \\
f_{\boldsymbol{c}\sigma}(\boldsymbol{s}) &= f_{c\sigma}(s_1) + f_{c_x \sigma}(s_2) + f_{c_y \sigma}(s_3) + f_{c_z \sigma}(s_4) + f_{c_w \sigma}(s_5) = 6,
\end{aligned} \tag{4.228}$$

and similarly that the nonzero components of $f_c(s')$ are given by

$$\begin{aligned}
f_{cx}(s') &= f_{cx}(s_1') + f_{c_x x}(s_2') + f_{c_y x}(s_3') + f_{c_z x}(s_4') + f_{c_w x}(s_5') = 1,\\
f_{cy}(s') &= f_{cy}(s_1') + f_{c_x y}(s_2') + f_{c_y y}(s_3') + f_{c_z y}(s_4') + f_{c_w y}(s_5') = 1,\\
f_{cz}(s') &= f_{cz}(s_1') + f_{c_x z}(s_2') + f_{c_y z}(s_3') + f_{c_z z}(s_4') + f_{c_w z}(s_5') = 1,\\
f_{cw}(s') &= f_{cw}(s_1') + f_{c_x w}(s_2') + f_{c_y w}(s_3') + f_{c_z w}(s_4') + f_{c_w w}(s_5') = 1,\\
f_{c\sigma}(s') &= f_{c\sigma}(s_1') + f_{c_x \sigma}(s_2') + f_{c_y \sigma}(s_3') + f_{c_z \sigma}(s_4') + f_{c_w \sigma}(s_5') = 6.
\end{aligned} \tag{4.229}$$

Hence if the profiles, $\{x, y, z, w, \sigma\}$, are all distinct then we see from (4.228) and (4.229) that $f_c(s) = f_c(s')$. Moreover, if two or more of these profiles are the same, then by combining the corresponding frequencies in (4.228) and (4.229) it follows that $f_c(s) = f_c(s')$ must hold in all cases. Thus if $\boldsymbol{P}$ satisfies A3° then we may conclude that $P_c(s) = P_c(s')$, which together with (4.109) and (4.202) implies by definition that:

$$\begin{aligned}
&P_c(s_1)P_{c_x}(s_2)P_{c_y}(s_3)P_{c_z}(s_4)P_{c_w}(s_5)\\
&\quad = P_c(s_1')P_{c_x}(s_2')P_{c_y}(s_3')P_{c_z}(s_4')P_{c_w}(s_5')\\
&\quad \Rightarrow \frac{P_c(s_1)}{P_c(s_1')} = \frac{[P_{c_x}(s_2')/P_{c_x}(s_2)]\cdot[P_{c_y}(s_3')/P_{c_y}(s_3)]}{[P_{c_z}(s_4)/P_{c_z}(s_4')]\cdot[P_{c_w}(s_5)/P_{c_w}(s_5')]}\\
&\quad \Rightarrow \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{\left[\frac{P_{c_x}(ij)P_{c_x}(gh)}{P_{c_x}(ih)P_{c_x}(gj)}\right]\cdot\left[\frac{P_{c_y}(ij)P_{c_y}(gh)}{P_{c_y}(ih)P_{c_y}(gj)}\right]}{\left[\frac{P_{c_x}(ij)P_{c_x}(gh)}{P_{c_x}(ih)P_{c_x}(gj)}\right]\cdot\left[\frac{P_{c_y}(ij)P_{c_y}(gh)}{P_{c_y}(ih)P_{c_y}(gj)}\right]}.
\end{aligned} \tag{4.230}$$

Finally, by combining (4.227) with (4.230) and recalling that by definition, $(c_{ij}, c_{gh}, c_{ih}, c_{gj}) = (x, y, z, w)$, we see that (4.211) must hold identically for F in (4.227).

(b) To verify that $\langle$MODEL G2$\rangle \subseteq \langle$A1, A2, A3°$\rangle$, consider any frequency process, $\boldsymbol{N} \in \langle$MODEL G2$\rangle$, and observe that if we now set $A_c^*(i) = A_c(i)/E(N^c)$ in (4.2), then again by (4.108) it follows that for each $c \in C$ and $s \in S$ the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ must satisfy

$$\begin{aligned}
P_c(s) &= P_c[N = N(s)]\prod_{ij}\{A_c^*(i)B_c(j)F(c_{ij})\}^{N_{ij}(s)}\\
&= P_c[N = N(s)]\left(\prod_i A_c^*(i)^{N_i(s)}\right)\left(\prod_j B_c(j)^{N_j(s)}\right)\\
&\quad \cdot\prod_v F(v)^{f_{cv}(s)}
\end{aligned} \tag{4.231}$$

so that for any $n \in Z_{++}$, and n-tuples of interaction patterns, $\boldsymbol{s} = (s_\alpha : \alpha = 1, \ldots, n) \in S^n$, and separation configurations, $\boldsymbol{c} = (c_\alpha : \alpha = 1, \ldots, n) \in C^n$, it follows at once from (4.202) and (4.231) that

$$
\begin{aligned}
P_c(s) &= \prod_{\alpha=1}^{n} P_{c_\alpha}(s_\alpha) \\
&= \prod_{\alpha=1}^{n} \left[P_{c_\alpha}[N = N(s_\alpha)] \left(\prod_i A^*_{c_\alpha}(i)^{N_i(s_\alpha)} \right) \right. \\
&\quad \left. \cdot \left(\prod_j B_{c_\alpha}(j)^{N_j(s_\alpha)} \right) \prod_v F(v)^{f_{c_\alpha v}(s_\alpha)} \right] \\
&= \prod_{\alpha=1}^{n} \left[P_{c_\alpha}[N = N(s_\alpha)] \left(\prod_i A^*_{c_\alpha}(i)^{N_i(s_\alpha)} \right) \right. \\
&\quad \left. \cdot \left(\prod_j B_{c\alpha}(j)^{N_j(s_\alpha)} \right) \cdot \prod_v F(v)^{f_{cv}(s)} \right].
\end{aligned}
\tag{4.232}
$$

But for any other n-tuple of interaction patterns, $s' = (s'_\alpha = 1, \ldots, n) \in S^n$, with $A(s) = A(s')$ we must have $N_i(s_\alpha) = N_i(s'_\alpha)$ and $N_j(s_\alpha) = N_j(s'_\alpha)$ for all $i, j \in I \times J$ and $\alpha = 1, \ldots, n$ [which in turn implies that $N(s_\alpha) = N(s'_\alpha)$ for all $\alpha = 1, \ldots, n$]. Hence it follows from (4.232) that for any $s, s' \in S^n$ with $A(s) = A(s')$ and any $c \in C^n$,

$$
P_c(s)/P_c(s') = \prod_{v \in V} F(v)^{[f_{cv}(s) - f_{cv}(s')]}, \tag{4.233}
$$

and we may conclude that

$$
\begin{aligned}
&[A(s) = A(s'), f_c(s) = f_c(s')] \Rightarrow P_c(s)/P_c(s') = 1 \\
&\quad \Rightarrow P_c(s) = P_c(s').
\end{aligned}
\tag{4.234}
$$

Thus $\boldsymbol{P_N}$ satisfies A3° whenever $\boldsymbol{N} \in$ ⟨MODEL G2⟩, and it follows that ⟨MODEL G2⟩ $\subseteq$ ⟨A1, A2, A3°⟩. □

Given this characterization of all deterrence-invariant gravity models, we next turn to the case of monotone deterrence-invariant gravity models:

Theorem 4.4 (G2* Characterization)
(i) *The class of frequency processes,* ⟨MODEL G2*⟩, *is characterized by the following properties of its generators:*

$$
\langle \text{MODEL G2}^* \rangle = \langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4} \rangle. \tag{4.235}
$$

(ii) *If in addition it is true that either* $|I| \geq 3$ *or* $|J| \geq 3$ *then*

$$
\langle \text{MODEL G2}^* \rangle = \langle \text{A1}, \text{A2}, \text{A4}' \rangle. \tag{4.236}
$$

PROOF: **(i).** (a) To establish that $\langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4}\rangle \subseteq \langle \text{MODEL G2}^*\rangle$ observe first from (4.209) that if an independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfies A3° then $\boldsymbol{N_P} \in \langle \text{MODEL G2}\rangle$, so that the representation in (4.2) must hold for some appropriate choice of deterrence function, $F : V \to R_{++}$, and origin and destination functions, $A_c : I \to R_{++}, B_c : J \to R_{++}$, for each $c \in C$. Thus it suffices to show that if $\boldsymbol{P}$ also satisfies A4 then the function, F, must be nonincreasing. To do so, observe first that if for any given distinct origins, $i, g \in I$, and distinct destinations, $j, h \in J$, we set $s = (ij, gh)$ and $s' = (ih, gj)$, then by definition, $A(s) = A(s')$. Hence, if for any separation profiles, $x, y \in V$, we choose separation configuration, $c \in C = V^{I\times J}$, with $c_{ij} = c_{gj} = c_{gh} = y$ and $c_{ih} = x$, then it follows from A4 together with (4.107), (4.108), and (4.2), that:

$$
\begin{aligned}
y \leq x \Rightarrow & (c_{ij} \leq c_{ih}, c_{gj} = c_{gh}) \Rightarrow c_s \leq c_{s'} \\
& \Rightarrow P_c(s) \geq P_c(s') \Rightarrow p_c(ij)p_c(gh) \geq p_c(ih)p_c(gj) \\
& \Rightarrow \frac{p_c(ij)}{p_c(ih)} \geq \frac{p_c(gj)}{p_c(gh)} \Rightarrow \frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ih}^c)} \geq \frac{\mathrm{E}(N_{gj}^c)}{\mathrm{E}(N_{gh}^c)} \\
& \Rightarrow \frac{A_c(i)B_c(j)F(y)}{A_c(i)B_c(h)F(x)} \geq \frac{A_c(g)B_c(j)F(y)}{A_c(g)B_c(h)F(y)} \\
& \Rightarrow F(y) \geq F(x),
\end{aligned}
\tag{4.237}
$$

so that F is seen to be nonincreasing on V. Hence $\boldsymbol{N_P} \in \langle \text{MODEL G2}^*\rangle$ whenever $\boldsymbol{P}$ satisfies $(\text{A1}, \text{A2}, \text{A3}^\circ, \text{A4})$, and we may conclude that

$$\langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4}\rangle \subseteq \langle \text{MODEL G2}^*\rangle.$$

(b) Conversely, to establish that $\langle \text{MODEL G2}^*\rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4}\rangle$, simply observe from Theorem 4.1 together with (4.14), (4.21), (4.186) and (4.209) above that $\langle \text{MODEL G2}^*\rangle \subseteq \langle \text{MODEL G2}\rangle \cap \langle \text{MODEL G1}^*\rangle = \langle \text{A1}, \text{A2}, \text{A3}^\circ\rangle \cap \langle \text{A1}, \text{A2}, \text{A4}\rangle = \langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4}\rangle$. Hence (4.210) must hold.

(ii). (a) Similarly, to establish (4.236) under the additional condition that either $|I| \geq 3$ or $|J| \geq 3$, observe from (4.138) and (4.210) that $\langle \text{A1}, \text{A2}, \text{A4}'\rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}'\rangle = \langle \text{MODEL G2}\rangle$, and hence that under these conditions, $\boldsymbol{N_P} \in \langle \text{MODEL G2}\rangle$ whenever $\boldsymbol{P}$ satisfies $(\text{A1}, \text{A2}, \text{A4}')$. Thus the same argument as in (4.237) now shows that $\boldsymbol{N_P} \in \langle \text{MODEL G2}^*\rangle$, and hence that $\langle \text{A1}, \text{A2}, \text{A4}'\rangle \subseteq \langle \text{MODEL G2}^*\rangle$.

(b) Finally, to establish that $\langle \text{MODEL G2}^*\rangle \subseteq \langle \text{A1}, \text{A2}, \text{A4}'\rangle$, consider any frequency process, $\boldsymbol{N} \in \langle \text{MODEL G2}^*\rangle$, and observe from the nonincreasing monotonicity of the function, F, in (4.2) that for any separation configurations, $c, c' \in C$, and spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n), s' = (i_r' j_r' : r = 1, \ldots, n), t = (g_r h_r : r = 1, \ldots, n)$, and

$t' = (g'_r h'_r : r = 1, \ldots, n)$, it follows by definition that

$$\begin{aligned}(c_s \leq c'_{s'}, c_t \geq c'_{t'}) &\Rightarrow (c_{i_r j_r} \leq c'_{i'_r j'_r}, c_{g_r h_r} \geq c'_{g'_r h'_r} : r = 1, \ldots, n) \\ &\Rightarrow [F(c_{i_r j_r}) \geq F(c'_{i'_r j'_r}), F(c_{g_r h_r}) \leq F(c'_{g'_r h'_r}) : r = 1, \ldots, n] \\ &\Rightarrow [F(c_{i_r j_r} / F(c_{g_r h_r}) \geq F(c'_{i'_r j'_r}) / F(c'_{g'_r h'_r}) : r = 1, \ldots, n].\end{aligned} \tag{4.238}$$

Hence, if it is also true that $A(s) = A(t)$ and $A(s') = A(t')$, then it follows that $N_i(s) = N_i(t), N_j(s) = N_j(t), N_i(s') = N_i(t'), N_j(s') = N_j(t')$ for all $ij \in I \times J$, so that by (4.238) together with (4.108) we see that:

$$\begin{aligned}\frac{P_c(s)}{P_c(t)} &= \frac{P_c(N = n) \prod_{ij} \{A_c(i) B_c(j) F(c_{ij})\}^{N_{ij}(s)}}{P_c(N = n) \prod_{ij} \{A_c(i) B_c(j) F(c_{ij})\}^{N_{ij}(t)}} \\ &= \frac{\left[\prod_i A_c(i)^{N_i(s)}\right] \left[\prod_j B_c(j)^{N_j(s)}\right]}{\left[\prod_i A_c(i)^{N_i(t)}\right] \left[\prod_j B_c(j)^{N_j(t)}\right]} \prod_{r=1}^{n} \left[\frac{F(c_{i_r j_r})}{F(c_{g_r h_r})}\right] \\ &= \prod_{r=1}^{n} [F(c_{i_r j_r}) / F(c_{g_r h_r})] \\ &\geq \prod_{r=1}^{n} \left[F(c'_{i'_r j'_r}) / F(c'_{g'_r h'_r})\right] \\ &= \frac{\left[\prod_i A_{c'}(i)^{N_i(s')}\right] \left[\prod_j B_{c'}(j)^{N_j(s')}\right]}{\left[\prod_i A_{c'}(i)^{N_i(t')}\right] \left[\prod_j B_{c'}(j)^{N_j(t')}\right]} \prod_{r=1}^{n} \left[\frac{F(c'_{i'_r j'_r})}{F(c'_{g'_r h'_r})}\right] \\ &= \frac{P_{c'}(s')}{P_{c'}(t')},\end{aligned} \tag{4.239}$$

which implies that $\boldsymbol{P_N}$ satisfies condition (4.123). Thus we also have

$$\langle \text{MODEL G2}^* \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A4}' \rangle,$$

and may conclude that (4.236) must hold. □

Before proceeding, it is of interest to observe that when $|I| = 2 = |J|$, the axiom systems $\langle \text{A1}, \text{A2}, \text{A3}^\circ \rangle$ and $\langle \text{A1}, \text{A2}, \text{A3}^\circ, \text{A4} \rangle$ are strictly stronger than the corresponding axiom systems $\langle \text{A1}, \text{A2}, \text{A3}' \rangle$ and $\langle \text{A1}, \text{A2}, \text{A4}' \rangle$, i.e., that relative separation dependence (efficiency) is generally not sufficient to guarantee representability of independent interaction processes by (monotone) deterrence-invariant gravity models. The following counterexample illustrates a Poisson frequency process which satisfies $\langle \text{A1}, \text{A2}, \text{A3}' \rangle$ but not $\langle \text{A1}, \text{A2}, \text{A3}^\circ \rangle$. [A more complex counterexample covering the monotone case will be developed in Example 4.3 of Section 4.5.3 below.]

EXAMPLE 4.1. NONSUFFICIENCY OF THE RELATIVE SEPARATION AXIOM

Let $I = \{a, b\}, J = \{d, h\}$, and assume that all (extensive) measures of spatial separation are nonnegative, so that $V = R_+^K$. Next recall that for each cost configuration, $c \in C = V^{I \times J}$, the set of separation profiles appearing in c is given by $V_c = \{c_{ij} \in V : ij \in I \times J\}$. Hence, for each component, $k \in K$, of separation profiles, $v = (v^k : k \in K)$, we may define the set of scalar component values, v^k, corresponding to profiles in configuration c by $V_c^k = \{v^k \in R_+ : v \in V_c\}$. Finally, if we denote the maximum and minimum values in V_c^k by $\max(V_c^k) = \max\{v^k : v^k \in V_c^k\}$ and $\min(V_c^k) = \min\{v^k : v^k \in V_c^k\}$, respectively, then for each cost configuration, $c \in C$, we may construct the following (monotone) deterrence function, $F_c : V_c \to R_{++}$, defined for all separation values, $v = (v^k : k \in K) \in V_c$, by

$$F_c(v) = \prod_{k \in K} [2 + \max(V_c^k) - \min(V_c^k)]^{-v^k}. \tag{4.240}$$

Given this deterrence function, we now consider the Poisson frequency process, $\boldsymbol{N} = \{N^c : c \in C\}$, with mean frequencies defined for each $c \in C$ and $ij \in I \times J$ by

$$\mathrm{E}_c(N_{ij}) = F_c(c_{ij}). \tag{4.241}$$

In this context, if we let $A_c(i) = 1 = B_c(j)$ for all $c \in C$ and $ij \in I \times J$, then it follows at once that $\boldsymbol{N} \in \langle\text{MODEL G1}^*\rangle$. However, $\boldsymbol{N}$ cannot possibly belong to $\langle\text{MODEL G2}\rangle$. To see this, observe first that, as in (4.223) above, if (4.2) holds for any function, $F : V \to R_{++}$, then by (4.241) we must have

$$\frac{F_c(c_{ij})F_c(c_{gh})}{F_c(c_{ih})F_c(c_{gj})} = \frac{\mathrm{E}(N_{ij}^c)\mathrm{E}(N_{gh}^c)}{\mathrm{E}(N_{ih}^c)\mathrm{E}(N_{gj}^c)} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gj})} \tag{4.242}$$

for all $c \in C$. Hence choosing any numbers $x \geq z \in R_+$ and letting $c \in C$ be defined by $c_{ad}^1 = x, c_{ah}^1 = c_{bd}^1 = c_{bh}^1 = z$, and by $c_{ij}^k = 0$ for all $k > 1$ and $ij \in I \times J$ (so that $V_c^1 = \{x, z\}$ and $V_c^k = \{0\}$ for $k > 1$) it follows from (4.240) that $F_c(x, 0, \ldots, 0) = (2 + x - z)^{-x} \prod_{k>1}(2)^0 = (2 + x - z)^{-x}$, and similarly, that $F_c(z, 0, \ldots, 0) = (2 + x - z)^{-z}$. Thus we see from (4.242) that

$$(2 + x - z)^{z-x} = \frac{F(x, 0, \ldots, 0)}{F(z, 0, \ldots, 0)} \tag{4.243}$$

must hold identically for all $x \geq z \in R_+$. In particular, for $z = 1$ we have $F(x, 0, \ldots, 0) = F(1, 0, \ldots, 0)(1 + x)^{1-x} > 0$, and for $z = 0$ we have $F(x, 0, \ldots, 0) = F(0, \ldots, 0)(2 + x)^{-x} > 0$, which together must imply that $F(1, 0, \ldots, 0)(1 + x)^{1-x} = F(0, \ldots, 0)(2 + x)^{-x}$ for all $x \geq 1$. But for $x = 1$ we obtain $F(1, 0, \ldots, 0) = F(0, \ldots, 0)/3 > 0$, and for $x = 2$ we obtain $F(1, 0, \ldots, 0) = F(0, \ldots, 0)(3/16) < F(0, \ldots, 0)/3$, which together yield a contradiction. Hence no function F can satisfy (4.243), and we may conclude that $\boldsymbol{N}$ fails to satisfy (4.2) for any function $F : V \to R_{++}$.

However, the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ satisfies (A1, A2, A3$'$). To see this, observe first that since $\boldsymbol{N} \in \langle$MODEL G1$\rangle$ by construction, it follows that $\boldsymbol{P_N}$ satisfies (A1, A2, A3). Hence it remains only to show that $\boldsymbol{P_N}$ satisfies A3$'$. To do so, observe first from (4.185) together with (4.241) that if for each spatial interaction pattern, $s \in S$, and profile component, $k \in K$, we denote the *total k-cost* in s by $c_k(s) = \sum_{ij} c_{ij}^k N_{ij}(s)$, then for any pair of (comparable) patterns, $s, t \in S$, we must have

$$\begin{aligned} \frac{P_c(s)}{P_c(t)} &= \prod_{ij} F_c(c_{ij})^{\{N_{ij}(s)-N_{ij}(t)\}} \\ &= \prod_{k\in K} [2 + \max(V_c) - \min(V_c)]^{\{\Sigma_{ij} c_{ij}^k N_{ij}(t) - \Sigma_{ij} c_{ij}^k N_{ij}(s)\}} \\ &= \prod_{k\in K} [2 + \max(V_c^k) - \min(V_c^k)]^{\{c_k(t)-c_k(s)\}}. \end{aligned} \tag{4.244}$$

Next we observe from (4.244) that $\boldsymbol{P_N}$ must trivially satisfy A3$'$ for any $c \in C$ and $s, t \in S$ satisfying $c_s = c_t$. For if any other configuration, $c' \in C$, and patterns, $s', t' \in S$, satisfy both $c_s = c'_{s'}$ and $c_t = c'_{t'}$, then we must have $c'_{t'} = c_t = c_s = c'_{s'}$, which together with the argument in (4.147) implies that both $c(s) = c(t)$ and $c'(s') = c'(t')$. Hence $c_k(s) - c_k(t) = 0 = c_k(s') - c_k(t')$ for all $k \in K$, and we may conclude from (4.244) that $P_c(s)/P_c(t) = 1 = P_c(s')/P_c(t')$. With this in mind, we may assume without loss of generality that $c_s \neq c_t$. But this, together with the hypothesis $|I| = 2 = |J|$, in turn implies that for all $s, t, s', t' \in S$ with $A(s) = A(t)$ and $A(s') = A(t')$ we may take the frequency differences $N_{ij}(s) - N_{ij}(t)$, and $N_{ij}(s') - N_{ij}(t')$ to be *nonzero* for each $i, j \in I \times J$. To see this, suppose for example that $N_{ad}(s) = N_{ad}(t)$. Then since $I = \{a, b\}$ and $J = \{d, h\}$, we must have $N_{bd}(s) = N_d(s) - N_{ad}(s) = N_d(t) - N_{ad}(t) = N_{bd}(t)$ and $N_{ah}(s) = N_a(s) - N_{ad}(s) = N_a(t) - N_{ad}(t) = N_{ah}(t)$, which in turn implies that $N_{bh}(s) = N_h(s) - N_{ah}(s) = N_h(t) - N_{ah}(t) = N_{bh}(t)$. Hence $N_{ij}(s) = N_{ij}(t)$ must hold for all $i, j \in I \times J$, and it follows (by relabelling the elements of pattern t if necessary) that $c_s = c_t$ must hold for all such cases. Similar arguments show that $c_s = c_t$ must hold when either of the frequency difference $N_{ij}(s) - N_{ij}(t)$ or $N_{ij}(s') - N_{ij}(t')$ are zero for at least one $i, j \in I \times J$, and hence that A3$'$ must hold for all such cases. But this in turn implies that each $ij \in I \times J$ must appear either in s or t and must appear in either s' or t' [for otherwise we would have either $N_{ij}(s) = 0 = N_{ij}(t)$ or $N_{ij}(s') = 0 = N_{ij}(t')$]. Thus, we see from the equalities $c_s = c'_{s'}$ and $c_t = c'_{t'}$ that for each c_{ij} there is some c'_{gf} with $c_{ij} = c'_{gf}$, and hence that $V_c \subseteq V_{c'}$. Similarly, for each c'_{ij} there is some c_{gf} with $c'_{ij} = c_{gf}$, so that $V_{c'} \subseteq V_c$ also holds, and we must have $V_c = V_{c'}$. But since this implies in particular that $V_c^k = V_{c'}^k$ for each $k \in K$, it then follows that both $\max(V_c^k) = \max(V_{c'}^k)$ and $\min(V_c^k) = \min(V_{c'}^k)$ must hold for all $k \in K$. Finally, since the argument in (4.147) together with the equalities

$c_s = c'_{s'}$ and $c_t = c'_{t'}$ also imply that $c_k(t) - c_k(s) = c_k(t') - c_k(s')$ for all $k \in K$, we may conclude from (4.244) that for all, $c, c' \in C$ and $s, t, s', t' \in S$ with $A(s) = A(t), A(s') = A(t'), c_s = c'_{s'}$ and $c_t = c'_{t'}$

$$\begin{aligned}\frac{P_c(s)}{P_c(t)} &= \prod_{k\in K} [2 + \max(V_c^k) - \min(V_c^k)]^{\{c_k(t)-c_k(s)\}} \\ &= \prod_{k\in K} [2 + \max(V_{c'}^k) - \min(V_{c'}^k)]^{\{c_k(t)-c_k(s)\}} \\ &= \prod_{k\in K} [2 + \max(V_{c'}^k) - \min(V_{c'}^k)]^{\{c'_k(t')-c'_k(s')\}} \\ &= \frac{P_{c'}(s')}{P_{c'}(t')}.\end{aligned} \tag{4.245}$$

Thus the generator $\boldsymbol{P_N}$ of $\boldsymbol{N}$ does satisfy A3$'$, and hence yields the desired counterexample. •

Remark 4.1. It is of interest to note that in addition to axiom A3$'$, the above counterexample 'almost' satisfies axiom A4$'$. In particular, this example satisfies A4$'$ for all interaction patterns, $s, t, s', t' \in S$, in which the interaction frequencies, $N_{ij}(s), N_{ij}(t), N_{ij}(s'), N_{ij}(t')$, are positive for each $i, j \in I \times J$. To see this, observe simply that under this condition, each $v \in V_c$ must appear at least once in both c_s and c_t, and each $w \in V_{c'}$ must appear at least once in both $c'_{s'}$ and $c'_{t'}$. But the inequality, $c_s \leq c'_{s'}$, then implies that for each $v = (v^k : k \in K) \in V_c$ there must be some $w = (w^k : k \in K) \in V'_c$ with $v \leq w$, so that for each $k \in K$ we must have both $\min(V_c^k) \leq \min(V_{c'}^k)$ and $\max(V_c^k) \leq \max(V_{c'}^k)$. Similarly, $c_t \geq c'_{t'}$ implies that both $\min(V_c^k) \geq \min(V_{c'}^k)$ and $\max(V_c^k) \geq \max(V_{c'}^k)$ for each $k \in K$. Hence for any $s, t, s', t' \in S$ with $c_s \leq c'_{s'}$ and $c_t \geq c'_{t'}$ we must again have $\max(V_c^k) = \max(V_{c'}^k)$ and $\min(V_c^k) = \min(V_{c'}^k)$ for each $k \in K$. Moreover, since $c_s \leq c'_{s'}$ and $c_t \geq c'_{t'}$ together with the argument in (4.147) also imply that $[c_k(s) \leq c'_k(s'), c_k(t) \geq c'_k(t')]$ and hence that $c_k(t) - c_k(s) \geq c'_k(t') - c'_k(s')$ for each $k \in K$, we may conclude from (4.244) together with the same argument as in (4.245) that for all, $c, c' \in C$ and $s, t, s', t' \in S$, the conditions $\{A(s) = A(t), A(s') = A(t'), c_s \leq c'_{s'}, c_t \geq c'_{t'}\}$ now imply that $P_c(s)/P_c(t) \geq P_c(s')/P_c(t')$. Hence A4$'$ must hold for all such interaction patterns. Note finally that if the subclass of all interaction patterns, $s \in S$, with positive frequencies, $N_{ij}(s)$, for each $i, j \in I \times J$ is denoted by S_+, and if axioms A3$'$ and A4$'$ are restricted to interaction patterns in S_+, then an examination of the proofs of Theorems 4.3(ii) and 4.4(ii) shows that these proofs can easily be modified to involve only interaction patterns in S_+. Hence, if axioms A3$'$ and A4$'$ in Theorems 4.3 and 4.4 are replaced by these weaker versions, then Example 4.1 is seen to yield a frequency process $\boldsymbol{N}$ satisfying this weaker form of A4$'$ which fails to be representable by Model G2. •

Characterizations of both origin-deterrence-invariant and destination deterrence-invariant gravity models will be developed in the following section on local characterizations. Hence we turn now to the characterization of relatively invariant gravity models (G5) in terms of the uniform separation dependence axiom (A3″) developed in Section 2.2.3(A).

Theorem 4.5 (G5 Characterization) *The class of frequency processes,* $\langle$MODEL G5$\rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G5} \rangle = \langle \text{A1}, \text{A2}, \text{A3}'' \rangle. \tag{4.246}$$

PROOF: (a) To establish that $\langle \text{A1}, \text{A2}, \text{A3}'' \rangle \subseteq \langle \text{MODEL G5} \rangle$, consider any interaction process $\boldsymbol{P}$ satisfying (A1, A2, A3″) and let the correspondences, ϕ_{ijgh}, from V^2 to R_{++} be defined for each distinct origin-destination pair $ij, gh \in I \times J$ and pair of separation profiles $(x, y) \in V^2$ by

$$\phi_{ijgh}(x, y) = \{P_c(ij)/P_c(gh) : c_{ij} = x, c_{gh} = y, c \in C\}. \tag{4.247}$$

Next observe that if for any $c, c' \in C$ with $c_{ij} = x = c'_{ij}$ and $c_{gh} = y = c'_{gh}$, we let the comparable interaction patterns $s, t, s', t' \in S$ be defined by $s = (ij) = s' \in S_1$ and $t = (gh) = t' \in S_1$, then $A(s) = A(s'), A(t) = A(t'), c_s = c'_{s'}$, and $c_t = c'_{t'}$ imply from A3″ together with (4.109) that

$$\frac{P_c(ij)}{P_c(gh)} = \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} = \frac{P_{c'}(ij)}{P_{c'}(gh)}, \tag{4.248}$$

and hence that (4.247) yields a well-defined function, $\phi_{ijgh} : V^2 \to R_{++}$, which by definition satisfies the following identity for all $c \in C$:

$$\frac{P_c(ij)}{P_c(gh)} = \phi_{ijgh}(c_{ij}, c_{gh}). \tag{4.249}$$

As a second consequence of axiom A3″, observe that if for any $x, y, z, w \in V$ and distinct origins, $i, g \in I$, and destinations, $j, h \in J$, we choose $c, c' \in C$ with $c_{ij} = x = c'_{ih}, c_{ih} = y = c'_{ij}, c_{gh} = z = c'_{gj}$, and $c_{gj} = w = c'_{gh}$, then by setting $s = (ij, gh) = t'$ and $t = (ih, gj) = s'$, we see that these comparable patterns, $s, t, s', t' \in S$, satisfy $A(s) = A(s'), A(t) = A(t'), c_s = c'_{s'}$, and $c_t = c'_{t'}$. Hence it follows from A3″, (4.109), and (4.249), together with the distinctness of the origin-destination pairs $\{ij, ih, gj, gh\}$, that

$$\begin{aligned} \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} &\Rightarrow \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_{c'}(ih)P_{c'}(gj)}{P_{c'}(ij)P_{c'}(gh)} \\ &\Rightarrow \phi_{ijih}(x, y) \cdot \phi_{ghgj}(z, w) = \phi_{ihij}(x, y) \cdot \phi_{gjgh}(z, w). \end{aligned} \tag{4.250}$$

In addition, observe by definition that if for any distinct origin-destination pairs, $ij, gh \in I \times J$, and any $x, y \in V$ we choose $c \in C$ with $c_{ij} = x$ and $c_{gh} = y$ then

$$\phi_{ijgh}(x, y) \cdot \phi_{ghij}(y, x) = \frac{P_c(ij)}{P_c(gh)} \cdot \frac{P_c(gh)}{P_c(ij)} = 1. \tag{4.251}$$

As a final property of these ϕ-functions, observe that if for any three distinct origin-destination pairs, $ij, gh, ab \in I \times J$, and any $x, y, z, \sigma \in V$ we choose configurations, $c, c' \in C$, with $c_{ij} = x = c'_{ij}, c_{gh} = y = c'_{gh}, c_{ab} = z$, and $c'_{ab} = \sigma$, then it also follows from (4.247) and (4.248) that

$$\begin{aligned} \frac{\phi_{ijab}(x,z)}{\phi_{ghab}(y,z)} &= \frac{P_c(ij)/P_c(ab)}{P_c(gh)/P_c(ab)} = \frac{P_c(ij)}{P_c(gh)} = \phi_{ijgh}(x,y) \\ &= \frac{P_{c'}(ij)}{P_{c'}(gh)} = \frac{P_{c'}(ij)/P_{c'}(ab)}{P_{c'}(gh)/P_{c'}(ab)} = \frac{\phi_{ijab}(x,\sigma)}{\phi_{ghab}(y,\sigma)}. \end{aligned} \tag{4.252}$$

Next, choosing any fixed elements, $a \in I, b \in J$ and $\sigma \in V$, and defining the functions, $\theta_{ij} : V \to R_{++}$, for each $i, j \in I \times J$ and $x \in V$ by

$$\theta_{ij}(x) = \begin{cases} \phi_{ijab}(x,\sigma), & ij \in I \times J - \{ab\} \\ 1, & ij = ab, \end{cases} \tag{4.253}$$

it follows from (4.252) and (4.253), together with (4.107), that the mean frequencies for the frequency process, $\boldsymbol{N_P} = \{\boldsymbol{N}_c : c \in C\}$, generated by $\boldsymbol{P}$ must satisfy the following relation for all $c \in C$ and distinct origin-destination pairs, $ij, gh \in I \times J - \{ab\}$:

$$\begin{aligned} \frac{\mathrm{E}(N^c_{ij})}{\mathrm{E}(N^c_{gh})} &= \frac{P_c(ij)}{P_c(gh)} = \frac{\phi_{ijab}(c_{ij}, c_{ab})}{\phi_{ghab}(c_{gh}, c_{ab})} \\ &= \frac{\phi_{ijab}(c_{ij}, \sigma)}{\phi_{ghab}(c_{gh}, \sigma)} = \frac{\theta_{ij}(c_{ij})}{\theta_{gh}(c_{gh})}. \end{aligned} \tag{4.254}$$

Given this representation of mean frequencies, we next establish the following additional properties of these θ-functions. First, for any choice of $ij \in I \times J$ with $i \neq a$ we must have $ij \neq ab$ and $ib \neq aj$, so that for each $x \in V$ there exist configurations, $c, c' \in C$, with $c_{ij} = x = c'_{ib}$ and $c_{ab} = \sigma = c'_{aj} = c'_{ab}$. Hence, letting $s = (ij, ab), s' = (ib, aj)$, and $t = (ab, ab) = t'$, we see by construction that the comparable patterns, $s, s', t, t' \in S$, satisfy $A(s) = A(s'), A(t) = A(t'), c_s = c'_{s'}$ and $c_t = c'_{t'}$, so that by A3$''$, $P_c(s)/P_c(t) = P_{c'}(s')/P_{c'}(t')$. Thus it follows from (4.252) and (4.254) together with (4.109) that for all $i \in I - \{a\}, j \in J - \{b\}$, and $x \in V$,

$$\begin{aligned} \theta_{ij}(x) = \phi_{ijab}(x,\sigma) &= \frac{P_c(ij)}{P_c(ab)}\frac{P_c(ab)}{P_c(ab)} = \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} \\ &= \frac{P_{c'}(ib)}{P_{c'}(ab)}\frac{P_{c'}(aj)}{P_{c'}(ab)} = \phi_{ibab}(x,\sigma)\phi_{ajab}(\sigma,\sigma) = \theta_{ib}(x)\theta_{aj}(\sigma). \end{aligned} \tag{4.255}$$

Moreover, since $\theta_{ab}(\sigma) = 1 \Rightarrow \theta_{ib}(x) = \theta_{ib}(x)\theta_{ab}(\sigma)$ by definition, we may conclude from (4.255) that for all $ij \in I \times J$ and $x \in V$,

$$i \neq a \Rightarrow \theta_{ij}(x) = \theta_{ib}(x)\theta_{aj}(\sigma). \tag{4.256}$$

Similarly, if we choose any $ij \in I \times J$ with $j \neq b$ and any $c' \in C$ with $c'_{aj} = x$ and $c'_{ib} = \sigma = c'_{ab}$, then the same argument now shows that for all $ij \in I \times J$ and $x \in V$,

$$j \neq b \Rightarrow \theta_{ij}(x) = \theta_{ib}(\sigma)\theta_{aj}(x). \tag{4.257}$$

Thus by combining (4.256) and (4.257), we see that for all $ij \in I \times J$ and $x \in V$,

$$\begin{aligned} [i \neq a, j \neq b] &\Rightarrow \theta_{ib}(x)\theta_{aj}(\sigma) = \theta_{ij}(x) = \theta_{ib}(\sigma)\theta_{aj}(x) \\ &\Rightarrow \frac{\theta_{ib}(x)}{\theta_{ib}(\sigma)} = \frac{\theta_{aj}(x)}{\theta_{aj}(\sigma)}. \end{aligned} \tag{4.258}$$

Given these preliminary observations, we are now ready to construct the desired representation of mean frequencies. In particular, if we choose any destination, $d \in J - \{b\}$, and define $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : C \to R_{++}$, respectively by

$$\lambda(c) = \frac{\mathrm{E}(N^c_{ad})}{\theta_{ad}(c_{ad})\theta_{ad}(\sigma)} > 0, \quad c \in C, \tag{4.259}$$

$$A(i) = \theta_{ib}(\sigma) > 0, \quad i \in I, \tag{4.260}$$

$$B(j) = \theta_{aj}(\sigma) > 0, \quad j \in J, \tag{4.261}$$

$$F(v) = \theta_{ad}(v) > 0, \quad v \in V, \tag{4.262}$$

then our objective is to show that for all $c \in C$ and $ij \in I \times J$,

$$\mathrm{E}(N^c_{ij}) = \lambda(c)A(i)B(j)F(c_{ij}). \tag{4.263}$$

To do so, we begin by considering any $ij \in I \times J$ with $i \neq a$. For this case, observe that by setting $j = d(\neq b)$ in (4.258), we must have

$$\frac{\theta_{ib}(x)}{\theta_{ib}(\sigma)} = \frac{\theta_{ad}(x)}{\theta_{ad}(\sigma)} \tag{4.264}$$

for all $x \in V$. Hence by setting $gh = ad$ in (4.254) and combining (4.254) with (4.256), (4.258) and (4.264) we obtain

$$\begin{aligned} \frac{\mathrm{E}(N^c_{ij})}{\mathrm{E}(N^c_{ad})} &= \frac{\theta_{ij}(c_{ij})}{\theta_{ad}(c_{ad})} = \frac{\theta_{ib}(c_{ij})\theta_{aj}(\sigma)}{\theta_{ad}(c_{ad})} \\ &= \frac{[\theta_{ib}(\sigma)\theta_{ad}(c_{ij})/\theta_{ad}(\sigma)]\theta_{aj}(\sigma)}{\theta_{ad}(c_{ad})} \\ \Rightarrow \mathrm{E}(N^c_{ij}) &= \left[\frac{\mathrm{E}(N^c_{ad})}{\theta_{ad}(c_{ad})\theta_{ad}(\sigma)}\right] \cdot \theta_{ib}(\sigma) \cdot \theta_{aj}(\sigma) \cdot \theta_{ad}(c_{ij}), \end{aligned} \tag{4.265}$$

and hence that (4.263) holds for this case. Next suppose that $j \neq b$, and choose any origin, $g \in I - \{a\}$. Then, as a parallel to (4.264), it now follows from (4.258) that for all $x \in V$,

$$\frac{\theta_{aj}(x)}{\theta_{aj}(\sigma)} = \frac{\theta_{gb}(x)}{\theta_{gb}(\sigma)} = \frac{\theta_{ad}(x)}{\theta_{ad}(\sigma)}. \tag{4.266}$$

Hence by (4.254), (4.257), (4.258), and (4.266) we see that

$$\begin{aligned}\frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ad}^c)} &= \frac{\theta_{ij}(c_{ij})}{\theta_{ad}(c_{ad})} = \frac{\theta_{ib}(\sigma)\theta_{aj}(c_{ij})}{\theta_{ad}(c_{ad})} \\ &= \frac{\theta_{ib}(\sigma)[\theta_{aj}(\sigma)\theta_{ad}(c_{ij})/\theta_{ad}(\sigma)]}{\theta_{ad}(c_{ad})} \\ \Rightarrow \mathrm{E}(N_{ij}^c) &= \left[\frac{\mathrm{E}(N_{ad}^c)}{\theta_{ad}(c_{ad})\theta_{ad}(\sigma)}\right] \cdot \theta_{ib}(\sigma)\cdot\theta_{aj}(\sigma)\cdot\theta_{ad}(c_{ij}),\end{aligned} \tag{4.267}$$

and hence that (4.263) also holds for this case. Finally, to establish (4.263) for the remaining case in which $ij = ab$, observe first that for the above choices of $g \in I - \{a\}$ and $d \in J - \{b\}$ the origin-destination pairs ab, ad, and gd are all distinct, so that by (4.252), (4.253), and (4.254),

$$\begin{aligned}\frac{\mathrm{E}(N_{ab}^c)}{\mathrm{E}(N_{ad}^c)} &= \frac{\phi_{abgd}(c_{ab},\sigma)}{\phi_{adgd}(c_{ad},\sigma)} = \frac{\phi_{abad}(c_{ab},\sigma)/\phi_{gdad}(\sigma,\sigma)}{\phi_{adab}(c_{ad},\sigma)/\phi_{gdab}(\sigma,\sigma)} \\ &= \frac{\phi_{abad}(c_{ab},\sigma)[\phi_{adab}(\sigma,\sigma)/\phi_{gdab}(\sigma,\sigma)]}{\theta_{ad}(c_{ad})/\phi_{gdab}(\sigma,\sigma)} \\ &= \frac{\phi_{abad}(c_{ab},\sigma)\theta_{ad}(\sigma)}{\theta_{ad}(c_{ad})}.\end{aligned} \tag{4.268}$$

But observe that if in (4.250) we now set $i = g = a, j = b, h = d$ and $y = z = w = \sigma$, then (4.250), (4.251), and (4.253) yield the following identity for all $x \in V$:

$$\begin{aligned}&\phi_{abad}(x,\sigma)\theta_{ad}(\sigma) = \phi_{abad}(x,\sigma)\phi_{adab}(\sigma,\sigma) \\ &= \phi_{adab}(x,\sigma)\phi_{abad}(\sigma,\sigma) = \phi_{adab}(x,\sigma)/\phi_{adab}(\sigma,\sigma) \\ &= \theta_{ad}(x)/\theta_{ad}(\sigma).\end{aligned} \tag{4.269}$$

Hence, by substituting (4.269) into (4.268) with $x = c_{ab}$, and recalling that $\theta_{ab}(\sigma) = 1$ by definition, we see that

$$\begin{aligned}\frac{\mathrm{E}(N_{ab}^c)}{\mathrm{E}(N_{ad}^c)} &= \frac{\theta_{ad}(c_{ab})/\theta_{ad}(\sigma)}{\theta_{ad}(\sigma)}\cdot[\theta_{ab}(\sigma)\theta_{ab}(\sigma)] \\ \Rightarrow \mathrm{E}(N_{ab}^c) &= \left[\frac{\mathrm{E}(N_{ad}^c)}{\theta_{ad}(c_{ad})\theta_{ad}(\sigma)}\right]\cdot\theta_{ab}(\sigma)\cdot\theta_{ab}(\sigma)\cdot\theta_{ad}(c_{ab}).\end{aligned} \tag{4.270}$$

Thus (4.263) is seen to hold in all cases, and we may conclude that

$$\langle \mathrm{A1, A2, A3''}\rangle \subseteq \langle \mathrm{MODEL\ G5}\rangle.$$

(b) To establish that $\langle$MODEL G5$\rangle \subseteq \langle$A1, A2, A3″$\rangle$, observe simply that if $\boldsymbol{N} \subseteq \langle$MODEL G5$\rangle$ then for all spatial interaction patterns $s, t, s', t' \in S_n \subseteq S$ with $A(s) = A(s')$ and $A(t) = A(t')$, and all separation configurations, $c, c' \in C$, with $c_s = c'_{s'}$ and $c_t = c'_{t'}$, it follows that $N_i(s) =$

$N_i(s'), N_j(s) = N_j(s'), N_i(t) = N_i(t'), N_j(t) = N_j(t')$ for all $ij \in I \times J$, and $f_{cv}(s) = f_{c'v}(s'), f_{cv}(t) = f_{c'v}(t')$ for all $v \in V$, so that by (4.5) and (4.109):

$$\begin{aligned}\frac{P_c(s)}{P_c(t)} &= \frac{\prod_{ij}\{\lambda(c)A(i)B(j)F(c_{ij})\}^{N_{ij}(s)}}{\prod_{ij}\{\lambda(c)A(i)B(j)F(c_{ij})\}^{N_{ij}(t)}} \\ &= \frac{\lambda(c)^n\left[\prod_i A(i)^{N_i(s)}\right]\left[\prod_j B(j)^{N_j(s)}\right]\prod_v F(v)^{f_{cv}(s)}}{\lambda(c)^n\left[\prod_i A(i)^{N_i(t)}\right]\left[\prod_j B(j)^{N_j(t)}\right]\prod_v F(v)^{f_{cv}(t)}} \\ &= \frac{\lambda(c')^n\left[\prod_i A(i)^{N_i(s')}\right]\left[\prod_j B(j)^{N_j(s')}\right]\prod_v F(v)^{f_{c'v}(s')}}{\lambda(c')^n\left[\prod_i A(i)^{N_i(t')}\right]\left[\prod_j B(j)^{N_j(t')}\right]\prod_v F(v)^{f_{c'v}(t')}} \\ &= \frac{P_{c'}(s')}{P_{c'}(t')}.\end{aligned} \tag{4.271}$$

Thus $\boldsymbol{P_N}$ satisfies A3$''$, and it follows that $\langle$MODEL G5$\rangle \subseteq \langle$A1, A2, A3$''\rangle$.□

Next we show that *monotone* relatively invariant gravity models (G5*) are characterized by uniform separation efficiency (A4$''$):

Theorem 4.6 (G5* Characterization) *The class of frequency processes, ⟨MODEL G5*⟩, is characterized by the following properties of its generators:*

$$\langle \text{MODEL G5}^* \rangle = \langle \text{A1, A2, A4}'' \rangle. \tag{4.272}$$

PROOF: (a) To establish that $\langle$A1, A2, A4$''\rangle \subseteq \langle$MODEL G5*$\rangle$ observe first from inclusion relation (4.139) in Proposition 4.10 that each independent interaction process, $\boldsymbol{P}$, satisfying A4$''$ also satisfies A3$''$. Hence it follows from (4.246) that $\boldsymbol{N_P} \in \langle$MODEL G5$\rangle$, so that the representation in (4.263) must hold for some appropriate choice of the functions, $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : V \to R_{++}$. Thus it suffices to show that the function, F, in the present case is nonincreasing. To do so, observe first that if for any given distinct origins, $i, g \in I$ and distinct destinations, $j, h \in J$, we set $s = (ij, gh) = t = t', s' = (ih, gj)$, then by definition, $A(s) = A(s') = A(t) = A(t')$. Hence, if for any separation profiles, $x, y \in V$, we choose a separation configuration, $c \in C = V^{I \times J}$, with $c_{ij} = c_{gj} = c_{gh} = y$ and $c_{ih} = x$, then it follows from A4$''$ [with $c' = c$]

together (4.263) that [in a manner paralleling (4.237)]:

$$
\begin{aligned}
y \leq x &\Rightarrow (c_{ij} \leq c_{ih}, c_{gj} = c_{gh}) \Rightarrow c_s \leq c_{s'} \\
&\Rightarrow P_c(s) \geq P_c(s') \Rightarrow p_c(ij)p_c(gh) \geq p_c(ih)p_c(gj) \\
&\Rightarrow \frac{p_c(ij)}{p_c(ih)} \geq \frac{p_c(gj)}{p_c(gh)} \Rightarrow \frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ih}^c)} \geq \frac{\mathrm{E}(N_{gj}^c)}{\mathrm{E}(N_{gh}^c)} \\
&\Rightarrow \frac{\lambda(c)A(i)B(j)F(y)}{\lambda(c)A(i)B(h)F(x)} \geq \frac{\lambda(c)A(g)B(j)F(y)}{\lambda(c)A(g)B(h)F(y)} \\
&\Rightarrow F(y) \geq F(x),
\end{aligned}
\tag{4.273}
$$

so that F is seen to be nonincreasing on V. Hence $\boldsymbol{N_P} \in \langle\text{MODEL G5}^*\rangle$ whenever $\boldsymbol{P}$ satisfies (A1, A2, A4″), and we may conclude that

$$\langle \text{A1, A2, A4}'' \rangle \subseteq \langle \text{MODEL G5}^* \rangle.$$

(b) Conversely, to establish that $\langle\text{MODEL G5}^*\rangle \subseteq \langle\text{A1, A2, A4}''\rangle$, consider any Poisson frequency process, $\boldsymbol{N} \in \langle\text{MODEL G5}^*\rangle$, and observe from the nonincreasing monotonicity of the function, F, in (4.5) that for any separation configurations, $c, c' \in C$, and spatial interaction patterns, $s = (i_r j_r : r = 1, \ldots, n), s' = (i'_r j'_r : r = 1, \ldots, n), t = (g_r h_r : r = 1, \ldots, n)$, and $t' = (g'_r h'_r : r = 1, \ldots, n)$, the argument in (4.238) continues to hold. Hence, if it is also true that $A(s) = A(s')$ and $A(t) = A(t')$, then by (4.238) and (4.271) we see that [in a manner paralleling the argument in (4.239)] the probability generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ must satisfy:

$$
\begin{aligned}
\frac{P_c(s)}{P_c(t)} &= \frac{\prod_{ij}\{\lambda(c)A(i)B(j)F(c_{ij})\}^{N_{ij}(s)}}{\prod_{ij}\{\lambda(c)A(i)B(j)F(c_{ij})\}^{N_{ij}(t)}} \\
&= \frac{\lambda(c)^n \left[\prod_i A(i)^{N_i(s)}\right]\left[\prod_j B(j)^{N_j(s)}\right]}{\lambda(c)^n \left[\prod_i A(i)^{N_i(t)}\right]\left[\prod_j B(j)^{N_j(t)}\right]} \prod_{r=1}^{n}\left[\frac{F(c_{i_r j_r})}{F(c_{g_r h_r})}\right] \\
&\geq \frac{\lambda(c')^n \left[\prod_i A(i)^{N_i(s')}\right]\left[\prod_j B(j)^{N_j(s')}\right]}{\lambda(c')^n \left[\prod_i A(i)^{N_i(t')}\right]\left[\prod_j B(j)^{N_j(t')}\right]} \prod_{r=1}^{n}\left[\frac{F(c'_{i'_r j'_r})}{F(c'_{g'_r h'_r})}\right] \\
&= \frac{P_{c'}(s')}{P_{c'}(t')},
\end{aligned}
\tag{4.274}
$$

which implies that $\boldsymbol{P_N}$ satisfies condition (4.124). Thus we also have

$$\langle \text{MODEL G5}^* \rangle \subseteq \langle \text{A1, A2, A4}'' \rangle,$$

and may conclude that (4.272) holds. □

Remark 4.2. Notice that the characterizations in Theorems 4.3 through 4.6 yield several implications among the axioms of Section 4.4.2 not listed in Proposition 4.10. In particular, since $\langle \text{MODEL G5} \rangle \subseteq \langle \text{MODEL G2} \rangle$, it follows from the proofs of Theorems 4.3 and 4.5 that $\langle \text{A1}, \text{A2}, \text{A3}'' \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}' \rangle$. Similarly, since $\langle \text{MODEL G5}^* \rangle \subseteq \langle \text{MODEL G2}^* \rangle$, it also follows from the proofs of Theorem 4.4 and 4.6 that $\langle \text{A1}, \text{A2}, \text{A3}'' \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}' \rangle$. While these relationships can be established directly from the axioms themselves, the proofs are essentially the same as those in Theorems 4.3 and 4.5. •

Finally we turn to the class of invariant gravity models (G6), and show that these models are characterized by uniform separation dependence (A3$''$) together with the additional property of sub-configuration dependence (A5):

Theorem 4.7 (G6 Characterization) *The class of frequency processes,* $\langle \text{MODEL G6} \rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G6} \rangle = \langle \text{A1}, \text{A2}, \text{A3}'', \text{A5} \rangle. \tag{4.275}$$

PROOF: Having established (4.246) and (4.272), it follows from an inspection of (4.5) and (4.6) that assertion (4.275) will follow if it can be shown that constancy of the function, $\lambda : C \to R_{++}$, in (4.6) is equivalent to axiom A5 [in the presence of axioms (A1, A2, A3$''$)].

(a) To do so, first consider any interaction process, $\boldsymbol{P}$ satisfying axioms (A1, A2, A3$''$, A5), and observe from (4.246) above that $\boldsymbol{N_P}$ satisfies (4.5) for some choice of functions, $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : V \to R_{++}$. Hence it follows from (4.125) and (3.69) that for all $i \in I$ and $c, c' \in C$,

$$\begin{aligned} c_i = c_i' &\Rightarrow F(c_{ij}) = F(c_{ij}'), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_{ij})/\mathrm{E}_{c'}(N_{ij}) = \lambda(c)/\lambda(c'), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_{ij}) = \mathrm{E}_{c'}(N_{ij})[\lambda(c)/\lambda(c')], \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_i) = \mathrm{E}_{c'}(N_i)[\lambda(c)/\lambda(c')] \\ &\Rightarrow \lambda(c) = \lambda(c') \end{aligned} \tag{4.276}$$

and we see that $\lambda(c) = \lambda(c')$ holds whenever $c_i = c_i'$ holds for at least one $i \in I$. This in turn implies that by choosing any distinct origins, $i, g \in I$, and any fixed spatial configuration, $c^\circ = (c_1^\circ, \ldots, c_i^\circ, \ldots, c_g^\circ, \ldots, c_L^\circ) \in C$ [with $L = |I|$], the equality,

$$\begin{aligned} \lambda(c) &= \lambda(c_1, \ldots, c_i, \ldots, c_g, \ldots, c_L) \\ &= \lambda(c_1^\circ, \ldots, c_i, \ldots, c_g^\circ, \ldots, c_L^\circ) \\ &= \lambda(c_1^\circ, \ldots, c_i^\circ, \ldots, c_g^\circ, \ldots, c_L^\circ) = \lambda(c^\circ) \end{aligned} \tag{4.277}$$

must hold identically for all $c \in C = V^{I \times J}$, and thus that λ must be constant on C. Hence, defining, $A^\circ : I \to R_{++}$, for all $i \in I$ by $A^\circ(i) =$

$\lambda(c^o)A(i)$, we see that $\mathrm{E}(N_{ij}^c) = A^o(i)B(j)F(c_{ij})$ must hold identically for all $ij \in I \times J$ and $c \in C$, so that $\boldsymbol{N_P} \in \langle \text{MODEL G6} \rangle$ by (4.6), and we must have $\langle \text{A1}, \text{A2}, \text{A3}'', \text{A5} \rangle \subseteq \langle \text{MODEL G6} \rangle$.

(b) Conversely, for any Poisson frequency process, $\boldsymbol{N} \in \langle \text{MODEL G6} \rangle$, with generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, it follows from condition (4.6) together with (3.69) that for all $c, c' \in C$ and $i \in I$,

$$\begin{aligned} c_i = c_i' &\Rightarrow F(c_{ij}) = F(c_{ij}'), \quad j \in J \\ &\Rightarrow \mathrm{E}(N_{ij}^c) = \mathrm{E}(N_{ij}^{c'}), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_{ij}) = \mathrm{E}_{c'}(N_{ij}), \quad j \in J \\ &\Rightarrow \mathrm{E}_c(N_i) = \mathrm{E}_{c'}(N_i), \end{aligned} \tag{4.278}$$

and hence that $\boldsymbol{P_N}$ must satisfy A5 as well as (A1, A2, A3$''$). Thus, we may also conclude that $\langle \text{MODEL G6} \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}'', \text{A5} \rangle$, and assertion (4.275) is established. □

Similarly, it follows by the same argument that *monotone* invariant gravity models are characterized by simply replacing (A3$''$) with uniform separation efficiency (A4$''$):

Theorem 4.8 (G6* Characterization) *The class of frequency processes,* $\langle \text{MODEL G6}^* \rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G6}^* \rangle = \langle \text{A1}, \text{A2}, \text{A4}'', \text{A5} \rangle. \tag{4.279}$$

PROOF: To establish assertion (4.279), simply observe that the arguments for $\langle \text{MODEL G6} \rangle$ above did not involve any properties of the function, F. Hence by replacing A3$''$ with A4$''$, the arguments establishing (4.275) are also seen to establish (4.279). □

(B) LOCAL BEHAVIORAL CHARACTERIZATIONS

Next we establish the local characterizations of gravity models developed in Section 2.2.3(B) of Chapter 2. To do so, let us first recall that the characterization of destination-deterrence-invariant gravity models (G3) in terms of axioms (A6,A8) required that $|J| \geq 3$, and similarly, that the characterization of origin-deterrence-invariant models (G4) in terms of axioms (A7,A9) required that $|I| \geq 3$. To see that these restrictions are essential, we now consider the following counterexample involving the destination proportionality and separability axioms (A6,A8). [A counterexample for the origin proportionality and separability axioms (A7,A9) can be constructed in a completely parallel manner, as in Smith (1984)]:

EXAMPLE 4.2. NONSUFFICIENCY OF THE PROPORTIONALITY AND SEPARATION AXIOMS

As in Example 4.1, suppose that $J = \{d, h\}$ and that $V = R_+^K$. In this context, consider the unique Poisson frequency process, $\boldsymbol{N} = \{N^c : c \in C\}$, with mean interaction frequencies, $\mathrm{E}(N_{ij}^c)$, given for all $c \in C(= V^{I \times J})$ and $ij \in I \times J$ by

$$\mathrm{E}(N_{ij}^c) = \begin{cases} \lambda \cdot [(1 + ||c_{id}||)^{-1} + ||c_{ih}||], & j = d \\ \lambda \cdot [(1 + ||c_{ih}||)^{-1} + ||c_{id}||], & j = h, \end{cases} \tag{4.280}$$

where $|| \cdot ||$ denotes the Euclidean norm, and where λ is any positive constant. Then by (4.280) together with (4.107) it follows that the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ satisfies the following identity for all $c \in C$ and $i \in I$:

$$\frac{P_c(id)}{P_c(ih)} = \frac{\mathrm{E}(N_{id}^c)}{\mathrm{E}(N_{ih}^c)} = \frac{(1 + ||c_{id}||)^{-1} + ||c_{ih}||}{(1 + ||c_{ih}||)^{-1} + ||c_{id}||}. \tag{4.281}$$

This in turn implies that for any $x, y \in V, i, g \in I$ and $c, c' \in C$ with $c_{id} = x = c'_{gd}$ and $c_{ih} = y = c'_{gh}$, we must have

$$\frac{P_c(id)}{P_c(ih)} = \frac{(1 + ||x||)^{-1} + ||y||}{(1 + ||y||)^{-1} + ||x||} = \frac{P_{c'}(gd)}{P_{c'}(gh)}, \tag{4.282}$$

which, together with the assumption that $J = \{d, h\}$, implies that $\boldsymbol{P_N}$ satisfies destination proportionality (A6). Similarly, if $c_{id} = c_{ih}$ and $c'_{id} = c'_{ih}$ for any $c, c' \in C$ and $i \in I$ then it again follows from (4.281) that $P_c(id)/P_c(ih) = 1 = P_{c'}(id)/P_{c'}(ih)$, and hence that $\boldsymbol{P_N}$ also satisfies destination separability (A8). However, it is not true that $\boldsymbol{N} \in \langle \text{MODEL G3} \rangle$. For if there were to exist a destination function, $B : J \to R_{++}$, and deterrence function, $F : V \to R_{++}$, together with origin functions, $A_c : I \to R_{++}$, for each $c \in C$, satisfying (4.3), then it would follow from (4.281) that

$$\frac{B(d)F(c_{id})}{B(h)F(c_{ih})} = \frac{\mathrm{E}(N_{id}^c)}{\mathrm{E}(N_{ih}^c)} = \frac{(1 + ||c_{id}||)^{-1} + ||c_{ih}||}{(1 + ||c_{ih}||)^{-1} + ||c_{id}||} \tag{4.283}$$

must hold identically for all $c \in C$ and $i \in I$. But by choosing any $i \in I$ and $c \in C$ with $c_{id} = c_{ih}$, we see from (4.283) that $B(d) = B(h)$, and hence that for all $x, y \in V$ the function F must satisfy the identity

$$\frac{F(x)}{F(y)} = \frac{(1 + ||x||)^{-1} + ||y||}{(1 + ||y||)^{-1} + ||x||}. \tag{4.284}$$

To see that this is not possible, choose any $x, y \in V$ with $||x|| = 1$ and $||y|| = 2$, and observe on the one hand that an evaluation of (4.284) yields

$$\frac{F(x)}{F(y)} = \frac{(1/2) + 2}{(1/3) + 1} = 15/8. \tag{4.285}$$

On the other hand, if we compare both x and y with the zero separation profile, $0 \in V(= R_+^K)$, then same calculation shows that $F(x)/F(0) = 1/4$ and $F(0)/F(y) = 9$. Hence we must also have

$$\frac{F(x)}{F(y)} = \frac{F(x)}{F(0)} \cdot \frac{F(0)}{F(y)} = 9/4 = 18/8 \tag{4.286}$$

which together with (4.285) yields a contradiction. Thus no such function can exist, and we may conclude that $\boldsymbol{N} \notin \langle \text{MODEL G3} \rangle$. •

Given this counterexample, it is of interest to develop stronger axiomatizations which characterize destination-deterrence-invariant and origin-deterrence-invariant gravity models under the more general conditions that $|J| \geq 2$ and $|I| \geq 2$. In particular, our approach here is to preserve the basic proportionality axioms (A6,A7), and to strengthen the separation axioms (A8,A9) in a manner which yields the desired characterizations. With this end in mind, we begin by observing that since $\langle \text{MODEL G3} \rangle \cup \langle \text{MODEL G4} \rangle \subseteq \langle \text{MODEL G2} \rangle$ by definition, it follows from Theorem 4.3 (i) together with (4.207) that the generators, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of frequency processes, $\boldsymbol{N} \in \langle \text{MODEL G3} \rangle \cup \langle \text{MODEL G4} \rangle$, must always satisfy the relative separation independence axiom (A3′). In particular, if for any $i, g \in I$ and $j, h \in J$ we let $s = (ij, gh) = s', t = (ih, gj)$, and $t' = (gj, ih)$, so that $A(s) = A(s') = A(t) = A(t')$, then it follows from A3′ and (4.109) that

$$\begin{aligned}
&(c_{ij}, c_{gh}, c_{ih}, c_{gj}) = (c'_{ij}, c'_{gh}, c'_{gj}, c'_{ih}) \\
&\Rightarrow (c_s = c'_{s'}, c_t = c'_{t'}) \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} \\
&\Rightarrow \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_{c'}(ij)P_{c'}(gh)}{P_{c'}(gj)P_{c'}(ih)} \\
&= \frac{P_{c'}(ij)P_{c'}(gh)}{P_{c'}(ih)P_{c'}(gj)}
\end{aligned} \tag{4.287}$$

must hold for all $P_c \in \boldsymbol{P_N}$. Hence if we write the equality, $(c_{ij}, c_{gh}, c_{ih}, c_{gj}) = (c'_{ij}, c'_{gh}, c'_{gj}, c'_{ih})$ in matrix form, then it follows that $\boldsymbol{P_N}$ must satisfy the following local version of the relative separation dependence axiom (A3′):

A3$^{\#}$. (Local Relative Separation Dependence) *For all origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$\begin{bmatrix} c_{ij} & c_{ih} \\ c_{gj} & c_{gh} \end{bmatrix} = \begin{bmatrix} c'_{ij} & c'_{gj} \\ c'_{ih} & c'_{gh} \end{bmatrix} \Rightarrow \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_{c'}(ij)P_{c'}(gh)}{P_{c'}(ih)P_{c'}(gj)}. \tag{4.288}$$

This axiom [which is seen to be equivalent to property P5 in Smith (1984)] may be viewed as a type of 'cross-product invariance' condition, similar to CPI-condition discussed in Section 2.2.3(C) of Chapter 2. Given this local axiom, we are now ready to establish the desired local characterizations of destination-deterrence-invariant and origin-deterrence-invariant gravity models. Turning first to destination-deterrence-invariant models we have:

Theorem 4.9 (G3 Characterization)
(i) *The class of frequency processes,* $\langle$MODEL G3$\rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G3} \rangle = \langle \text{A1}, \text{A2}, \text{A3}^{\#}, \text{A6} \rangle. \tag{4.289}$$

(ii) *If in addition it is true that* $|J| \geq 3$, *then*

$$\langle \text{MODEL G3} \rangle = \langle \text{A1}, \text{A2}, \text{A6}, \text{A8} \rangle. \tag{4.290}$$

PROOF: As in the proof of Theorem 4.3, it is convenient to begin by proving part (ii), and then employ this result to prove the more general assertion in part (i).

(ii). (a) To establish that $\langle \text{A1}, \text{A2}, \text{A6}, \text{A8} \rangle \subseteq \langle \text{MODEL G3} \rangle$ when $|J| \geq 3$, consider any independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying $(\text{A6}, \text{A8})$, and for each pair of separation profiles, $x, y \in V$, and distinct pair of destinations, $j, h \in J$, let

$$\phi_{jh}(x, y) = \{P_c(ij)/P_c(ih) : i \in I, c \in C, (c_{ij}, c_{ih}) = (x, y)\}. \tag{4.291}$$

Then as in (4.212) it follows [from $C = V^{I \times J}$] that $\phi_{jh}(x, y) \neq \emptyset$. Moreover, for any $g \in I$ and $c' \in C$ with $(c'_{gj}, c'_{gh}) = (x, y)$, it follows from A6 that $P_c(ij)/P_c(ih) = P_{c'}(gj)/P_{c'}(gh)$, and hence that $|\phi_{jh}(x, y)| = 1$. Thus for each distinct pair, $j, h \in J$, expression (4.291) yields a well defined function, $\phi_{jh} : V^2 \rightarrow R_{++}$, satisfying the identity

$$\phi_{jh}(c_{ij}, c_{ih}) = \frac{P_c(ij)}{P_c(ih)}, \tag{4.292}$$

for all $i \in I$ and $c \in C$, as well as the identity

$$\phi_{jh}(x, y)\phi_{hj}(y, x) = \frac{P_c(ij)}{P_c(ih)} \cdot \frac{P_c(ih)}{P_c(ij)} = 1, \tag{4.293}$$

for all $x, y \in V$. In addition, since $|J| \geq 3$, it follows that for any choice of distinct destinations, $j, h, d \in J$, and separation profiles, $x, y, z \in V$, we choose any configuration, $c \in C$, satisfying the condition that $(c_{ij}, c_{ih}, c_{id}) = (x, y, z)$ for some $i \in I$, then (4.292) implies that

$$\phi_{jh}(x, y) = \frac{P_c(ij)}{P_c(ih)} = \frac{P_c(ij)/P_c(id)}{P_c(ih)/P_c(id)} = \frac{\phi_{jd}(x, z)}{\phi_{hd}(y, z)}. \tag{4.294}$$

Finally, since $\boldsymbol{P}$ also satisfies A8, it follows that for any $c, c' \in C$ with $c_{ij} = x = c_{ih}$ and $c'_{ij} = y = c'_{ih}$,

$$\phi_{jh}(x, x) = \frac{P_c(ij)}{P_c(ih)} = \frac{P_{c'}(ij)}{P_{c'}(ih)} = \phi_{jh}(y, y). \tag{4.295}$$

With these observations, we next show that for all profiles, $x, y \in V$, and destinations, $j, h, a, b \in J$, with $j \neq h$ and $a \neq b$,

$$\phi_{jh}(x,y)\phi_{hj}(y,y) = \phi_{ab}(x,y)\phi_{ba}(y,y). \tag{4.296}$$

To do so, observe first that for any distinct destinations, $j, h, a \in J$, it follows from (4.294) and (4.293) that

$$\begin{aligned} \phi_{hj}(y,y) &= \frac{\phi_{ha}(y,y)}{\phi_{ja}(y,y)} = \frac{\phi_{aj}(y,y)}{\phi_{ah}(y,y)} \\ &\Rightarrow \phi_{jh}(x,y)\phi_{hj}(y,y) = \frac{\phi_{jh}(x,y)}{\phi_{ah}(y,y)} \cdot \phi_{aj}(y,y) \\ &\Rightarrow \phi_{jh}(x,y)\phi_{hj}(y,y) = \phi_{ja}(x,y)\phi_{aj}(y,y), \end{aligned} \tag{4.297}$$

which together with (4.295) also shows that

$$\begin{aligned} \phi_{hj}(y,x)\phi_{jh}(y,y) &= \phi_{hj}(y,x)\phi_{jh}(x,x) \\ &= \phi_{ha}(y,x)\phi_{ah}(x,x) = \phi_{ha}(y,x)\phi_{ah}(y,y) \\ &\Rightarrow \frac{1}{\phi_{jh}(x,y)} \cdot \frac{1}{\phi_{hj}(y,y)} = \frac{1}{\phi_{ah}(x,y)} \cdot \frac{1}{\phi_{ha}(y,y)} \\ &\Rightarrow \phi_{jh}(x,y)\phi_{hj}(y,y) = \phi_{ah}(x,y)\phi_{ha}(y,y). \end{aligned} \tag{4.298}$$

Hence, by combining (4.297) and (4.298) we see that for all distinct destinations $j, h, a \in J$,

$$\begin{aligned} \phi_{jh}(x,y)\phi_{hj}(y,y) &= \phi_{ja}(x,y)\phi_{aj}(y,y) \\ &= \phi_{ha}(x,y)\phi_{ah}(y,y) = \phi_{hj}(x,y)\phi_{jh}(y,y), \end{aligned} \tag{4.299}$$

so that the value $\phi_{jh}(x,y)\phi_{hj}(y,y)$ is independent of the ordering of j and h. Thus (4.296) must hold whenever the *unordered* sets, $\{j,h\}$ and $\{a,b\}$, are the same, and it remains only to consider cases in which $\{j,h\} \neq \{a,b\}$. But by symmetry, we may then assume without loss of generality that $b \notin \{j,h\}$, so that (j,h,b) are distinct. Hence for this case, it follows from (4.297) and (4.298) that if $a \neq j$ [so that (j,a,b) are also distinct] then

$$\phi_{jh}(x,y)\phi_{hj}(y,y) = \phi_{jb}(x,y)\phi_{bj}(y,y) = \phi_{ab}(x,y)\phi_{ba}(y,y). \tag{4.300}$$

Finally, by observing that the last equality in (4.300) holds trivially if $a = j$, we may conclude that (4.296) holds in all cases. Given this identity, it then follows that for any fixed choice of distinct destinations, $a, b \in J$, and fixed separation profile, $\sigma \in V$, the relation

$$F(v) = \phi_{ab}(v,\sigma)\phi_{ba}(\sigma,\sigma), \quad v \in V, \tag{4.301}$$

yields a well-defined function, $F : V \to R_{++}$, which is independent of the choice of $a, b \in J$. Thus by employing (4.292), (4.293), (4.294), (4.296), and (4.301), we see that for all $c \in C, i \in I$, and distinct destination pairs, $j, h \in J - \{a\}$,

$$\begin{aligned}\frac{P_c(ij)}{P_c(ia)} &= \phi_{ja}(c_{ij}, c_{ia}) = \frac{\phi_{jh}(c_{ij}, \sigma)}{\phi_{ah}(c_{ia}, \sigma)} \\ &= \frac{\phi_{jh}(c_{ij}, \sigma)}{\phi_{ah}(c_{ia}, \sigma)} \cdot \frac{\phi_{hj}(\sigma, \sigma)\phi_{jh}(\sigma, \sigma)}{\phi_{ah}(\sigma, \sigma)\phi_{ha}(\sigma, \sigma)} \\ &= \frac{\phi_{jh}(\sigma, \sigma)}{\phi_{ah}(\sigma, \sigma)} \cdot \frac{\phi_{jh}(c_{ij}, \sigma)\phi_{hj}(\sigma, \sigma)}{\phi_{ah}(c_{ia}, \sigma)\phi_{ha}(\sigma, \sigma)} \\ &= \frac{\phi_{jh}(\sigma, \sigma)}{\phi_{ah}(\sigma, \sigma)} \cdot \frac{\phi_{ab}(c_{ij}, \sigma)\phi_{ba}(\sigma, \sigma)}{\phi_{ab}(c_{ia}, \sigma)\phi_{ba}(\sigma, \sigma)} \\ &= \phi_{ja}(\sigma, \sigma)\frac{F(c_{ij})}{F(c_{ia})},\end{aligned} \tag{4.302}$$

so that by letting the function, $B: J \rightarrow R_{++}$, be defined for all $j \in J$ by

$$B(j) = \begin{cases} \phi_{ja}(\sigma, \sigma), & j \neq a \\ 1, & j = a, \end{cases} \tag{4.303}$$

we obtain the following identity for all $c \in C, i \in I$, and $j \in J - \{a\}$,

$$\frac{P_c(ij)}{P_c(ia)} = \frac{B(j)F(c_{ij})}{B(a)F(c_{ia})}. \tag{4.304}$$

Hence to obtain the desired representation of mean interaction frequencies for the frequency process, $\boldsymbol{N_P} = \{N^c : c \in C\}$, generated by $\boldsymbol{P}$, let the function, $A_c: I \rightarrow R_{++}$, be defined for each $c \in C$ by

$$A_c(i) = \frac{\mathrm{E}(N_{ia}^c)}{B(a)F(c_{ia})}, \quad i \in I, \tag{4.305}$$

and observe from (4.304) and (4.305) together with (4.107) that

$$\begin{aligned}&\frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ia}^c)} = \frac{P_c(ij)}{P_c(ia)} = \frac{B(j)F(c_{ij})}{B(a)F(c_{ia})} \\ &\Rightarrow \mathrm{E}(N_{ij}^c) = \frac{\mathrm{E}(N_{ia}^c)}{B(a)F(c_{ia})} \cdot B(j)F(c_{ij}) = A_c(i)B(j)F(c_{ij})\end{aligned} \tag{4.306}$$

for all $c \in C$ and $ij \in I \times J$. Thus $\boldsymbol{N_P} \in \langle\text{MODEL G3}\rangle$, and it follows that $\langle \text{A1, A2, A6, A8} \rangle \subseteq \langle \text{MODEL G3} \rangle$.

(b) To establish the converse, we may simply observe from (4.309) below that $\langle \text{A1, A2, A3}^{\#}\text{, A6} \rangle \subseteq \langle \text{A1, A2, A6, A8} \rangle$ and hence that [as in part ii(b) of Theorem 4.3] it suffices to show that $\langle \text{MODEL G3} \rangle \subseteq \langle \text{A1, A2, A3}^{\#}\text{, A6} \rangle$. With this in mind, we turn to the general characterization of $\langle \text{MODEL G3} \rangle$ in part (i).

(i). (a) To establish that $\langle \text{A1, A2, A3}^{\#}, \text{A6} \rangle \subseteq \langle \text{MODEL G3} \rangle$, consider any independent interaction process $\boldsymbol{P}$ satisfying (A3$^{\#}$, A6), and observe that since the argument preceding (4.292) did not require the condition, $|J| \geq 3$, it again follows from A6 that for each pair of distinct destinations, $j, h \in J$, there is a well defined function, $\phi_{jh} : V^2 \to R_{++}$, satisfying the identity in (4.292). Hence if for any profiles, $x, y, z \in V$, and distinct $i, g \in I$ and $j, h \in J$ we choose $c, c' \in C$ with $c_{ij} = x = c'_{ij}, c_{ih} = y = c'_{gj}$, and $c_{gj} = c_{gh} = c'_{ih} = c'_{gh} = z$, then by A3$^{\#}$ and (4.292) it follows that

$$
\begin{aligned}
&\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_{c'}(ij)P_{c'}(gh)}{P_{c'}(ih)P_{c'}(gj)} \\
&\Rightarrow \frac{P_c(ij)/P_c(ih)}{P_c(gj)/P_c(gh)} = \frac{P_{c'}(ij)/P_{c'}(ih)}{P_{c'}(gj)/P_{c'}(gh)} \\
&\Rightarrow \frac{\phi_{jh}(x,y)}{\phi_{jh}(z,z)} = \frac{\phi_{jh}(x,z)}{\phi_{jh}(y,z)} \\
&\Rightarrow \phi_{jh}(x,y) = \phi_{jh}(z,z)\frac{\phi_{jh}(x,z)}{\phi_{jh}(y,z)}.
\end{aligned}
\tag{4.307}
$$

In particular, if $x = y$ then we obtain the identity, $\phi_{jh}(x,x) = \phi_{jh}(z,z)$ for all $x, z \in V$, which in turn implies that for all $i \in I$ and $c, c' \in C$ with $c_{ij} = x = c_{ih}$ and $c'_{ij} = z = c'_{ih}$ we must have

$$
\frac{P_c(ij)}{P_c(ih)} = \phi_{jh}(x,x) = \phi_{jh}(z,z) = \frac{P_{c'}(ij)}{P_{c'}(ih)}. \tag{4.308}
$$

Thus we see that $A3^{\#}$ and A6 together imply A8, and may conclude that

$$
\langle \text{A1, A2, A3}^{\#}, \text{A6} \rangle \subseteq \langle \text{A1, A2, A6, A8} \rangle. \tag{4.309}
$$

Hence if $|J| \geq 3$ then the desired result follows from the proof of part ii(a) above, and it suffices to assume that $|J| = 2$, say with $J = \{a, b\}$. But if for any fixed profile, $\sigma \in V$, we define the function, $F : V \to R_{++}$, for all $v \in V$ by

$$
F(v) = \phi_{ab}(v, \sigma) \tag{4.310}
$$

and define the function, $B : J \to R_{++}$, by

$$
B(j) = \begin{cases} \phi_{ab}(\sigma,\sigma), & j = a \\ 1, & j = b, \end{cases} \tag{4.311}
$$

then it follows at once from (4.307), (4.310) and (4.311) that for all $c \in C$ and $i \in I$,

$$
\frac{P_c(ia)}{P_c(ib)} = \phi_{ab}(c_{ia}, c_{ib}) = \phi_{ab}(\sigma,\sigma)\frac{\phi_{ab}(c_{ia},\sigma)}{\phi_{ab}(c_{ib},\sigma)} = \frac{B(a)F(c_{ia})}{B(b)F(c_{ib})}, \tag{4.312}
$$

and the desired result again follows from the same argument as in (4.305) and (4.306) above.

(b) Finally, to establish that $\langle\text{MODEL G3}\rangle \subseteq \langle\text{A1, A2, A3}^{\#}, \text{A6}\rangle$, we may recall that by construction $\langle\text{A3}'\rangle \subseteq \langle\text{A3}^{\#}\rangle$, and hence that

$$\langle\text{MODEL G3}\rangle \subseteq \langle\text{MODEL G2}\rangle \subseteq \langle\text{A1, A2, A3}'\rangle \subseteq \langle\text{A1, A2, A3}^{\#}\rangle.$$

Moreover, if $\boldsymbol{N} \in \langle\text{MODEL G3}\rangle$ then by (4.3) it follows that for all $c, c' \in C, i, g \in I$ and $j, h \in J$ with $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$, the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ satisfies,

$$\frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)}. \tag{4.313}$$

Hence $\langle\text{MODEL G3}\rangle \subseteq \langle\text{A6}\rangle$ as well, and thus the desired result is established. □

Remark 4.3. An alternative proof of part ii(a) of Theorem 4.9 is given in Proposition 4.15 below [see also Theorems 3.3 and 3.4 in Smith (1984)]. The present proof is somewhat more general, in that it makes no appeal to the finiteness of $|J|$ (or $|I|$). •

As a parallel to Theorem 4.9 we have the following characterizations of monotone destination-deterrence-invariant gravity models:

Theorem 4.10 (G3* Characterization)
(i) *The class of frequency processes,* $\langle\text{MODEL G3}^*\rangle$, *is characterized by the following properties of its generators:*

$$\langle\text{MODEL G3}^*\rangle = \langle\text{A1, A2, A3}^{\#}, \text{A10}\rangle. \tag{4.314}$$

(ii) *If in addition it is true that* $|J| \geq 3$, *then*

$$\langle\text{MODEL G3}^*\rangle = \langle\text{A1, A2, A8, A10}\rangle. \tag{4.315}$$

PROOF: **(i).** (a) To establish that $\langle\text{A1, A2, A3}^{\#}, \text{A10}\rangle \subseteq \langle\text{MODEL G3}^*\rangle$, observe first from (4.141) that $\langle\text{A10}\rangle \subseteq \langle\text{A6}\rangle$ and hence that for each $\boldsymbol{P}$ satisfying $(\text{A1, A2, A3}^{\#}, \text{A10})$ there exist functions, $F : V \rightarrow R_{++}$ and $B : J \rightarrow R_{++}$, together with functions, $A_c : I \rightarrow R_{++}$, for each $c \in C$ such that (4.306) holds for all $c \in C$ and $ij \in I \times J$. Hence it suffices to show that F is nonincreasing. But if for any $x, y \in V$ with $x \geq y$ we choose distinct origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C (= V^{I \times J})$, with $c_{ij} = y$ and $c_{gj} = c'_{gh} = c'_{ih} = x$, then

$$\begin{aligned} x \geq y &\Rightarrow (c_{ij} \leq c'_{gj}, c_{ih} = c'_{gh}) \\ &\Rightarrow \frac{B(j)F(y)}{B(h)F(x)} = \frac{P_c(ij)}{P_c(ih)} \geq \frac{P_{c'}(gj)}{P_{c'}(gh)} = \frac{B(j)F(x)}{B(h)F(x)} \\ &\Rightarrow F(x) \leq F(y). \end{aligned} \tag{4.316}$$

(b) Conversely, to see that $\langle$MODEL G3*$\rangle \subseteq \langle$A1, A2, A3$^{\#}$, A10$\rangle$, observe first that since $\langle$MODEL G3*$\rangle \subseteq \langle$MODEL G3$\rangle \subseteq \langle$A1, A2, A3$^{\#}\rangle$, it remains only to show that $\langle$MODEL G3*$\rangle \subseteq \langle$A10$\rangle$. But for any $\boldsymbol{N} \in \langle$MODEL G3*$\rangle$, it follows that (4.3) must hold for some destination function, $B: J \rightarrow R_{++}$, and nonincreasing deterrence function, $F: V \rightarrow R_{++}$. Hence for the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ we see from (4.107) together with the monotonicity of F that for all $i \in I, j, h \in J$, and $c, c' \in C$,

$$\begin{aligned}(c_{ij} \leq c'_{gj}, c_{ih} \geq c'_{gh}) &\Rightarrow [F(c_{ij}) \geq F(c'_{gj}), F(c_{ih}) \leq F(c'_{gh})] \\ &\Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} \geq \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)},\end{aligned} \tag{4.317}$$

and hence that $\boldsymbol{P_N}$ satisfies A10.

(ii). To establish part (ii), simply observe that if $|J| \geq 3$ and if axiom A3$^{\#}$ is replaced by A8, then exactly the same arguments in part (i) now hold for part (ii). □

These characterizations of destination-deterrence-invariant gravity models in turn yield the following dual characterizations of origin-deterrence-invariant gravity models:

Theorem 4.11 (G4 Characterization)
(i) *The class of frequency processes,* $\langle$MODEL G4$\rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G4} \rangle = \langle \text{A1, A2, A3}^{\#}, \text{A7} \rangle. \tag{4.318}$$

(ii) *If in addition it is true that* $|I| \geq 3$, *then*

$$\langle \text{MODEL G4} \rangle = \langle \text{A1, A2, A7, A9} \rangle. \tag{4.319}$$

PROOF: The following proof closely parallels the proof of Theorem 4.9 [and indeed amounts simply to interchanging the roles of origins and destinations]. However, to facilitate the proof of Theorem 4.13 below, it is convenient to sketch main elements of the proof, and in particular, to give a formal definition of the functions for origin pairs which parallel those for destination pairs in (4.291) above.

(ii). (a) To establish that $\langle$A1, A2, A7, A9$\rangle \subseteq \langle$MODEL G4$\rangle$ when $|I| \geq 3$, consider any independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying (A7, A9), and for each pair of separation profiles, $x, y \in V$, and distinct pair of origins, $i, g \in I$, let

$$\psi_{ig}(x, y) = \{P_c(ij)/P_c(gj) : j \in J, c \in C, (c_{ij}, c_{gj}) = (x, y)\}. \tag{4.320}$$

Then as in (4.212) it follows that $\psi_{ji}(x, y) \neq \emptyset$. Moreover, for any $h \in J$ and $c' \in C$ with $(c'_{ih}, c'_{gh}) = (x, y)$, it follows from A7 that $P_c(ij)/P_c(gj) =$

$P_{c'}(ih)/P_{c'}(gh)$, and hence that $|\psi_{ig}(x,y)| = 1$. Thus for each distinct pair, $i, g \in I$, expression (4.320) yields a well defined function, $\psi_{ig}: V^2 \rightarrow R_{++}$, satisfying the identity

$$\psi_{ig}(c_{ij}, c_{gj}) = \frac{P_c(ij)}{P_c(gj)}, \tag{4.321}$$

for all $j \in J$ and $c \in C$, as well as the identity

$$\psi_{ig}(x,y)\psi_{gi}(y,x) = \frac{P_c(ij)}{P_c(gj)} \cdot \frac{P_c(gj)}{P_c(ij)} = 1, \tag{4.322}$$

for all $x, y \in V$ and distinct $i, g \in I$. In addition, since $|I| \geq 3$, it follows that for any choice of distinct origins, $i, g, a \in I$, and separation profiles, $x, y, z \in V$, we choose any configuration, $c \in C$, satisfying the condition that $(c_{ij}, c_{gj}, c_{aj}) = (x, y, z)$ for some $j \in J$, then (4.321) implies that

$$\psi_{ig}(x,y) = \frac{P_c(ij)}{P_c(gj)} = \frac{P_c(ij)/P_c(aj)}{P_c(gj)/P_c(aj)} = \frac{\psi_{ia}(x,z)}{\psi_{ga}(y,z)}. \tag{4.323}$$

Finally, since $\boldsymbol{P}$ also satisfies A9, it follows that for any $c, c' \in C$ with $c_{ij} = x = c_{gj}$ and $c'_{ij} = y = c'_{gj}$,

$$\psi_{ig}(x,x) = \frac{P_c(ij)}{P_c(gj)} = \frac{P_{c'}(ij)}{P_{c'}(gj)} = \psi_{ig}(y,y). \tag{4.324}$$

Next, in a manner paralleling (4.296) through (4.300), these relations are seen to imply that for all separation profiles, $x, y \in V$, and origins, $i, g, a, b \in I$, with $i \neq g$ and $a \neq b$,

$$\psi_{ig}(x,y)\psi_{gi}(y,y) = \psi_{ab}(x,y)\psi_{ba}(y,y). \tag{4.325}$$

Hence if for any fixed $\sigma \in V$ and $a, b \in I$ we now let the functions, $F: V \rightarrow R_{++}$ and $A: I \rightarrow R_{++}$, be defined respectively for all $v \in V$ and $i \in I$ by

$$F(v) = \psi_{ab}(v,\sigma)\psi_{ba}(\sigma,\sigma), \tag{4.326}$$

$$A(i) = \begin{cases} \psi_{ia}(\sigma,\sigma), & i \neq a \\ 1, & i = a, \end{cases} \tag{4.327}$$

then as a parallel to (4.302) we now see that for all $c \in C, j \in J$, and distinct origin pairs, $i, g \in I - \{a\}$,

$$\begin{aligned}\frac{P_c(ij)}{P_c(aj)} &= \psi_{ia}(c_{ij}, c_{aj}) = \frac{\psi_{ig}(c_{ij}, \sigma)}{\psi_{ag}(c_{aj}, \sigma)} \\ &= \frac{\psi_{ig}(c_{ij}, \sigma)}{\psi_{ag}(c_{aj}, \sigma)} \cdot \frac{\psi_{gi}(\sigma, \sigma)\psi_{ig}(\sigma, \sigma)}{\psi_{ag}(\sigma, \sigma)\psi_{ga}(\sigma, \sigma)} \\ &= \frac{\psi_{ig}(\sigma, \sigma)}{\psi_{ag}(\sigma, \sigma)} \cdot \frac{\psi_{ig}(c_{ij}, \sigma)\psi_{gi}(\sigma, \sigma)}{\psi_{ag}(c_{aj}, \sigma)\psi_{ga}(\sigma, \sigma)} \\ &= \psi_{ia}(\sigma, \sigma) \cdot \frac{\psi_{ab}(c_{ij}, \sigma)\psi_{ba}(\sigma, \sigma)}{\psi_{ab}(c_{aj}, \sigma)\psi_{ba}(\sigma, \sigma)} \\ &= \frac{A(i)F(c_{ij})}{A(a)F(c_{aj})},\end{aligned} \tag{4.328}$$

so that the desired representation follows by setting

$$B_c(j) = \frac{\mathrm{E}(N_{aj}^c)}{A(a)F(c_{aj})} \tag{4.329}$$

and employing (4.328) together with (4.107), as in (4.306) to obtain

$$\mathrm{E}(N_{ij}^c) = A(i)B_c(j)F(c_{ij}), \text{ for all } c \in C \text{ and } ij \in I \times J.$$

(b) As in Theorem 4.9 above, part ii(b) follows from part i(b) together with (4.331) below.

(i). (a) To establish that $\langle \mathrm{A1, A2, A3^{\#}, A7} \rangle \subseteq \langle \mathrm{MODEL\ G4} \rangle$, consider any independent interaction process $\boldsymbol{P}$ satisfying $(\mathrm{A3^{\#}, A7})$, and observe that since the argument following (4.320) did not require the condition, $|I| \geq 3$, it again follows from A7 that for each pair of distinct origins, $i, g \in I$, there is a well defined function, $\psi_{ig} : V^2 \to R_{++}$, satisfying the identity in (4.321). Hence if for any profiles, $x, y, z \in V$, and distinct $i, g \in I$ and $j, h \in J$ we choose $c, c' \in C$ with $c_{ij} = x = c'_{ij}, c_{gj} = y = c'_{ih}$, and $c_{ih} = c_{gh} = c'_{gj} = c'_{gh} = z$, then by $\mathrm{A3^{\#}}$ and (4.321) it follows that

$$\begin{aligned}&\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{P_{c'}(ij)P_{c'}(gh)}{P_{c'}(ih)P_{c'}(gj)} \\ &\Rightarrow \frac{P_c(ij)/P_c(gj)}{P_c(ih)/P_c(gh)} = \frac{P_{c'}(ij)/P_{c'}(gj)}{P_{c'}(ih)/P_{c'}(gh)} \\ &\Rightarrow \frac{\psi_{ig}(x, y)}{\psi_{ig}(z, z)} = \frac{\psi_{ig}(x, z)}{\psi_{ig}(y, z)} \\ &\Rightarrow \psi_{ig}(x, y) = \psi_{ig}(z, z)\frac{\psi_{ig}(x, z)}{\psi_{ig}(y, z)},\end{aligned} \tag{4.330}$$

so that by setting $x = y$ we may now conclude from the argument in (4.308) that

$$\langle \mathrm{A1, A2, A3^{\#}, A7} \rangle \subseteq \langle \mathrm{A1, A2, A7, A9} \rangle. \tag{4.331}$$

Hence if $|I| \geq 3$ then the desired result follows from the proof of part ii(a) above, and it suffices to assume that $|I| = 2$, say with $I = \{a, b\}$. But if for any fixed profile, $\sigma \in V$, we define the functions, $F: V \to R_{++}$ and $A: I \to R_{++}$, for all $v \in V$ and $i \in I$ by

$$F(v) = \psi_{ab}(v, \sigma), \tag{4.332}$$

$$A(i) = \begin{cases} \psi_{ab}(\sigma, \sigma), & i = a \\ 1, & i = b, \end{cases} \tag{4.333}$$

then it follows at once from (4.330), (4.332) and (4.333) that for all $c \in C$ and $j \in J$,

$$\frac{P_c(aj)}{P_c(bj)} = \psi_{ab}(c_{aj}, c_{bj}) = \psi_{ab}(\sigma, \sigma)\frac{\psi_{ab}(c_{aj}, \sigma)}{\psi_{ab}(c_{bj}, \sigma)} = \frac{A(a)F(c_{aj})}{A(b)F(c_{bj})}, \tag{4.334}$$

and the desired result again follows from the construction in (4.329).

(b) Finally, to establish that $\langle \text{MODEL G4} \rangle \subseteq \langle \text{A1, A2, A3}^{\#}, \text{A7} \rangle$, we again recall that by construction $\langle \text{A3}' \rangle \subseteq \langle \text{A3}^{\#} \rangle$, and hence that

$$\langle \text{MODEL G4} \rangle \subseteq \langle \text{MODEL G2} \rangle \subseteq \langle \text{A1, A2, A3}' \rangle \subseteq \langle \text{A1, A2, A3}^{\#} \rangle.$$

Moreover, if $\boldsymbol{N} \in \langle \text{MODEL G4} \rangle$ then by (4.4) it follows that for all $c, c' \in C, i, g \in I$ and $j, h \in J$ with $c_{ij} = c'_{ih}$ and $c_{gj} = c'_{gh}$, the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ satisfies,

$$\frac{P_c(ij)}{P_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} = \frac{A(i)F(c'_{ih})}{A(g)F(c'_{gh})} = \frac{P_{c'}(ih)}{P_{c'}(gh)}. \tag{4.335}$$

Hence $\langle \text{MODEL G4} \rangle \subseteq \langle \text{A7} \rangle$ as well, and the desired result is established. $\square$

Finally, we have the following dual to Theorem 4.10 for monotone origin-deterrence-invariant gravity models:

Theorem 4.12 (G4* Characterization)
(i) *The class of frequency processes,* $\langle \text{MODEL G4}^* \rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL G4}^* \rangle = \langle \text{A1, A2, A3}^{\#}, \text{A11} \rangle. \tag{4.336}$$

(ii) *If in addition it is true that* $|I| \geq 3$, *then*

$$\langle \text{MODEL G4}^* \rangle = \langle \text{A1, A2, , A9, A11} \rangle. \tag{4.337}$$

PROOF: **(i).** (a) To establish that $\langle A1, A2, A3^{\#}, A11\rangle \subseteq \langle \text{MODEL G4}^*\rangle$, observe again from (4.141) that $\langle A11\rangle \subseteq \langle A7\rangle$ and hence from Theorem 4.11(i) that for each $\boldsymbol{P}$ satisfying $(A1, A2, A3^{\#}, A11)$ there exist functions, $F : V \to R_{++}$ and $A : I \to R_{++}$, together with functions, $B_c : I \to R_{++}$, for each $c \in C$ such that (4.4) holds for all $c \in C$ and $ij \in I \times J$. Hence it suffices to show that F is nonincreasing. But if for any $x, y \in V$ with $x \geq y$ we choose distinct origins, $i, g \in I$, destinations, $j, h \in J$, and separation configurations, $c, c' \in C(= V^{I\times J})$, with $c_{ij} = y$ and $c_{gj} = c'_{gh} = c'_{ih} = x$, then [as a parallel to (4.316)] we now have

$$\begin{aligned} x \geq y &\Rightarrow (c_{ij} \leq c'_{ih}, c_{gj} = c'_{gh}) \\ &\Rightarrow \frac{A(i)F(y)}{A(g)F(x)} = \frac{P_c(ij)}{P_c(gj)} \geq \frac{P_{c'}(ih)}{P_{c'}(gh)} = \frac{A(i)F(x)}{A(g)F(x)} \\ &\Rightarrow F(x) \leq F(y). \end{aligned} \tag{4.338}$$

(b) Conversely, to see that $\langle \text{MODEL G4}^*\rangle \subseteq \langle A1, A2, A3^{\#}, A11\rangle$, observe first that since $\langle \text{MODEL G4}^*\rangle \subseteq \langle \text{MODEL G4}\rangle \subseteq \langle A1, A2, A3^{\#}\rangle$, it remains only to show that $\langle \text{MODEL G4}^*\rangle \subseteq \langle A11\rangle$. But for any $\boldsymbol{N} \in \langle \text{MODEL G4}^*\rangle$, it follows that (4.4) must hold for some origin function, $A : I \to R_{++}$, and nonincreasing deterrence function, $F : V \to R_{++}$. Hence for the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ we again see from (4.107) together with the monotonicity of F that for all $i, g \in I, j, h \in J$, and $c, c' \in C$,

$$\begin{aligned} (c_{ij} \leq c'_{ih}, c_{gj} \geq c'_{gh}) &\Rightarrow [F(c_{ij}) \geq F(c'_{ih}), F(c_{gj}) \leq F(c'_{gh})] \\ &\Rightarrow \frac{P_c(ij)}{P_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} \geq \frac{A(i)F(c'_{ih})}{A(g)F(c'_{gh})} = \frac{P_{c'}(ih)}{P_{c'}(gh)}, \end{aligned} \tag{4.339}$$

and hence that $\boldsymbol{P_N}$ satisfies A11.

(ii). To establish part (ii), observe that if $|I| \geq 3$ and if axiom $A3^{\#}$ is now replaced by A9, then exactly the same arguments in part (i) now hold for part (ii). □

Before proceeding to a more specialized analysis of exponential gravity models, it is of interest to observe that Models G5 and G5* above can also be characterized in terms of local properties of interaction behavior. In particular, we now have the following local characterization of Model G5 in terms of the biproportionality axioms (A6,A7):

Theorem 4.13 (G5 Characterization) *The class of frequency processes, $\langle \text{MODEL G5}\rangle$, is characterized by the following local properties of its generators:*

$$\langle \text{MODEL G5}\rangle = \langle A1, A2, A6, A7\rangle. \tag{4.340}$$

PROOF: (a) To establish that $\langle \text{A1, A2, A6, A7} \rangle \subseteq \langle \text{MODEL G5} \rangle$, consider any independent interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfying (A6,A7). Then for any pair of distinct destinations, $j, h \in J$, it follows from A6 together with the argument following (4.291) that there exists a function, $\phi_{jh} : V^2 \to R_{++}$, such that for all $c \in C$ and $i \in I$,

$$\frac{P_c(ij)}{P_c(ih)} = \phi_{jh}(c_{ij}, c_{ih}). \tag{4.341}$$

Similarly, for each pair of distinct origins, $i, g \in I$, it follows from A7 together with the argument following (4.320) there exists a function, $\psi_{ig} : V^2 \to R_{++}$, such that for all $c \in C$ and $j \in J$,

$$\frac{P_c(ij)}{P_c(gj)} = \psi_{ig}(c_{ij}, c_{gj}). \tag{4.342}$$

Next, observe that for any distinct origins, $i, g \in I$, distinct destinations, $j, h \in J$, and quadruple of separation profiles, $(x, y, z, w) \in V^4$, there always exists some configuration, $c \in C(= V^{I \times J})$, with $c_{ij} = x, c_{gj} = y, c_{ih} = z$, and $c_{gh} = w$. Hence from (4.341) and (4.342) it follows that

$$\frac{\psi_{ig}(x, y)}{\psi_{ig}(z, w)} = \frac{P_c(ij)P_c(gh)}{P_c(gj)P_c(ih)} = \frac{\phi_{jh}(x, z)}{\phi_{jh}(y, w)} \tag{4.343}$$

holds identically for all distinct origins, $i, g \in I$, destinations, $j, h \in J$, and (possibly nondistinct) separation profiles, $x, y, z, w \in V$. But (4.343) implies in particular that if we choose any *fixed* pair of distinct origins, $a, b \in I$, then

$$\frac{\psi_{ig}(x, y)}{\psi_{ig}(z, w)} = \frac{\phi_{jh}(x, z)}{\phi_{jh}(y, w)} = \frac{\psi_{ab}(x, y)}{\psi_{ab}(z, w)} \tag{4.344}$$

must hold identically for all choices of distinct origins, $i, g \in I$, and all profiles, $x, y, z, w \in V$. In addition, (4.343) also implies that for all triples of profiles, $(x, y, z) \in V^3$,

$$\begin{aligned} \frac{\psi_{ab}(x, y)}{\psi_{ab}(z, z)} &= \frac{\phi_{jh}(x, z)}{\phi_{jh}(y, z)} = \frac{\phi_{jh}(x, z)}{\phi_{jh}(z, z)} \left[\frac{\phi_{jh}(y, z)}{\phi_{jh}(z, z)}\right]^{-1} \\ &= \frac{\psi_{ab}(x, z)}{\psi_{ab}(z, z)} \left[\frac{\psi_{ab}(y, z)}{\psi_{ab}(z, z)}\right]^{-1} = \frac{\psi_{ab}(x, z)}{\psi_{ab}(y, z)}. \end{aligned} \tag{4.345}$$

Employing these relations, we next show that for any fixed choice of a destination, $d \in J$, and fixed choice of a separation profile, $\sigma \in V$, the following relation holds for all $ij \in I \times J$ and $c \in C$:

$$\frac{P_c(ij)}{P_c(ad)} = [\psi_{ab}(c_{ad}, \sigma)]^{-1} \psi_{ia}(\sigma, \sigma) \phi_{jd}(\sigma, \sigma) \psi_{ab}(c_{ij}, \sigma). \tag{4.346}$$

To establish this result, suppose first that $i \neq a$ and $j \neq d$. Then, observing from (4.341) and (4.342) that $\psi_{aa}(\sigma,\sigma)$ and $\psi_{dd}(\sigma,\sigma)$ are by definition equal to one, it follows from (4.341), (4.342), and (4.344) [with $h = d$] that

$$\begin{aligned}\frac{P_c(ij)}{P_c(ad)} &= \frac{P_c(ij)}{P_c(aj)} \cdot \frac{P_c(aj)}{P_c(ad)} = \psi_{ia}(c_{ij}, c_{aj})\phi_{jd}(c_{aj}, c_{ad}) \\ &= \left[\psi_{ia}(\sigma,\sigma)\frac{\psi_{ab}(c_{ij},\sigma)}{\psi_{ab}(c_{aj},\sigma)}\right] \cdot \left[\phi_{jd}(\sigma,\sigma)\frac{\psi_{ab}(c_{aj},\sigma)}{\psi_{ab}(c_{ad},\sigma)}\right] \\ &= [\psi_{ab}(c_{ad},\sigma)]^{-1}\psi_{ia}(\sigma,\sigma)\phi_{jd}(\sigma,\sigma)\psi_{ab}(c_{ij},\sigma),\end{aligned} \tag{4.347}$$

and (4.346) is seen to hold for this case. Next, suppose that $i = a$ and $j \neq d$. Then by (4.341) and (4.344), together with $\psi_{aa}(\sigma,\sigma) = 1$, we have

$$\begin{aligned}\frac{P_c(aj)}{P_c(ad)} &= \phi_{jd}(c_{aj}, c_{ad}) = \phi_{jd}(\sigma,\sigma)\frac{\psi_{ab}(c_{aj},\sigma)}{\psi_{ab}(c_{ab},\sigma)} \\ &= [\psi_{ab}(c_{ab},\sigma)]^{-1}\psi_{aa}(\sigma,\sigma)\phi_{jd}(\sigma,\sigma)\psi_{ab}(c_{aj},\sigma)\end{aligned} \tag{4.348}$$

and (4.346) again holds. Moreover, if $i \neq a$ and $j = d$, then by (4.341), (4.344), and (4.345), together with $\psi_{dd}(\sigma,\sigma) = 1$, we have:

$$\begin{aligned}\frac{P_c(id)}{P_c(ad)} &= \psi_{ia}(c_{id}, c_{ad}) = \psi_{ia}(\sigma,\sigma)\frac{\psi_{ab}(c_{id},c_{ad})}{\psi_{ab}(\sigma,\sigma)} \\ &= \psi_{ia}(\sigma,\sigma)\frac{\psi_{ab}(c_{id},\sigma)}{\psi_{ab}(c_{ad},\sigma)} \\ &= [\psi_{ab}(c_{ad},\sigma)]^{-1}\psi_{ia}(\sigma,\sigma)\phi_{dd}(\sigma,\sigma)\psi_{ab}(c_{id},\sigma).\end{aligned} \tag{4.349}$$

Finally, if $i = a$ and $j = d$, then both sides of (4.346) are trivially equal to one, and we may conclude from (4.347) through (4.349) that (4.346) holds identically for all $ij \in I \times J$ and $c \in C$. Given this result, observe next that since $\boldsymbol{P}$ also satisfies A1 and A2, it follows from (4.107) that the frequency process, $\boldsymbol{N_P} = \{N^c : c \in C\}$, generated by $\boldsymbol{P}$ satisfies

$$\frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ad}^c)} = \frac{\mathrm{E}(N_{ij}^c)/\mathrm{E}(N^c)}{\mathrm{E}(N_{ad}^c)/\mathrm{E}(N^c)} = \frac{P_c(ij)/P_c(S_1)}{P_c(ad)/P_c(S_1)} = \frac{P_c(ij)}{P_c(ad)}, \tag{4.350}$$

for all $ij \in I \times J$ and $c \in C$. Thus, if for any fixed choice of $\sigma \in V, d \in J$, and $a, b \in I$ with $a \neq b$, we now define the functions, $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : V \to R_{++}$, by:

$$\lambda(c) = \mathrm{E}(N_{ad}^c)/\psi_{ab}(c_{ad},\sigma), \quad c \in C, \tag{4.351}$$

$$A(i) = \psi_{ia}(\sigma,\sigma), \quad i \in I, \tag{4.352}$$

$$B(j) = \phi_{jd}(\sigma,\sigma), \quad j \in J, \tag{4.353}$$

$$F(v) = \psi_{ab}(v,\sigma), \quad v \in V, \tag{4.354}$$

then it follows at once from (4.346) and (4.350) through (4.354) that

$$\mathrm{E}(N_{ij}^c) = \lambda(c)A(i)B(j)F(c_{ij}) \tag{4.355}$$

holds identically for all $ij \in I \times J$ and $c \in C$. Hence $\boldsymbol{N_P} \in \langle \text{MODEL G5} \rangle$ whenever $\boldsymbol{P}$ satisfies (A1, A2, A6, A7), and we may conclude that

$$\langle \text{A1, A2, A6, A7} \rangle \subseteq \langle \text{MODEL G5} \rangle.$$

(b) To establish the converse, consider any $\boldsymbol{N} \in \langle \text{MODEL G5} \rangle$, and observe from (4.350) that with respect to the representation in (4.5) the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ must satisfy the following identity for all $i, g \in I, j, h \in J$, and $c \in C$:

$$\frac{P_c(ij)}{P_c(gj)} = \frac{A(i)B(j)F(c_{ij})}{A(g)B(h)F(c_{gh})}. \tag{4.356}$$

Hence for all $i, g \in I, j, h \in J$, and $c, c' \in C$, it follows from (4.356) that

$$\begin{aligned} &(c_{ij} = c'_{ih}, c_{gj} = c'_{gh}) \Rightarrow \\ &\frac{P_c(ij)}{P_c(gj)} = \frac{A(i)F(c_{ij})}{A(g)F(c_{gj})} = \frac{A(i)F(c'_{ih})}{A(g)F(c'_{gh})} = \frac{P_{c'}(ih)}{P_{c'}(gh)}, \end{aligned} \tag{4.357}$$

and similarly that,

$$\begin{aligned} &(c_{ij} = c'_{gj}, c_{ih} = c'_{gh}) \Rightarrow \\ &\frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)}, \end{aligned} \tag{4.358}$$

so that $\boldsymbol{P_N}$ is seen to satisfy both axioms A6 and A7. Thus we must also have $\langle \text{MODEL G5} \rangle \subseteq \langle \text{A1, A2, A6, A7} \rangle$, and may conclude that assertion (4.340) holds. □

As a parallel to this result, we have the following local characterization of monotone relatively-invariant gravity models in terms of the bimonotonicity axioms (A10,A11):

Theorem 4.14 (G5* Characterization) *The class of frequency processes, $\langle \text{MODEL G5}^* \rangle$, is characterized by the following local properties of its generators:*

$$\langle \text{MODEL G5}^* \rangle = \langle \text{A1, A2, A10, A11} \rangle. \tag{4.359}$$

Proof: (a) To establish that $\langle \text{A1, A2, A10, A11} \rangle \subseteq \langle \text{MODEL G5}^* \rangle$, we first recall from (4.141) that each independent interaction process, $\boldsymbol{P}$, satisfying (A10,A11) also satisfies (A6,A7). Hence it follows from (4.340) that $\boldsymbol{N_P} \in \langle \text{MODEL G5} \rangle$, so that the representation in (4.355) must hold for some appropriate choice of the functions, $\lambda : C \to R_{++}, A : I \to R_{++}, B : J \to R_{++}$, and $F : V \to R_{++}$. Moreover, precisely the same argument as in (4.316) [or (4.338)] also shows that F must be nonincreasing on V. Hence $\boldsymbol{N_P} \in \langle \text{MODEL G5}^* \rangle$ whenever $\boldsymbol{P}$ satisfies (A1,A2,A10,A11), and we may conclude that $\langle \text{A1, A2, A10, A11} \rangle \subseteq \langle \text{MODEL G5}^* \rangle$.

(b) Conversely, for any $\boldsymbol{N} \in \langle \text{MODEL G5}^* \rangle$, it follows at once from the arguments for Model G5 above that the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ satisfies (4.356) for the functions, $A : I \to R_{++}$, $B : J \to R_{++}$, and $F : V \to R_{++}$, in (4.5). Moreover, since F is nonincreasing, it then follows that for all $i, g \in I$, $j, h \in J$, and $c, c' \in C$,

$$\begin{aligned}(c_{ij} \leq c'_{gj},\ c_{ih} \geq c'_{gh}) &\Rightarrow [F(c_{ij}) \geq F(c'_{gj}),\ F(c_{ih}) \leq F(c'_{gh})] \\ &\Rightarrow \frac{P_c(ij)}{P_c(ih)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} \geq \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}(gj)}{P_{c'}(gh)},\end{aligned} \tag{4.360}$$

and hence that $\boldsymbol{P_N}$ satisfies A10. A parallel argument shows that $\boldsymbol{P_N}$ also satisfies A11. Thus we see that $\langle \text{MODEL G5}^* \rangle \subseteq \langle \text{A1, A2, A10, A11} \rangle$, and may conclude that assertion (4.359) holds. □

Remark 4.4. It should be clear from a comparison of the aggregate characterizations in Theorems 4.5 and 4.6 with the local characterizations in Theorems 4.11 and 4.12 that [as in Remark 4.2 above] there exist certain equivalence relations among the axioms not already mentioned in Proposition 4.1. In particular, it follows at once from (4.246) and (4.340) that $\langle \text{A1, A2, A3}'' \rangle = \langle \text{A1, A2, A6, A7} \rangle$, and similarly, from (4.272) and (4.359) that $\langle \text{A1, A2, A4}'' \rangle = \langle \text{A1, A2, A10, A11} \rangle$. In particular, the equivalence between aggregate axiom, A3$''$, and the local axioms (A6,A7) [in the presence of A1 and A2], suggests that our formulation of axiom A3$''$ can be weakened considerably. Indeed, an examination of the argument in expressions (4.247) through (4.262) in the proof of Theorem 4.5 shows that the construction of the relatively invariant gravity representation in expression (4.263) involved only interaction patterns in S_1 and S_2. Hence from a formal viewpoint it is important to observe that axiom A3$''$ can be weakened to require only that (4.121) hold for all activity-equivalent interaction patterns of size two or less. However, since the present statement of axiom A3$''$ serves to emphasize the parallel structure of axioms A3 and A3$''$, we choose to employ this stronger form. Finally, it should be noted that the above equivalences can of course be proved directly from the axioms themselves, without any appeal to gravity models. However, since these direct proofs are virtually the same as those above, they add little to the present development. •

Remark 4.5. Note from the arguments in (4.276) and (4.277) of Theorem 4.7 that Model G5 together with A5 is equivalent to Model G6, and similarly, that Model G5* together with A5 is equivalent to Model G6*. Hence, as alternative characterizations of these models, we also see from Theorem 4.11 that

$$\langle \text{MODEL G6} \rangle = \langle \text{A1, A2, A5, A6, A7} \rangle$$

and

$$\langle \text{MODEL G6}^* \rangle = \langle \text{A1, A2, A5, A10, A11} \rangle.$$

Finally, since sub-configuration dependence (A5) is seen to be a local property of interaction behavior at each origin, these characterizations of Model G6 and Model G6* can also be regarded as local in nature. •

4.5.3 Characterizations of Exponential Gravity Models

Finally we establish the characterizations of the exponential gravity models in Section 4.2.2 as discussed in Section 2.4.3 above. Hence we now assume that spatial separation is extensively measurable in terms of *cost configurations*, and begin by establishing the characterization in Section 2.4.3 of the most general class of exponential gravity models (E1) in terms of cost dependency (A12). The notational conventions and proof format mentioned at the beginning of Section 4.5.2 will continue to apply here.

Theorem 4.15 (E1 Characterization) *The class of frequency processes, ⟨MODEL E1⟩, is characterized by the following properties of it generators:*

$$\langle \text{MODEL E1} \rangle = \langle \text{A1}, \text{A2}, \text{A12} \rangle. \tag{4.361}$$

Proof: (a) Assertion (4.361) can be viewed as a stronger form of assertion (4.169) in which axiom, A3, is replaced by axiom, A12. In particular, it is clear from a comparison of (4.119) and (4.132) that for any spatial interaction patterns, $s, s' \in S$, the equality of separation arrays, $c_s = c_{s'}$, in (4.119) is now replaced by the corresponding equality of total-cost profiles, $c(s) = c(s')$, in (4.132). Hence, the basic outline of the proof of (4.361) closely parallels that of (4.169). To begin with, let the indicator functions, $\delta_g, \sigma_h \in (0,1)^{I\times J}$, be defined for all $g \in I$ and $h \in J$ as in (4.170) and (4.171), respectively. If for any fixed choice of a rational-valued separation configuration, $c \in C(Q)$, expression (4.175) is now replaced by the definition of total-cost profiles, i.e., by (4.116), for each $s \in S$, and expression (4.178) is now replaced by the corresponding condition that

$$c^k(s) = c^k(s') \Leftrightarrow \sum_{ij} c_{ij}^k z(ij) = 0, \quad k \in K, \tag{4.362}$$

then axiom A12 implies that whenever the equalities on the right hand sides of (4.176), (4.177), and (4.362) hold, the equality in (4.180) must also hold. Hence, if in Lemma 4.4 we now set $T = (I+J) \cup K$, $W = I \times J$, and employ (4.107) to define the functions, $b_c\colon I \times J \to R_{++}$ and $a_c\colon T \times (I \times J) \to Q$, respectively, for all $ij \in I \times J$ and $c \in C$ by $b_c(ij) = \log[p_c(ij)]$ and

$$a_{ct}(ij) = \begin{cases} \delta_g(ij), & t = (g,1) \in I \times \{1\} \\ \sigma_h(ij), & t = (h,2) \in J \times \{2\} \\ c_{ij}^t, & t \in K, \end{cases} \tag{4.363}$$

then it follows from part (ii) of Lemma 4.4 that for each rational configuration, $c \in C(Q)$, there must exist a function, $\lambda_c : T \to R$, such that (4.153) holds as an equality, which in this case takes the form:

$$\begin{aligned}\log[p_c(ij)] = &\sum_{g\in I} \lambda_c(g,1)\delta_g(ij) + \sum_{h\in J} \lambda_c(h,2)\sigma_h(ij) \\ &+ \sum_{k\in K} \lambda_c(k)c_{ij}^k.\end{aligned} \tag{4.364}$$

This representation may now be employed in Lemma 4.5 as follows. First, set $Y = I + J$ and let the vector-valued function, $a : Y \to Q^{I\times J}$, be defined in terms of (4.363) by

$$\begin{aligned} a_y &= (a_y(ij) : ij \in I \times J) \\ &= \begin{cases} (\delta_g(ij) : ij \in I \times J), & y = (g,1) \in I \times \{1\} \\ (\sigma_h(ij) : ij \in I \times J), & y = (h,2) \in J \times \{2\}. \end{cases}\end{aligned} \tag{4.365}$$

Then, defining the vector-valued function, $b : C \to R^{I\times J}$, for all $c \in C$ by $b_c = (\log[p_c(ij)] : ij \in I \times J)$, and choosing $w : C \times Y \to R$ and $\theta : C \times K \to R$ to be any functions satisfying $w_c(y) = \lambda_c(y)$ and $\theta_c(k) = -\lambda_c(k)$ for all $y \in I + J \subseteq T$, $k \in K \subseteq T$, and $c \in C(Q)$, it follows from (4.364) that the hypothesis in (4.154) holds for this choice of functions, i.e., that

$$\log[p_c(ij)] = \sum_{y\in Y} w_c(y)a_y(ij) - \sum_{k\in K} \theta_c(k)c_{ij}^k, \quad ij \in I \times J \tag{4.366}$$

for all $c \in C(Q)$. Next, recall from the continuity condition (R3) for the interaction process, $\boldsymbol{P}$, [together with the resulting continuity property (3.29) of the frequency process, $\boldsymbol{N_P}$, generated by $\boldsymbol{P}$] that for each $ij \in I \times J$, the interaction probability, $p_c(ij)$, in (4.107) is a continuous function of c. Hence, the function, $b : C \to R^{I\times J}$, must be *continuous,* and it follows from part (*i*) of Lemma 4.5 that the functions, w and θ , can be chosen so that the equality in (4.366) holds identically for all $c \in C$. Thus, the same argument as in (4.182) and (4.183) above now shows that

$$p_c(ij) = \exp[w_c(i,1) + w_c(j,2) - \sum_{k\in K} \theta_c(k)c_{ij}^k] \tag{4.367}$$

for all $c \in C$ and $ij \in I \times J$. Finally, if one defines the functions, $A : C \times I \to R_{++}$, $B : C \times J \to R_{++}$, and $\theta : C \to R^k$, by $A_c(i) = \mathrm{E}(N^c)\exp[w_c(i)]$, $B_c(j) = \exp[w_c(j)]$, and $\theta_c = (\theta_c(k) : k \in K)$, respectively, then it follows from (4.367), together with (4.107), that for this choice of functions,

$$\mathrm{E}(N_{ij}^c) = A_c(i)B_c(j)\exp[-\theta_c^t c_{ij}] \tag{4.368}$$

for all $ij \in I \times J$ and $c \in C$. Thus, for each $\boldsymbol{P}$ satisfying (A1,A2,A12), one may conclude that $\boldsymbol{N_P} \in \langle\text{MODEL E1}\rangle$, and hence that

$$\langle \text{A1, A2, A12} \rangle \subseteq \langle \text{MODEL E1} \rangle.$$

(b) To establish the converse, observe first that, as in (4.185), the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of each $\boldsymbol{N} \in \langle\text{MODEL E1}\rangle$ must satisfy

$$\begin{aligned} P_c(s) &= P_c[N = N(s)] \prod_{ij} \{A^*(i)B_c(j)\exp[-\theta_c^t c_{ij}]\}^{N_{ij}(s)} \\ &= P_c[N = N(s)] \left(\prod_i A^*(i)^{N_i(s)}\right) \\ &\quad \cdot \left(\prod_j B_c(j)^{N_j(s)}\right) \exp[-\theta_c^t c(s)] \end{aligned} \tag{4.369}$$

for all $s \in S$ and $c \in C$. Thus, if $A(s) = A(s')$ and $c(s) = c(s')$ both hold for any $s, s' \in S$, then $P_c(s) = P_c(s')$ follows at once from (4.369), and $\boldsymbol{P_N}$ is seen to satisfy A12. Hence $\langle\text{MODEL E1}\rangle \subseteq \langle\text{A1, A2, A12}\rangle$, and assertion (4.361) must hold. □

Next we establish the corresponding characterization of monotone exponential gravity models (E1*) in terms of cost efficiency (A13):

Theorem 4.16 (E1* Characterization) *The class of frequency processes, $\langle\text{MODEL E1}^*\rangle$, is characterized by the following properties of its generators:*

$$\langle \text{MODEL E1}^* \rangle = \langle \text{A1, A2, A13} \rangle. \tag{4.370}$$

PROOF: (a) To establish that $\langle\text{A1, A2, A13}\rangle \subseteq \langle\text{MODEL E1}^*\rangle$, consider any $\boldsymbol{P}$ satisfying (A1,A2,A13), together with any rational-valued separation configuration, $c \in C(Q)$, and any spatial interaction pattern, $s^\circ \in S$, with positive interaction frequencies, $N_{ij}(s^\circ) > 0$, for all origin-destination pairs, $ij \in I \times J$. Then, recalling the definitions of $N_i(s^\circ), N_j(s^\circ), c(s^\circ)$, and $p_c(ij)$ in (4.112), (4.113), (4.116), and (4.107), respectively, we may now construct the following *linear programming problem*:

$$\text{minimize:} \sum_{ij} x_{ij} \log[p_c(ij)], \tag{4.371}$$

subject to the conditions that $x_{ij} \in R_+$, $ij \in I \times J$, and that:

$$\sum_{ij} \delta_g(ij) x_{ij} = N_g(s^\circ), \quad g \in I, \tag{4.372}$$

$$\sum_{ij} \sigma_h(ij) x_{ij} = N_h(s^\circ), \quad h \in J, \tag{4.373}$$

$$\sum_{ij} c_{ij} x_{ij} \leq c(s^\circ). \tag{4.374}$$

To analyze the solution properties of this programming problem, observe first from (4.372) together with the nonnegativity conditions that $0 \leq x_{ij} \leq N_i(s^\circ)$ must hold for all $ij \in I \times J$, and hence that the set of feasible solutions is bounded. Moreover, since the (positive) vector,

$$x^\circ = (x^\circ_{ij} : ij \in I \times J) = (N_{ij}(s^\circ) : ij \in I \times J) \in Z^{I \times J}_{++} \tag{4.375}$$

is feasible by definition, it then follows from (4.372), (4.373), and (4.374) that the set of feasible solutions is nonempty and compact. Hence there must always exist at least one solution to this programming problem. Given this observation, our first objective is to show that if $\boldsymbol{P}$ satisfies (A1,A2,A13), then x° in (4.375) is *always a solution.* To do so, we begin by reformulating (4.371) through (4.374) in terms of the notation in Lemmas 4.6 and 4.7 above. In particular, if we now let $Y = I + J$, $p = |Y|$, $q = |K|$, $m = |I \times J|$, and let

$$a_y = \begin{cases} (\delta_g(ij) : ij \in I \times J) \in \{0,1\}^m, & y = (g,1) \in I \times \{1\} \\ (\sigma_h(ij) : ij \in I \times J) \in \{0,1\}^m, & y = (h,2) \in J \times \{2\}, \end{cases} \tag{4.376}$$

$$A^t = [a_y : y \in Y] \in \{0,1\}^{m \times p} \subseteq Q^{m \times p}, \tag{4.377}$$

$$B_c = [-c_{ij} \in Q^q : ij \in I \times J] \in Q^{q \times m}, \tag{4.378}$$

$$b_y = \begin{cases} N_g(s^\circ), & y = (g,1) \in I \times \{1\} \\ N_h(s^\circ), & y = (h,2) \in J \times \{2\}, \end{cases} \tag{4.379}$$

$$b = (b_y : y \in Y) \in Z^p_{++} \subseteq Q^p, \tag{4.380}$$

$$d_c = -c(s^\circ) \in Q^q, \tag{4.381}$$

$$f_c = (\log[p_c(ij)] : ij \in I \times J) \in R^m, \tag{4.382}$$

then the programming problem defined by (4.371), (4.372), (4.373), and (4.374) can be rewritten in vector form as:

$$\text{minimize:} f^t_c x \qquad \text{subject to:} Ax = b,\ B_c x \geq d_c,\ x \in R^m_+. \tag{4.383}$$

In this form, it follows at once from (4.377), (4.378), (4.380), and (4.381) that all the conditions of Lemma 4.7 are satisfied. Hence, by Lemma 4.7, the programming problem in (4.383) must possess a *rational solution,* say, $x^* = (x^*_{ij} : ij \in I \times J) \in Q^m_+$. Moreover, since (4.372) implies that at least one x^*_{ij} must be positive, it follows that these rational numbers can be expressed as integer ratios, $x^*_{ij} = N^*_{ij}/M$, with a common positive denominator, $M \in Z_{++}$, i.e.,

$$x^* = (N^*_{ij}/M : ij \in I \times J) \in Q^m_+. \tag{4.384}$$

But if we now choose any spatial interaction patterns, $s_1, s_2 \in S$, with $N_{ij}(s_1) = N^*_{ij}$ and $N_{ij}(s_2) = M \cdot N_{ij}(s^\circ)$ for all $ij \in I \times J$, then since x^*

satisfies (4.372) and (4.373), it follows that for all $i \in I$ and $j \in J$,

$$N_i(s_1) = \sum_j N_{ij}^* = M(\sum_j N_{ij}^*/M) = M \cdot N_i(s^\circ) = N_i(s_2), \tag{4.385}$$

$$N_j(s_1) = \sum_i N_{ij}^* = M(\sum_i N_{ij}^*/M) = M \cdot N_j(s^\circ) = N_j(s_2). \tag{4.386}$$

Hence by (4.114) it follows that $A(s_1) = A(s_2)$. Moreover, since x^* also satisfies (4.374), it follows that

$$\begin{aligned} c(s_1) &= \sum_{ij} c_{ij} N_{ij}(s_1) = \sum_{ij} c_{ij} N_{ij}^* = M \cdot \sum_{ij} c_{ij}(N_{ij}^*/M) \\ &\leq M \cdot \sum_{ij} c_{ij} N_{ij}(s^\circ) = \sum_{ij} c_{ij}[M \cdot N_{ij}(s^\circ)] = c(s_2). \end{aligned} \tag{4.387}$$

Finally, since $\boldsymbol{P}$ satisfies A13, we see from (4.135) and (4.387) together with $A(s_1) = A(s_2)$ that $P_c(s_1) \geq P_c(s_2)$. But since $\boldsymbol{P}$ also satisfies A1 and A2, this together with the equality, $N(s_1) = N(s_2)$, then implies from (4.108) that

$$\prod_{ij} p_c(ij)^{N_{ij}(s_1)} \geq \prod_{ij} p_c(ij)^{N_{ij}(s_2)}, \tag{4.388}$$

so that by taking logs in (4.388) we obtain,

$$\begin{aligned} &\sum_{ij} N_{ij}(s_1) \log[p_c(ij)] \geq \sum_{ij} N_{ij}(s_2) \log[p_c(ij)] \\ &\Rightarrow M \cdot \sum_{ij} (N_{ij}^*/M) \log[p_c(ij)] \geq M \cdot \sum_{ij} N_{ij}(s^\circ) \log[p_c(ij)] \\ &\Rightarrow \sum_{ij} x_{ij}^* \log[p_c(ij)] \geq \sum_{ij} x_{ij}^\circ \log[p_c(ij)] \\ &\Rightarrow f_c^t x^* \geq f_c^t x^\circ. \end{aligned} \tag{4.389}$$

Hence, the feasibility of x° and the optimality of x^* in (4.383) implies that x° must also be an optimal solution to (4.383).

Given this result, it follows at once from Lemma 4.6, together with the positivity of $x^\circ \in R^m_{++}$ that there must exist vectors, $w_c = (w_c(y) : y \in Y) \in R^p$ and $\theta_c = (\theta_c(k) : k \in K) \in R^q_+$ satisfying (4.157). But if we evaluate (4.157) at each component of $f_c = (\log[p_c(ij)] : ij \in I \times J)$, then by employing (4.377) and (4.378) we obtain

$$\log[p_c(ij)] = \sum_{y \in Y} w_c(y) a_y(ij) - \sum_{k \in K} \theta_c(k) c_{ij}^k, \tag{4.390}$$

for all $ij \in I \times J$, and all rational configurations, $c \in C(Q)$. Thus, by the continuity argument following (4.366), it may now be seen from part (ii) of Lemma 4.5 that the functions, $w : C \times Y \to R$ and $\theta : C \times K \to R_+$, can be chosen to satisfy (4.390) for all $c \in C$. This in turn implies from the argument following (4.367) that for this choice of functions, (4.368) must now hold with $\theta_c \in R_+^K$ for all $c \in C$. Hence $\boldsymbol{N_P} \in \langle$MODEL E1*$\rangle$ for each $\boldsymbol{P}$ satisfying (A1,A2,A13), and it follows that $\langle$A1, A2, A13$\rangle \subseteq$ MODEL E1*$\rangle$.

(b) To establish the converse, consider any $\boldsymbol{N} \in \langle$MODEL E1*$\rangle$ and observe from (4.22) that $\boldsymbol{N} \in \langle$MODEL E1$\rangle$. Hence if $\boldsymbol{P_N} = \{P_c : c \in C\}$ is the generator of $\boldsymbol{N}$, then it follows at once that for each $c \in C$ and $s \in S$, $P_c(s)$ satisfies (4.369) for the functions, $A_c : I \to R_{++}$, $B_c : J \to R_{++}$, and nonnegative vector, $\theta_c \in R_+^K$, in MODEL E1*. This in turn implies that for any $c \in C$ and $s, s' \in S$ with $A(s) = A(s')$,

$$\begin{aligned} P_c(s)/P_c(s') &= \exp[-\theta_c^t c(s)]/\exp[-\theta_c^t c(s')] \\ &= \exp[\theta_c^t\{c(s') - c(s)\}] \end{aligned} \tag{4.391}$$

which, together with the nonnegativity of θ_c, shows that for all $s, s' \in S$ with $A(s) = A(s')$,

$$\begin{aligned} c(s) \leq c(s') &\Rightarrow \theta_c^t\{c(s') - c(s)\} \geq 0 \\ &\Rightarrow \exp[\theta_c^t\{c(s') - c(s)\}] \geq 1 \\ &\Rightarrow P_c(s) \geq P_c(s'), \end{aligned} \tag{4.392}$$

and hence that $\boldsymbol{P_N}$ must satisfy A13 for each $\boldsymbol{N} \in \langle$MODEL E1*$\rangle$. Thus, we may conclude that $\langle$MODEL E1*$\rangle \subseteq \langle$A1, A2, A13$\rangle$ holds as well, and assertion (4.370) is established. □

Given these general characterizations of exponential gravity models, we next consider the class of deterrence-invariant exponential gravity models (E2). As in the case of general deterrence-invariant models (G2) it is necessary to distinguish the special case in which $|I| = 2 = |J|$. To obtain characterizations covering this case, we now introduce the following strengthening of cost dependency. In particular, if for any n-tuple of cost configurations, $\boldsymbol{c} = (c_\alpha : \alpha = 1, \ldots, n) \in C^n$, we define the corresponding vector generalization of total-cost profiles by:

$$\boldsymbol{c}(\boldsymbol{s}) = \sum_{\alpha=1}^{n} c_\alpha(s_\alpha), \tag{4.393}$$

then we have the following strong cost dependency axiom [paralleling the *strong separation dependence* axiom (A3°)]:

A12°. (Strong Cost Dependence) *For all $n \in Z_{++}$, and all n-tuples of cost configurations, $\boldsymbol{c}, \boldsymbol{c}' \in C^n$, and spatial interaction patterns, $\boldsymbol{s}, \boldsymbol{s}' \in S^n$, with $A(\boldsymbol{s}) = A(\boldsymbol{s}')$,*

$$\boldsymbol{c}(\boldsymbol{s}) = \boldsymbol{c}'(\boldsymbol{s}') \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}'}(\boldsymbol{s}'). \tag{4.394}$$

By setting $n = 1$ it is clear that A12° implies A12. In fact, we have the following parallel to (4.207):

$$\langle \text{A}12^\circ \rangle \subseteq \langle \text{A}12' \rangle. \tag{4.395}$$

To see this, observe that if for any comparable patterns $s, t, s', t' \in S$ it is true that $A(s) = A(t)$ and $A(s') = A(t')$, then letting $\boldsymbol{s} = (s, t')$ and $\boldsymbol{s}' = (t, s')$, it follows from (4.202) that $A(\boldsymbol{s}) = A(\boldsymbol{s}')$. Hence if A12° holds, then by letting $\boldsymbol{c} = (c, c') = \boldsymbol{c}'$, it follows from (4.202), (4.393) and (4.394) that

$$\begin{aligned}
&[c(s) = c'(s'),\ c(t) = c'(t')] \\
&\quad \Rightarrow \boldsymbol{c}(\boldsymbol{s}) = c(s) + c'(t') = c(t) + c'(s') = \boldsymbol{c}'(\boldsymbol{s}') \\
&\quad \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}'}(\boldsymbol{s}') \Rightarrow P_c(s)P_{c'}(t') = P_c(t)P_{c'}(s') \\
&\quad \Rightarrow \frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}.
\end{aligned} \tag{4.396}$$

Moreover, this strong cost dependency axiom also implies strong separation dependency, i.e.,

$$\langle \text{A}12^\circ \rangle \subseteq \langle \text{A}3^\circ \rangle. \tag{4.397}$$

To see this, recall first from (4.116) together with (4.164) and (4.165) that for any $c \in C$ and $s = (i_r j_r : r = 1, \ldots, n) \in S_n$, $n \in Z_{++}$, we may equivalently write the total-cost profile, $c(s)$, as

$$c(s) = \sum_{r=1}^{n} c_{i_r j_r} = \sum_{v \in V} v f_{cv}(s). \tag{4.398}$$

[Note again from the comment following (4.165) that the summation in (4.398) over the uncountably infinite set V is always well defined.] Next observe from (4.398) together with (4.204), and (4.393) that for all $\boldsymbol{c} = (c_\alpha : \alpha = 1, \ldots, m) \in C^m$ and $\boldsymbol{s} = (s_\alpha : \alpha = 1, \ldots, m) \in S^m$,

$$\begin{aligned}
\boldsymbol{c}(\boldsymbol{s}) &= \sum_{\alpha=1}^{m} c_\alpha(s_\alpha) = \sum_{\alpha} \Big[\sum_{v} v f_{c_\alpha v}(s_\alpha) \Big] \\
&= \sum_{v} v \Big[\sum_{\alpha} f_{c_\alpha v}(s_\alpha) \Big] = \sum_{v} v f_{\boldsymbol{c}v}(\boldsymbol{s}).
\end{aligned} \tag{4.399}$$

Hence if a given interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfies A12°, then it follows from (4.399) together with (4.398) that for all $\boldsymbol{c}, \boldsymbol{c}' \in C^m$ and $\boldsymbol{s}, \boldsymbol{s}' \in S^m$,

$$\begin{aligned}
f_{\boldsymbol{c}}(\boldsymbol{s}) = f_{\boldsymbol{c}'}(\boldsymbol{s}') &\Rightarrow f_{\boldsymbol{c}v}(\boldsymbol{s}) = f_{\boldsymbol{c}'v}(\boldsymbol{s}'), \quad v \in V \\
&\Rightarrow \boldsymbol{c}(\boldsymbol{s}) = \sum_{v} v f_{\boldsymbol{c}v}(\boldsymbol{s}) = \sum_{v} v f_{\boldsymbol{c}'v}(\boldsymbol{s}') = \boldsymbol{c}'(\boldsymbol{s}') \\
&\Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}'}(\boldsymbol{s}').
\end{aligned} \tag{4.400}$$

Thus $\boldsymbol{P}$ also satisfies A3°, and it follows that (4.397) must hold. Given these relationships, we now have the following characterizations of deterrence-invariant exponential gravity models:

Theorem 4.17 (E2 Characterization)
(i) *The class of frequency processes,* $\langle$MODEL E2$\rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL E2} \rangle = \langle \text{A1}, \text{A2}, \text{A12}^\circ \rangle. \tag{4.401}$$

(ii) *If in addition it is true that either* $|I| \geq 3$ *or* $|J| \geq 3$ *then*

$$\langle \text{MODEL E2} \rangle = \langle \text{A1}, \text{A2}, \text{A12}' \rangle. \tag{4.402}$$

PROOF: **(i).** (a) To establish that $\langle \text{A1}, \text{A2}, \text{A12}^\circ \rangle \subseteq \langle \text{MODEL E2} \rangle$, consider any interaction process $\boldsymbol{P}$ satisfying (A1, A2, A12°), and observe first from (4.397) together with Theorem 4.3 that $\boldsymbol{N_P} \in \langle \text{MODEL G2} \rangle$. Hence there must exist a deterrence function, $F: V \to R_{++}$, together with origin and destination functions, $A_c: I \to R_{++}$ and $B_c: J \to R_{++}$, for each $c \in C$ such that for all $ij \in I \times J$

$$\mathrm{E}(N_{ij}^c) = A_c(i) B_c(j) F(c_{ij}). \tag{4.403}$$

But (4.403) in turn implies that for all $c \in C$ and $ij, gh \in I \times J$,

$$\frac{\mathrm{E}_c(N_{ij})\mathrm{E}_c(N_{gh})}{\mathrm{E}_c(N_{ih})\mathrm{E}_c(N_{gj})} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gh})}. \tag{4.404}$$

By employing (4.404) we next show that the function, F, must be *continuous.* To see this, choose any fixed elements, $i^\circ j^\circ \in I \times J$ and $c^\circ \in C$, and for each profile $v \in V$ define the configuration, $c(v) = [c_{ij}(v): ij \in I \times J] \in C$, by $c_{i^\circ j^\circ}(v) = v$ and $c_{ij}(v) = c_{ij}^\circ$ for all $ij \neq i^\circ j^\circ$. Then by (4.404) it follows that for any $gh \neq i^\circ j^\circ$ and $v \in V$,

$$\begin{aligned} \frac{\mathrm{E}_{c(v)}(N_{i^\circ j^\circ})\mathrm{E}_{c(v)}(N_{gh})}{\mathrm{E}_{c(v)}(N_{i^\circ h})\mathrm{E}_{c(v)}(N_{gj^\circ})} &= \frac{F(v)F(c_{gh}^\circ)}{F(c_{i^\circ h}^\circ)F(c_{gj^\circ}^\circ)} \\ \Rightarrow F(v) &= \left[\frac{F(c_{i^\circ h}^\circ)F(c_{gj^\circ}^\circ)}{F(c_{gh}^\circ)}\right] \frac{\mathrm{E}_{c(v)}(N_{i^\circ j^\circ})\mathrm{E}_{c(v)}(N_{gh})}{\mathrm{E}_{c(v)}(N_{i^\circ h})\mathrm{E}_{c(v)}(N_{gj^\circ})}. \end{aligned} \tag{4.405}$$

But since the expression in brackets is constant, it follows at once from the continuity of $c(v)$ in v together with regularity condition R3 for interaction processes that F is continuous. Next we show that F must be a *log linear* function. To do so, we begin by again observing from (4.404) together with (4.107) that for all $c \in C$ and $ij, gh \in I \times J$

$$\frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{\mathrm{E}_c(N_{ij})\mathrm{E}_c(N_{gh})}{\mathrm{E}_c(N_{ih})\mathrm{E}_c(N_{gj})} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gj})}. \tag{4.406}$$

Next recall from (4.142) and (4.395) that $\langle A12^{\circ}\rangle \subseteq \langle A12'\rangle \subseteq \langle A12\rangle$, so that $\boldsymbol{P}$ satisfies A12. Hence if for any distinct origins, $i, g \in I$, and destinations, $g, h \in J$, we set $s = (ij, gh)$ and $s' = (ih, gj)$ in axiom A12 [so that $A(s) = A(s')$ by construction] and observe that for each possible choice of profiles, $x, y, z, w \in V$, there exists some configuration, $c \in C(= V^{I\times J})$, with $c_{ij} = x$, $c_{gh} = y$, $c_{ih} = z$, and $c_{gj} = w$, it then follows from A12 together with (4.107) and (4.108) that whenever $x + y = z + w$, we must have

$$\begin{aligned} x + y = z + w &\Rightarrow c(s) = c(s') \\ &\Rightarrow P_c(s) = P_c(s') \\ &\Rightarrow P_c(ij)P_c(gh) = P_c(ih)P_c(gj) \\ &\Rightarrow 1 = \frac{P_c(ij)P_c(gh)}{P_c(ih)P_c(gj)} = \frac{F(c_{ij})F(c_{gh})}{F(c_{ih})F(c_{gj})} \\ &\Rightarrow F(x)F(y) = F(z)F(w). \end{aligned} \tag{4.407}$$

Thus, if we now define the function, $q : V \to R$, for all $v \in V$ by $q(v) = \log F(v)$, then it follows from (4.407) that the relation

$$x + y = z + w \Rightarrow q(x) + q(y) = q(z) + q(w) \tag{4.408}$$

holds identically for all $x, y, z, w \in V$. But since (1.6) implies that $0 \in V$ and that V is an additively closed subset of R^K (i.e., $x+y \in V$ for all $x, y \in V$), it then follows from (4.408) together with the identity, $(x + y) + (0) = (x) + (y)$, that $q(x + y) + q(0) = q(x) + q(y)$ for all $x, y \in V$. Hence the associated continuous function, $f : V \to R$, defined for all $x \in V$ by

$$f(x) = q(x) - q(0) \tag{4.409}$$

must satisfy Cauchy's equation [Lemma 4.8], i.e., must satisfy

$$\begin{aligned} f(x + y) &= q(x + y) - q(0) = [q(x + y) + q(0)] - 2q(0) \\ &= [q(x) + q(y)] - 2q(0) = [q(x) - q(0)] + [q(y) - q(0)] \\ &= f(x) + f(y), \end{aligned} \tag{4.410}$$

for all $x, y \in V$, and we may conclude from Lemma 4.8 that there exists some constant vector, $\theta \in R^K$, such that

$$f(v) = -\theta^t v \tag{4.411}$$

for all $v \in V$. Thus, F is seen to be log linear on V, and in particular, can be written for all $v \in V$ as $F(v) = \exp[q(v)] = \exp[f(v) + q(0)] = \exp[q(0)]\exp[-\theta^t v]$. Finally, substituting this expression into (4.403) and defining the new origin functions, $A_c^* : I \to R_{++}$, for each $c \in C$ and $i \in I$ by $A_c^*(i) = A_c(i)\exp[q(0)] > 0$, it follows from (4.403) and (4.411) that the relation

$$\mathrm{E}(N_{ij}^c) = A_c^*(i)B_c(j)\exp[-\theta^t c_{ij}] \tag{4.412}$$

holds identically for all $c \in C$ and $ij \in I \times J$. Thus, $\boldsymbol{N_P} \in \langle\text{MODEL E2}\rangle$ whenever $\boldsymbol{P}$ satisfies (A1, A2, A12°), and we must have $\langle \text{A1}, \text{A2}, \text{A12}^\circ\rangle \subseteq \langle\text{MODEL E2}\rangle$.

(b) The converse follows from an argument paralleling that for Model G2 above. In particular, if $\boldsymbol{N} \in \langle\text{MODEL E2}\rangle$ and if we set $A_c^*(i) = A_c(i)/\mathrm{E}(N^c)$ in (4.8), then (as a parallel to (4.231)), it now follows from (4.116) that for any spatial interaction $c \in C$ and $s \in S$ the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ must satisfy

$$\begin{aligned} P_c(s) &= P_c[N = N(s)] \prod_{ij} \{A_c^*(i) B_c(j) \exp[-\theta^t c_{ij}]\}^{N_{ij}(s)} \\ &= P_c[N = N(s)] \left(\prod_i A_c^*(i)^{N_i(s)} \right) \\ &\quad \cdot \left(\prod_j B_c(j)^{N_j(s)} \right) \exp[-\theta^t \sum_{ij} c_{ij} N_{ij}(s)] \\ &= P_c[N = N(s)] \left(\prod_i A_c^*(i)^{N_i(s)} \right) \\ &\quad \cdot \left(\prod_j B_c(j)^{N_j(s)} \right) \exp[-\theta^t c(s)], \end{aligned} \tag{4.413}$$

so that for any $n \in Z_{++}$, and n-tuples of interaction patterns, $\boldsymbol{s} = (s_\alpha : \alpha = 1, \ldots, n) \in S^n$, and separation configurations, $\boldsymbol{c} = (c_\alpha : \alpha = 1, \ldots, n)$, it follows at once from (4.393) and(4.413) that

$$\begin{aligned} P_{\boldsymbol{c}}(\boldsymbol{s}) &= \prod_{\alpha=1}^{n} P_{c_\alpha}(s_\alpha) \\ &= \prod_{\alpha=1}^{n} \left\{ P_{c_\alpha}[N = N(s_\alpha)] \left(\prod_i A_{c_\alpha}^*(i)^{N_i(s_\alpha)} \right) \left(\prod_j B_{c_\alpha}(j)^{N_j(s_\alpha)} \right) \right. \\ &\qquad \left. \cdot \exp[-\theta^t c_\alpha(s_\alpha)] \right\} \\ &= \prod_{\alpha=1}^{n} \left[P_{c_\alpha}[N = N(s_\alpha)] \left(\prod_i A_{c_\alpha}^*(i)^{N_i(s_\alpha)} \right) \left(\prod_j B_{c_\alpha}(j)^{N_j(s_\alpha)} \right) \right] \\ &\qquad \cdot \exp[-\theta^t \boldsymbol{c}(\boldsymbol{s})]. \end{aligned} \tag{4.414}$$

But for any other n-tuple of interaction patterns, $\boldsymbol{s}' = (s'_\alpha : \alpha = 1, \ldots, n) \in S^n$, with $A(\boldsymbol{s}) = A(\boldsymbol{s}')$ we must have $N_i(s_\alpha) = N_i(s'_\alpha)$ and $N_j(s_\alpha) = N_j(s'_\alpha)$ for all $ij \in I \times J$ and $\alpha = 1, \ldots, n$ [which in turn implies that

$N(s_\alpha) = N(s'_\alpha)$ for all $\alpha = 1, \ldots, n$]. Thus, for any $\boldsymbol{s}, \boldsymbol{s}' \in S^n$ with $A(\boldsymbol{s}) = A(\boldsymbol{s}')$ and any $\boldsymbol{c} \in C^n$, it follows from (4.414) that

$$P_{\boldsymbol{c}}(\boldsymbol{s})/P_{\boldsymbol{c}}(\boldsymbol{s}') = \exp\{-\theta^t[\boldsymbol{c}(\boldsymbol{s}) - \boldsymbol{c}(\boldsymbol{s}')]\}, \tag{4.415}$$

and hence that

$$\begin{aligned}[A(\boldsymbol{s}) = A(\boldsymbol{s}'), \boldsymbol{c}(\boldsymbol{s}) = \boldsymbol{c}(\boldsymbol{s}')] &\Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s})/P_{\boldsymbol{c}}(\boldsymbol{s}') = 1 \\ &\Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}}(\boldsymbol{s}').\end{aligned} \tag{4.416}$$

Thus $\boldsymbol{P_N}$ satisfies A12° whenever $\boldsymbol{N} \in \langle$MODEL E2$\rangle$, and we may conclude that $\langle$MODEL E2$\rangle \subseteq \langle$A1, A2, A12°$\rangle$.

(ii). (a) To establish (4.402) under the additional condition that either $|I| \geq 3$ or $|J| \geq 3$, recall from (4.144) together with part (ii) of Theorem 4.3 that $\boldsymbol{N_P} \in \langle$MODEL G2$\rangle$. Hence (4.403) again holds for $\boldsymbol{N_P}$ so that the same argument in part i(a) leading to (4.412) now shows that $\boldsymbol{N_P} \in \langle$MODEL E2$\rangle$, and hence that $\langle$A1, A2, A12$'\rangle \subseteq \langle$MODEL E2$\rangle$.

(b) Finally, to establish the converse, observe simply from (4.395) and (4.401) that $\langle$MODEL E2$\rangle \subseteq \langle$A1, A2, A12°$\rangle \subseteq \langle$A1, A2, A12$'\rangle$. □

In addition to this result, the special *linear* structure of exponential gravity models [as exemplified by the use of Cauchy's equation above] implies that for the case of *scalar* separation measures (i.e., with $|K| = 1$), the characterization in (4.402) continues to holds. Because of the close parallel between Theorem 4.17 and Theorem 4.3 for general gravity models, it is appropriate to state this additional result as a separate theorem:

Theorem 4.18 (E2 Characterization with Scalar Costs) *If* $|K| = 1$, *the class of frequency processes,* $\langle$MODEL E2$\rangle$, *is characterized by the following properties of its generators:*

$$\langle \text{MODEL E2} \rangle = \langle \text{A1, A2, A12}' \rangle. \tag{4.417}$$

PROOF: Observe first that the proof of part ii(b) in Theorem 4.17 holds without restriction, so that we need only show that

$$\langle \text{A1, A2, A12}' \rangle \subseteq \langle \text{MODEL E2} \rangle$$

whenever $|K| = 1$. Moreover, since the proof of part ii(a) in Theorem 4.17 established this result for all cases with either $|I| > 2$ or $|J| > 2$, it suffices to assume that $|I| = 2 = |J|$, say $I = \{a, b\}$ and $J = \{d, h\}$. Finally, since $\langle$A12$'\rangle \subseteq \langle$A12$\rangle$ by (4.142), it follows from Theorem 4.15 that $\langle$A1, A2, A12$'\rangle \subseteq \langle$MODEL E1$\rangle$, and hence that for each $\boldsymbol{P}$ satisfying (A1, A2, A12$'$) and cost configuration, $c \in C$, there exists a scalar exponent, θ_c, and positive functions, $A_c : I \to R_{++}$, $B_c : J \to R_{++}$, such that the frequency process, $\boldsymbol{N_P} = \{N^c : c \in C\}$, generated by $\boldsymbol{P}$ satisfies,

$$\mathrm{E}(N^c_{ij}) = A_c(i) B_c(j) \exp[-\theta_c c_{ij}], \tag{4.418}$$

for all $c \in C$ and $ij \in I \times J$. Thus it suffices to show that the scalar exponent, θ_c, in (4.418) can always be chosen to be a constant independent of c. To do so, recall first from (1.6) that for this scalar case the set of cost values, V, may be either R or R_+. Next observe that since each cost configuration is of the form, $c = (c_{ad}, c_{ah}, c_{bd}, c_{bh}) \in C = V^{I \times J}$, it follows from (4.418) that

$$\frac{\mathrm{E}(N_{ad}^c)\mathrm{E}(N_{bh}^c)}{\mathrm{E}(N_{ah}^c)\mathrm{E}(N_{bd}^c)} = \exp\{-\theta_c[c_{ad} + c_{bh} - c_{ah} - c_{bd}]\}. \tag{4.419}$$

Hence for each cost configuration, $c \in C$, the quantity

$$\Delta_c = c_{ad} + c_{bh} - c_{ah} - c_{bd} \tag{4.420}$$

is of special importance in the present analysis. In particular, it is convenient to distinguish the subclass of cost configurations

$$C^\circ = \{c \in C : \Delta_c \neq 0\} \tag{4.421}$$

with nonzero values of Δ_c. For such configurations, notice that [as in Example 4.1] if we let $\min(V_c) = \min\{c_{ij} : ij \in I \times J\}$ and $\max(V_c) = \max\{c_{ij} : ij \in I \times J\}$, then by definition,

$$c \in C^\circ \Rightarrow \min(V_c) < \max(V_c). \tag{4.422}$$

Notice also that since the complement set, $C - C^\circ = \{c \in C : \Delta_c = 0\}$, is closed in $C(= R^{I \times J}$ or $R_+^{I \times J})$, it follows that C° is an open set in C. In order to simplify the analysis to follow, it is convenient to introduce one further restriction on configurations in C°, namely, that all configuration values be *distinct*. Hence we now denote this subclass of distinct-valued configurations by

$$D^\circ = \{c \in C^\circ : c_{ij} \neq c_{gk} \text{ for all distinct } ij, gk \in I \times J\}. \tag{4.423}$$

Notice that since the complement set, $C^\circ - D^\circ = \{c \in C^\circ : c_{ij} = c_{gk}$ for some distinct $ij, gk \in I \times J\}$, is closed in C°, it follows that D° is an open subset of C° (and hence of C). With this definition, we begin [as in the proof of Theorem 4.15] by considering the *rational-valued* configurations, $D^\circ(Q) = D^\circ \cap Q^{I \times J}$, in D°. To analyze this subclass, observe first that since D° is open in C and since $Q^{I \times J}$ is dense in $C(= R^{I \times J}$ or $R_+^{I \times J})$, it follows that $D^\circ(Q)$ is *dense* in D°, and in particular that $D^\circ(Q)$ is nonempty. With these observations, our first objective is to show that if $\boldsymbol{P}$ satisfies A12′, then for any two configurations, $c, c' \in D^\circ(Q)$, with

$$\min(V_{c'}) \leq \min(V_c), \quad \max(V_c) \leq \max(V_{c'}), \tag{4.424}$$

the scalar exponents, θ_c and $\theta_{c'}$, in (4.418) must be the same. To do so, we begin by constructing spatial interaction patterns, $s, t, s', t' \in S$, satisfying all the conditions of A12′ for these configurations. With this end in mind, choose any rational value, r, with $\min(V_c) < r < \max(V_c)$ [which is always possible in view of (4.422) together with the denseness of Q in R], and observe that r may always be expressed as a *positive convex combination* of the cost values, $\{c_{ij} : ij \in I \times J\}$, i.e., that there must exist positive numbers, $\{\alpha_{ij} : ij \in I \times J\} \subseteq R_{++}$, satisfying the following conditions:

$$\sum_{ij} \alpha_{ij} c_{ij} = r, \quad \sum_{ij} \alpha_{ij} = 1. \tag{4.425}$$

[In particular, if the set of distinct scalar cost values for any configuration, $c \in D^\circ$, is denoted by $V_c = \{x < y < z < w\}$, then $x < r < w$ implies the existence of $\sigma, \beta \in (0,1)$ with $[\sigma x + (1-\sigma)y] < r < [\beta z + (1-\beta)w]$, so that for some $\lambda \in (0,1)$ we must have $\lambda[\sigma x + (1-\sigma)y] + (1-\lambda)[\beta z + (1-\beta)w] = r$, and may take the appropriate α-values to be $\{\lambda\sigma,\ \lambda(1-\sigma),\ (1-\lambda)\beta,\ (1-\lambda)(1-\beta)\} \subseteq R_{++}$]. Next observe from (4.424) that we must also have $\min(V_{c'}) < r < \max(V_{c'})$, and hence by the same argument that there must also exist a positive set of numbers, $\{\alpha'_{ij} : ij \in I \times J\}$, satisfying

$$\sum_{ij} \alpha'_{ij} c'_{ij} = r, \quad \sum_{ij} \alpha'_{ij} = 1. \tag{4.426}$$

Given these positive solutions to (4.425) and (4.426), recall from part (v) of Lemma 4.2, together with the rationality of both r and $c, c' \in D^\circ(Q)$, that there must in fact exist *positive rational* solutions to (4.425) and (4.426). Hence we may assume that $\{\alpha_{ij} : ij \in I \times J\} \cup \{\alpha'_{ij} : ij \in I \times J\} \subseteq Q_{++}$. Next, we employ these solutions to construct a second pair of positive rational solutions $\{\beta_{ij} : ij \in I \times J\}$ and $\{\beta'_{ij} : ij \in I \times J\}$ as follows. Let Δ_c and $\Delta_{c'}$ be defined as in (4.420), and observe $c, c' \in D^\circ(Q)$ imply that both Δ_c and $\Delta_{c'}$ are nonzero rational numbers, so that the ratio, $\sigma = \Delta_c/\Delta_{c'} \in Q$, is well defined and satisfies

$$\Delta_c = \sigma \Delta_{c'} \neq 0. \tag{4.427}$$

Next observe from the positivity of all α_{ij} and α'_{ij} that there is some positive rational, $\varepsilon \in Q_{++}$, sufficiently small to ensure both that $\varepsilon < \min\{\alpha_{ij} : ij \in I \times J\}$ and $\varepsilon|\sigma| < \min\{\alpha'_{ij} : ij \in I \times J\}$. For this choice of ε, note from (4.425) and (4.426) that $\sum_{ij} \alpha_{ij} c_{ij} = \sum_{ij} \alpha'_{ij} c'_{ij}$, and hence from (4.427) that

$$\begin{aligned} &\sum_{ij} \alpha_{ij} c_{ij} + \varepsilon \Delta_c = \sum_{ij} \alpha'_{ij} c'_{ij} + \varepsilon\sigma\Delta_{c'} \\ &\Rightarrow c_{ad}(\alpha_{ad} + \varepsilon) + c_{ah}(\alpha_{ah} - \varepsilon) + c_{bd}(\alpha_{bd} - \varepsilon) + c_{bh}(\alpha_{bh} + \varepsilon) \\ &= c'_{ad}(\alpha'_{ad} + \varepsilon\sigma) + c'_{ah}(\alpha'_{ah} - \varepsilon\sigma) + c'_{bd}(\alpha'_{bd} - \varepsilon\sigma) \\ &\qquad + c'_{bh}(\alpha'_{bh} + \varepsilon\sigma). \end{aligned} \tag{4.428}$$

Thus letting

$$\begin{aligned}(\beta_{ad}, \beta_{ah}, \beta_{bd}, \beta_{bh}) &= (\alpha_{ad} + \varepsilon,\ \alpha_{ah} - \varepsilon,\ \alpha_{bd} - \varepsilon,\ \alpha_{bh} + \varepsilon) \\ (\beta'_{ad}, \beta'_{ah}, \beta'_{bd}, \beta'_{bh}) &= (\alpha'_{ad} + \varepsilon\sigma,\ \alpha'_{ah} - \varepsilon\sigma,\ \alpha'_{bd} - \varepsilon\sigma,\ \alpha'_{bh} + \varepsilon\sigma),\end{aligned} \tag{4.429}$$

it may be verified by inspection that $\{\beta_{ij} : ij \in I \times J\}$ and $\{\beta'_{ij} : ij \in I \times J\}$ are also positive rational solutions to (4.425) and (4.426), respectively. Hence observing that the finite set of positive rational numbers $\{\alpha_{ij} : ij \in I \times J\} \cup \{\alpha'_{ij} : ij \in I \times J\} \cup \{\beta_{ij} : ij \in I \times J\} \cup \{\beta'_{ij} : ij \in I \times J\}$ must have a representation with common denominator, $M \in Z_{++}$, say

$$\alpha_{ij} = n_{ij}/M, \quad \alpha'_{ij} = n'_{ij}/M, \quad \beta_{ij} = m_{ij}/M, \quad \beta'_{ij} = m'_{ij}/M, \tag{4.430}$$

we see from (4.425) through (4.430) that

$$\sum_{ij} n_{ij} c_{ij} = \sum_{ij} n'_{ij} c'_{ij}, \tag{4.431}$$

$$\sum_{ij} m_{ij} c_{ij} = \sum_{ij} m'_{ij} c'_{ij}, \tag{4.432}$$

and from the normalization conditions in (4.425) and (4.426) that

$$\sum_{ij} n_{ij} = M = \sum_{ij} n'_{ij} = \sum_{ij} m_{ij} = \sum_{ij} m'_{ij}. \tag{4.433}$$

With these observations, we may now choose the desired spatial interaction patterns, $s, t, s', t' \in S_{M'}$, be any patterns with origin-destination frequencies

$$N_{ij}(s) = n_{ij},\ N_{ij}(t) = m_{ij},\ N_{ij}(s') = n'_{ij},\ N_{ij}(t') = m'_{ij}. \tag{4.434}$$

[Note that the conditions, $c \in D^\circ$ and $c' \in D^\circ$, ensure that all cost values in the respective sets, V_c and $V_{c'}$, are distinct. Hence there is a one-to-one correspondence between each $ij \in I \times J$ and the corresponding cost values, $c_{ij} \in V_c$ and $c'_{ij} \in V_{c'}$, so that the cost-value frequencies $(n_{ij}), (m_{ij}), (n'_{ij})$ and (m'_{ij}) in (4.434) must yield well defined origin-destination frequencies.] For this choice of spatial interaction patterns, it follows at once from (4.431) and (4.432) that $c(s) = c'(s')$ and $c(t) = c'(t')$, and from (4.433) that all patterns are comparable. Moreover, the above construction also ensures that $A(s) = A(t)$ and $A(s') = A(t')$. To see this, observe from (4.429), (4.430), and (4.434) that

$$\begin{aligned}N_{ad}(s) + N_{ah}(s) &= M[\alpha_{ad} + \alpha_{ah}] \\ &= M[(\alpha_{ad} + \varepsilon) + (\alpha_{ah} - \varepsilon)] \\ &= M[\beta_{ad} + \beta_{ah}] \\ &= N_{ad}(t) + N_{ah}(t)\end{aligned} \tag{4.435}$$

and hence that the origin activity levels $N_a(s)$ and $N_a(t)$ are the same. In exactly the same manner, it may also be verified from (4.429), (4.430), and (4.434) that all origin and destination activity levels for s and t are identical, and similarly for s' and t'. Hence $A(s) = A(t)$ and $A(s') = A(t')$, which together with $c(s) = c'(s')$ and $c(t) = c'(t')$ imply from A12′ that

$$\frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.436}$$

But if we now observe from (4.428), (4.429), (4.430), and (4.434) that

$$\begin{aligned} c(s) - c(t) &= \sum_{ij} N_{ij}(s)c_{ij} - \sum_{ij} N_{ij}(t)c_{ij} \\ &= \sum_{ij}(M\alpha_{ij})c_{ij} - \sum_{ij}(M\beta_{ij})c_{ij} \\ &= M[\sum_{ij}\alpha_{ij}c_{ij}] - M[\sum_{ij}\alpha_{ij}c_{ij} + \varepsilon\Delta_c] \\ &= -M\varepsilon\Delta_c, \end{aligned} \tag{4.437}$$

and similarly that

$$c'(s') - c'(t') = -M\varepsilon\sigma\Delta_{c'}, \tag{4.438}$$

then by combining (4.436), (4.437), and (4.438) with (4.418), (4.369) and (4.391), we see that

$$\begin{aligned} &\frac{P_c(s)}{P_c(t)} = \frac{P_{c'}(s')}{P_{c'}(t')} \\ &\quad \Rightarrow \exp\{-\theta_c[c(s) - c(t)]\} = \exp\{-\theta_{c'}[c'(s') - c'(t')]\} \\ &\quad \Rightarrow \theta_c[c(t) - c(s)] = \theta_{c'}[c'(t') - c'(s')] \\ &\quad \Rightarrow \theta_c[M\varepsilon\Delta_c] = \theta_{c'}[M\varepsilon\sigma\Delta_{c'}]. \end{aligned} \tag{4.439}$$

Finally, since (4.427) together with the positivity of M and ε imply that $M\varepsilon\Delta_c = M\varepsilon\sigma\Delta_{c'} \neq 0$, we may conclude from (4.439) that $\theta_c = \theta_{c'}$ must hold for all $c, c' \in D^\circ(Q)$ satisfying (4.424). Given this basic identity, observe that if we choose any given $c^\circ \in D^\circ(Q)$ and set $\theta = \theta_{c^\circ}$, then it must be true that $\theta_{c'} = \theta$ for all $c' \in D^\circ(Q)$. For since there must always be some $c'' \in D^\circ(Q)$ with $\min(V_{c''}) \leq \min\{\min(V_{c^\circ}),\ \min(V_{c'})\}$ and $\max(V_{c''}) \geq \max\{\max(V_{c^\circ}), \max(V_{c'})\}$, it follows at once that (4.424) holds for both (c', c'') and (c°, c''), so that by the above result, $\theta_{c'} = \theta_{c''} = \theta_{c^\circ} = \theta$.

Hence it remains to show that this constant value, θ, can be extended to all configurations, $c \in C$. To do so, observe first that since $\Delta_c \neq 0$ for all $c \in C^\circ$, it follows from (4.419) and (4.420) that

$$\theta_c = -\frac{1}{\Delta_c}\log\left[\frac{\mathrm{E}(N^c_{ad})\mathrm{E}(N^c_{bh})}{\mathrm{E}(N^c_{ah})\mathrm{E}(N^c_{bd})}\right] \tag{4.440}$$

for all $c \in C^{\circ}$. Moreover, since Δ_c is continuous in c on C°, and since the regularity condition (R3) together with (3.69) implies that the quantity in brackets is continuous on all of C [for positive processes], we see that θ_c must be a continuous function of c on all of C°. In particular this implies that θ_c is continuous on D°. But since $D^{\circ}(Q)$ is dense in $D^{\circ} \subseteq C^{\circ}$, it follows that for each $c \in D^{\circ}$ there is some sequence (c_n) in $D^{\circ}(Q)$ with $c_n \rightarrow c$, so that by continuity, $\theta_{c_n} \equiv \theta$ must imply that $\theta_c = \theta$ on all of D°. Similarly, since for any $c \in C^{\circ} - D^{\circ}$ there is clearly a sequence (c_n) in D° (i.e., a sequence with all cost values distinct) which converges to c, the same argument shows that $\theta_c = \theta$ on all of C°. Hence it remains only to show that for any $c \in C - C^{\circ}$, we may always set $\theta_c = \theta$. To do so, observe simply that since $\Delta_c = 0$, it now follows from (4.419) and (4.420) that

$$\begin{aligned}
\frac{\mathrm{E}(N_{ad}^c)\mathrm{E}(N_{bh}^c)}{\mathrm{E}(N_{ah}^c)\mathrm{E}(N_{bd}^c)} &= \exp\{-\theta_c[0]\} = \exp\{-\theta[0]\} = \exp\{-\theta\Delta_c\} \\
&= \exp\{-\theta[c_{ad} + c_{bh} - c_{ah} - c_{bd}]\} \qquad (4.441) \\
&= \frac{\exp[-\theta c_{ad}] \cdot \exp[-\theta c_{bh}]}{\exp[-\theta c_{ah}] \cdot \exp[-\theta c_{bd}]}.
\end{aligned}$$

Thus by employing the construction in (4.223) through (4.226) with $F(v) = \exp[-\theta v]$, we may conclude that (4.418) also holds for $c \in C - C^{\circ}$ with $\theta_c = \theta$. Hence we have found a representation with θ_c constant on all of C, and the desired result is established. □

Next we characterize the class of *monotone* deterrence-invariant exponential gravity models. While such models can be characterized by combining strong cost dependence (A12°) with cost efficiency (A13), the following strengthening of cost efficiency incorporates both of these axioms:

A13°. (Strong Cost Efficiency) *For all $n \in Z_{++}$, and all n-tuples of cost configurations, $c, c' \in C^n$, and spatial interaction patterns, $s, s' \in S^n$, with $A(s) = A(s')$,*

$$\boldsymbol{c}(\boldsymbol{s}) \le \boldsymbol{c}'(\boldsymbol{s}') \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) \ge P_{\boldsymbol{c}'}(\boldsymbol{s}'). \qquad (4.442)$$

Observe that if $\boldsymbol{P}$ satisfies A13°, then $\boldsymbol{c}(\boldsymbol{s}) = \boldsymbol{c}(\boldsymbol{s}') \Rightarrow [\boldsymbol{c}(\boldsymbol{s}) \le \boldsymbol{c}(\boldsymbol{s}'), \boldsymbol{c}(\boldsymbol{s}) \ge \boldsymbol{c}(\boldsymbol{s}')] \Rightarrow [P_{\boldsymbol{c}}(\boldsymbol{s}) \ge P_{\boldsymbol{c}}(\boldsymbol{s}'), P_{\boldsymbol{c}}(\boldsymbol{s}) \le P_{\boldsymbol{c}}(\boldsymbol{s}')] \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) = P_{\boldsymbol{c}}(\boldsymbol{s}')$, so that $\boldsymbol{P}$ must also satisfy A12°. Moreover, if for any comparable $s, t, s', t' \in S$ and $c, c' \in C$ it is true that $A(s) = A(t)$, $A(s') = A(t')$, $c(s) \le c'(s')$, and $c(t) \ge c'(t')$, then letting $\boldsymbol{s} = (s, t')$, $\boldsymbol{s}' = (t, s')$, and $\boldsymbol{c} = (c, c') = \boldsymbol{c}'$, it follows from A13° [in a manner paralleling (4.396)] that

$$\begin{aligned}
&[c(s) \le c'(s'),\ c(t) \ge c'(t')] \\
&\quad \Rightarrow \boldsymbol{c}(\boldsymbol{s}) = c(s) + c'(t') \le c(t) + c'(s') = \boldsymbol{c}'(\boldsymbol{s}') \\
&\quad \Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) \ge P_{\boldsymbol{c}'}(\boldsymbol{s}') \Rightarrow P_c(s)P_{c'}(t') \ge P_c(t)P_{c'}(s') \qquad (4.443) \\
&\quad \Rightarrow \frac{P_c(s)}{P_c(t)} \ge \frac{P_{c'}(s')}{P_{c'}(t')},
\end{aligned}$$

so that A13° implies A13′ as well, and we may conclude that

$$\langle \text{A13}^\circ \rangle \subseteq \langle \text{A12}^\circ, \text{A13}' \rangle. \tag{4.444}$$

With these observations we now have the following monotone versions of both Theorems 4.17 and 4.18 above:

Theorem 4.19 (E2* Characterization)
(i) *The class of frequency processes,* ⟨MODEL E2*⟩, *is characterized by the following properties of it generators:*

$$\langle \text{MODEL E2}^* \rangle = \langle \text{A1}, \text{A2}, \text{A13}^\circ \rangle. \tag{4.445}$$

(ii) *If in addition it is true that either* $|I| \geq 3$, $|J| \geq 3$, *or* $|K| = 1$, *then*

$$\langle \text{MODEL E2}^* \rangle = \langle \text{A1}, \text{A2}, \text{A13}' \rangle. \tag{4.446}$$

Proof: **(i).** (a) To establish that $\langle \text{A1}, \text{A2}, \text{A13}^\circ \rangle \subseteq \langle \text{MODEL E2}^* \rangle$, consider any interaction process, $\boldsymbol{P}$, satisfying (A1, A2, A13°) and observe first from the relation $\langle \text{A13}^\circ \rangle \subseteq \langle \text{A12}^\circ \rangle$ together with part (i) of Theorem 4.17 that $\boldsymbol{N_P} \in \langle \text{MODEL E2} \rangle$, and hence that (4.8) holds for some $\theta \in R^K$ and pairs of functions, $A_c : I \to R_{++}$ and $B_c : J \to R_{++}$, for each $c \in C$. Thus it suffices to show that $\theta \in R_+^K$. But if for any separation profiles, $x = (x^k : k \in K)$, $y = (y^k : k \in K) \in V$, and distinct origins, $i, g \in I$, and destinations, $j, h \in J$, we choose interaction patterns, $s, t, s', t' \in S$, and separation configuration, $c \in C$, as in (4.237), then it follows from (4.8) together with $\langle \text{A13}^\circ \rangle \subseteq \langle \text{A13}' \rangle \subseteq \langle \text{A13} \rangle$ and the argument in (4.237) that

$$\begin{aligned}
x - y \geq 0 &\Rightarrow c(s) \leq c(s') \Rightarrow P_c(s) \geq P_c(s') \\
&\Rightarrow \frac{\mathrm{E}(N_{ij}^c)}{\mathrm{E}(N_{ih}^c)} \geq \frac{\mathrm{E}(N_{gj}^c)}{\mathrm{E}(N_{gh}^c)} \\
&\Rightarrow \frac{A_c(i)B_c(j)\exp[-\theta^t y]}{A_c(i)B_c(h)\exp[-\theta^t y]} \geq \frac{A_c(g)B_c(j)\exp[-\theta^t x]}{A_c(g)B_c(h)\exp[-\theta^t y]} \\
&\Rightarrow 1 \geq \exp[-\theta^t(x-y)] \Rightarrow \theta^t(x-y) \geq 0 \\
&\Rightarrow \sum_k \theta^k (x^k - y^k) \geq 0.
\end{aligned} \tag{4.447}$$

Finally, since for any choice of $k \in K$, the definition of V allows x and y to be chosen so that $x^k > y^k$ and $x^{k'} = y^{k'}$ for all $k' \in K$ with $k' \neq k$, it follows that $\theta^k \geq 0$ must hold for each $k \in K$. Hence $\theta = (\theta^k : k \in K) \in R_+^K$, and it follows that $\langle \text{A1}, \text{A2}, \text{A13}^\circ \rangle \in \langle \text{MODEL E2}^* \rangle$.

(b) To establish the converse, simply observe from the argument in (4.414) that if $\boldsymbol{N} \in \langle \text{MODEL E2}^* \rangle$ then for any $\boldsymbol{s}, \boldsymbol{s}' \in S^n$ with $A(\boldsymbol{s}) = A(\boldsymbol{s}')$ and any $\boldsymbol{c} \in C^n$, (4.415) holds for some $\theta \in R_+^K$. Hence it follows that

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{s}) \leq \boldsymbol{c}(\boldsymbol{s}') &\Rightarrow \boldsymbol{c}(\boldsymbol{s}') - \boldsymbol{c}(\boldsymbol{s}) \in R_+^K \\ &\Rightarrow \exp\{-\theta^t[\boldsymbol{c}(\boldsymbol{s}) - \boldsymbol{c}(\boldsymbol{s}')]\} \geq 1 \\ &\Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s})/P_{\boldsymbol{c}}(\boldsymbol{s}') \geq 1 \\ &\Rightarrow P_{\boldsymbol{c}}(\boldsymbol{s}) \geq P_{\boldsymbol{c}}(\boldsymbol{s}'). \end{aligned} \tag{4.448}$$

Thus $\boldsymbol{P_N}$ satisfies A13° whenever $\boldsymbol{N} \in \langle$MODEL E2*$\rangle$, and it follows that $\langle$MODEL E2*$\rangle \subseteq \langle$A1, A2, A13°$\rangle$.

(ii). (a) To establish (4.446) under the additional condition that either $|I| \geq 3$, $|J| \geq 3$, or $|K| = 1$ consider any interaction process, $\boldsymbol{P}$, satisfying (A1, A2, A13$'$). Observe first from the relation $\langle$A13$'\rangle \subseteq \langle$A12$'\rangle$ together with part (ii) of Theorem 4.17 that if either $|I| \geq 3$ or $|J| \geq 3$ then it must be true that $\boldsymbol{N_P} \in \langle$MODEL E2$\rangle$. Similarly from Theorem 4.18, if $|K| = 1$ then again $\boldsymbol{N_P} \in \langle$MODEL E2$\rangle$, so that in all cases (4.8) holds for some vector, $\theta \in R^K$, and pairs of functions, $A_c : I \to R_{++}$ and $B_c : J \to R_{++}$, for each $c \in C$. Hence it suffices to show that $\theta \in R_+^K$. But since $\langle$A13$'\rangle \subseteq \langle$A13$\rangle$ the argument in (4.447) again shows that $\theta \in R_+^K$, and thus that $\langle$A1, A2, A13$'\rangle \subseteq \langle$MODEL E2*$\rangle$.

(b) Conversely, (4.445) together with (4.444) shows at once that

$$\langle \text{MODEL E2}^* \rangle \subseteq \langle \text{A1, A2, A13°} \rangle \subseteq \langle \text{A1, A2, A13}' \rangle$$

This proves the theorem. □

As in the case of Models G2 and G2*, the axiom systems $\langle$A1, A2, A12°$\rangle$ and $\langle$A1, A2, A13°$\rangle$ are strictly stronger than the respective axiom systems $\langle$A1, A2, A12$'\rangle$ and $\langle$A1, A2, A13$'\rangle$, when the conditions of part (ii) fail, i.e., when $|I| = 2 = |J|$ and $|K| \geq 2$. This can be illustrated by means of the following counterexample [which also yields an example satisfying $\langle$A1, A2, A4$'\rangle$ but not $\langle$A1, A2, A3°, A4$\rangle$]:

EXAMPLE 4.3. Nonsufficiency of the Relative Cost Axioms

Suppose that $I = \{a, b\}$, $J = \{d, h\}$, and $V = R^K$ with $|K| \geq 2$, and observe that if we can construct a Poisson frequency process,

$$\boldsymbol{N} \notin \langle \text{MODEL G2} \rangle$$

for which the corresponding generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, satisfies A13$'$, then since $\langle$A13$'\rangle \subseteq \langle$A4$'$, A12$'\rangle$ and $\langle$MODEL G2*$\rangle \cup \langle$MODEL E2*$\rangle \cup \langle$MODEL E2$\rangle \subseteq \langle$MODEL G2$\rangle$, this will also yield a counterexample satisfying both axiom systems (A1, A2, A4$'$) and (A1, A2, A12$'$), but failing to be representable by Models G2*, E2*, or E2. To do so, let $\min(x)$ be defined as in Lemma 4.9 above, and let the k^{th} component of the function,

$\tau : R^K \rightarrow R^K_{++}$, be defined for each positive vector, $x = (x_h : h \in K) \in R^K_{++}$, by

$$\tau_k(x) = \begin{cases} 2, & k = \min\{h \in K : x_h = \min(x)\} \\ 1, & \text{otherwise.} \end{cases} \tag{4.449}$$

Similarly, for each negative vector, x (i.e., with $-x \in R^n_{++}$), let

$$\tau_k(x) = \begin{cases} 2, & k = \min\{h \in K : x_h = \max(x)\} \\ 1, & \text{otherwise,} \end{cases} \tag{4.450}$$

and for all other vectors, $x \in R^K$, let $\tau_k(x) = 1$. With this definition, it may then be verified by inspection that for all $x \in R^K$,

$$\tau(x)^t x = \begin{cases} 1^t x + \min(x), & x \in R^K_{++} \\ 1^t x + \max(x), & -x \in R^K_{++} \\ 1^t x, & \text{otherwise,} \end{cases} \tag{4.451}$$

and hence that the function, $\ell : R^K \rightarrow R$, in (4.163) is given precisely by the inner product:

$$\ell(x) = \tau(x)^t x, \quad x \in R^K. \tag{4.452}$$

With these preliminary observations, we may construct the desired frequency process, $\boldsymbol{N}$, as follows. For each cost configuration, $c \in C = V^{I \times J}$, let

$$\Delta_c = c_{ad} + c_{bh} - c_{ah} - c_{bd} \tag{4.453}$$

[as in (4.420) above], and let the function, $\theta : C \rightarrow R^K_{++}$, be defined for all $c \in C$ by

$$\theta_c = \tau(\Delta_c). \tag{4.454}$$

Given these positive coefficient vectors, we now consider the unique Poisson frequency process, $\boldsymbol{N} = \{N^c : c \in C\}$, with mean frequencies defined for each $c \in C$ and $ij \in I \times J$ by

$$\mathrm{E}(N^c_{ij}) = \exp[-\theta^t_c c_{ij}]. \tag{4.455}$$

If we again set $A_c(i) = 1 = B_c(j)$ for all $c \in C$ and $ij \in I \times J$ (as in Example 4.1), then it follows from (4.455) that $\boldsymbol{N} \in \langle$MODEL E1*$\rangle$. However, $\boldsymbol{N}$ cannot possibly belong to $\langle$MODEL G2$\rangle$. For if there were a function, $F : V \rightarrow R_{++}$, such that (4.2) holds for some choice of origin functions, A_c, and destination functions, B_c, for each $c \in C$, then by combining (4.454) and (4.455) together with (4.242) and (4.441), it would follow that the relation

$$\exp[-\tau(\Delta_c)^t \Delta_c] = \frac{\mathrm{E}(N^c_{ad})\mathrm{E}(N^c_{bh})}{\mathrm{E}(N^c_{ah})\mathrm{E}(N^c_{bd})} = \frac{F(c_{ad})F(c_{bh})}{F(c_{ah})F(c_{bd})} \tag{4.456}$$

must hold identically for all $c \in C$. But if for any $x, y, z \in R_{++}$ we let $c_{ad} = (x, 0, \ldots, 0)$, $c_{bh} = (0, y, 0, \ldots, 0)$, $c_{ah} = (0, 0, -z, \ldots, -z)$, and $c_{bd} = (0,\ldots,0)$ [where $|K| \geq 2$ by hypothesis], so that by definition, $\Delta_c = (x,y,z,\ldots,z) \in R^K_{++}$, then it follows from (4.451) that $\tau(\Delta_c)^t\Delta_c = 1^t\Delta_c + \min(\Delta_c) = [x+y+(k-2)z] + \min(x,y,z)$, and hence from (4.456) that

$$\begin{aligned}&\frac{F(x,0,\ldots,0)F(0,y,0,\ldots,0)}{F(0,0,-z,\ldots,-z)F(0,\ldots,0)} \\ &= \exp\{-[x+y+(k-2)z+\min(x,y,z)]\}.\end{aligned} \tag{4.457}$$

Thus if for each $x \in R_{++}$ we choose any $y,z > x$ [so that $x = \min(x,y,z)$], then (4.457) implies that,

$$F(x,0,\ldots,0) = \exp(-2x)\phi(y,z), \tag{4.458}$$

where

$$\begin{aligned}\phi(y,z) &= \exp\{-[y+(k-2)z]\} \\ &\cdot [F(0,0,-z,\ldots,-z)F(0,\ldots,0)/F(0,y,0,\ldots,0)].\end{aligned}$$

But since the left hand side of (4.458) involves only x, we see that $\phi(y,z)$ must be equal to a constant, which together with the arbitrary choice of $x \in R_{++}$ implies that for some $\alpha > 0$, the relation

$$F(x,0,\ldots,0) = \alpha\exp(-2x) \tag{4.459}$$

must hold identically for all $x \in R_{++}$. Similarly, if for each $y \in R_{++}$ we choose any $x,z > y$ [so that $y = \min(x,y,z)$], then by the same argument it follows that there is some constant, $\beta > 0$, such that the relation

$$F(0,y,0,\ldots,0) = \beta\exp(-2y) = (\beta/\alpha)F(y,0,\ldots,0) \tag{4.460}$$

holds identically for all $y \in R_{++}$. Hence, if for any $x,y,z \in R_{++}$ with $x = \min(x,y,z)$ we now let $\psi(z) = \exp[-(k-2)z]F(0,0,-z,\ldots,-z)F(0,\ldots,0)$, then (4.457) may be rewritten in terms of (4.459) and (4.460) as

$$\begin{aligned}&F(x,0,\ldots,0)(\alpha/\beta)F(y,0,\ldots,0) = \exp(-2x)\exp(-y)\psi(z) \\ &\Rightarrow [\alpha\exp(-2x)](\alpha/\beta)F(y,0,\ldots,0) = \exp(-2x)exp(-y)\psi(z) \\ &\Rightarrow F(y,0,\ldots,0) = [\beta\psi(z)/\alpha^2]\exp(-y).\end{aligned} \tag{4.461}$$

But since the left hand side of (4.461) is independent of z, it again follows that $\psi(z)$ is constant, and hence that for the constant, $\lambda = [\beta\psi(z)/\alpha^2] > 0$, we must have

$$F(y,0,\ldots,0) = \lambda\exp(-y) \tag{4.462}$$

for all $y \in R_{++}$. Finally, (4.462) together with (4.459) is then seen to imply that $\alpha\exp(-2x) = \lambda\exp(-x)$ must hold identically for all $x \in R_{++}$, which together with the positivity of α and λ clearly yields a contradiction. Hence no such function F can exist, and we may conclude that $\boldsymbol{N} \notin \langle\text{MODEL G2}\rangle$.

However, the generator, $\boldsymbol{P_N} = \{P_c : c \in C\}$, of $\boldsymbol{N}$ does satisfy A13$'$. To see this, observe first from the monotonicity and linear homogeneity properties of the function, ℓ, in Lemma 4.9 that for all $\alpha, \beta \in R$ and $x, y \in R^K$

$$\begin{aligned} \alpha x \geq \beta y &\Rightarrow \ell(\alpha x) \geq \ell(\beta y) \\ &\Rightarrow \alpha\ell(x) \geq \beta\ell(y). \end{aligned} \tag{4.463}$$

With this in mind, let us now consider any cost configurations, $c, c' \in C$, and comparable spatial interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(t)$, $A(s') = A(t')$, $c(s) \leq c'(s')$, and $c(t) \geq c'(t')$. First observe that by definition,

$$\begin{aligned} c(s) - c(t) &= \sum_{ij} N_{ij}(s)c_{ij} - \sum_{ij} N_{ij}(t)c_{ij} \\ &= \{N_{ad}(s)c_{ad} + N_{ah}(s)c_{ah} + N_{bd}(s)c_{bd} + N_{bh}(s)c_{bh}\} \\ &\quad - \{N_{ad}(t)c_{ad} + N_{ah}(t)c_{ah} + N_{bd}(t)c_{bd} + N_{bh}(t)c_{bh}\}. \end{aligned} \tag{4.464}$$

Next observe that the origin activity-equivalence conditions,

$$N_{ad}(s) + N_{ah}(s) = N_{ad}(t) + N_{ah}(t) \text{ and } N_{bd}(s) + N_{bh}(s) = N_{bd}(t) + N_{bh}(t)$$

imply that

$$N_{ad}(s) - N_{ad}(t) = N_{ah}(t) - N_{ah}(s) \text{ and } N_{bd}(s) - N_{bd}(t) = N_{bh}(t) - N_{bh}(s),$$

which together with (4.464) yields the relation

$$\begin{aligned} c(s) - c(t) &= [N_{ad}(s) - N_{ad}(t)](c_{ad} - c_{ah}) \\ &\quad + [N_{bd}(s) - N_{bd}(t)](c_{bd} - c_{bh}). \end{aligned} \tag{4.465}$$

But the destination activity-equivalence condition,

$$N_{ad}(s) + N_{bd}(s) = N_{ad}(t) + N_{bd}(t),$$

then implies that

$$N_{bd}(s) - N_{bd}(t) = -[N_{ad}(s) - N_{ad}(t)],$$

which together with (4.453) and (4.465) yields

$$\begin{aligned} c(s) - c(t) &= [N_{ad}(s) - N_{ad}(t)](c_{ad} + c_{bh} - c_{ah} - c_{bd}) \\ &= [N_{ad}(s) - N_{ad}(t)]\Delta_c. \end{aligned} \tag{4.466}$$

In a similar manner, the cost difference, $c'(s') - c'(t')$, can be written as

$$c'(s') - c'(t') = [N_{ad}(s') - N_{ad}(t')]\Delta_{c'}. \tag{4.467}$$

Finally, letting $\alpha = N_{ad}(s) - N_{ad}(t)$, $\beta = N_{ad}(s') - N_{ad}(t')$, $x = \Delta_c$ and $y = \Delta_{c'}$ in (4.463) and recalling that $\boldsymbol{N} \in \langle\text{MODEL E1}^*\rangle$ by construction, we may conclude from (4.391) together with (4.452), (4.454), and (4.463) that

$$\begin{aligned}
&[c(s) \leq c'(s'),\ c(t) \geq c'(t')] \\
&\quad \Rightarrow c(s) - c(t) \leq c'(s') - c'(t') \\
&\quad \Rightarrow [N_{ad}(s) - N_{ad}(t)]\Delta_c \leq [N_{ad}(s') - N_{ad}(t')]\Delta_{c'} \\
&\quad \Rightarrow [N_{ad}(s) - N_{ad}(t)]\ell(\Delta_c) \leq [N_{ad}(s') - N_{ad}(t')]\ell(\Delta_{c'}) \\
&\quad \Rightarrow [N_{ad}(s) - N_{ad}(t)]\theta_c^t\Delta_c \leq [N_{ad}(s') - N_{ad}(t')]\theta_{c'}^t\Delta_{c'} \\
&\quad \Rightarrow \theta_c^t\{[N_{ad}(s) - N_{ad}(t)]\Delta_c\} \leq \theta_{c'}^t\{[N_{ad}(s') - N_{ad}(t')]\Delta_{c'}\} \\
&\quad \Rightarrow \theta_c^t[c(s) - c(t)] \leq \theta_{c'}^t[c'(s') - c'(t')] \\
&\quad \Rightarrow \exp\{-\theta_c^t[c(s) - c(t)]\} \geq \exp\{-\theta_{c'}^t[c'(s') - c'(t')]\} \\
&\quad \Rightarrow \frac{P_c(s)}{P_c(t)} \geq \frac{P_{c'}(s')}{P_{c'}(t')}.
\end{aligned} \tag{4.468}$$

Hence the generator $\boldsymbol{P_N}$ of $\boldsymbol{N}$ does satisfy A13′, and we see that the frequency process $\boldsymbol{N}$ yields the desired counterexample. •

Remark 4.6. Given this counterexample, it is also of interest to note that the frequency process $\boldsymbol{N}$ in Example 4.1 is representable by a monotone *exponential* gravity model, i.e., that $\boldsymbol{N} \in \langle\text{MODEL E1}^*\rangle$. To see this, observe that if for each $c \in C$ and $k \in K$ we let $\theta_c^k = \log[2+\max(V_c^k)-\min(V_c^k)] > 0$, and denote the corresponding coefficient vectors by $\theta_c = (\theta_c^k : k \in K) \in R_{++}^K$, then it follows at once from (4.240) that

$$F_c(v) = \exp[-\sum_{k\in K} \theta_c^k v^k] = \exp[-\theta_c^t v]. \tag{4.469}$$

Hence $\boldsymbol{N} \in \langle\text{MODEL E1}^*\rangle$, and it follows from Theorem 4.16 that the generator $\boldsymbol{P_N}$ of $\boldsymbol{N}$ satisfies A13. However, $\boldsymbol{P_N}$ does not satisfy A13′. To see this, let $|K| = 1$, $c = (c_{ad}, c_{ah}, c_{bd}, c_{bh}) = (2,2,2,2)$, $c' = (c'_{ad}, c'_{ah}, c'_{bd}, c'_{bh}) = (1,2,1,4)$, $s = (ad, bh) = s'$, and $t = (ah, bd) = t'$, so that by construction, $A(s) = A(t)$, $A(s') = A(t')$, $c(s) = 4 < 5 = c'(s')$, and $c(t) = 4 > 3 = c'(t')$. But $\max(V_c) - \min(V_c) = 2 - 2 = 0$ and $\max(V_{c'}) - \min(V_{c'}) = 5 - 1 = 4$ imply from (4.244) that

$$\frac{P_c(s)}{P_c(t)} = [2+0]^{4-4} = 1 < 36 = [2+4]^{5-3} = \frac{P_{c'}(s')}{P_{c'}(t')}. \tag{4.470}$$

Hence the generator $\boldsymbol{P_N}$ of $\boldsymbol{N}$ fails to satisfy A13′, and we see that $\boldsymbol{P_N}$ yields an explicit example of an interaction process satisfying $A13$ but not A13′. •

Given these characterizations of deterrence-invariant exponential gravity models, we next consider the stronger properties of destination-deterrence invariance and origin-deterrence invariance with respect to exponential gravity models. The appropriate characterizations here amount simply to combining the cost dependency and cost efficiency axioms with the corresponding characterizations of destination-deterrence-invariant and origin-deterrence-invariant gravity models for the general case. Turning first to destination-deterrence-invariant exponential gravity models (together with the monotone versions of these models), we have

Theorem 4.20 (E3 and E3* Characterizations)
(i) *The classes of frequency processes,* $\langle \text{MODEL E3} \rangle$ *and* $\langle \text{MODEL E3}^* \rangle$*, are each characterized by the following properties of their generators:*

$$\langle \text{MODEL E3} \rangle = \langle \text{A1}, \text{A2}, \text{A6}, \text{A12}' \rangle, \tag{4.471}$$
$$\langle \text{MODEL E3}^* \rangle = \langle \text{A1}, \text{A2}, \text{A6}, \text{A13}' \rangle. \tag{4.472}$$

(ii) *If in addition it is true that* $|J| \geq 3$*, then*

$$\langle \text{MODEL E3} \rangle = \langle \text{A1}, \text{A2}, \text{A6}, \text{A8}, \text{A12} \rangle, \tag{4.473}$$
$$\langle \text{MODEL E3}^* \rangle = \langle \text{A1}, \text{A2}, \text{A6}, \text{A8}, \text{A13} \rangle. \tag{4.474}$$

PROOF: **(i).** (a) To establish that $\langle \text{A1}, \text{A2}, \text{A6}, \text{A12}' \rangle \subseteq \langle \text{MODEL E3} \rangle$, we begin by recalling that $\langle \text{A12}' \rangle \subseteq \langle \text{A3}' \rangle \subseteq \langle \text{A3}^{\#} \rangle$, and hence that $\langle \text{A1}, \text{A2}, \text{A6}, \text{A12}' \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}^{\#}, \text{A6} \rangle$. Thus by Theorem 4.9(i) we see that for each $\boldsymbol{P}$ satisfying (A1, A2, A6, A12′), the frequency process $\boldsymbol{N_P} = \{N^c : c \in C\}$, generated by $\boldsymbol{P}$ must be an element of $\langle \text{MODEL G3} \rangle$, and hence may conclude that there exist functions, $B : I \to R_{++}$, $F : V \to R_{++}$, and $A_c : I \to R_{++}$ for each $c \in C$ such that the relation

$$\text{E}(N_{ij}^c) = A_c(i)B(j)F(c_{ij}) \tag{4.475}$$

holds for all $c \in C$ and $ij \in I \times J$. But, as with the representation in (4.403), this in turn implies that (4.404) must hold for all $c \in C$ and $ij, gh \in I \times J$. Moreover, since $\boldsymbol{P}$ satisfies A12′ and $\langle \text{A12}' \rangle \subseteq \langle \text{A12} \rangle$, the argument in (4.405) through (4.411) continues to hold for $\boldsymbol{P}$. Hence replacing (4.403) by (4.475) above, and employing the same definitions of θ and $A_c^*(i)$ as (4.412), it now follows that the frequency process $\boldsymbol{N_P}$ generated by $\boldsymbol{P}$ must satisfy

$$\text{E}(N_{ij}^c) = A_c^*(i)B(j)\exp[-\theta^t c_{ij}], \tag{4.476}$$

for all $c \in C$ and $ij \in I \times J$, so that $\boldsymbol{N_P} \in \langle \text{MODEL E3} \rangle$. Similarly, to establish that $\langle \text{A1}, \text{A2}, \text{A6}, \text{A13}' \rangle \subseteq \langle \text{MODEL E3}^* \rangle$, we need simply observe that $\langle \text{A13}' \rangle \subseteq \langle \text{A12}' \rangle \Rightarrow \langle \text{A1}, \text{A2}, \text{A6}, \text{A13}' \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A6}, \text{A12}' \rangle$ so that (4.476) continues to hold. Moreover, since $\langle \text{A13}' \rangle \subseteq \langle \text{A13} \rangle$, the argument in (4.447) with (4.476) replacing (4.8) again shows that $\theta \in R_+^K$, and hence that for each $\boldsymbol{P}$ satisfying $\langle \text{A1}, \text{A2}, \text{A6}, \text{A13}' \rangle$ we must have $\boldsymbol{N_P} \in \langle \text{MODEL E3}^* \rangle$.

(b) Finally, to establish that $\langle\text{MODEL E3}\rangle \subseteq \langle\text{A1, A2, A6, A12}'\rangle$ we need only observe from (4.289) and (4.401) together with (4.395) that

$$\begin{aligned}\langle\text{MODEL E3}\rangle &\subseteq \langle\text{MODEL G3}\rangle \cap \langle\text{MODEL E2}\rangle \\ &\subseteq \langle\text{A1, A2, A6}\rangle \cap \langle\text{A1, A2, A12}^\circ\rangle \subseteq \langle\text{A1, A2, A6, A12}'\rangle.\end{aligned}$$

Similarly, it follows from (4.289) and (4.445) together with (4.444) that

$$\begin{aligned}\langle\text{MODEL E3}^*\rangle &\subseteq \langle\text{MODEL G3}\rangle \cap \langle\text{MODEL E2}^*\rangle \\ &\subseteq \langle\text{A1, A2, A6}\rangle \cap \langle\text{A1, A2, A13}^\circ\rangle \subseteq \langle\text{A1, A2, A6, A13}'\rangle.\end{aligned}$$

(ii). (a) To establish that $\langle\text{A1, A2, A6, A8, A12}\rangle \subseteq \langle\text{MODEL E3}\rangle$ whenever $|J| \geq 3$, observe from Theorem 4.9(ii) that for each $\boldsymbol{P}$ satisfying (A1, A2, A6, A8), it must be true that $\boldsymbol{N_P} \in \langle\text{MODEL G3}\rangle$, and hence that (4.475) continues to hold for $\boldsymbol{N_P}$. Moreover, since $\boldsymbol{P}$ satisfies A12, the argument following (4.475) also continues to hold, and we may again conclude that $\boldsymbol{N_P} \subseteq \langle\text{MODEL E3}\rangle$. Similarly, to establish that

$$\langle\text{A1, A2, A6, A8, A13}\rangle \subseteq \langle\text{MODEL E3}^*\rangle$$

when $|J| \geq 3$, observe that

$$\langle\text{A13}\rangle \subseteq \langle\text{A12}\rangle \Rightarrow \langle\text{A1, A2, A6, A8, A13}\rangle \subseteq \langle\text{MODEL E3}\rangle.$$

Hence (4.9) holds for $\boldsymbol{N_P}$, and it follows from A13 and the argument in (4.447)(with (4.9) replacing (4.8)) that $\theta \in R_+^K$. Thus, for each $\boldsymbol{P}$ satisfying $\langle\text{A1, A2, A6, A8, A13}\rangle$ we must have $\boldsymbol{N_P} \in \langle\text{MODEL E3}^*\rangle$.

(b) To establish that $\langle\text{MODEL E3}\rangle \subseteq \langle\text{A1, A2, A6, A8, A12}\rangle$, recall from (4.142) together with the development of $\text{A3}^\#$ that $\langle\text{A12}'\rangle \subseteq \langle\text{A3}', \text{A12}\rangle \subseteq \langle\text{A3}^\#, \text{A12}\rangle$ and hence from (4.309) and (4.471) that

$$\begin{aligned}\langle\text{MODEL E3}\rangle &= \langle\text{A1, A2, A6, A12}'\rangle \\ &\subseteq \langle\text{A1, A2, A3}^\#\text{, A6, A12}\rangle \subseteq \langle\text{A1, A2, A6, A8, A12}\rangle.\end{aligned}$$

Moreover, since this together with $\langle\text{A13}'\rangle \subseteq \langle\text{A12}'\rangle$ implies that

$$\begin{aligned}\langle\text{MODEL E3}^*\rangle &= \langle\text{A1, A2, A6, A13}'\rangle \subseteq \langle\text{A1, A2, A3}^\#\text{, A6, A13}\rangle \\ &\subseteq \langle\text{A1, A2, A6, A8, A13}\rangle,\end{aligned}$$

we also have $\langle\text{MODEL E3}^*\rangle \subseteq \langle\text{A1, A2, A6, A8, A13}\rangle$. □

Remark 4.7. It should be noted from the proof of part (i) above that certain weaker axiomatizations are possible. In particular, the proof shows that

$$\langle\text{MODEL E3}\rangle = \langle\text{A1, A2, A3}^\#\text{, A6, A12}\rangle$$

and $\langle\text{MODEL E3}^*\rangle = \langle \text{A1}, \text{A2}, \text{A3}^{\#}, \text{A6}, \text{A13}\rangle$. •

Given this characterization of destination-deterrence-invariant exponential gravity models, it should be clear from the parallel between Theorems 4.9 and 4.11 (and between Theorems 4.10 and 4.12) above that the characterization of origin-deterrence-invariant exponential gravity models (together with the monotone versions of these models) amounts essentially to replacing J by I and replacing (A6,A10) by (A7,A11):

Theorem 4.21 (E4 and E4* Characterizations)
(i) *The classes of frequency processes,* $\langle\text{MODEL E4}\rangle$ *and* $\langle\text{MODEL E4}^*\rangle$, *are each characterized by the following properties of their generators:*

$$\langle\text{MODEL E4}\rangle \;= \langle \text{A1}, \text{A2}, \text{A7}, \text{A12}'\rangle, \tag{4.477}$$

$$\langle\text{MODEL E4}^*\rangle \;= \langle \text{A1}, \text{A2}, \text{A7}, \text{A13}'\rangle. \tag{4.478}$$

(ii) *If in addition it is true that* $|I| \geq 3$, *then*

$$\langle\text{MODEL E4}\rangle = \langle \text{A1}, \text{A2}, \text{A7}, \text{A9}, \text{A12}\rangle, \tag{4.479}$$

$$\langle\text{MODEL E4}^*\rangle = \langle \text{A1}, \text{A2}, \text{A7}, \text{A9}, \text{A13}\rangle. \tag{4.480}$$

PROOF: **(i).** (a) As in part i(a) of Theorem 4.20, $\langle \text{A1}, \text{A2}, \text{A7}, \text{A12}'\rangle \subseteq \langle \text{A1}, \text{A2}, \text{A3}^{\#}, \text{A7}\rangle$, so that by Theorem 4.11(i) we see that for each $\boldsymbol{P}$ satisfying $(\text{A1}, \text{A2}, \text{A7}, \text{A12}')$, the frequency process $\boldsymbol{N_P} = \{N_c : c \in C\}$, generated by $\boldsymbol{P}$ must be an element of $\langle\text{MODEL G4}\rangle$, and hence may conclude that there exist functions, $A : I \to R_{++}$, $F : V \to R_{++}$, and $B_c : J \to R_{++}$ for each $c \in C$ such that the relation

$$\mathrm{E}(N_{ij}^c) = A(i)B_c(j)F(c_{ij}) \tag{4.481}$$

holds for all $c \in C$ and $ij \in I \times J$. But, as with the representation in (4.475) above, this in turn implies that (4.404) again holds for all $c \in C$ and $ij, gh \in I \times J$. Hence, by the same argument, if we now set $B_c^*(j) = B_c(j)\exp[q(0)]$ and employ the representation in (4.412), then it follows that the frequency process $\boldsymbol{N_P}$ generated by $\boldsymbol{P}$ must satisfy

$$\mathrm{E}(N_{ij}^c) = A(i)B_c^*(j)\exp[-\theta^t c_{ij}], \tag{4.482}$$

for all $c \in C$ and $ij \in I \times J$, so that $\boldsymbol{N_P} \in \langle\text{MODEL E4}\rangle$. Similarly, to establish that

$$\langle \text{A1}, \text{A2}, \text{A7}, \text{A13}'\rangle \subseteq \langle\text{MODEL E4}^*\rangle,$$

the same argument as part i(a) of Theorem 4.20 now shows that

$$\langle \text{A1}, \text{A2}, \text{A7}, \text{A13}'\rangle \subseteq \langle\text{MODEL E4}\rangle.$$

Moreover, since $\langle \text{A13}'\rangle \subseteq \langle \text{A13}\rangle$, the argument in (4.447) with (4.10) replacing (4.8) again shows that $\theta \in R_+^K$, and hence that for each $\boldsymbol{P}$ satisfying $\langle \text{A1}, \text{A2}, \text{A7}, \text{A13}'\rangle$ we must have $\boldsymbol{N_P} \in \langle\text{MODEL E4}^*\rangle$.

(b) To establish that $\langle\text{MODEL E4}\rangle \subseteq \langle\text{A1, A2, A7, A12}'\rangle$ we need only observe from (4.318) and (4.401) together with (4.395) that

$$\langle\text{MODEL E4}\rangle \subseteq \langle\text{MODEL G4}\rangle \cap \langle\text{MODEL E2}\rangle$$
$$\subseteq \langle\text{A1, A2, A7}\rangle \cap \langle\text{A1, A2, A12}^\circ\rangle \subseteq \langle\text{A1, A2, A7, A12}'\rangle.$$

Similarly, it follows from (4.318) and (4.445) together with (4.444) that

$$\langle\text{MODEL E4}^*\rangle \subseteq \langle\text{MODEL G4}\rangle \cap \langle\text{MODEL E2}^*\rangle$$
$$\subseteq \langle\text{A1, A2, A7}\rangle \cap \langle\text{A1, A2, A13}^\circ\rangle \subseteq \langle\text{A1, A2, A7, A13}'\rangle.$$

(ii). (a) To establish that $\langle\text{A1, A2, A7, A9, A12}\rangle \subseteq \langle\text{MODEL E4}\rangle$ whenever $|I| \geq 3$, observe from Theorem 4.11(ii) that for each $\boldsymbol{P}$ satisfying (A1,A2,A7,A9), it must be true that $\boldsymbol{N_P} \in \langle\text{MODEL G4}\rangle$, and hence that (4.481) continues to hold for $\boldsymbol{N_P}$. Moreover, since $\boldsymbol{P}$ satisfies A12, the argument following (4.481) also continues to hold, and we may again conclude that $\boldsymbol{N_P} \in \langle\text{MODEL E4}\rangle$. Similarly, to establish that

$$\langle\text{A1, A2, A7, A9, A13}\rangle \subseteq \langle\text{MODEL E4}^*\rangle$$

when $|I| \geq 3$, observe from the above result that $\langle\text{A13}\rangle \subseteq \langle\text{A12}\rangle \Rightarrow \langle\text{A1, A2, A7, A9, A13}\rangle \subseteq \langle\text{MODEL E4}\rangle$. Hence (4.10) holds for $\boldsymbol{N_P}$, and it again follows from A13 and the argument in (4.447) [with (4.10) replacing (4.8)] that $\theta \in R_+^K$. Therefore, for each $\boldsymbol{P}$ satisfying $\langle\text{A1, A2, A7, A9, A13}\rangle$ we must have $\boldsymbol{N_P} \in \langle\text{MODEL E4}^*\rangle$.

(b) Finally, to establish that $\langle\text{MODEL E4}\rangle \subseteq \langle\text{A1, A2, A7, A9, A12}\rangle$, recall from part ii(b) of Theorem 4.20 that $\langle\text{A12}'\rangle \subseteq \langle\text{A3}^\#\text{, A12}\rangle$ and hence from (4.331) and (4.477) that

$$\langle\text{MODEL E4}\rangle = \langle\text{A1, A2, A7, A12}'\rangle \subseteq \langle\text{A1, A2, A3}^\#\text{, A7, A12}\rangle$$
$$\subseteq \langle\text{A1, A2, A7, A9, A12}\rangle.$$

Moreover, since this together with $\langle\text{A13}'\rangle \subseteq \langle\text{A12}'\rangle$ implies that

$$\langle\text{MODEL E4}^*\rangle = \langle\text{A1, A2, A7, A13}'\rangle \subseteq \langle\text{A1, A2, A3}^\#\text{, A7, A13}\rangle$$
$$\subseteq \langle\text{A1, A2, A7, A9, A13}\rangle,$$

we also have $\langle\text{MODEL E4}^*\rangle \subseteq \langle\text{A1, A2, A7, A9, A13}\rangle$. □

Remark 4.8. As in Remark 4.7 above, this proof also shows that

$$\langle\text{MODEL E4}\rangle = \langle\text{A1, A2, A3}^\#\text{, A7, A12}\rangle$$

and $\langle\text{MODEL E4}^*\rangle = \langle\text{A1, A2, A3}^\#\text{, A7, A13}\rangle$. •

Given these characterizations of exponential gravity models which are relatively invariant with respect to either origins or destinations, we turn now to the case of exponential gravity models which are relatively invariant with respect to both origins and destinations:

Theorem 4.22 (E5 and E5* Characterizations) *The classes of frequency processes,* $\langle$MODEL E5$\rangle$ *and* $\langle$MODEL E5*$\rangle$*, are each characterized by the following properties of their generators:*

$$\langle \text{MODEL E5} \rangle = \langle \text{A1}, \text{A2}, \text{A12}'' \rangle, \tag{4.483}$$

$$\langle \text{MODEL E5}^* \rangle = \langle \text{A1}, \text{A2}, \text{A13}'' \rangle. \tag{4.484}$$

PROOF: (a) To establish that $\langle \text{A1}, \text{A2}, \text{A12}'' \rangle \subseteq \langle \text{MODEL E5} \rangle$, consider any process $\boldsymbol{P}$ satisfying $\langle \text{A1}, \text{A2}, \text{A12}'' \rangle$, and recall first from (4.142) that $\langle \text{A12}'' \rangle \subseteq \langle \text{A3}'' \rangle$, and hence from expression (4.246) in Theorem 4.5 that there must exist functions $\eta : C \to R_{++}$, $A : I \to R_{++}$, $B : J \to R_{++}$, and $F : V \to R_{++}$ such that for all $c \in C$ and $ij \in I \times J$

$$\text{E}_c(N_{ij}) = \eta(c)A(i)B(j)F(c_{ij}). \tag{4.485}$$

But, as in Theorem 4.17, (4.485) implies that (4.404) again holds for all $c \in C$ and $ij, gh \in I \times J$. Moreover since $\langle \text{A12}'' \rangle \subseteq \langle \text{A12} \rangle$ it again follows that $\boldsymbol{P}$ satisfies A12, and thus that the argument in expressions (4.405) through (4.411) continue to hold. Hence if the function, $\lambda : C \to R_{++}$, is defined for all $c \in C$ by $\lambda(c) = \eta(c)\exp[q(0)] > 0$, then expression (4.412) now takes the form

$$\text{E}(N_{ij}^c) = \lambda(c)A(i)B(j)\exp[-\theta^t c_{ij}] \tag{4.486}$$

for all $c \in C$ and $ij \in I \times J$. Thus, $\boldsymbol{N_P} \in \langle \text{MODEL E5} \rangle$ whenever $\boldsymbol{P}$ satisfies $(\text{A1}, \text{A2}, \text{A12}'')$, and we must have $\langle \text{A1}, \text{A2}, \text{A12}'' \rangle \subseteq \langle \text{MODEL E5} \rangle$. Similarly, to establish that $\langle \text{A1}, \text{A2}, \text{A13}'' \rangle \subseteq \langle \text{MODEL E5}^* \rangle$, recall from (4.145) that $\langle \text{A13}'' \rangle \subseteq \langle \text{A12}'' \rangle$ and hence that $\langle \text{A1}, \text{A2}, \text{A13}'' \rangle \subseteq \langle \text{MODEL E5} \rangle$. Moreover, since $\langle \text{A13}'' \rangle \subseteq \langle \text{A13} \rangle$, the argument in (4.447) with (4.11) replacing (4.8) again shows that $\theta \in R_+^K$, and hence that for each $\boldsymbol{P}$ satisfying $\langle \text{A1}, \text{A2}, \text{A13}'' \rangle$ we must have $\boldsymbol{N_P} \in \langle \text{MODEL E5}^* \rangle$.

(b) Conversely, to establish that $\langle \text{MODEL E5} \rangle \subseteq \langle \text{A1}, \text{A2}, \text{A12}'' \rangle$, observe first that with respect to expression (4.11), the argument in (4.369) now shows that for all $c \in C$ and comparable $s, t \in S$,

$$\frac{P_c(s)}{P_c(t)} = \frac{\lambda(c)^{N(s)} \left(\prod_i A(i)^{N_i(s)}\right) \left(\prod_j B(j)^{N_j(s)}\right) \exp[-\theta^t c(s)]}{\lambda(c)^{N(t)} \left(\prod_i A(i)^{N_i(t)}\right) \left(\prod_j B(j)^{N_j(t)}\right) \exp[-\theta^t c(t)]}. \tag{4.487}$$

Hence for any comparable interaction patterns, $s, t, s', t' \in S$, with $A(s) = A(s')$, $A(t) = A(t')$, and any separation configurations, $c, c' \in C$, with

$c(s) = c'(s')$ and $c(t) = c'(t')$, we see from (4.487) that

$$\frac{P_c(s)}{P_c(t)} = \frac{\lambda(c)^{N(s)} \left(\prod_i A(i)^{N_i(s)}\right) \left(\prod_j B(j)^{N_j(s)}\right) \exp[-\theta^t c(s)]}{\lambda(c)^{N(t)} \left(\prod_i A(i)^{N_i(t)}\right) \left(\prod_j B(j)^{N_j(t)}\right) \exp[-\theta^t c(t)]}$$
$$= \frac{\lambda(c)^{N(s')} \left(\prod_i A(i)^{N_i(s')}\right) \left(\prod_j B(j)^{N_j(s')}\right) \exp[-\theta^t c'(s')]}{\lambda(c)^{N(t')} \left(\prod_i A(i)^{N_i(t')}\right) \left(\prod_j B(j)^{N_j(t')}\right) \exp[-\theta^t c'(t')]} \tag{4.488}$$
$$= \frac{P_{c'}(s')}{P_{c'}(t')}.$$

Thus $\boldsymbol{P_N}$ satisfies A12″, and it follows that $\boldsymbol{N} \in \langle$MODEL E5$\rangle$. Similarly, to establish that $\langle$A1, A2, A13″$\rangle \subseteq \langle$MODEL E5*$\rangle$, observe simply that if axiom A12″ is replaced by A13″, then the same argument which was employed to show that $\langle$MODEL E5$\rangle \subseteq \langle$A1, A2, A12″$\rangle$ now shows that $\langle$MODEL E5*$\rangle \subseteq \langle$A1, A2, A13″$\rangle$. □

Given these characterizations of relatively invariant exponential gravity models, we turn finally to the class of invariant exponential gravity models. As in Theorems 4.7 and 4.8, the appropriate characterizations of these models are obtained by simply adding the sub-configuration dependence axiom (A5):

Theorem 4.23 (E6 and E6* Characterizations) *The classes of frequency processes,* $\langle$MODEL E6$\rangle$ *and* $\langle$MODEL E6*$\rangle$, *are each characterized by the following properties of their generators:*

$$\langle \text{MODEL E6} \rangle = \langle \text{A1, A2, A5, A12}'' \rangle, \tag{4.489}$$

$$\langle \text{MODEL E6}^* \rangle = \langle \text{A1, A2, A5, A13}'' \rangle. \tag{4.490}$$

Proof: (a) To establish that $\langle$A1, A2, A5, A12″$\rangle \subseteq \langle$MODEL E6$\rangle$, consider any $\boldsymbol{P}$ satisfying (A1, A2, A5, A12″), and observe in particular that the identity in (4.276) holds for the function, $\lambda : C \to R_{++}$, in (4.486). Hence we may now conclude from A5 together with the argument following (4.276) that $\langle$A1, A2, A5, A12″$\rangle \subseteq \langle$MODEL E6$\rangle$. Similarly, if A12″ is replaced by A13″, then the same argument also shows that $\langle$A1, A2, A5, A13″$\rangle \subseteq \langle$MODEL E6*$\rangle$.

(b) To establish that $\langle$MODEL E6$\rangle \subseteq \langle$A1, A2, A5, A12″$\rangle$, simply observe from (4.19), (4.22), (4.275) and (4.483) that

$$\langle \text{MODEL E6} \rangle \subseteq \langle \text{MODEL G6} \rangle \cap \langle \text{MODEL E5} \rangle$$
$$\subseteq \langle \text{A1, A2, A5} \rangle \cap \langle \text{A1, A2, A12}'' \rangle = \langle \text{A1, A2, A5, A12}'' \rangle.$$

Similarly, it follows from (4.19), (4.21), (4.275) and (4.484) that

$$\begin{aligned}\langle \text{MODEL E6}^* \rangle &\subseteq \langle \text{MODEL G6} \rangle \cap \langle \text{MODEL E5}^* \rangle \\ &\subseteq \langle \text{A1, A2, A5} \rangle \cap \langle \text{A1, A2, A13}'' \rangle = \langle \text{A1, A2, A5, A13}'' \rangle .\end{aligned}$$

This proves the theorem. □

Remark 4.9. In view of Remarks 4.4 and 4.5 (following Theorem 4.14) it should be clear that a number of equivalent characterizations of exponential gravity can be also obtained. For example, since Theorem 4.13 shows that the representation in (4.485) is also implied by (A1,A2,A6,A7), and since the argument following (4.485) depends only on A12, we see on the one hand that $\langle \text{A1, A2, A6, A7, A12} \rangle \subseteq \langle \text{MODEL E5} \rangle$. Moreover, since (4.142) and (4.145) together with Remark 4.4 imply that $\langle \text{A1, A2, A12}'' \rangle \subseteq \langle \text{A1, A2, A3}'', \text{A12} \rangle = \langle \text{A1, A2, A6, A7, A12} \rangle$, we may conclude from (4.483) that $\langle \text{MODEL E5} \rangle = \langle \text{A1, A2, A6, A7, A12} \rangle$. Similarly, it should also be clear from the arguments in Remarks 4.4 and 4.5 that

$$\begin{aligned}\langle \text{MODEL E5}^* \rangle &= \langle \text{A1, A2, A6, A7, A13} \rangle , \\ \langle \text{MODEL E6} \rangle &= \langle \text{A1, A2, A5, A6, A7, A12} \rangle ,\end{aligned}$$

and $\langle \text{MODEL E6}^* \rangle = \langle \text{A1, A2, A5, A6, A7, A13} \rangle$. •

4.6 Generalizations of Gravity Models

In this final section we establish the results of Section 2.5.2 and 2.5.3. To do so, we begin by formulating a general class of 'two-stage' interaction processes which will serve as a framework for all processes considered in this section. Basically these processes involve the selection (or perception) of a nonempty set of possible destinations (opportunities) by an actor, followed by an interaction with exactly one of these destinations. To formalize such processes, let the finite universe of relevant *destinations* be denoted by J (possibly with $J = B$ as in Section 2.5.2 and 2.5.3), and let the class of nonempty subsets of J be denoted by $\boldsymbol{D} = \{D \in J : D \neq \phi\}$. Then we now take the relevant *individual interaction space* to be given by the product $\Omega_1 = I \times \boldsymbol{D} \times J$, with typical elements, $(iDj) \in \Omega_1$, denoting the *origin*, i, of an individual actor, the *destination set*, D, considered by the actor, and the *destination choice*, j, by the actor [where it is implicitly assumed that $j \in D$, as in expression (4.493) below]. Hence, as parallel to the attribute functions of Section 3.3.2, we may now define a *destination-set attribute function*, $\boldsymbol{d} : \Omega_1 \to \boldsymbol{D}$, with values, $\boldsymbol{d}(\omega) = D$, representing the relevant set of destinations for the actor in individual interaction, $\omega \in \Omega_1$. Next, if we augment $\boldsymbol{D}$ by the null element, o, and let $\boldsymbol{D}_o = \boldsymbol{D} \cup \{o\}$, then [in a manner paralleling (3.2) and (3.3)] we may define the corresponding family of *destination set functions*, $\boldsymbol{\Gamma} = \{\Gamma_r : \Omega \to \boldsymbol{D}_o | r \in Z_{++}\}$, for all $r \in Z_{++}$, $n \in Z_+$, and $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ by

$$\Gamma_r(\omega) = \begin{cases} \boldsymbol{d}(\omega_r), & r \leq n \\ o, & r > n. \end{cases} \tag{4.491}$$

Since each r^{th} individual interaction within any interaction pattern, $\omega \in \Omega$, is completely defined by $\omega_r = (i_r, D_r, j_r) = [\boldsymbol{i}(\omega_r), \boldsymbol{d}(\omega_r), \boldsymbol{j}(\omega_r)]$, it follows that the relevant class of measurable events, $\overline{M}$, is simply the σ-field generated by these attribute functions, i.e.,

$$\overline{\boldsymbol{M}} = \sigma(\boldsymbol{I} \cup \boldsymbol{\Gamma} \cup \boldsymbol{J}), \tag{4.492}$$

where by definition, $M = \sigma(\boldsymbol{I} \cup \boldsymbol{J}) \subseteq \overline{\boldsymbol{M}}$. Next we assume that $J \subseteq I$ [as in Sections 2.5.2 and 2.5.3], so that the relevant class of separation configurations, $C = V^{I \times J}$, by definition satisfies the condition that separation profiles, c_{jh}, between each destination pair, $j, h \in J$, are well defined. Finally, observe from the finiteness of $I \times \boldsymbol{D} \times J$ that the universe of interaction patterns, Ω, is a countable set, and hence that each probability measure, P, on $\langle \Omega, \overline{\boldsymbol{M}} \rangle$ is representable as a *probability function*, $P : \Omega \to R_+$. With these observations and conventions, we now say that

Definition 4.10 An interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{\boldsymbol{M}} \rangle$ is designated as a *two-stage interaction process* iff for $c \in C$, $n \in Z_{++}$, and $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$,

$$P_c(\omega) > 0 \Rightarrow j_r \in D_r, \quad r = 1, \ldots, n. \tag{4.493}$$

Here the intuitive notion of a 'two-stage' interaction is formalized by the consistency condition that the realized destination, j_r, in each r^{th} interaction if ω is always an element of the destination set, D_r. We shall employ the following notation throughout. For each $iDj \in I \times \boldsymbol{D} \times J$ let the (iDj)-*frequency*, $N_{iDj} : \Omega \to Z_+$, be defined for all $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$ by $N_{iDj}(\omega) = |\{r : i_r D_r j_r = iDj\}|$. The associated marginal frequencies are then defined by $N_{iD} = \sum_j N_{iDj}$, $N_{ij} = \sum_D N_{iDj}$, $N_i = \sum_j N_{ij}$, $N_D = \sum_i N_{iD}$, and $N = \sum_i N_i$. Within this general framework we now formalize the class of hierarchical interaction processes in Section 2.5.2.

4.6.1 Interaction Processes with Hierarchical Destinations

To begin with, we postulate the existence of a family of *cluster partitions*, $\{\boldsymbol{D}_c : c \in C\}$, of the destination set, J, where by definition each partition, $\boldsymbol{D}_c$, consists of a finite collection of mutually disjoint destination sets, $\{D_1, \ldots, D_n\} \subseteq \boldsymbol{D}$, with $\cup_{k=1}^n D_k = J$. Each set, $D \in \boldsymbol{D}_c$, is designated as a *destination cluster*. In this context we now say that

Definition 4.11 A two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{\boldsymbol{M}} \rangle$ is designated as a *hierarchical interaction process* iff there exists a family of cluster partitions, $\{\boldsymbol{D}_c : c \in C\}$, such that for all $c \in C$, $n \in Z_{++}$, and $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$,

$$P_c(\omega) > 0 \Leftrightarrow D_r \in \boldsymbol{D}_c, \qquad r = 1, \ldots, n. \tag{4.494}$$

Hence for any given separation configuration, $c \in C$, it is postulated that the only destination sets considered by actors are destination clusters in $\boldsymbol{D}_c$. If we now denote the set of *feasible* (iDj)-*realizations* by

$$\Omega_{1c} = \{iDj \in \Omega_1 : j \in D \in \boldsymbol{D}_c\} \tag{4.495}$$

and similarly, let $\Omega_{nc} = (\Omega_{1c})^n$ and employ the conditional notation of (3.24), then as a parallel to the independence axioms A1 and A2 for spatial interaction processes, we now have the following notion of *independence* for hierarchical interaction processes:

Definition 4.12 A hierarchical interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be *independent* iff for all $c \in C$, $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega_{nc}$, $n \in Z_{++}$, and $(n_{iDj} \in Z_{+} : iDj \in \Omega_{1c})$:

$$P_c^n(\omega) = \prod_{r=1}^{n} P_c^n(I_r = i_r,\ \Gamma_r = D_r,\ J_r = j_r), \tag{4.496}$$

$$P_c(N_{iDj} = n_{iDj} : iDj \in \Omega_{1c}) = \prod_{iDj \in \Omega_{1c}} P_c(N_{iDj} = n_{iDj}). \tag{4.497}$$

Given this special class of two-stage processes, we next formalize the cluster-level separation dependence axiom A3.1 of Section 2.5.2 within the present framework. First, for each $i \in I$ and $D \in \boldsymbol{D}_c$ let the individual (iD)-*event* be denoted by $iD = \{i\} \times \{D\} \times D \subseteq \Omega_{1c}$, and similarly, let each finite collection of such events be designated as a *cluster-level interaction pattern*, $\boldsymbol{\sigma} = (i_r D_r : r = 1, \ldots, n) \in (I \times \boldsymbol{D}_c)^n$. Equivalently, if for each $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$ we let $\boldsymbol{\sigma}(\omega) = (i_r D_r : r = 1, \ldots, n)$, then the cluster-level interaction pattern, $\boldsymbol{\sigma} \in (I \times \boldsymbol{D}_c)^n$, corresponds to the $\overline{M}$-measurable event $\Omega(\boldsymbol{\sigma}) = \{\omega \in \Omega_{nc} : \boldsymbol{\sigma}(\omega) = \boldsymbol{\sigma}\}$. Hence, if the class of such patterns is denoted by $\boldsymbol{\Sigma}_c$, then it follows that the associated *cluster-level pattern probabilities*, $P_c(\boldsymbol{\sigma}) = \Sigma_{\omega \in \Omega(\boldsymbol{\sigma})} P_c(\omega)$, are well defined for each $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_c$. Next, if for each $iD \in I \times \boldsymbol{D}_c$, and $\boldsymbol{\sigma} = (i_r D_r : r = 1, \ldots, n) \in \boldsymbol{\Sigma}_c$ we denote the (iD)-*frequency* in pattern $\boldsymbol{\sigma}$ by $N_{iD}(\boldsymbol{\sigma}) = \sum_{\omega \in \Omega(\boldsymbol{\sigma})} N_{iD}(\omega)$, and set $N_i(\boldsymbol{\sigma}) = \sum_D N_{iD}(\boldsymbol{\sigma})$ and $N_D(\boldsymbol{\sigma}) = \sum_i N_{iD}(\boldsymbol{\sigma})$, then the corresponding *cluster-level activity profile* is given [as in (2.200)] by

$$A_c(\boldsymbol{\sigma}) = [(N_i(\boldsymbol{\sigma}) : i \in I),\ (N_D(\boldsymbol{\sigma}) : D \in \boldsymbol{D}_c)]. \tag{4.498}$$

Finally, if for each $i \in I$ and $D \in \boldsymbol{D}_c$, we let

$$c_{iD} = (c_{ij} : j \in D) \tag{4.499}$$

denote the relevant *(iD)-separation profile* in c, and again define the corresponding *cluster-level separation array* for each interaction pattern, $\boldsymbol{\sigma} = (i_r D_r : r = 1, \ldots, n)$, by

$$c_{\boldsymbol{\sigma}} = (c_{i_r D_r} : r = 1, \ldots, n), \tag{4.500}$$

then the following version of the separation dependence axiom (A3) at the cluster level is well defined for each hierarchical interaction process, $\boldsymbol{P}$:

A3.1. (Cluster-Level Separation Dependence) *For all $c \in C$ and $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \boldsymbol{\Sigma}_c$ with $A_c(\boldsymbol{\sigma}) = A_c(\boldsymbol{\sigma}')$,*

$$c_{\boldsymbol{\sigma}} = c_{\boldsymbol{\sigma}'} \Rightarrow P_c(\boldsymbol{\sigma}) = P_c(\boldsymbol{\sigma}'). \tag{4.501}$$

Given this behavioral axiom at the cluster level, we now have the following representation theorem for the mean cluster-level interaction frequencies of *independent* hierarchical interaction processes, where for each separation configuration, $c \in C$, the corresponding set of (iD)-separation profiles in (4.499) is denoted by $c(I \times \boldsymbol{D}_c)$:

Proposition 4.11 *An independent hierarchical interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with cluster partition family, $\{\boldsymbol{D}_c : c \in C\}$, satisfies axiom* A3.1 *iff for each $c \in C$ there exist functions, $A_c : I \to R_{++}$, $B_c : \boldsymbol{D}_c \to R_{++}$, and $F_c : c(I \times \boldsymbol{D}_c) \to R_{++}$, such that for all $iD \in I \times \boldsymbol{D}_c$,*

$$\begin{aligned} \mathrm{E}_c(N_{iD}) &= \mathrm{E}_c(N) P_c^1(I_1 = i, \Gamma_1 = D) \\ &= A_c(i) B_c(D) F_c(c_{iD}). \end{aligned} \tag{4.502}$$

Proof: First observe from independence condition (4.496) that for all $c \in C$, and all $i_1, \ldots, i_n \in I$ and $D_1, \ldots, D_n \in \boldsymbol{D}_c$ with $n \in Z_{++}$,

$$\begin{aligned} &P_c^n(I_r = i_r, \Gamma_r = D_r : r = 1, \ldots, n) \\ &= \sum_{(j_r : r=1,\ldots,n) \in \Pi_r D_r} P_c^n(I_r = i_r, \Gamma_r = D_r,\ J_r = j_r : r = 1, \ldots, n) \\ &= \sum_{(j_r : r=1,\ldots,n) \in \Pi_r D_r} \prod_{r=1}^{n} P_c^n(I_r = i_r,\ \Gamma_r = D_r,\ J_r = j_r) \\ &= \prod_{r=1}^{n} \sum_{j \in D_r} P_c^n(I_r = i_r,\ \Gamma_r = D_r,\ J_r = j_r) \\ &= \prod_{r=1}^{n} P_c^n(I_r = i_r,\ \Gamma_r = D_r). \end{aligned} \tag{4.503}$$

Similarly, if for each iD-frequency, n_{iD}, we denote the set of iDj frequencies consistent with n_{iD} by

$$Z(n_{iD}) = \{(n_{iDj} : j \in D) \mid \sum_{j \in D} n_{iDj} = n_{iD}\},$$

then it also follows from condition (4.497) that for all frequency profiles $(n_{iD} : iD \in I \times \boldsymbol{D}_c)$:

$$\begin{aligned} P_c(N_{iD} &= n_{iD} : iD \in I \times \boldsymbol{D}_c) \\ &= \sum_{(n_{iDj}\,:\,iDj \in \Omega_{1c}) \in \Pi_{iD} Z(n_{iD})} P_c(N_{iDj} = n_{iDj} : iDj \in \Omega_{1c}) \\ &= \sum_{(n_{iDj}\,:\,iDj \in \Omega_{1c}) \in \Pi_{iD} Z(n_{iD})} \prod_{iDj \in \Omega_{1c}} P_c(N_{iDj} = n_{iDj}) \\ &= \prod_{iD \in I \times \boldsymbol{D}_c} \sum_{(n_{iDj}\,:\,j \in J) \in Z(n_{iD})} \prod_{j \in D} P_c(N_{iDj} = n_{iDj}) \\ &= \prod_{iD \in I \times \boldsymbol{D}_c} \sum_{(n_{iDj}\,:\,j \in J) \in Z(n_{iD})} P_c(N_{iDj} = n_{iDj} : j \in J) \\ &= \prod_{iD \in I \times \boldsymbol{D}_c} P_c(N_{iD} = n_{iD}). \end{aligned} \tag{4.504}$$

Hence if we now consider the interaction subprocess, $\boldsymbol{P}^* = \{P_c^* : c \in C\}$, defined by replacing the destination set, J, with the cluster partition, $\boldsymbol{D}_c$, for each $c \in C$, then it follows at once from an examination of (4.498) through (4.504) that axioms (A1,A2,A3) hold for each probability measure, P_c^*. Thus by axioms A1 and A2, it follows from (3.69) and (4.107) that $\mathrm{E}_c(N_{iD}) = \mathrm{E}_c(N) P_c^1(I_1 = i,\ \Gamma_1 = D)$, and hence that the first equality in (4.502) must hold. Moreover, if for each $c \in C$ we also replace J by $\boldsymbol{D}_c$ and V_c by $c(I \times \boldsymbol{D}_c)$ in Model G1, then the formal arguments establishing (4.169) of Theorem 4.1 also show that, in the presence of independence conditions (4.496) and (4.497), axiom A3.1 is both necessary and sufficient for the existence of functions, $A_c : I \to R_{++}$, $B_c : \boldsymbol{D}_c \to R_{++}$, and $F_c : c(I \times \boldsymbol{D}_c) \to R_{++}$, satisfying the second equality in expression (4.502). □

Next we formalize the destination-level separation dependence axiom A3.2 of Section 2.5.2 within the present framework. First, if for any fixed cluster set, $D \in \boldsymbol{D}_c$, we let the *D-domain* in Ω be denoted by

$$\begin{aligned} \Omega_D = \{\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega \mid j_r \in D_r = D, \\ r = 1, \ldots, n \in Z_{++}\} \cup \Omega_o, \end{aligned} \tag{4.505}$$

where Ω_0 is the *0-interaction space* [in (3.1)], then each non-null element, $\omega = (i_r D j_r : r = 1, \ldots, n) \in \Omega_D - \Omega_0$, is now designated as a *D-interaction* *pattern*. As a parallel to (4.498), the corresponding *D-activity profile* is denoted by

$$A(\omega) = [(N_i(\omega) : i \in I),\ (N_j(\omega) : j \in D)], \tag{4.506}$$

and similarly, as a parallel to (4.500), the *D-separation array* is denoted by

$$c_\omega = (c_{i_r j_r} : r = 1, \ldots, n). \tag{4.507}$$

With these definitions, the following destination-level version of separation dependence (A3) is well defined for each hierarchical interaction process, $\boldsymbol{P}$:

A3.2. (Destination-Level Separation Dependence) *For all separation configurations, $c \in C$, destination clusters, $D \in \boldsymbol{D}_c$, and D-interaction patterns, $\omega, \omega' \in \Omega_D$, with $A(\omega) = A(\omega')$,*

$$c_\omega = c_{\omega'} \Rightarrow P_c(\omega) = P_c(\omega'). \tag{4.508}$$

If for each $D \in \boldsymbol{D}_c$ and $ij \in I \times D$ we now let $\mathrm{E}_c(N_{ij} \mid \Omega_D)$ denote the conditional mean (ij)-interaction frequency given the $\overline{M}$-measurable event, Ω_D, and let $V_{cD} = \{c_{ij} \in V : ij \in I \times D\}$, then [as a parallel to Proposition 4.11 above] we now have the following representational consequence of this axiom:

Proposition 4.12 *An independent hierarchical interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, with cluster partition family, $\{\boldsymbol{D}_c : c \in C\}$, satisfies axiom* A3.2 *iff for each $c \in C$ and $D \in \boldsymbol{D}_c$ there exist functions, $A_{cD} : I \to R_{++}$, $B_{cD} : D \to R_{++}$ and $F_{cD} : V_{cD} \to R_{++}$, such that for all $ij \in I \times D$,*

$$\begin{aligned} \mathrm{E}_c(N_{ij} \mid \Omega_D) &= \mathrm{E}_c(N \mid \Omega_D) P_c^1(I_1 = i,\ J_1 = j \mid \Gamma_1 = D) \\ &= A_{cD}(i) B_{cD}(j) F_{cD}(c_{ij}). \end{aligned} \tag{4.509}$$

PROOF: First observe from independence condition (4.496) that for each $c \in C$, $D \in \boldsymbol{D}_c$, and $n \in Z_{++}$,

$$\begin{aligned} P_c^n(\Omega_D) &= P_c^n(I_r \in I,\ \Gamma_r = D,\ J_r \in D : r = 1, \ldots, n) \\ &= \prod_{r=1}^{n} P_c^n(I_r \in I,\ \Gamma_r = D,\ J_r \in D). \end{aligned} \tag{4.510}$$

Thus, if we denote the conditional probability measure given the event, Ω_D, by $P_c(\cdot \mid \Omega_D)$, then it also follows from independence condition (4.496) together with (4.510) that for all $i_1, \ldots, i_n \in I$ and $j_1, \ldots, j_n \in D$ with

$n \in Z_{++}$,

$$\begin{aligned} P_c^n(I_r &= i_r,\ J_r = j_r : r = 1,\ldots,n \,|\, \Omega_D) \\ &= P_c^n(I_r = i_r,\ \Gamma_r = D,\ J_r = j_r : r = 1,\ldots,n)/P_c^n(\Omega_D) \\ &= [\prod_{r=1}^n P_c^n(I_r = i_r,\ \Gamma_r = D,\ J_r = j_r)]/P_c^n(\Omega_D) \\ &= \prod_{r=1}^n \left[\frac{P_c^n(I_r = i_r,\ \Gamma_r = D,\ J_r = j_r)}{P_c^n(I_r \in I,\ \Gamma_r = D,\ J_r \in D)}\right] \\ &= \prod_{r=1}^n P_c^n(I_r = i_r,\ J_r = j_r | \Omega_D). \end{aligned} \tag{4.511}$$

Similarly, since by definition

$$P_c(\Omega_D) = P_c(N_{iHj} = 0 : j \in H \in \boldsymbol{D}_c - \{D\}), \tag{4.512}$$

it follows from independence condition (4.497) that for each frequency profile, $(n_{ij} \in Z_+ : ij \in I \times D)$,

$$\begin{aligned} P_c(N_{iDj} &= n_{ij} : ij \in I \times D \,|\, \Omega_D) \\ &= P_c[(N_{iDj} = n_{ij} : ij \in I \times D),\ (N_{iHj} = 0 : j \in H \in \boldsymbol{D}_c - \{D\})] \\ &\quad \cdot [P_c(\Omega_D)]^{-1} \\ &= [\prod_{ij \in I \times D} P_c(N_{iDj} = n_{ij})] \cdot P_c(N_{iHj} = 0 : j \in H \in \boldsymbol{D}_c - \{D\}) \\ &\quad \cdot [P_c(\Omega_D)]^{-1} \\ &= \prod_{ij \in I \times D} P_c(N_{iDj} = n_{ij}). \end{aligned} \tag{4.513}$$

Hence it follows in particular (by summing over all values of n_{gh}, $gh \neq ij$) that $P_c(N_{iDj} = n_{ij} \,|\, \Omega_D) = P_c(N_{iDj} = n_{ij})$, so that

$$P_c(N_{iDj} = n_{ij} : ij \in I \times D \,|\, \Omega_D) = \prod_{ij \in I \times D} P_c(N_{iDj} = n_{ij} \,|\, \Omega_D). \tag{4.514}$$

Thus, if we now define the class of *spatial D-interaction patterns* by $S_D = \{o\} \cup \{\cup_{n>0}(i \times D)^n\}$, and consider the spatial interaction process, $\boldsymbol{P}^\circ = \{P_c : c \in C\}$, on S_D defined for all $c \in C$ and $s = (i_r j_r : r = 1,\ldots,n) \in S_D$, by $P_c^\circ(s) = P_c(I_r = i_r,\ J_r = j_r : r = 1,\ldots,n \,|\, \Omega_D)$ together with $P_c^\circ(o) = P_c(\Omega_0 \mid \Omega_D)$, then it follows at once from (4.511) and (4.514) that $\boldsymbol{P}^\circ$ must satisfy independence axioms A1 and A2. Hence by (3.69) and (4.107) we see that $\mathrm{E}_c(N_{ij} \,|\, \Omega_D) = \mathrm{E}_c(N \,|\, \Omega_D) P_c^1(I_1 = i,\ J_1 = j \,|\, \Omega_D) = \mathrm{E}_c(N \,|\, \Omega_D) P_c^1(I_1 = i,\ J_1 = j \,|\, \Gamma_1 = D)$, and thus that the first equality in (4.509) must hold. Moreover, if for each $s = (i_r j_r : r = 1,\ldots,n) \in S_D$ we let $\omega(s) \stackrel{\bullet}{=} (i_r D j_r : r = 1, \ldots, n)$, then by definition it follows that for all non-null $s, s' \in S_D$, $c_s = c_{s'} \Leftrightarrow c_{\omega(s)} = c_{\omega(s')}$ and similarly that

$$\begin{aligned} P_c^\circ(s) = P_c^\circ(s') &\Leftrightarrow P_c[\omega(s) \,|\, \Omega_D] = P_c[\omega(s') \,|\, \Omega_D] \\ &\Leftrightarrow P_c[\omega(s)]/P_c(\Omega_D) = P_c[\omega(s')]/P_c(\Omega_D) \\ &\Leftrightarrow P_c[\omega(s)] = P_c[\omega(s')], \end{aligned} \tag{4.515}$$

so that by axiom A3.2 for $\boldsymbol{P}$, we see that $\boldsymbol{P}^\circ$ must also satisfies A3. Thus if we replace J by D and V_c by V_{cD} in Model G1, then [as in Proposition 4.11 above], it follows from the formal arguments establishing (4.169) of Theorem 4.1 that for each independent hierarchical interaction processes, $\boldsymbol{P}$, axiom A3.2 is both necessary and sufficient for the existence of functions, $A_{cD} : I \to R_{++}$, $B_{cD} : D \to R_{++}$, and $F_{cD} : V_{cD} \to R_{++}$ satisfying the second equality in (4.509). □

Given these two formal representations, we may now derive the desired representation of mean interaction levels, $\mathrm{E}_c(N_{ij})$, for independent hierarchical interaction processes, $\boldsymbol{P}$, as follows.

Proposition 4.13 *If an independent hierarchical interaction process,* $\boldsymbol{P} = \{P_c : c \in C\}$, *with cluster partition family,* $\{\boldsymbol{D}_c : c \in C\}$, *satisfies both axioms* A3.1 *and* A3.2 *then for each* $c \in C$ *and* $D \in \boldsymbol{D}_c$ *there exist functions,* $A_c : I \to R_{++}$, $B_c : \boldsymbol{D}_c \to R_{++}$, $F_c : c(I \times \boldsymbol{D}_c) \to R_{++}$, $B_{cD} : D \to R_{++}$, $F_{cD} : V_{cD} \to R_{++}$, *such that for all* $ij \in I \times D$,

$$\mathrm{E}_c(N_{ij}) = A_c(i) B_c(D) F_c(c_{iD}) \frac{B_{cD}(j) F_{cD}(c_{ij})}{\sum_{h \in D} B_{cD}(h) F_{cD}(c_{ih})}. \tag{4.516}$$

PROOF: First observe that if for any $c \in C$ and $j \in J$ we let $D(j)$ denote the unique cluster set in $\boldsymbol{D}_c$ containing j, then the event (ij) is equivalent to the event $(iD(j)j)$. Hence it follows from independence condition (4.496) that for all $i_1, \ldots, i_n \in I$ and $j_1, \ldots, j_n \in J$ with $n \in Z_{++}$,

$$\begin{aligned} P_c^n(i_r j_r : r = 1, \ldots, n) &= P_c^n(i_r D(j_r) j_r : r = 1, \ldots, n) \\ &= \prod_{r=1}^n P_c^n(i_r D(j_r) j_r) \\ &= \prod_{r=1}^n P_c^n(i_r j_r), \end{aligned} \tag{4.517}$$

and similarly from independence condition (4.497) that for all frequency profiles, $(n_{ij} \in Z_+ : ij \in I \times J)$,

$$\begin{aligned} P_c^n(N_{ij} = n_{ij} : ij \in I \times J) &= P_c(N_{iD(j)j} = n_{ij} : ij \in I \times J) \\ &= \prod_{ij \in I \times J} P_c(N_{iD(j)j} = n_{ij}) \\ &= \prod_{ij \in I \times J} P_c(N_{ij} = n_{ij}). \end{aligned} \tag{4.518}$$

Thus we see from (4.517) and (4.518) that the *spatial interaction process*, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, M \rangle$ generated by each independent hierarchical interaction process is also independent, i.e., satisfies axioms A1 and A2. This in turn implies from (3.69) and (4.107) that for all $c \in C$ and $iDj \in \Omega_{1c}$,

$$\begin{aligned} \mathrm{E}_c(N_{ij}) &= \mathrm{E}_c(N) P_c^1(I_1 = i,\ J_1 = j) \\ &= \mathrm{E}_c(N) P_c^1(I_1 = i,\ \Gamma_1 = D,\ J_1 = j) \\ &= \mathrm{E}_c(N) P_c^1(I_1 = i,\ \Gamma_1 = D) P_c^1(J_1 = j |\ I_1 = i,\ \Gamma_1 = D). \end{aligned} \tag{4.519}$$

But by (4.509) we must have

$$\begin{aligned} &P_c^1(J_1 = j \mid I_1 = i,\ \Gamma_1 = D) \\ &= \frac{P_c^1(I_1 = i,\ J_1 = j | \Gamma_1 = D)}{\sum_{h \in D} P_c^1(I_1 = i,\ J_1 = h | \Gamma_1 = D)} \\ &= \frac{B_{cD}(j) F_{cD}(c_{ij})}{\sum_{h \in D} B_{cD}(h) F_{cD}(c_{ih})}, \end{aligned} \tag{4.520}$$

and hence may conclude by substituting (4.502) and (4.520) into (4.519) that (4.516) must hold for this choice of functions. □

4.6.2 Interaction Processes with Random Destination Sets

To establish the results in Section 2.5.3, we begin by formalizing the class of *independent two-stage interaction processes* in terms of the general class of two-stage interaction processes in Definition 4.10 above:

Definition 4.13
(i) A two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, on $\langle \Omega, \overline{M} \rangle$ is said to be *regular* iff $|J| \geq 3$, and in addition, it is true for all $c \in C$, $n \in Z_{++}$, and $\omega = (i_r D_r j_r : r = 1, \ldots, n) \in \Omega$, that

$$P_c(\omega) > 0 \Leftrightarrow j_r \in D_r,\ r = 1, \ldots, n. \tag{4.521}$$

(ii) A regular two-stage interaction process, $\boldsymbol{P}$, on $\langle\Omega,\overline{M}\rangle$ is designated as an *independent two-stage interaction process* iff the associated spatial interaction process on $\langle\Omega,\overline{M}\rangle$ is independent, i.e., satisfies A1 and A2.

The cardinality condition, $|J|\geq 3$, avoids the need to consider separate special cases. The positivity condition, (4.521) requires that all meaningful outcomes be possible. In particular, if for each $c\in C$ we now denote the *conditional set probabilities* by $P_c^1(D\,|\,i)\equiv P_c^1(\Gamma_1=D\,|\,I_1=i)$, and denote the *conditional destination probabilities* by $P_c^1(j\,|\,iD)=P_c^1(J_1=j\,|\,I_1=i,\ \Gamma_1=D)$, then it follows from (4.521) that $P_c^1(D\,|\,i)>0$ and $P_c^1(j\,|\,iD)>0$ for all $i\in I$, $D\in\boldsymbol{D}$, and $j\in D$. Finally, if we again let $\boldsymbol{D}_j=\{D\in\boldsymbol{D}: j\in D\}$, then the independence condition, together with (3.69) and (4.107), implies that for all $c\in C$ and $ij\in I\times J$,

$$\begin{aligned}
\mathrm{E}_c(N_{ij}) &= \mathrm{E}_c(N)P_c^1(I_1=i,\ J_1=j)\\
&=[\mathrm{E}_c(N)P_c^1(I_1=i)]P_c^1(J_1=j\,|\,I_1=i)\\
&=\mathrm{E}_c(N_i)P_c^1(J_1=j\,|\,I_1=i)\\
&=\mathrm{E}_c(N_i)\sum_{D\in\boldsymbol{D}_j}P_c^1(J_1=j,\ \Gamma_1=D|I_1=i)\\
&=\mathrm{E}_c(N_i)\sum_{D\in\boldsymbol{D}_j}P_c^1(J_1=j|I_1=i,\ \Gamma_1=D)P_c^1(\Gamma_1=D\,|\,I_1=i)\\
&=\mathrm{E}_c(N_i)\sum_{D\in\boldsymbol{D}_j}P_c^1(j\,|\,iD)P_c^1(D\,|\,i).
\end{aligned}\tag{4.522}$$

Hence the mean frequencies, $\mathrm{E}_c(N_{ij})$, for spatial interaction processes generated by independent two-stage interaction processes, $\boldsymbol{P}=\{P_c: c\in C\}$, are seen to be representable entirely in terms of the conditional probabilities $P_c^1(D\,|\,i)$ and $P_c^1(j\,|\,iD)$ [together with $\mathrm{E}_c(N_i)$].

Given these general observations, we now consider representations for each of these probabilities in turn. Turning first to the conditional set probabilities, $P_c^1(D\,|\,i)$, we formulate the following set-level version of the destination proportionality axiom (A6) stated in Section 2.5.3:

A6.1. (Conditional Set Proportionality) *For all origins, $i,g\in I$, destination sets, $D,H\in D$, and separation configurations, $c,c'\in C$,*

$$(c_{iD}=c'_{gD},\ c_{iH}=c'_{gH})\Rightarrow\frac{P_c^1(D\,|\,i)}{P_c^1(H\,|\,i)}=\frac{P_{c'}^1(D\,|\,g)}{P_{c'}^1(H\,|\,g)}.\tag{4.523}$$

In terms of this axiom, we then have the following representational consequence for conditional set probabilities:

Proposition 4.14 *An independent two-stage interaction process, $\boldsymbol{P}=\{P_c: c\in C\}$, satisfies axiom* A6.1 *iff for all destination sets, $D\in\boldsymbol{D}$, there exist* *functions,* $F_D : V^D \rightarrow R_{++}$, *such that for each* $c \in C$, $i \in I$, *and* $D \in \boldsymbol{D}$,

$$P_c^1(D \mid i) = \frac{F_D(c_{iD})}{\sum_{H \in \boldsymbol{D}} F_H(c_{iH})}. \tag{4.524}$$

PROOF: If (4.524) holds, then it follows at once [as in (2.231)] that for any origins, $i, g \in I$, destination sets, $D, H \in \boldsymbol{D}$, and separation configurations, $c, c' \in C$, with $c_{iD} = c'_{gD}$ and $c_{iH} = c'_{gH}$, we must have

$$\frac{P_c^1(D \mid i)}{P_c^1(H \mid i)} = \frac{F_D(c_{iD})}{F_H(c_{iH})} = \frac{F_D(c'_{gD})}{F_H(c'_{gH})} = \frac{P_{c'}^1(D \mid g)}{P_{c'}^1(H \mid g)}, \tag{4.525}$$

and hence may conclude that A6.1 holds. To establish the converse, we begin by considering any pair of singleton destination sets, $\{j\}, \{h\} \in \boldsymbol{D}$. In particular, if for any fixed origin, $a \in I$, and pair of separation profiles, $x, y \in V$, we choose any separation configuration, $c^\circ \in C(= V^{I \times J})$, with $c^\circ_{aj} = x$ and $c^\circ_{ah} = y$, then it follows from axiom A6.1 that [in a manner paralleling (4.291)] the relation

$$\phi_{jh}(x, y) = \frac{P_{c^\circ}^1(\{j\} \mid a)}{P_{c^\circ}^1(\{h\} \mid a)} \tag{4.526}$$

yields a well defined function, $\phi_{jh} : V^2 \rightarrow R_{++}$, which satisfies the following identity for all $i \in I$ and $c \in C$,

$$\phi_{jh}(c_{ij}, c_{ih}) = \frac{P_c^1(\{j\} \mid i)}{P_c^1(\{h\} \mid i)}. \tag{4.527}$$

Next, for any *fixed* destination, $b \in J$, and *fixed* separation profile, $\sigma \in V$, let the functions, $F_{\{j\}} : V \rightarrow R_{++}$, $j \in J - \{b\}$ be defined for all $x \in V$ by

$$F_{\{j\}}(x) = \phi_{jb}(x, \sigma), \tag{4.528}$$

and similarly, choosing any *fixed* destination, $d \in J - \{b\}$, let the function, $F_{\{b\}} : V \rightarrow R_{++}$, be defined for all $x \in V$ by

$$F_{\{b\}}(x) = \frac{\phi_{bd}(x, \sigma)}{\phi_{bd}(\sigma, \sigma)}. \tag{4.529}$$

With these definitions, it follows that if for any profiles, $x, y, z \in V$, and distinct destinations, $j, h, g \in J$ [which exist by the regularity condition $|J| \geq 3$], we choose any separation configuration, $c \in C(= V^{I \times J})$, with $c_{aj} = x$, $c_{ah} = y$, and $c_{ag} = z$, then by (4.527),

$$\phi_{jh}(x, y) = \frac{P_c^1(\{j\} \mid a)}{P_c^1(\{h\} \mid a)} = \frac{P_c^1(\{j\} \mid a)/P_c^1(\{g\} \mid a)}{P_c^1(\{h\} \mid a)/P_c^1(\{g\} \mid a)} = \frac{\phi_{jg}(x, z)}{\phi_{hg}(y, z)}. \tag{4.530}$$

Hence for any $j, h \in J - \{b\}$ and $c \in C$ it follows from (4.527) and (4.530) that

$$\frac{P_c^1(\{j\} \mid i)}{P_c^1(\{h\} \mid i)} = \phi_{jh}(c_{ij}, c_{ih}) = \frac{\phi_{jb}(c_{ij}, \sigma)}{\phi_{hb}(c_{ih}, \sigma)} = \frac{F_{\{j\}}(c_{ij})}{F_{\{h\}}(c_{ih})}. \tag{4.531}$$

Similarly, observe from (4.530) and the argument in (4.293) that for any $j \in J - \{b, d\}$,

$$\begin{aligned} \frac{P_c^1(\{j\} \mid i)}{P_c^1(\{b\} \mid i)} &= \phi_{jb}(c_{ij}, c_{ib}) = \frac{\phi_{jd}(c_{ij}, \sigma)}{\phi_{bd}(c_{ib}, \sigma)} \\ &= \frac{\phi_{jb}(c_{ij}, \sigma)/\phi_{db}(\sigma, \sigma)}{\phi_{bd}(c_{ib}, \sigma)} = \frac{\phi_{jb}(c_{ij}, \sigma)}{\phi_{bd}(c_{ib}, \sigma)\phi_{db}(\sigma, \sigma)} \\ &= \frac{\phi_{jb}(c_{ij}, \sigma)}{\phi_{bd}(c_{ib}, \sigma)/\phi_{bd}(\sigma, \sigma)} = \frac{F_{\{j\}}(c_{ij})}{F_{\{b\}}(c_{ib})}. \end{aligned} \tag{4.532}$$

Finally, again choosing any $j \in J - \{b, d\}$, we also see from (4.531) and (4.532) that

$$\begin{aligned} \frac{P_c^1(\{d\} \mid i)}{P_c^1(\{b\} \mid i)} &= \frac{P_c^1(\{d\} \mid i)}{P_c^1(\{j\} \mid i)} \cdot \frac{P_c^1(\{j\} \mid i)}{P_c^1(\{b\} \mid i)} \\ &= \frac{F_{\{d\}}(c_{id})}{F_{\{j\}}(c_{ij})} \cdot \frac{F_{\{j\}}(c_{ij})}{F_{\{b\}}(c_{ib})} = \frac{F_{\{d\}}(c_{id})}{F_{\{b\}}(c_{ib})}. \end{aligned} \tag{4.533}$$

Hence we may conclude from (4.531), (4.532) and (4.533) that for all singleton sets, $\{j\}, \{h\} \in \boldsymbol{D}$, and all $i \in I$ and $c \in C$,

$$\frac{P_c^1(\{j\} \mid i)}{P_c^1(\{h\} \mid i)} = \frac{F_{\{j\}}(c_{ij})}{F_{\{h\}}(c_{ih})}. \tag{4.534}$$

To extend this family of F-functions to all of $\boldsymbol{D}$, choose a fixed origin, $a \in I$, and for each destination set, $D \in \boldsymbol{D}$, and choose a *fixed representative* element, $d \in D$. Then if for each separation profile array $x = (x_j : j \in D) \in V^D$, we choose any configuration $c^\circ \in C$ with $c^\circ_{aD} = x$, it follows from axiom A6.1 that the relation

$$F_D(x) = F_{\{d\}}(c^\circ_{ad}) \frac{P^1_{c^\circ}(D \mid a)}{P^1_{c^\circ}(\{d\} \mid a)} = F_{\{d\}}(x_d) \frac{P^1_{c^\circ}(D \mid a)}{P^1_{c^\circ}(\{d\} \mid a)} \tag{4.535}$$

yields a well defined function, $F_D : V^D \to R_{++}$, which is independent of the choices of $a \in I$, $d \in D$, and $c^\circ \in C$, and which satisfies the following identity for all $i \in I$ and $c \in C$,

$$F_D(c_{iD}) = F_{\{d\}}(c_{id}) \frac{P_c^1(D \mid i)}{P_c^1(\{d\} \mid i)}. \tag{4.536}$$

Hence, by (4.534) and (4.536), we may conclude that for all $c \in C$, $i \in I$ and destination sets, $D, H \in \boldsymbol{D}$, with representatives, $d \in D$ and $h \in H$,

$$\begin{aligned} \frac{P_c^1(D \mid i)}{P_c^1(H \mid i)} &= \frac{P_c^1(D \mid i)/P_c^1(\{d\} \mid i)}{P_c^1(H \mid i)/P_c^1(\{h\} \mid i)} \cdot \frac{P_c^1(\{d\} \mid i)}{P_c^1(\{h\} \mid i)} \\ &= \frac{F_D(c_{iD})/F_{\{d\}}(c_{id})}{F_H(c_{iH})/F_{\{h\}}(c_{ih})} \cdot \frac{F_{\{d\}}(c_{id})}{F_{\{h\}}(c_{ih})} = \frac{F_D(c_{iD})}{F_H(c_{iH})}, \end{aligned} \tag{4.537}$$

which together with the normalization condition, $1 = \sum_{D \in \boldsymbol{D}} P_c^1(D \mid i)$, implies that (4.524) must hold identically for this family of functions, $\{F_D : D \in \boldsymbol{D}\}$. □

Turning now to the conditional destination probabilities, $P_c^1(j \mid iD)$, we begin by formulating the following versions of the destination proportionality axiom (A6) and destination separability axiom (A8) stated in Section 2.5.3:

A6.2. (Conditional Destination Proportionality) *For all origins, $i, g \in I$, separation configurations, $c, c' \in C$, destination sets, $D, H \in \boldsymbol{D}$, and destinations, $j, h \in D \cap H$,*

$$(c_{ij} = c'_{gj},\ c_{ih} = c'_{gh}) \Rightarrow \frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{P_{c'}^1(j \mid gH)}{P_{c'}^1(h \mid gH)}. \tag{4.538}$$

A8.2. (Conditional Destination Separability) *For all origins, $i \in I$, destination sets, $D \in \boldsymbol{D}$, destinations, $j, h \in D$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c_{ih},\ c'_{ij} = c'_{ih}) \Rightarrow \frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{P_{c'}^1(j \mid iD)}{P_{c'}^1(h \mid iD)}. \tag{4.539}$$

We next establish the following representational consequence of these axioms for conditional destination probabilities of independent two-stage interaction processes:

Proposition 4.15 *An independent two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, satisfies axioms (A6.2,A8.2) iff there exist functions, $B : J \to R_{++}$, and, $F : V \to R_{++}$, such that for all $i \in I, D \in \boldsymbol{D}, j \in D$, and $c \in C$,*

$$P_c^1(j \mid iD) = \frac{B(j)F(c_{ij})}{\sum_{h \in D} B(h)F(c_{ij})}. \tag{4.540}$$

PROOF: First observe that if (4.540) holds then it follows [as in (2.234)] that for all $i, g \in I$, $D, H \in \boldsymbol{D}$, $j, h \in D \cap H$, and $c, c' \in C$ with $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$,

$$\frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}^1(j \mid gH)}{P_{c'}^1(h \mid gH)}, \tag{4.541}$$

and similarly [as in (2.235)] that for all $c, c' \in C$ with $c_{ij} = c_{ih}$ and $c'_{ij} = c'_{ih}$,

$$\frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{B(j)}{B(h)} = \frac{P_{c'}^1(j \mid iH)}{P_{c'}^1(h \mid iH)}. \tag{4.542}$$

Hence $\boldsymbol{P}$ satisfies both axioms A6.2 and A8.2. To establish the converse, observe that if for any fixed origin, $a \in I$, and any pair of destinations, $j, h \in J$, and separation profiles, $x, y \in V$, we choose any separation configuration, $c^\circ \in C(= V^{I \times J})$, with $c^\circ_{aj} = x$ and $c^\circ_{ah} = y$, then it follows from axiom A6.2 that [as in (4.526)] the relation

$$\phi_{jh}(x, y) = \frac{P_{c^\circ}^1(j|aJ)}{P_{c^\circ}^1(h \mid aJ)} \tag{4.543}$$

yields a well defined function, $\phi : V^2 \to R_{++}$, which satisfies the following identity for all $i \in I$, $c \in C$, and $D \in \boldsymbol{D}$ with $j, h \in D$,

$$\phi_{jh}(c_{ij}, c_{ih}) = \frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)}. \tag{4.544}$$

Next, if for any distinct destinations, $j, h, d \in J$, and any separation profiles, $x, y, z \in V$, we choose some configuration, $c \in C$, with $c_{aj} = x$, $c_{ah} = y$, and $c_{ad} = z$, then [as in (4.530)] we obtain the identity

$$\phi_{jh}(x, y) = \frac{P_c^1(j \mid aJ)/P_c^1(d \mid aJ)}{P_c^1(h \mid aJ)/P_c^1(d \mid aJ)} = \frac{\phi_{id}(x, z)}{\phi_{hd}(y, z)}. \tag{4.545}$$

Similarly,

$$\phi_{jd}(x, z) = \phi_{jh}(x, z)/\phi_{dh}(z, z) \text{ and } \phi_{hd}(y, z) = \phi_{hj}(y, z)/\phi_{dj}(z, z),$$

which together with (4.545) imply that

$$\phi_{jh}(x, y) = \frac{\phi_{jh}(x, z)/\phi_{dh}(z, z)}{\phi_{hj}(y, z)/\phi_{dj}(z, z)} = \frac{\phi_{jh}(x, z)}{\phi_{hj}(y, z)} \cdot \frac{\phi_{dj}(z, z)}{\phi_{dh}(z, z)}. \tag{4.546}$$

Thus if we now let $M = |J| - 1$, and choose any fixed profile, $\sigma \in V$, then for each distinct destination pair, $j, h \in J$, and choice of $d \in J - \{j, h\}$, it follows from (4.545) and (4.546) that for all profiles, $x, y \in V$,

$$\begin{aligned} \phi_{jh}(x, y)^M &= \frac{\phi_{jh}(x, \sigma)}{\phi_{hj}(y, \sigma)} \cdot \frac{\phi_{dj}(\sigma, \sigma)}{\phi_{dh}(\sigma, \sigma)} \prod_{k \in J - \{j, h\}} \left[\frac{\phi_{jk}(x, \sigma)}{\phi_{hk}(y, \sigma)}\right] \\ &= \frac{\phi_{dj}(\sigma, \sigma)}{\phi_{dh}(\sigma, \sigma)} \cdot \frac{\prod_{k \in J - \{j\}} \phi_{jk}(x, \sigma)}{\prod_{k \in J - \{h\}} \phi_{hk}(y, \sigma)}. \end{aligned} \tag{4.547}$$

Hence, by letting

$$r_{jh} = \frac{\phi_{dj}(\sigma, \sigma)^{1/M}}{\phi_{dh}(\sigma, \sigma)^{1/M}}, \tag{4.548}$$

and defining the functions, $\alpha_j : V \to R_{++}$, for all $j \in J$ and $x \in V$ by

$$\alpha_j(x) = \prod_{k \in J - \{j\}} \phi_{jk}(x, \sigma)^{1/M}, \tag{4.549}$$

we obtain

$$\begin{aligned} \phi_{jh}(x, y) &= \frac{\phi_{dj}(\sigma, \sigma)^{1/M}}{\phi_{dh}(\sigma, \sigma)^{1/M}} \cdot \frac{\prod_{k \in J - \{j\}} \phi_{jk}(x, \sigma)^{1/M}}{\prod_{k \in J - \{h\}} \phi_{hk}(y, \sigma)^{1/M}} \\ &= r_{jh} \left[\frac{\alpha_j(x)}{\alpha_h(y)} \right]. \end{aligned} \tag{4.550}$$

But by axiom A8.2 it also follows that if for any $j, h \in J$ and $x, y \in V$ we choose $c, c' \in C$ with $c_{aj} = c_{ah} = x$ and $c'_{aj} = c'_{ah} = y$, then

$$\begin{aligned} \phi_{jh}(x, x) = \frac{P_c^1(j \mid aJ)}{P_c^1(h \mid aJ)} &= \frac{P_{c'}^1(j \mid aJ)}{P_{c'}^1(h \mid aJ)} = \phi_{jh}(y, y) \\ \Rightarrow r_{jh} \left[\frac{\alpha_j(x)}{\alpha_h(x)} \right] &= r_{jh} \left[\frac{\alpha_j(y)}{\alpha_h(y)} \right] \\ \Rightarrow \alpha_j(x) &= \alpha_j(y)[\alpha_h(x)/\alpha_h(y)]. \end{aligned} \tag{4.551}$$

Hence, if for any fixed destination, $b \in J$, we let the functions, $B : J \to R_{++}$, and, $F : V \to R_{++}$, be defined respectively for all $j \in J$ and $x \in V$ by

$$B(j) = \begin{cases} r_{jb}\alpha_j(\sigma), & j \in J - \{b\} \\ \alpha_b(\sigma), & j = b, \end{cases} \tag{4.552}$$

$$F(x) = \alpha_b(x), \tag{4.553}$$

then it follows from (4.544) and (4.550) through (4.553) that for all $c \in C$, $i \in I$, $D \in \boldsymbol{D}$, and $j \in D$,

$$\begin{aligned} \frac{P_c^1(j \mid iD)}{P_c^1(b \mid iD)} &= r_{jb} \left[\frac{\alpha_j(c_{ij})}{\alpha_b(c_{ib})} \right] = r_{jb} \left[\frac{\alpha_b(c_{ij})[\alpha_j(\sigma)/\alpha_b(\sigma)]}{\alpha_b(c_{ib})} \right] \\ &= \frac{r_{jb}\alpha_j(\sigma)}{\alpha_b(\sigma)} \cdot \frac{\alpha_b(c_{ij})}{\alpha_b(c_{ib})} \\ &= \frac{B(j)F(c_{ij})}{B(b)F(c_{ib})}, \end{aligned} \tag{4.554}$$

which together with the identity, $1 = \sum_{j \in D} P_c^1(j \mid iD)$, implies that (4.540) must hold for this choice of functions. □

4.6.3 Interaction Processes with Prominence Effects

Finally we consider the class of *prominence models* of conditional destination probabilities developed in Section 2.5.3(b) of Chapter 2. To do so, we first formalize the appropriate class of separation configurations for such processes, which include dissimilarity relations among alternative destinations as well as separation relations between origins and destinations. Hence, again letting $J = B$, we postulate the existence of a nonempty finite set of *origin-destination separation measures*, $c^k : I \times J \to R$, $k \in K$, yielding *(ij)-separation profiles*, $c_{ij} = (c_{ij}^k : k \in K) \in V \subseteq R^K$, for each origin-destination pair, $ij \in I \times J$. In addition, we now postulate the existence of a nonempty finite set of *destination dissimilarity measures*, $c^w : J \times J \to R$, $w \in W$, satisfying the *symmetry condition*, $c_{jh}^w = c_{hj}^w$, for all $j, h \in J$ and $w \in W$. If the corresponding *(jh)-separation profiles* are denoted by $c_{jh} = (c_{jh}^w : w \in W)$, then these two types of profiles are together taken to define the relevant *composite separation configuration*, $c = [(c_{ij} : ij \in I \times J),\ (c_{jh} : jh \in J \times J)]$, for such two-stage processes. If the domain of feasible dissimilarity profiles is denoted by $U \subseteq R^W$, then it is postulated that the relevant *configuration* class for this case is given by

$$C = \{c \in (V^{I \times J}) \times (U^{J \times J}) : c_{jh} = c_{hj},\ j, h \in J\}. \tag{4.555}$$

[In addition, if $I \cap J \neq \phi$, then (as in the introductory remarks to Section 4.5.2) we implicitly replace I and J by *disjoint copies*, $I \cup \{1\}$ and $J \cup \{2\}$, in order to avoid any ambiguities in the labeling of components of each composite separation configuration.] Next, if the class of all destination sets with k or more elements is denoted by $\boldsymbol{D}_k = \{D \in \boldsymbol{D} : |D| \geq k\}$, and if we write $D/j = D - \{j\}$, then for each $j \in D \in \boldsymbol{D}_2$ and $c \in C$ the *(jD)-separation profile* in c is denoted by

$$c_{jD} = (c_{jh} : h \in D/j). \tag{4.556}$$

Finally, if the class of permutations, $\pi = (\pi_1, \ldots, \pi_n)$, of the integers, $(1, \ldots, n)$, is again denoted by Π_n, and if for each vector, $x = (x_1, \ldots, x_n) \in R^n$, we denote the corresponding permutation of x by x_π [as in (3.17) of Chapter 3], then for all $x, y \in R^n$ we now write

$$x =_s y \Leftrightarrow x = y_\pi \quad \text{for some} \quad \pi \in \Pi_n, \tag{4.557}$$

$$x \geq_s y \Leftrightarrow x \geq y_\pi \quad \text{for some} \quad \pi \in \Pi_n, \tag{4.558}$$

and similarly, write $x \leq_s y \Leftrightarrow y \geq_s x$. These definitions are easily seen to imply that

$$x =_s y \Leftrightarrow (x \geq_s y,\ x \leq_s y). \tag{4.559}$$

In terms of this notation, we may now state the following five axioms for regular two-stage processes:

P1. (Pairwise Destination Proportionality) *For all origins, $i, g \in I$, distinct destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c'_{gj},\ c_{ih} = c'_{gh}) \Rightarrow \frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{P_{c'}^1(j \mid g\{jh\})}{P_{c'}^1(h \mid g\{jh\})}. \quad (4.560)$$

P2. (Pairwise Destination Separability) *For all origins, $i \in I$, distinct destinations, $j, h \in J$, and separation configurations, $c, c' \in C$,*

$$(c_{ij} = c_{ih}, c'_{ij} = c'_{ih}) \Rightarrow \frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{P_{c'}^1(j \mid i\{jh\})}{P_{c'}^1(h \mid i\{jh\})}. \quad (4.561)$$

P3. (Pairwise Destination Independence) *For all origins, $i \in I$, distinct destinations, $j, h, d \in J$, and separation configurations, $c \in C$,*

$$\frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{P_c^1(j \mid i\{jd\})/P_c^1(d \mid i\{jd\})}{P_c^1(h \mid i\{hd\})/P_c^1(d \mid i\{hd\})}. \quad (4.562)$$

P4. (Prominence Monotonicity) *For all $i, g \in I$, $D \in \boldsymbol{D}_3$, $\{jh\}, \{bd\} \subseteq D$, and separation configurations, $c, c' \in C$, if $c_{jD} \geq_s c'_{bD}$ and $c_{hD} \leq_s c'_{dD}$, then*

$$\frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{P_c^1(j \mid i\{jh\})/P_c^1(h \mid i\{jh\})} \geq \frac{P_{c'}^1(b \mid gD)/P_{c'}^1(d \mid gD)}{P_{c'}^1(b \mid g\{bd\})/P_{c'}^1(d \mid g\{bd\})}. \quad (4.563)$$

P5. (Pairwise Prominence Independence) *For all $i \in I$, $D \in \boldsymbol{D}_3$, $j, h, d \in D$, and all configurations, $c, c', c'' \in C$ with $c_{iD} = c'_{iD} = c''_{iD}$, $c_{jD} =_s c'_{jD}$, $c_{hD} =_s c''_{hD}$, and $c'_{dD} =_s c''_{dD}$,*

$$\frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{P_{c'}^1(j \mid iD)/P_{c'}^1(d \mid iD)}{P_{c''}^1(h \mid iD)/P_{c''}^1(d \mid iD)}. \quad (4.564)$$

In terms of these axioms, we now have the following *prominence theory* of conditional destination interactions for regular two-stage interaction processes:

Definition 4.14 A regular two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is said to be consistent with the *prominence theory* of conditional destination interactions iff $\boldsymbol{P}$ satisfies $(P1, P2, P3, P4, P5)$.

Next, we formalize the following class of *prominence functions* for representing prominence effects in conditional destination interactions:

Definition 4.15 For any given destination set, $D \in \boldsymbol{D}_2$, a function, $\Phi_D : U^{|D|-1} \to R_{++}$, is designated as a *D-prominence function* iff the following symmetric monotonicity condition holds for all $x, y \in U^{|D|-1}$,

$$x \geq_s y \Rightarrow \Phi_D(x) \geq \Phi_D(y). \quad (4.565)$$

Observe in particular from the symmetry condition, $c_{jh} = c_{hj}$, that each D-prominence function for binary set, $D = \{jh\}$, satisfies the identity $\Phi_D(c_{jD}) \equiv \Phi_D(c_{hD})$ for all $c \in C$. With these definitions, we are now ready to establish the following representational consequence of the prominence theory:

Proposition 4.16 *A regular two-stage interaction process, $\boldsymbol{P} = \{P_c : c \in C\}$, is consistent with the prominence theory of conditional destination interactions iff there exist functions, $B : J \to R_{++}$, $F : V \to R_{++}$, together with D-prominence functions, $\Phi_D : U^{|D|-1} \to R_{++}$, for each $D \in \boldsymbol{D}_2$ such that for all $c \in C$, $i \in I$, $D \in \boldsymbol{D}_2$, and $j, h \in D$,*

$$p_c(j \mid iD) = \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\sum_{h\in D} \Phi_D(c_{hD})B(h)F(c_{ih})}. \tag{4.566}$$

PROOF: (i) To establish necessity, observe that if (4.566) holds then [as in (2.256)], it must be true for all $i, g \in I$, $j, h \in J$, and $c, c' \in C$, that the equalities $c_{ij} = c'_{gj}$ and $c_{ih} = c'_{gh}$ imply

$$\frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c'_{gj})}{B(h)F(c'_{gh})} = \frac{P_{c'}^1(j \mid g\{jh\})}{P_{c'}^1(h \mid g\{jh\})}, \tag{4.567}$$

and similarly [as in (2.257)] that the equalities $c_{ij} = c_{ih}$ and $c'_{ij} = c'_{ih}$ imply

$$\frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{B(j)}{B(h)} = \frac{P_{c'}^1(j \mid i\{jh\})}{P_{c'}^1(h \mid i\{jh\})}. \tag{4.568}$$

Moreover, it also follows [as in (2.260)] that for all $i \in I$, $c \in C$, and distinct $j, h, d \in J$

$$\begin{aligned} \frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} &= \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})} = \frac{B(j)F(c_{ij})/B(d)F(c_{id})}{B(h)F(c_{ih})/B(d)F(c_{id})} \\ &= \frac{P_c^1(j \mid i\{jd\})/P_c^1(d \mid i\{jd\})}{P_c^1(h \mid i\{hd\})/P_c^1(d \mid i\{hd\})}. \end{aligned} \tag{4.569}$$

Next observe that [as in (2.263)] the representation in (4.566) together with (4.565) yields the following implication for all $c, c' \in C$, $i, g \in I$, $D \in \boldsymbol{D}_3$, and $\{jh\}, \{bd\} \subseteq D$,

$$\begin{aligned} &(c_{jD} \geq_s c'_{bD},\ c_{hD} \leq_s c'_{dD}) \\ &\quad \Rightarrow [\Phi_D(c_{jD}) \geq \Phi_D(c'_{bD}),\ \Phi_D(c_{hD}) \leq \Phi_D(c'_{dD})] \\ &\quad \Rightarrow \frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{P_c^1(j \mid i\{jh\})/P_c^1(h \mid i\{jh\})} = \frac{\Phi_D(c_{jD})}{\Phi_D(c_{hD})} \geq \frac{\Phi_D(c'_{bD})}{\Phi_D(c'_{dD})} \\ &\quad = \frac{P_{c'}^1(b \mid gD)/P_{c'}^1(d \mid gD)}{P_{c'}^1(b \mid g\{bd\})/P_{c'}^1(d \mid g\{bd\})}. \end{aligned} \tag{4.570}$$

Finally, for any $i \in I$, $j,h,d \in D \in \boldsymbol{D}_3$, and $c,c',c'' \in C$, with $c_{iD} = c'_{iD} = c''_{iD}$, $c_{jD} =_s c'_{jD}$, $c'_{dD} =_s c''_{dD}$, and $c''_{hD} =_s c_{hD}$, conditions (4.559) and (4.565) imply that $\Phi_D(c_{jD}) = \Phi_D(c'_{jD})$, $\Phi_D(c'_{dD}) = \Phi_D(c''_{dD})$, and $\Phi_D(c''_{hD}) = \Phi_D(c_{hD})$, which in turn implies from (4.566) that

$$\begin{aligned}\frac{P^1_c(j \mid iD)}{P^1_c(h \mid iD)} &= \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\Phi_D(c_{hD})B(h)F(c_{ih})} = \frac{\Phi_D(c'_{jD})B(j)F(c'_{ij})}{\Phi_D(c''_{hD})B(h)F(c''_{ih})} \\ &= \frac{\Phi_D(c'_{jD})B(j)F(c'_{ij})}{\Phi_D(c''_{hD})B(h)F(c''_{ih})} \cdot \frac{\Phi_D(c''_{dD})B(d)F(c''_{id})}{\Phi_D(c'_{dD})B(d)F(c'_{id})} \\ &= \frac{P^1_{c'}(j \mid iD)/P^1_{c'}(d \mid iD)}{P^1_{c''}(h \mid iD)/P^1_{c''}(d \mid iD)}.\end{aligned} \tag{4.571}$$

Hence the necessity of (P1,P2,P3,P4,P5) follows at once from (4.567), (4.568), (4.569), (4.570) and (4.571).

(ii) To establish sufficiency, it must be shown that axiom system (P1,P2, P3,P4,P5) implies the existence of functions, $B: J \to R_{++}$, $F: V \to R_{++}$, together with D-prominence functions, $\Phi_D: U^{|D|-1} \to R_{++}$, $D \in \boldsymbol{D}_2$, such that (4.566) holds. To do so, we begin by considering *binary* destination sets in $\boldsymbol{D}_2$. Observe first that if for any distinct destinations, $j,h \in J$, and any separation profiles, $x,y \in V$, we choose a fixed origin, $a \in I$, and separation configuration, $c^\circ \in C$ with $c^\circ_{aj} = x$ and $c^\circ_{ah} = y$, then [in a manner similar to (4.543) and (4.544) above] it follows from axiom P1 that the relation

$$\phi_{jh}(x,y) = \frac{P^1_{c^\circ}(j \mid a\{jh\})}{P^1_{c^\circ}(h \mid a\{jh\})} \tag{4.572}$$

yields a well defined function, $\phi_{jh}: V^2 \to R_{++}$, which satisfies the following identity for all $i \in I$, $c \in C$, and $j,h \in J$,

$$\phi_{jh}(c_{ij}, c_{ih}) = \frac{P^1_c(j \mid i\{jh\})}{P^1_c(h \mid i\{jh\})}. \tag{4.573}$$

Next (recalling that $|J| \geq 3$ for regular two-stage processes) it follows that if for any distinct destinations, $j,h,d \in J$ and separation profiles, $x,y,z \in V$, we choose $i \in I$ and $c \in C$ with $c_{ij} = x$, $c_{ih} = y$ and $c_{id} = z$, then by axiom P3 together with (4.573) we must have,

$$\phi_{jh}(x,y) = \frac{P^1_c(j \mid i\{jd\})/P^1_c(d \mid i\{jd\})}{P^1_c(h \mid i\{hd\})/P^1_c(d \mid i\{hd\})} = \frac{\phi_{jd}(x,z)}{\phi_{hd}(y,z)}. \tag{4.574}$$

Similarly,

$$\phi_{jd}(x,z) = \phi_{jh}(x,z)/\phi_{dh}(z,z) \text{ and } \phi_{hd}(y,z) = \phi_{hj}(y,z)/\phi_{dj}(z,z),$$

which together with (4.574) imply that

$$\phi_{jh}(x,y) = \frac{\phi_{jh}(x,z)/\phi_{dh}(z,z)}{\phi_{hj}(y,z)/\phi_{dj}(z,z)} = \frac{\phi_{jh}(x,z)}{\phi_{hj}(y,z)} \cdot \frac{\phi_{dj}(z,z)}{\phi_{dh}(z,z)}. \tag{4.575}$$

Hence if we now let $M = |J| - 1$, and for any fixed profile, $\sigma \in V$, and choice of $d \in J - \{j, h\}$ let

$$r_{jh} = \frac{\phi_{dj}(\sigma,\sigma)^{1/M}}{\phi_{dh}(\sigma,\sigma)^{1/M}} \tag{4.576}$$

and define the functions, $\alpha_j : V \to R_{++}$, for all $j \in J$ and $x \in V$ by

$$\alpha_j(x) = \prod_{k \in J - \{j\}} \phi_{jk}(x,\sigma)^{1/M}, \tag{4.577}$$

then by exactly the same arguments in expressions (4.545) through (4.550) in the proof of Proposition 4.15 it follows that for all $j, h \in J$ and $x, y \in V$,

$$\phi_{jh}(x,y) = r_{jh}[\alpha_j(x)/\alpha_h(y)]. \tag{4.578}$$

Moreover, by axiom P2 it also follows [as in (4.551)] that if for any distinct $j, h \in J$ and any $x, y \in V$ we choose $i \in I$ and $c, c' \in C$ with $c_{aj} = c_{ah} = x$ and $c'_{aj} = c'_{ah} = y$, then

$$\begin{aligned} \phi_{jh}(x,x) = \frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} &= \frac{P_{c'}^1(j \mid i\{jh\})}{P_{c'}^1(h \mid i\{jh\})} = \phi_{jh}(y,y) \\ \Rightarrow r_{jh}\left[\frac{\alpha_j(x)}{\alpha_h(x)}\right] &= r_{jh}\left[\frac{\alpha_j(y)}{\alpha_h(y)}\right] \\ \Rightarrow \alpha_j(x) &= \alpha_j(y)[\alpha_h(x)/\alpha_h(y)]. \end{aligned} \tag{4.579}$$

Hence, if for any fixed destination, $b \in J$, we again let the functions, $B : J \to R_{++}$, and, $F : V \to R_{++}$, be defined respectively for all $j \in J$ and $x \in V$ by

$$B(j) = \begin{cases} r_{jb}\alpha_j(\sigma), & j \in J - \{b\} \\ \alpha_b(\sigma), & j = b, \end{cases} \tag{4.580}$$

$$F(x) = \alpha_b(x), \tag{4.581}$$

then it follows from (4.573) and (4.579) through (4.581) that for all $c \in C$, $i \in I$, $j \in D$,

$$\begin{aligned} \frac{P_c^1(j \mid i\{j,b\})}{P_c^1(b \mid i\{j,b\})} &= r_{jb}\left[\frac{\alpha_j(c_{ij})}{\alpha_b(c_{ib})}\right] = r_{jb}\left[\frac{\alpha_b(c_{ij})[\alpha_j(\sigma)/\alpha_b(\sigma)]}{\alpha_b(c_{ib})}\right] \\ &= \frac{r_{jb}\alpha_j(\sigma)}{\alpha_b(\sigma)} \cdot \frac{\alpha_b(c_{ij})}{\alpha_b(c_{ib})} = \frac{B(j)F(c_{ij})}{B(b)F(c_{ib})}, \end{aligned} \tag{4.582}$$

which in turn implies from axiom P3 that for all $j, h \in J$,

$$\frac{P_c^1(j \mid i\{jh\})}{P_c^1(h \mid i\{jh\})} = \frac{P_c^1(j \mid i\{jb\})/P_c^1(b \mid i\{jb\})}{P_c^1(h \mid i\{hb\})/P_c^1(b \mid i\{hb\})} = \frac{B(j)F(c_{ij})}{B(h)F(c_{ih})}. \tag{4.583}$$

Finally, the symmetry of (jh)-separation profiles implies that for each binary set, $D = \{j, h\}$, the function, $\Phi_D : U \to R_{++}$, defined for all $x \in U$ by $\Phi_D(x) = 1$ is automatically a D-prominence function. Hence for this choice of Φ_D, we see that the equality

$$\frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\Phi_D(c_{hD})B(h)F(c_{hD})} \tag{4.584}$$

holds identically and may thus conclude from the normalization condition, $1 = P_c^1(j \mid iD) + P_c^1(h \mid iD)$, that (4.566) holds for all binary distribution sets D.

Next we consider any destination set, $D \in \boldsymbol{D}_3$. For each such set, observe from (4.559) that for all $c, c' \in C$, $i, g \in I$, and $j, h \in D$, if $(c_{jD} =_s c'_{bD},\ c_{hD} =_s c'_{dD})$ then both $(c_{jD} \geq_s c'_{bD},\ c_{hD} \leq_s c'_{dD})$ and $(c_{jD} \leq_s c'_{bD},\ c_{hD} \geq_s c'_{dD})$ hold. Hence by two applications of axiom P4 we see that (4.563) holds as an equality for this case, so that by substituting (4.583) into (4.563) we obtain the equality

$$\frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{B(j)F(c_{ij})/B(h)F(c_{ih})} = \frac{P_{c'}^1(j \mid gD)/P_{c'}^1(h \mid gD)}{B(j)F(c'_{gj})/B(h)F(c'_{gh})}. \tag{4.585}$$

Next, if for each (jD)-separation profile, c_{jD}, and $h \in D/j$ we write $c_{jD} = (c_{jh},\ c_{jD/h})$, and choose any *fixed* origin, $a \in I$, and distinct *fixed* destinations, $b, d \in D$, then it follows from the definition of C in (4.555) that for each profile, $u \in U$, and pair of profile arrays, $x, y \in U^{|D|-2}$, there exists a separation configuration, $c^\circ \in C$, with $c^\circ_{bd} = u\ (= c^\circ_{db}), c^\circ_{bD/d} =_s x$, and $c^\circ_{dD/b} =_s y$. But since for any other choice of origin, $i \in I$, destinations, $j, h \in J$, and configuration, $c \in C$, with $c_{jh} = u\ (= c_{hj})$, $c_{jD/h} =_s x$, and $c_{hD/j} =_s y$, it must by definition be true that $c_{jD} =_s c^\circ_{bD}$ and $c_{hD} =_s c^\circ_{dD}$, we may conclude from (4.585) that the relation,

$$\Psi_D(u, x, y) = \frac{P_{c^\circ}^1(b \mid aD)/P_{c^\circ}^1(d \mid aD)}{B(b)F(c_{ab})/B(d)F(c_{ad})} \tag{4.586}$$

yields a well defined function, $\Psi_D : U \times (U^{|D|-2})^2 \to R_{++}$, which is independent of the choices of $a \in I$, $b, d \in J$, and $c^\circ \in C$, and which by construction satisfies the following identity for all $c \in C$ and $i \in I$:

$$\Psi_D(c_{jh}, c_{jD/h}, c_{hD/j}) = \frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{B(j)F(c_{ij})/B(h)F(c_{ih})}. \tag{4.587}$$

Next observe that since $|D| \geq 3$ implies the existence of three distinct destinations, $j, h, d \in D$, it follows from (4.555) that for any choice of origin, $i \in I$, and profiles, $u \in U$ and $x, y, z \in U^{|D|-2}$, there must exist configurations, $c, c', c'' \in C$, with $c_{iD} = c'_{iD} = c''_{iD}$, $c_{jh} = c'_{jd} = c''_{hd} =$

u, $c_{jD/h} =_s c'_{jD/d} =_s x$, $c'_{dD/j} =_s c''_{dD/h} =_s z$, and $c''_{hD/d} =_s c_{hD/j} =_s y$. In particular this implies that

$$\begin{aligned}
(c_{jh}, c_{jD/h}) &=_s (u,x) =_s (c'_{jd}, c'_{jD/d}) \Rightarrow c_{jD} =_s c'_{jD} \\
(c_{hj}, c_{hD/j}) &=_s (u,y) =_s (c''_{hd}, c''_{hD/d}) \Rightarrow c_{hD} =_s c''_{hD} \\
(c'_{dj}, c'_{dD/j}) &=_s (u,z) =_s (c''_{dh}, c''_{dD/h}) \Rightarrow c'_{dD} =_s c''_{dD},
\end{aligned} \tag{4.588}$$

so that by (4.587) and axiom P5 we obtain the identity:

$$\begin{aligned}
\Psi_D(u,x,y) &= \frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{B(j)F(c_{ij})/B(h)F(c_{ih})} \\
&= \frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} \cdot \frac{B(h)F(c_{ih})}{B(j)F(c_{ij})} \\
&= \frac{P_{c'}^1(j \mid iD)/P_{c'}^1(d \mid iD)}{P_{c''}^1(h \mid iD)/P_{c''}^1(d \mid iD)} \cdot \frac{B(h)F(c''_{ih})/B(d)F(c''_{id})}{B(j)F(c'_{ij})/B(d)F(c'_{id})} \\
&= \frac{P_{c'}^1(j \mid iD)/P_{c'}^1(d \mid iD)}{B(j)F(c'_{ij})/B(d)F(c'_{id})} \cdot \frac{B(h)F(c''_{ih})/B(d)F(c''_{id})}{P_{c''}^1(h \mid iD)/P_{c''}^1(d \mid iD)} \\
&= \Psi_D(u,x,z)/\Psi_D(u,y,z).
\end{aligned} \tag{4.589}$$

Hence, if for any profile array, $\sigma \in U^{|D|-2}$, we define the function, $\Phi_D : U^{|D|-1} \to R_{++}$, for all $(u,x) \in U \times (U^{|D|-2}) = U^{|D|-1}$ by

$$\Phi_D(u,x) = \psi_D(u,x,\sigma), \tag{4.590}$$

then it follows from (4.587), (4.589) and (4.590) that for all $c \in C$, $i \in I$, and $j,h \in D$,

$$\begin{aligned}
&\frac{P_c^1(j \mid iD)/P_c^1(h \mid iD)}{B(j)F(c_{ij})/B(h)F(c_{ih})} = \Psi_D(c_{jh}, c_{jD/h}, c_{hD/j}) \\
&= \frac{\Psi_D(c_{jh}, c_{jD/h}, \sigma)}{\Psi_D(c_{hj}, c_{hD/j}, \sigma)} = \frac{\Phi_D(c_{jh}, c_{jD/h})}{\Phi_D(c_{hj}, c_{hD/j})} = \frac{\Phi_D(c_{jD})}{\Phi_D(c_{hD})}.
\end{aligned} \tag{4.591}$$

Thus we see that for all $i \in I$, $c \in C$, and $j,h \in D \in \boldsymbol{D}_3$,

$$\frac{P_c^1(j \mid iD)}{P_c^1(h \mid iD)} = \frac{\Phi_D(c_{jD})B(j)F(c_{ij})}{\Phi_D(c_{hD})B(h)F(c_{ih})}, \tag{4.592}$$

and may conclude from the normalization condition, $1 = \sum_{j \in D} P_c^1(j \mid iD)$, that (4.566) must hold for this choice of functions, Φ_D. Finally, to establish that Φ_D satisfies the symmetric monotonicity condition (4.565), observe first that if for any $x,y \in U^{|D|-2}$ and $u \in U$ we choose any origin, $a \in I$, destinations, $b,d \in J$, and configuration, $c \in C$, with $c_{bd} = c_{db} = u$, $c_{bD/d} = x$, and $c_{dD/b} = y$, then it follows from (4.592) and condition

(4.563) of P4 [with $c' = c$ and $b = d = h$] that

$$\begin{aligned} x \geq_s y &\Rightarrow (u,x) \geq_s (u,y) \Rightarrow (c_{bD} \geq_s c_{dD},\ c_{bD} \leq_s c_{bD}) \\ &\Rightarrow \frac{\Phi_D(u,x)}{\Phi_D(u,y)} = \frac{P_c^1(b\,|\,aD)/P_c^1(d\,|\,aD)}{B(b)F(c_{ab})/B(d)F(c_{ad})} \\ &\qquad\qquad\qquad \geq \frac{P_c^1(b\,|\,aD)/P_c^1(b\,|\,aD)}{B(b)F(c_{ab})/B(b)F(c_{ab})} = 1 \\ &\Rightarrow \Phi_D(u,x) \geq \Phi_D(u,y). \end{aligned} \tag{4.593}$$

Moreover, it also follows from (4.559) that for any $z, w \in U^{|D|-1}$,

$$z =_s w \Rightarrow \Phi_D(z) = \Phi_D(w). \tag{4.594}$$

Hence, by employing these relations, we may now establish symmetric monotonicity as follows. Choose any

$$x = (x_1, x_2, \ldots, x_n) \text{ and } y = (y_1, y_2, \ldots, y_n)$$

in $U^{|D|-1}$ (where by hypothesis $|D| \geq 3 \Rightarrow n \geq 2$), and observe from (4.558) that

$$\begin{aligned} x \geq_s y &\Rightarrow (x_1, x_2, \ldots, x_n) \geq_s (y_1, y_2, \ldots, y_n) \\ &\Rightarrow (x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}) \geq (y_1, y_2, \ldots, y_n) \end{aligned} \tag{4.595}$$

for some permutation $\pi = (\pi_1, \pi_2, \ldots, \pi_n) \in \Pi_n$. But by employing (4.593), (4.594) and (4.595) together with (4.557) we see that

$$\begin{aligned} &(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}) \geq (x_{\pi_1}, y_2, \ldots, y_n) =_s (y_2, x_{\pi_1}, \ldots, y_n) \\ &\geq (y_2, y_1, \ldots, y_n) =_s y \\ &\Rightarrow \Phi_D(x) = \Phi_D(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}) \geq \Phi_D(x_{\pi_1}, y_2, \ldots, y_n) \\ &\quad = \Phi_D(y_2, x_{\pi_1}, \ldots, y_n) \geq \Phi_D(y_2, y_1, \ldots, y_n) = \Phi_D(y) \\ &\Rightarrow \Phi_D(x) \geq \Phi_D(y), \end{aligned} \tag{4.596}$$

so that (4.565) follows by combining (4.595) and (4.596). Hence Φ_D is a D-prominence function for each $D \in \boldsymbol{D}_3$, and the result is established. □

4.7 Notes and References

In this final section, it is appropriate to relate the more important concepts and results of this chapter to those existing in the literature. As in Section 3.10 above, we shall discuss the concepts and results of each subsection in turn.

SECTION 4.2

The classification of models in this section extends the classification in Smith (1987a) to include Models G2, G3, and G4 (together with their associated exponential and monotone versions). This classification includes almost all gravity models appearing in the literature. Indeed, [as noted by Sen and Sööt (1981)], most models are in fact instances of Model E2, together with a sufficiently broad interpretation of the relevant components of separation configurations. The focus in the present classification scheme is primarily on the degree of dependency between model forms and separation configurations. For additional discussion of such model sensitivity questions, see for example, Choukroun (1975) and Weber and Sen (1985).

SECTION 4.3

The Carroll-Bevis processes in Section 4.3.1 are of course inspired by the original work of Carroll and Bevis (1957). However, all of the definitions and results presented here are essentially new. The relevant literature for threshold interaction processes in Section 4.3.2 and simple search processes in Section 4.3.3 were discussed in Section 3.10 above.

SECTION 4.4

Most of the axioms in this section have appeared in the literature in various forms and, for the most part are all contained in Smith (1984,1987a). The relevant literature for axioms A1 and A2 was discussed in Section 3.10 above. Axioms A4 is inspired by the general 'efficiency principle' studied in Smith (1978b), and is in part inspired by the broader principle of 'least effort' first articulated by Zipf (1949). The more general version of this axiom in A3 first appeared in Smith (1987a), and is simply an extension to the case of non-monotonic deterrence functions. Axiom A5 is a non-monotonic type of functional 'separability' axiom [as discussed more generally Eichhorn (1978) (1978, Section 6.1), for example]. In the context of gravity models, this axiom first appeared in Smith (1987a), and seems to have no previous applications. Axioms (A6,A7,A8,A9,A10,A11) first appeared in Smith (1984). However, axioms A6 and A7 are closely related to previous work by Kirby (1970) and Cesario (1973) on the proportional invariance properties of gravity models. Axioms A13 represents a multidimensional version of the 'cost efficiency' principle first studied in Smith (1978a), and later extended by Erlander (1985), Smith (1987a), and Erlander and Smith (1990). The weaker version of this axiom in A12 is closely related to the Boltzmann hypothesis ('maximum entropy' principle) studied by Wilson (1967) and others [as discussed in Smith (1978a, Section 4.1) and in Section 2.4.3(A) above]. Finally, the configuration-independent versions of axioms (A3,A4,A12,A13), namely (A3$'$, A4$'$, A12$'$, A13$'$) and (A3$''$, A4$''$, A12$''$, A13$''$), are all new, and serve primarily to unify the characterization of all model forms in terms of aggregate population behavior.

SECTION 4.5

The analytical definitions and results in Section 4.5.1 are all standard, except for Lemmas 4.4, 4.5, and 4.9. The present proof of Lemma 4.4 unifies the results of Lemma 3.4 in Smith (1978a) and Lemma 4 in Smith (1978b). As mentioned in the text, a full development of Lemma 4.5 is given in Smith (1986b). Lemma 4.9 is needed only for the special construction in Example 4.3 (which is new). Turning to the aggregate characterizations of general gravity models in Section 4.5.2(A), the proofs of Theorems 4.1 and 4.2 are essentially reformulations (within the present more general framework) of the arguments given in Smith (1987a). [With respect to Theorem 4.2, it is of interest to note that this result extends the 'efficiency principle' in Smith (1978b) to a wider class of proximity relations. In particular, the proximity relations, $\wp_c$, in (4.187) may in general fail to yield a *connected* relation on $I \times J$, since neither $(ij, kh) \in \wp_c$ or $(kh, ij) \in \wp_c$ may hold for certain origin-destination pairs, $ij, kh \in I \times J$. Hence the proof of Theorem 4.2 shows that the 'efficiency principle' in Smith (1978b) is in fact extendable to notions of spatial proximity which are representable only by reflexive, transitive relations (i.e., *preorders*) on $I \times J$.] Theorems 4.3 through 4.6 are new, and in particular are based on the new aggregate behavioral axioms (A3$'$, A4$'$, A3$''$, A4$''$). The local characterizations of general gravity models in Section 4.5.2(B) were first established in Smith (1984). However, (as mentioned in Remark 4.3) the present proof of Theorem 4.9(ii) is new. In addition, the present proof of Theorem 4.13 is new, and is more directly tailored to the combined biproportionality axioms (A6,A7). Turning finally to the characterizations of exponential gravity models in Section 4.5.3, Theorems 4.15 and 4.16 are again reformulations of the corresponding results in Smith (1987a). [The linear programming argument employed in the proof of assertion (4.370) is due to Erlander. The present continuity assumption, R3, allows this result to be extended to all configurations, $c \in C$. However, more recent results established in Erlander and Smith (1990) show that this continuity assumption can in fact be dispensed with in most cases.] Finally, Theorems 4.17 through 4.23 are all new and, as with Theorems 4.3 through 4.6, are based on the new aggregate behavioral axioms (A12$'$, A13$'$, A12$''$, A13$''$). Of particular interest here is the proof of assertion (4.402) in Theorem 4.17, which is based on Cauchy's functional equation. Indeed, an examination of all proofs in Theorems 4.1 through 4.23 shows that this is the *only* argument requiring that the set, V, of profiles be a additively-closed product set in R^n containing the origin. Hence there remains an interesting open question as to whether assertion (4.402) can be established under more general conditions.

SECTION 4.6

The results of this final section are essentially all new. However, the representations of *hierarchical processes* in Section 4.6.1 follow directly from the aggregate characterizations of general gravity models in Theorem 4.1 above. The representations of *regular two-stage processes* in Section 4.6.2 are also based on the local characterization of G3 Models in Theorem 4.9 above. However, a close examination of the proof of Proposition 4.14 shows that the overlapping nature of destination sets (in contrast to individual destinations) requires an essentially different type of argument. Finally, the representation results for the *prominence theory* in Section 4.6.3 essentially combine the local characterization results in Theorem 4.9 with the results of Smith and Yu (1982). However, it is of interest to note that the requirement of symmetry in the definition of dissimilarity measures significantly complicates the analysis here. In particular, this *symmetry* condition implies that the class of feasible dissimilarity configurations *fail* to form a cartesian product set. Hence the argument establishing the existence of multiplicatively separable representations for this case is significantly more complex.

Part II

Methods

CHAPTER 5

Maximum Likelihood

5.1 Introduction

As mentioned in the introduction to the book, the key use for the gravity model is forecasting and for forecasting, it is important to know which aspects of the base period remain unchanged into the forecast period. Therefore, in the context of the gravity model, we would usually need to know which aspects of the model are configuration-free. In Part I of this book, we saw the conditions under which one or more of the functions $A_c(i)$, $B_c(j)$ and $F_c(c_{ij})$ are configuration-free and therefore can be assumed to remain invariant from base to forecast period. However, we still need to get numerical estimates of the parameters in these functions. This chapter and the next are devoted to this topic.

Several methods for parameter estimation have been studied in the statistics literature. Maximum likelihood and least squares — both linear and nonlinear — are among those most commonly used. In this chapter, the application of maximum likelihood procedures to gravity model parameter estimation will be examined. Chapter 6 will be devoted to least squares procedures. Bayesian methods are not considered in this book, although such procedures have been applied to gravity models [see West (1994)].

In Section 5.1.1 below, we present some notation and also specify the form of the model that will be considered in this chapter as well as in Chapter 6. Section 5.1.2 will discuss maximum likelihood in general and as it applies to the gravity model. Section 5.1.3 will then present a preview of the rest of the chapter.

5.1.1 PRELIMINARIES

In practical situations where the gravity model is used, there is a single observation for each N_{ij}. Therefore, no confusion need arise, and some simplicity is achieved, if we use the same symbol N_{ij} for a random variable and its observed value. Further, we assume that when these observations are collected, a single configuration $c \in C$ exists. Thus, for purposes of parameter estimation no explicit mention need be made of a configuration.

In this chapter and the next, we largely [although not entirely] confine our attention to the model

$$\boldsymbol{N} \in < \text{POISSON} >, \tag{5.1}$$

$$E(N_{ij}) = T_{ij} = A(i)B(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] : i \in I, j \in J. \tag{5.2}$$

As described in Section 2.3, the form (5.2) is general enough for most practical purposes, particularly since it is able to at least approximate the most general gravity model specification:

$$E(N_{ij}) = T_{ij} = A(i)B(j)F(\boldsymbol{c}_{ij}). \tag{5.3}$$

We also reiterate that the model above is not necessarily configuration-free; only that configurations are irrelevant when only one configuration is being considered. Perhaps also worth pointing out is the fact that, although we have chosen to focus on the model (5.2), maximum likelihood methods can be applied to more general forms of $F(\boldsymbol{c}_{ij})$. However, then some of the properties established in this chapter might not hold true and the procedures developed might be inappropriate. Nevertheless, it can be shown that some of the asymptotic properties of maximum likelihood estimates would continue to hold true under fairly general specifications of $F(\boldsymbol{c}_{ij})$.

Notice that, in a departure from the practice in previous chapters, no negative sign is placed in front of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^t$ in (5.2). Thus, one might see the θ_k's of this chapter and the next to be negatives of the θ_k's of the previous chapters. This change makes the notation more natural in discussions of parameter estimation and also simplifies presentation. Notice also that the c_{ij} and θ of previous chapters have been replaced by a $\boldsymbol{c}_{ij}$ and a $\boldsymbol{\theta}$. Since we shall be moving back and forth between vectors and their components, this change in notation is convenient. A boldfaced symbol will usually stand for a vector with components which are not bold. Thus, $\boldsymbol{c}_{ij} = ({c_{ij}}^{(1)}, \ldots, {c_{ij}}^{(K)})^t$ is a separation profile, and the individual ${c_{ij}}^{(k)}$'s are the values of individual separation measures. Notice also that, unless otherwise indicated, a vector will always be a *column* vector. Thus, $\boldsymbol{\theta}$ is a column vector and its transpose, $\boldsymbol{\theta}^t$, is a row vector.

These ${c_{ij}}^{(k)}$'s are assumed known in the present context. The $A(i)$'s, $B(j)$'s and θ_k's are the parameters to be estimated. In much of the algorithmic work presented in this chapter, we focus on the θ_k's. This is for several reasons:

1. As we shall see later, once suitable estimates of θ_k's are found, maximum likelihood estimates of $A(i)$'s and $B(j)$'s can be obtained easily using the DSF procedure (Section 5.3.1).

2. In many applications, particularly in transportation planning, an estimate of $\boldsymbol{\theta}$ is all that is required.

3. As we shall see in Section 5.3, the maximum likelihood estimate of $\boldsymbol{\theta}$ is unique, while those for $A(i)$'s and $B(j)$'s are not.

We shall also assume that the set I of origins is the finite set $I = \{1, \ldots, I\}$ with I members, so that we can write $i \in I$ and $i = 1, \ldots, I$ interchangeably. Using the same notation for a set and its cardinality need not cause any confusion. Similarly, we let $J = \{1, \ldots, J\}$ and $K = \{1, \ldots, K\}$, so that $j \in J$ and $k \in K$ are interchangeable with $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Moreover the elements of the vectors

$$\boldsymbol{N} = (N_{11}, \ldots, N_{1J}, \ldots\ldots, N_{I1}, \ldots, N_{IJ})^t \tag{5.4}$$

and

$$\boldsymbol{T} = (T_{11}, \ldots, T_{1J}, \ldots\ldots, T_{I1}, \ldots, T_{IJ})^t \tag{5.5}$$

will be assumed to be always in the specific order shown above. Also, in order to avoid somewhat awkward notation, $P(N_{ij})$ will stand for the probability that the r.v. N_{ij} takes the specific value N_{ij} (instead of saying $P[N_{ij} = N_{ij}]$).

5.1.2 Maximum Likelihood Estimation

Since each N_{ij} has a Poisson distribution and the N_{ij}'s are independent,

$$\begin{aligned} P(\boldsymbol{N}) &= \prod_{ij} P(N_{ij}) = \prod_{ij} \exp[-T_{ij}]\, T_{ij}^{N_{ij}} / N_{ij}! \\ &= \prod_{ij} \{\exp(-A(i)B(j)\, \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}])(A(i)B(j)\, \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}])^{N_{ij}} / N_{ij}!\}. \end{aligned} \tag{5.6}$$

This probability can readily be computed when T_{ij}'s are known, or alternatively if $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta})$, the concatenation $(\boldsymbol{A}^t, \boldsymbol{B}^t, \boldsymbol{\theta}^t)^t$ of

$$\begin{aligned} &\boldsymbol{A} = (A(1), \ldots, A(I))^t,\ \boldsymbol{B} = (B(1), \ldots, B(J))^t \text{ and} \\ &\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^t, \end{aligned} \tag{5.7}$$

is known. If, instead, observations N_{ij} are known and (5.6) is seen as a function of the $A(i)$'s, $B(j)$'s and θ_k's, it is called their likelihood function, and the values $\hat{A}(i)$, $\hat{B}(j)$ and $\hat{\theta}$ of $A(i)$, $B(j)$ and θ_k that maximize it are called Maximum Likelihood (ML) estimates. Let $\hat{\boldsymbol{A}} = (\hat{A}(1), \ldots, \hat{A}(I))^t$, $\hat{\boldsymbol{B}} = (\hat{B}(1), \ldots, \hat{B}(J))^t$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^t$ and let $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ be their concatenation. If the likelihood function is maximized by $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$, so will its logarithm, the log-likelihood function, which is seen from (5.6) to be

$$\begin{aligned} \mathcal{L} &= \sum_{ij} \{-A(i)B(j) \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] \\ &\quad + N_{ij}(\log A(i) + \log B(j) + \boldsymbol{\theta}^t \boldsymbol{c}_{ij}) - \log(N_{ij}!)\} \\ &= \sum_{ij} -A(i)B(j)\, \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] + \sum_i N_{i\oplus} \log A(i) \\ &\quad + \sum_j N_{\oplus j} \log B(j) + \sum_k \theta_k [c_{ij}{}^{(k)} N_{ij}] - \sum_{ij} \log[N_{ij}!], \end{aligned} \tag{5.8}$$

where the replacement of a subscript by a $\oplus$ indicates that we have summed with respect to that subscript (e.g., $T_{i\oplus} = \sum_j T_{ij}$, $T_{\oplus j} = \sum_i T_{ij}$; note that in this notation we would represent N_i used in earlier chapters as $N_{i\oplus}$). The partial derivatives of $\mathcal{L}$ are

$$\begin{aligned} \partial\mathcal{L}/\partial A(i) &= \sum_j \{-B(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] + N_{ij}/A(i)\} \\ &= [N_{i\oplus} - T_{i\oplus}][A(i)]^{-1} \quad \text{for } i \in I, \end{aligned} \tag{5.9}$$

$$\begin{aligned} \partial\mathcal{L}/\partial B(j) &= \sum_i \{-A(i)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] + N_{ij}/B(j)\} \\ &= [N_{\oplus j} - T_{\oplus j}][B(j)]^{-1} \quad \text{for } j \in J \end{aligned} \tag{5.10}$$

and

$$\begin{aligned} \partial\mathcal{L}/\partial\theta_k &= \sum_{ij} \{-c_{ij}^{(k)} A(i)B(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] + N_{ij}c_{ij}^{(k)}\} \\ &= \sum_{ij} c_{ij}^{(k)}[N_{ij} - T_{ij}] \quad \text{for } k \in K. \end{aligned} \tag{5.11}$$

ML estimates may be found by maximizing $\mathcal{L}$ directly by, say, the gradient search procedure. Alternatively, one could set the partial derivatives of $\mathcal{L}$ — (5.9), (5.10) and (5.11) — equal to zero and solve the resultant equations:

$$T_{i\oplus} = N_{i\oplus} \text{ for } i \in I,\ T_{\oplus j} = N_{\oplus j} \text{ for } j \in J \tag{5.12}$$

$$\text{and} \quad \sum_{ij} c_{ij}^{(k)} T_{ij} = \sum_{ij} c_{ij}^{(k)} N_{ij} \quad \text{for } k \in K. \tag{5.13}$$

In this chapter we assume that, if necessary, rows and columns of the N_{ij} matrix have been deleted so that all $N_{i\oplus} > 0$ and all $N_{\oplus j} > 0$, since if (say) $N_{1\oplus} = 0$, then the ML estimate of $T_{1\oplus}$ is zero, and corresponding N_{ij}'s will add nothing to the estimation of parameters—except to say that the ML estimate of $A(1)$ is zero.

Notice that the N_{ij}'s enter the computation of ML estimates only through $N_{i\oplus}$, $N_{\oplus j}$ and $c_{ij}{}^{(k)}N_{ij}$. These are called sufficient statistics (Rao, 1973, p. 130–132, Kendall and Stuart, 1967, p. 22 *et seq.*). One definition of sufficiency is that it be possible to write the likelihood function as a product of two functions g_1 and g_2, where g_1 depends only on the parameters and the sufficient statistics, and g_2 is independent of the parameters. It is easy to see from (5.6) that

$$\begin{aligned} P(\boldsymbol{N}) = &\Bigg[\Big(\prod_{ij} \exp(-A(i)B(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}])\Big)\Big(\prod_i [A(i)]^{N_{i\oplus}}\Big) \\ &\cdot \Big(\prod_j [B(j)]^{N_{\oplus j}}\Big)\Big(\exp\Big[\sum_{k=1}^K \theta_k\Big[\sum_{i=1}^I \sum_{j=1}^J c_{ij}^{(k)} N_{ij}\Big]\Big]\Big)\Bigg] \left([N_{ij}!]^{-1}\right) \end{aligned} \tag{5.14}$$

and hence, by setting $g_2 = [N_{ij}!]^{-1}$, we see that $N_{i\oplus}$, $N_{\oplus j}$ and $c_{ij}{}^{(k)}N_{ij}$, for $i \in I$, $j \in J$ and $k \in K$ form a set of $I + J + K$ sufficient statistics. A practical consequence of the existence of these sufficient statistics stems from the fact that they are usually quite large, even when individual N_{ij}'s are not. Since estimates based on larger samples are usually 'better', the fact that ML estimates are based on these statistics, endows them with very good small sample properties.

Of course, maximum likelihood estimation is not restricted to the exponential gravity model. Any continuously differentiable deterrence function $F[c_{ij}, \boldsymbol{\theta}]$, where $\boldsymbol{\theta}$ is a vector of parameters will do, although now the partial derivatives with respect to θ_k,

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_{ij} \left[-A(i)B(j)\frac{\partial F[c_{ij}, \boldsymbol{\theta}]}{\partial \theta_k} + \frac{N_{ij}}{F[c_{ij}, \boldsymbol{\theta}]}\frac{\partial F[c_{ij}, \boldsymbol{\theta}]}{\partial \theta_k} \right],$$

would not necessarily lead to the simple equations akin to (5.12) and (5.13), nor need all the pleasant properties mentioned later in this chapter hold true.

5.1.3 A Preview of this Chapter

The existence and uniqueness of maximum likelihood estimates of the parameters in (5.2) are examined in Section 5.2. It is shown that, under some mild conditions, the ML estimate of $\boldsymbol{A} = (A(1), \ldots A(I))^t$, $\boldsymbol{B} = (B(1), \ldots B(J))^t$ and $\boldsymbol{\theta}$ exist and that of $\boldsymbol{\theta}$ is unique. The estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ are not unique; however, if one $A(i)$ or $B(j)$ is chosen to be an arbitrary positive number, the remaining $A(i)$'s and $B(j)$'s are unique — under the mild conditions mentioned above.

That there is a single value of $\boldsymbol{\theta}$ corresponding to the maximum of the likelihood function is very useful, since then we can obtain ML estimates by solving (5.12) and (5.13) or by setting up a procedure which raises the value of $\mathcal{L}$ at every iteration. If such a procedure were not to stop or substantially slow down before (5.12) and (5.13) are achieved, it would yield the unique ML estimate of $\boldsymbol{\theta}$. The readers' attention is also drawn to conditions ML1 and ML2, because it is when these conditions are close to being violated that computational procedures slow down.

In Section 5.3, the computation of ML estimates for three frequently used special cases are discussed. One of the special cases is where each $F_{ij} = F(c_{ij})$ in the model (5.3) is known and we need to estimate T_{ij}'s or equivalently $A(i)$'s and $B(j)$'s. The method of computation of estimates is a procedure that we call the DSF procedure but has also been called the row-column balancing algorithm. This procedure is particularly important because apart from the fact that it has important applications in forecasting, the DSF procedure is imbedded within several other procedures.

Another special case considered in Section 5.3 is that where $F(c_{ij})$ is a step function — a case for which Evans and Kirby (1974) have presented a very efficient procedure. This algorithm is possibly the most commonly used method of estimating $F(c_{ij})$. While our presentation is more general, in actual use, a single measure c_{ij} (e.g., travel time) is typically used and its range divided into intervals (e.g., 0–1 minute, 1–2 minute, etc.). For each interval a value of $F(c_{ij})$ is estimated. However, the procedure can not only accommodate multiple separation measures, some or all of them can be ordinal, categorical or nominal. The major drawback of this procedure is its enormous data appetite [which is possibly the underlying cause for sample sizes to be so very large in surveys used for transportation studies].

The third case studied in Section 5.3 is that where we have only one component in $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}$ is a scalar θ_1. The three computational methods mentioned above are each illustrated by means of numerical examples. The purpose is to provide some feel for each method as well as to present some practical details.

Sections 5.4 through Section 5.6 are devoted to methods for computing ML estimates for the general model given by (5.2) and (5.1). In Section 5.4, a procedure called the LDSF procedure is presented. This procedure, which is a linearization of the DSF mentioned above, has a wide array of uses. Of particular importance is its use in Section 5.5 to develop several computational methods for estimating parameters for the general case of the model given by (5.1) and (5.2).

Some numerical applications and comparisons of these last-mentioned parameter estimation methods are then presented in Section 5.6. One of the procedures developed in Section 5.5, which we have called the modified scoring procedure, emerges as the clear winner and an excellent performer. In fact, the performance of this procedure is so good that ML estimation of gravity model parameters need no longer be considered computationally arduous.

In Section 5.7, methods are given for computing covariances of estimates of gravity model parameters. The computation of covariances of forecasted T_{ij}'s is also addressed. We draw the readers' attention to Section 5.7.3, where different methods of obtaining forecasts are summarized along with methods for obtaining corresponding covariance matrices. These covariances are the most appropriate method of assessing parameter and forecast sensitivities, and are indispensable in computing standard errors of estimates and running certain tests of significance. They can also be useful in determining sample sizes of surveys which gather data for gravity model parameter estimation. The methods presented are not difficult computationally.

Section 5.8 discusses the measurement of goodness of fit of the gravity model. It also contains a discussion of residuals and their use in the diagnosis of deficiencies in model formulation.

Maximum likelihood estimation is so much in favor among analysts because ML estimates frequently possess highly desirable asymptotic properties. These properties are consistency, efficiency and asymptotic normality. It is shown in Section 5.9 that the properties hold for the model we are considering. In addition, ML estimation procedures for the gravity model appear to be robust. Small sample properties of the maximum likelihood estimates of gravity model parameters are also investigated in Section 5.9. It appears that even when the size of the T_{ij}'s is very small and the components of $\boldsymbol{N}$ are mostly zeros, maximum likelihood procedures can continue to yield good results.

Thus, ML estimates of gravity model parameters are highly desirable from just about every point of view. Because of the modified scoring procedure, they are also relatively easy to compute. This chapter provides a reasonably complete coverage of the use of ML estimation in analyses involving the gravity model.

5.2 Existence and Uniqueness of ML Estimates

Using the convention that small case letters stand for the logarithms of corresponding capital letters (e.g., $t_{ij} = \log[T_{ij}]$, $a(i) = \log[A(i)]$), Model (5.2) may be written as

$$t_{ij} = a(i) + b(j) + \sum_{k=1}^{K} \theta_k c_{ij}^{(k)} : i \in I, j \in J. \tag{5.15}$$

Let M denote the coefficient matrix of the right side of the system of equations (5.15). Such a matrix for $I = J = 3$ and $K = 2$ is illustrated below:

$$M = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & c_{11}^{(1)} & c_{11}^{(2)} \\ 1 & 0 & 0 & 0 & 1 & 0 & c_{12}^{(1)} & c_{12}^{(2)} \\ 1 & 0 & 0 & 0 & 0 & 1 & c_{13}^{(1)} & c_{13}^{(2)} \\ 0 & 1 & 0 & 1 & 0 & 0 & c_{21}^{(1)} & c_{21}^{(2)} \\ 0 & 1 & 0 & 0 & 1 & 0 & c_{22}^{(1)} & c_{22}^{(2)} \\ 0 & 1 & 0 & 0 & 0 & 1 & c_{23}^{(1)} & c_{23}^{(2)} \\ 0 & 0 & 1 & 1 & 0 & 0 & c_{31}^{(1)} & c_{31}^{(2)} \\ 0 & 0 & 1 & 0 & 1 & 0 & c_{32}^{(1)} & c_{32}^{(2)} \\ 0 & 0 & 1 & 0 & 0 & 1 & c_{33}^{(1)} & c_{33}^{(2)} \end{pmatrix}. \tag{5.16}$$

The transpose M^t of M is the coefficient matrix of the T_{ij}'s in the system of $I + J + K$ equations represented by (5.12) and (5.13), i.e., (5.12) and (5.13) may be written as

$$M^t \boldsymbol{T} = M^t \boldsymbol{N}. \tag{5.17}$$

Now we can state the following theorem on the existence and uniqueness of ML estimates of gravity model parameters. [Notice that Conditions ML1 and ML2, which we shall refer to repeatedly in this chapter, are defined in the statement of Theorem 5.1.]

Theorem 5.1 *The two conditions,*

ML1. *The rank of M is $I+J+K-1$, and*

ML2. *There exist IJ positive numbers $y_{ij}{}^{(0)}$ such that*

$$y_{i\oplus}{}^{(0)} = N_{i\oplus}, \quad y_{\oplus j}{}^{(0)} = N_{\oplus j},$$
$$\sum_{ij} c_{ij}{}^{(k)} y_{ij}{}^{(0)} = \sum_{ij} c_{ij}{}^{(k)} N_{ij} \tag{5.18}$$

for all $i \in I$, $j \in J$ and $k \in K$ (i.e., there is a solution to (5.12) and (5.13) although not necessarily of the gravity model form),

are necessary and sufficient for the existence of a unique vector

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^t,$$

which is the ML estimate of $\boldsymbol{\theta}$ and for which numbers $\hat{A}(1), \ldots, \hat{A}(I)$ and $\hat{B}(1), \ldots, \hat{B}(J)$ can be found so as to satisfy (5.2), (5.12) and (5.13). Moreover, the corresponding estimates $\hat{T}_{ij}$ of T_{ij} are also unique.

Notice that we do not assert that the ML estimates of $A(i)$'s and $B(j)$'s are unique. In fact, if $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ is an ML estimate of $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta})$ so is $(\gamma\hat{\boldsymbol{A}}, \gamma^{-1}\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$, for any $\gamma > 0$. However, it may be seen from the discussion below that $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ and $(\gamma\hat{\boldsymbol{A}}, \gamma^{-1}\hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}})$ yield the same set of $\hat{T}_{ij}$'s. It is noteworthy that $\hat{\boldsymbol{\theta}}$ is unique even though $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ are not.

Both Conditions ML1 and ML2 perhaps deserve some attention. We examine these before proving the theorem in Section 5.2.3.

5.2.1 Condition ML1

Let the columns of the matrix M be called $\boldsymbol{m}_s$ where $s = 1, \cdots, I+J+K$, i.e.,

$$M = (\boldsymbol{m}_1, \cdots, \boldsymbol{m}_I, \boldsymbol{m}_{I+1}, \cdots, \boldsymbol{m}_{I+J}, \boldsymbol{m}_{I+J+1}, \cdots, \boldsymbol{m}_{I+J+K}), \tag{5.19}$$

and let $M_{(1)}$ be the matrix consisting of the first $I+J$ columns of M:

$$M_{(1)} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{I+J})^t. \tag{5.20}$$

Lemma 5.1 *The rank of $M_{(1)}$ is $I+J-1$ and the deletion of any column of $M_{(1)}$ results in a matrix of rank $I+J-1$.*

PROOF: Since $\sum_{i=1}^{I} \boldsymbol{m}_i = \sum_{j=1}^{J} \boldsymbol{m}_{I+j}$, the rank of $M_{(1)}$ is at most $I+J-1$. We now show that it is at least $I + J - 1$.

Delete one of the columns of $M_{(1)}$ — say $\boldsymbol{m}_1$ — and consider a linear combination

$$\boldsymbol{d} = \sum_{s=2}^{I+J} d_s \boldsymbol{m}_s, \tag{5.21}$$

of the remaining columns with $(d_2, \ldots, d_{I+J})^t \neq \mathbf{o}$, where, as in the rest of the book, $\mathbf{o}$ is a zero vector $\mathbf{o} = (0, \ldots, 0)^t$.

Consider the two possibilities: either

$$d_{I+1} = \ldots = d_{I+J} = 0, \tag{5.22}$$

or for some p such that $I < p \leq I + J$, $d_p \neq 0$. In the latter case, since $\boldsymbol{m}_p$ contains a 1 in some position at which all other $\boldsymbol{m}_s$'s such that $2 \leq s \leq I+J$ have zeros, $\boldsymbol{d} \neq \mathbf{o}$ [this 1 would be in a position where $\boldsymbol{m}_1$ has a 1; the reader might find it useful to examine the first 6 columns of (5.16)]. On the other hand, if (5.22) holds, then $\boldsymbol{d}$ cannot be zero since $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_I$ are obviously independent. Thus, in either case, $\boldsymbol{d} \neq \mathbf{o}$. Hence the matrix $(\boldsymbol{m}_2, \ldots, \boldsymbol{m}_{I+J})$ is of full rank. [We define a matrix of dimension $r \times s$ of full rank if it is of rank r when $r \leq s$ or of rank s when $r \geq s$.]

If any other column had been deleted, the result would have been similar. Therefore, the rank of $M_{(1)}$ is at least $I + J - 1$. The lemma follows. □

Thus the rank of M depends on its last K columns. If the rank is less than $I + J + K - 1$, these columns would be mutually linearly dependent or linearly dependent on the first $I + J$ columns. In that case it would obviously be too much to expect a unique estimate of $\boldsymbol{\theta}$.

An example is the following: let $c_{ij}{}^{(2)}$ be parking cost $c_{ij}{}^{(2)} = c_j{}^{(2)}$ which is fully determined by the destination j. Then $\boldsymbol{m}_{I+J+2}$, the vector consisting of $c_{ij}{}^{(2)}$'s, may be written as

$$\boldsymbol{m}_{I+J+2} = \sum_{j=1}^{J} c_j{}^{(2)} \boldsymbol{m}_{I+j},$$

showing linear dependence of the $\boldsymbol{m}_s$'s. That is, for purposes of estimation, the factor $\exp[\theta^{(2)} c_j{}^{(2)}]$ cannot be distinguished from the B_js, making unique estimation of parameters impossible.

5.2.2 Condition ML2

Denote by $((a_{ij}))$ a matrix, the ij-th element of which is a_{ij}. Using this notation, set

$$((c_{ij}{}^{(1)})) = \begin{pmatrix} 1 & 50 \\ 50 & 1 \end{pmatrix} \quad \text{and} \quad ((N_{ij})) = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}. \tag{5.23}$$

These matrices illustrate a situation where Condition ML2 does not hold. To see this, first notice that, whatever the values of ${y_{ij}}^{(0)}$ might be, because of the first line in (5.18),

$$\sum_{ij} c_{ij}^{(1)} N_{ij} = 15 \quad \text{and} \quad \sum_{ij} {y_{ij}}^{(0)} = N_{ij} = 15. \tag{5.24}$$

Moreover, because

$$\begin{pmatrix} 1 & 50 \\ 50 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 49 \\ 49 & 0 \end{pmatrix},$$

it follows that

$$\sum_{ij} c_{ij}^{(1)} {y_{ij}}^{(0)} = \sum_{ij} {y_{ij}}^{(0)} + 49({y_{12}}^{(0)} + {y_{21}}^{(0)}).$$

Hence,

$$\sum_{ij} c_{ij}^{(1)} {y_{ij}}^{(0)} - \sum_{ij} c_{ij}^{(1)} N_{ij} = 49({y_{12}}^{(0)} + {y_{21}}^{(0)}) > 0, \tag{5.25}$$

which violates the last equation in (5.18). Thus, in this case, suitable ${y_{ij}}^{(0)}$'s do not exist and Condition ML2 does not hold.

But such situations are rare. In order to investigate them further, first set $K = 1$. Let

$$\boldsymbol{y} = (y_{11}, \ldots, y_{1J}, \ldots\ldots, y_{I1}, \ldots, y_{IJ})^t \tag{5.26}$$

and consider the region $\mathcal{R}^{(0)}$ of $\boldsymbol{y}$'s such that

$$y_{i\oplus} = N_{i\oplus}, \quad y_{\oplus j} = N_{\oplus j}, \quad y_{ij} > 0, \tag{5.27}$$

for all $i \in I$ and $j \in J$.

We now show that $\mathcal{R}^{(0)}$ is non-empty. If all N_{ij}'s are positive, set $y_{ij} = N_{ij}$. If some $N_{ij} = 0$, there must be a positive N_{iv} and a positive N_{uj} (since $N_{i\oplus} > 0$ and $N_{\oplus j} > 0$). Then for a small enough δ, one can set

$$y_{ij} = \delta, \quad y_{iv} = N_{iv} - \delta, \quad y_{uj} = N_{uj} - \delta \text{ and } y_{uv} = N_{uv} + \delta \tag{5.28}$$

and the first two conditions in (5.27) would not be violated. Repeating a similar procedure for each zero-valued N_{ij}, a point in $\mathcal{R}^{(0)}$ would be reached.

Now assume that ML2 does not hold and further assume that there are two points $\boldsymbol{y}^{(1)} \in \mathcal{R}^{(0)}$ and $\boldsymbol{y}^{(2)} \in \mathcal{R}^{(0)}$ such that

$$\sum_{ij} {c_{ij}}^{(1)} {y_{ij}}^{(1)} < \sum_{ij} {c_{ij}}^{(1)} N_{ij} < \sum_{ij} {c_{ij}}^{(1)} {y_{ij}}^{(2)}. \tag{5.29}$$

Then for all α, such that $0 < \alpha < 1$, $\boldsymbol{y}^{(\alpha)} = \alpha\boldsymbol{y}^{(1)} + (1-\alpha)\boldsymbol{y}^{(2)}$ is in $\mathcal{R}^{(0)}$, and for some α

$$\sum_{ij} c_{ij}{}^{(1)} y_{ij}{}^{(\alpha)} = \sum_{ij} c_{ij}{}^{(1)} N_{ij}. \tag{5.30}$$

This violates the non-existence of $y_{ij}{}^{(0)}$'s meeting condition (5.18). Hence, the failure of Condition ML2 to hold implies that either,

$$\sum_{ij} c_{ij}{}^{(1)} y_{ij} > \sum_{ij} c_{ij}{}^{(1)} N_{ij} \qquad \text{for all } \boldsymbol{y} \in \mathcal{R}^{(0)} \tag{5.31}$$

or

$$\sum_{ij} c_{ij}{}^{(1)} y_{ij} < \sum_{ij} c_{ij}{}^{(1)} N_{ij} \quad \text{for all } \boldsymbol{y} \in \mathcal{R}^{(0)}, \tag{5.32}$$

i.e., the N_{ij}'s are in an extremal configuration with respect to costs $c_{ij}{}^{(1)}$. [Notice that, in the case of (5.31), the N_{ij}'s constitute a solution to the (linear programming) transportation problem: minimize $\sum_{ij} c_{ij}{}^{(1)} y_{ij}$ subject to $y_{ij} \geq 0$, $y_{i\oplus} = O_i$ and $y_{\oplus j} = D_j$, where O_i's and D_j's are constants and $O_\oplus = D_\oplus$.]

This result is not entirely startling. Evans (1971) has shown that for fixed $T_{i\oplus}$ and $T_{\oplus j}$, θ_1 is a monotonic function of $\sum_{ij} c_{ij}{}^{(1)} T_{ij}$ and maximum and minimum total costs (i.e., $\sum_{ij} c_{ij}{}^{(1)} T_{ij}$) occur when $\theta_1 \to \infty$ or $\theta_1 \to -\infty$. [Evans's proof may be extended to cover $K > 1$ quite directly. Such an extension follows. Let $D_k g$ stand for the partial derivative of g with respect to θ_k. Then from (5.15), since $t_{ij} = \log[T_{ij}]$, $D_k T_{ij} = [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}]T_{ij}$. But since $T_{i\oplus}$ is a constant,

$$D_k T_{i\oplus} = 0 = \sum_j [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}] T_{ij}. \tag{5.33}$$

Hence,

$$0 = \sum_{ij} D_k a(i) [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}] T_{ij}. \tag{5.34}$$

Similarly,

$$0 = \sum_{ij} D_k b(j) [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}] T_{ij}. \tag{5.35}$$

If $C_k = \sum_{ij} c_{ij}{}^{(k)} T_{ij}$, then, from (5.34) and (5.35), we have

$$\begin{aligned} D_k C_k &= \sum_{ij} c_{ij}{}^{(k)} D_k T_{ij} \\ &= \sum_{ij} T_{ij} c_{ij}{}^{(k)} [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}] \\ &= \sum_{ij} T_{ij} [D_k a(i) + D_k b(j) + c_{ij}{}^{(k)}]^2 \geq 0. \end{aligned} \tag{5.36}$$

The last expression before the inequality in (5.36) follows from the expression preceding it by the addition of the last terms in (5.34) and (5.35). In (5.36) the equality occurs when $D_k a(i) + D_k b(j) + c_{ij}{}^{(k)} = 0$, i.e., when $c_{ij}{}^{(k)}$ is of the form $c_{ij}{}^{(k)} = c_{i,1}{}^{(k)} + c_{i,2}{}^{(k)}$ violating condition ML1.]

The above result, that N_{ij}'s are in an extremal configuration, can be extended to $K \geq 1$ in the following way:

Lemma 5.2 *Let $\mathcal{R}^{(K_1)}$ denote a non-empty region given by*

$$y_{i\oplus} = N_{i\oplus}, \quad y_{\oplus j} = N_{\oplus j}, \quad y_{ij} > 0 \quad \textit{for all } i \in I \textit{ and } j \in J \tag{5.37}$$

and

$$\sum_{ij} c_{ij}{}^{(k)} y_{ij} = \sum_{ij} c_{ij}{}^{(k)} N_{ij} \textit{ for all } k \leq K_1. \tag{5.38}$$

If

$$\sum_{ij} c_{ij}{}^{(n)} y_{ij} \neq \sum_{ij} c_{ij}{}^{(n)} N_{ij} \tag{5.39}$$

for all $\boldsymbol{y} \in \mathcal{R}^{(K_1)}$ and all n such that $K_1 \leq n \leq K$, then for each such n we must have either

$$\sum_{ij} c_{ij}{}^{(n)} y_{ij} < \sum_{ij} c_{ij}{}^{(n)} N_{ij} \quad \textit{for all } \boldsymbol{y} \in \mathcal{R}^{(K_1)} \tag{5.40}$$

or

$$\sum_{ij} c_{ij}{}^{(n)} y_{ij} > \sum_{ij} c_{ij}{}^{(n)} N_{ij} \quad \textit{for all } \boldsymbol{y} \in \mathcal{R}^{(K_1)}. \tag{5.41}$$

The proof is similar to the discussion given above.

5.2.3 Proof of Theorem 5.1

Lemma 5.3 *Condition ML2 is necessary and sufficient for the existence of a unique set of estimates $\hat{T}_{ij}$ of the form*

$$\hat{T}_{ij} = \hat{A}(i)\hat{B}(j) \exp[\hat{\boldsymbol{\theta}}^t c_{ij}] \tag{5.42}$$

which solve (5.12) and (5.13).

The lemma asserts that $\hat{T}_{ij}$'s are unique; in the next lemma we show that for a unique set of $\hat{T}_{ij}$'s, the θ_k's are unique. The result given in Lemma 5.3 has been proved by Haberman (1974, pp. 35–37) and others under conditions essentially equivalent to ML2. However, a proof is presented below for completeness.

Proof of Lemma 5.3: Denote as Problem I that of obtaining a local minimum of

$$\mathcal{S} = \sum_{ij} [y_{ij} \log(y_{ij}) - y_{ij}], \tag{5.43}$$

subject to the constraints

$$y_{i\oplus} = O_i, \quad y_{\oplus j} = D_j, \quad y_{ij} \geq 0 \tag{5.44}$$

and

$$\sum_{ij} c_{ij}{}^{(k)} y_{ij} = \sum_{ij} c_{ij}{}^{(k)} N_{ij}. \tag{5.45}$$

Using Lagrange multipliers, a necessary condition for an interior point solution to this problem is seen to be

$$\log(y_{ij}) = \lambda_i^A + \lambda_j^B + c_{ij}{}^{(k)}\lambda_k. \tag{5.46}$$

Thus, setting $\hat{A}(i) = \exp(\lambda_i^A)$, $\hat{B}(j) = \exp(\lambda_j^B)$, $\lambda_k = \theta_k$, we see that a value of y_{ij} which is an interior point solution to Problem I, is of the form (5.42) and must obey (5.44) and (5.45). Moreover, any y_{ij} meeting these requirements is an interior point solution to Problem I. Because $\partial^2 \mathcal{S}/\partial y_{ij}^2 = y_{ij}^{-1} > 0$ and $\partial^2 \mathcal{S}/\partial y_{ij}\partial y_{i'j'} = 0$ if $i \neq i'$ or $j \neq j'$, (5.43) can be readily seen to be convex in the interior of the feasible region which is also convex. Therefore, an interior point solution to Problem I is unique. Thus, there is at most one set of numbers of the form (5.42) which obey (5.27) and (5.45).

Now we show that such a set of numbers exists. First note that since, from (5.44), $y_{ij} \leq O_i$ for all i and j, the feasible region is bounded. It is also closed and is hence compact. Moreover, on noting that

$$\lim_{h \to 0} h \log(h) - h = 0, \tag{5.47}$$

we see that (5.43) can be extended to a continuous function over the feasible region. Thus it has a minimum in the region.

Notice that since the positivity constraints on y_{ij}'s are the only inequality constraints, a boundary point of the feasible region must have $y_{ij} = 0$ for some i and j. For that y_{ij},

$$\frac{\partial \mathcal{S}}{\partial y_{ij}} = \log[y_{ij}] \to -\infty$$

as $y_{ij} \to 0$. Therefore, if we moved from a boundary point in the direction of an interior point, the function (5.43) would decline since its derivative is negative near the boundary. Such a direction always exists since the existence of an interior point is assured by Condition ML2. Therefore the minimum value of (5.43) cannot occur at the boundary of the feasible region.

But a minimum of (5.43) does occur in the region and must then occur in its interior. Therefore, an interior point solution to Problem I exists. We have already seen that such a solution is unique and has the required properties. □

Note that while the above lemma has been stated in terms of ML conditions, the proof is actually more general in that it would hold if the N_{ij}'s were replaced by any set of non-negative numbers with $N_{i\oplus} \neq 0$ and $N_{\oplus j} \neq 0$. The following corollary will be needed later.

Corollary 5.1 *For all $i \in I$ and all $j \in J$, let F_{ij} be any set of non-negative numbers and O_i and D_j be any set of positive numbers. Let $\delta_{ij} = 0$ when $F_{ij} = 0$ and $\delta_{ij} = 1$ when $F_{ij} > 0$. Then the condition,*

ML3. *There exists a set of positive numbers ${y_{ij}}^{(0)}$ such that*

$$\sum_j \delta_{ij} {y_{ij}}^{(0)} = O_i, \quad \sum_i \delta_{ij} {y_{ij}}^{(0)} = D_j, \tag{5.48}$$

is necessary and sufficient for the existence and uniqueness of IJ numbers $\hat{T}_{ij}$ of the form

$$\hat{T}_{ij} = \hat{A}(i)\hat{B}(j)F_{ij} \tag{5.49}$$

and obeying

$$\hat{T}_{i\oplus} = O_i \quad \text{and} \quad \hat{T}_{\oplus j} = D_j. \tag{5.50}$$

The proof is similar to that for Lemma 5.3 except that instead of (5.43), the objective function is

$$\sum_{ij} \delta_{ij} \left[y_{ij} \log(y_{ij}/F_{ij}) - y_{ij}\right] \tag{5.51}$$

and the only constraints are (5.44). Notice that (5.48) assures that $O_\oplus = D_\oplus$.

Lemma 5.4 *For all $i \in I$ and $j \in J$, let $\hat{T}_{ij}$'s be positive numbers of the form given by (5.42). Then Condition ML1 is necessary and sufficient for the uniqueness of $\boldsymbol{\theta}$.*

Proof: Since, by Lemma 5.1, $M_{(1)}$ has rank $I+J-1$, any column — say $\boldsymbol{m}_1$ — of $M_{(1)}$ can be expressed in terms of the remainder of the columns. Therefore, by (5.15) and the definition of the $\boldsymbol{m}_s$'s (Section 5.2.1),

$$\hat{\boldsymbol{t}} = (\hat{t}_{11}, \ldots, \hat{t}_{1J}, \ldots\ldots, \hat{t}_{I1}, \ldots, \hat{t}_{IJ})^t, \tag{5.52}$$

where $\hat{t}_{ij} = \log(\hat{T}_{ij})$, may be written as

$$\hat{\boldsymbol{t}} = \sum_{i=2}^{I} a(i)\boldsymbol{m}_i + \sum_{j=1}^{J} b(j)\boldsymbol{m}_{I+j} + \sum_{k=1}^{K} \theta_k \boldsymbol{m}_{I+J+k}. \tag{5.53}$$

Suppose there is another $\boldsymbol{\theta}$, call it $\tilde{\boldsymbol{\theta}}$, such that

$$\hat{\boldsymbol{t}} = \sum_{i=2}^{I} \tilde{a}(i)\boldsymbol{m}_i + \sum_{j=1}^{J} \tilde{b}(j)\boldsymbol{m}_{I+j} + \sum_{k=1}^{K} \tilde{\theta}_k \boldsymbol{m}_{I+J+k}. \tag{5.54}$$

Then, by subtraction,

$$0=\sum_{i=2}^{I}(a(i)-\tilde{a}(i))\boldsymbol{m}_i+\sum_{j=1}^{J}(b(j)-\tilde{b}(j))\boldsymbol{m}_{I+j}+\sum_{k=1}^{K}(\theta_k-\tilde{\theta}_k)\boldsymbol{m}_{I+J+k}. \quad (5.55)$$

Since $\boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}$, (5.55) shows that $\boldsymbol{m}_s$ for $2 \leq s \leq I+J+K$ are linearly dependent, implying that M has rank less than $I+J+K-1$, which violates condition ML1. This establishes the sufficiency of ML1.

To prove necessity, notice that from Lemma 5.1, if M has rank less than $I+J+K-1$, the vectors $\boldsymbol{m}_{I+J+k} : k = 1, \ldots, K$ must be mutually linearly dependent or one or more of them must be linearly dependent on the columns of $M_{(1)}$. Either way, we can get a $\breve{\boldsymbol{\theta}} \neq \mathbf{o}$, such that

$$0=\sum_{i=2}^{I}\breve{a}(i)\boldsymbol{m}_i+\sum_{j=1}^{J}\breve{b}(j)\boldsymbol{m}_{I+j}+\sum_{k=1}^{K}\breve{\theta}_k\boldsymbol{m}_{I+J+k}. \quad (5.56)$$

Adding (5.53) to (5.56), we see that $\boldsymbol{\theta}$ then cannot be unique. □

Theorem 5.1 follows from Lemmas 5.3 and 5.4.

Notice that the proof of Lemma 5.4 also shows that if $a(1)$ is set equal to zero, the remaining $a(i)$'s or $b(j)$'s are unique; i.e., if $A(1)$ is set equal to 1 the remaining $A(i)$'s or $B(j)$'s are unique. In fact, if any $A(i)$ or $B(j)$ is set equal to any positive number, the remaining $A(i)$'s or $B(j)$'s are unique.

We have shown that there is a unique solution to (5.2), (5.12) and (5.13). We still need to assure ourselves that this solution maximizes the likelihood function (5.6). Notice that (5.6) being a product of probabilities is nonnegative and bounded. Consider its values over the set $\{\boldsymbol{T} : 0 \leq T_{ij} \text{ for all } i,j\}$. It is quite clear from (5.6) that if any $T_{ij} \to \infty$, $P(\boldsymbol{N}) \to 0$. If any $T_{ij} = 0$, either $A(i) = 0$ or $B(j) = 0$ and it follows from (5.14) that $P(\boldsymbol{N}) = 0$. Hence the maximum of (5.6) as a function of T_{ij}'s must occur at an interior point — where (5.12) and (5.13) must hold. But there is only one such point under the conditions of Theorem 5.1. It follows that the maximum of the likelihood function (5.6) must occur at that point. Moreover, as we have seen, this point corresponds to a unique value of $\boldsymbol{\theta}$.

5.2.4 ML Estimation for Multinomial Gravity Models

Had we assumed a multinomial distribution for $\boldsymbol{N}$, the estimate of $\boldsymbol{\theta}$ would have been the same. This is shown below. While the treatment in the book is mainly confined to independent Poisson N_{ij}'s, the fact that one gets the same ML estimates under a multinomial assumption is important for two reasons. First, there could be situations where the multinomial distribution is appropriate. The multinomial framework makes weaker assumptions [see

Section 3.9.1]. In fact, the multinomial assumption for N_{ij}'s was commonplace until about a decade ago. Second, some contingency table results in the statistics literature have only been established for the multinomial case. The treatment below allows their use in this book (see Section 5.9) and elsewhere.

We may rewrite (5.6) as follows:

$$\begin{aligned} P(\boldsymbol{N}) &= \prod_{ij} \exp[-T_{ij}]\, T_{ij}^{N_{ij}}/N_{ij}! \\ &= \left(N_{\oplus\oplus}! \prod_{ij} \frac{P_{ij}^{N_{ij}}}{N_{ij}!} \right) \left(\frac{T_{\oplus\oplus}{}^{N_{\oplus\oplus}} \exp[-T_{\oplus\oplus}]}{N_{\oplus\oplus}!} \right), \end{aligned} \tag{5.57}$$

where $P_{ij} = T_{ij}/T_{\oplus\oplus}$. The quantity within the first pair of brackets on the last line is the multinomial probability function which is also the likelihood function assuming a multinomial distribution for the N_{ij}'s. The second pair of brackets enclose a Poisson probability function for $N_{\oplus\oplus}$. The result (5.57) is well known and has been used elsewhere in the book (e.g., in (3.135)). It also shows that using a given set of N_{ij}'s, the P_{ij}'s and the $T_{\oplus\oplus}$ can be estimated independently of each other.

Let us start with the multinomial part. Absorbing the $T_{\oplus\oplus}$'s into, say, the $A(i)$'s, we can write, using (5.2), $P_{ij} = \alpha(i)\beta(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}]$, where $\alpha(i) = A(i)/T_{\oplus\oplus}$ and $\beta(j) = B(j)$. Making this substitution and ignoring the factorials, the multinomial part of (5.57) can be written as (by regrouping terms):

$$\prod_{ij} P_{ij}^{N_{ij}} = \prod_{i=1}^{I} \alpha(i)^{N_{i\oplus}} \prod_{j=1}^{J} \beta(j)^{N_{\oplus j}} \exp[\sum_{k=1}^{K} (\theta_k \sum_{ij} c_{ij}{}^{(k)} N_{ij})]. \tag{5.58}$$

Taking logarithms and maximizing subject to the natural constraint

$$\sum_{ij} P_{ij} = \sum_{ij} \alpha(i)\beta(j) \exp[\sum_{k=1}^{K} \theta_k c_{ij}{}^{(k)}] = 1 \tag{5.59}$$

we get

$$N_{i\oplus} = \lambda P_{i\oplus}, N_{\oplus j} = \lambda P_{\oplus j} \text{ and } \sum_{ij} c_{ij}{}^{(k)} N_{ij} = \lambda \sum_{ij} c_{ij}{}^{(k)} P_{ij} \tag{5.60}$$

for all $i \in I$, $j \in J$, $k \in K$. From (5.59) and either of the first two equations in (5.60), we get $\lambda = [N_{\oplus\oplus}]^{-1}$.

The second part of the last line in (5.57) yields $N_{\oplus\oplus}$ as the estimate of $T_{\oplus\oplus}$. Thus, the final estimate of the T_{ij}'s are exactly the same as those obtained, for Poisson N_{ij}'s, from (5.6). This is just as we should expect.

Because the P_{ij}'s are parameterized in the same way as T_{ij}'s, the existence and uniqueness results proved above remain the same. In particular, the matrix M multiplies the vector of the logarithms of $\alpha(i)$, $\beta(j)$ and θ_k to yield the vector $\boldsymbol{p}$ of the logarithms $p_{ij} = \log[P_{ij}]$ of the P_{ij}'s. Define the matrix $M_{(2)}$ as one obtained by deleting one of the first $I+J$ columns of M. Then $M_{(2)}$ multiplies the vector $\boldsymbol{\xi}$ of the logs of $\alpha(i)$, $\beta(j)$ and θ_k (with one $\alpha(i)$ or $\beta(j)$ — say $\beta(J)$ — left out) to yield the same vector $\boldsymbol{p}$, i.e., $\boldsymbol{p} = M_{(2)}\boldsymbol{\xi}$. Moreover, if Condition ML1 holds, then $M_{(2)}$ is of full rank, and it may be shown as in the proof of Lemma 5.4 that $\boldsymbol{\xi}$ is unique.

Now we show that the estimate of $\boldsymbol{\theta}$ is the same whether we make a Poisson or a multinomial assumption. Let $\log[\boldsymbol{x}] = (\log x_1, \ldots, \log x_n)^t$ when $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ and consider the equation $\log[\boldsymbol{T}] = \log[T_{\oplus\oplus}\boldsymbol{P}] = M_{(2)}\boldsymbol{\xi}^{(0)}$, where $\boldsymbol{P}$ is the vector of P_{ij}s. One solution for $\boldsymbol{\xi}^{(0)}$ is obtained from $\boldsymbol{\xi}$ by multiplying each $\alpha(i)$ in the first I components $\log[\alpha(i)]$ of $\boldsymbol{\xi}$ by $T_{\oplus\oplus}$. Because of the uniqueness of $\boldsymbol{\xi}^{(0)}$ (see proof of Lemma 5.4), this is the only solution. Consequently, we get the same estimate of $\boldsymbol{\theta}$ whether we assume a multinomial distribution for $\boldsymbol{N}$ or a Poisson distribution.

5.3 ML Estimation Algorithms: Special Cases

In this section, three useful and frequently used subcases are considered: the cases when

1. $F(\boldsymbol{c}_{ij}) = F_{ij}$ is known,
2. $F(\boldsymbol{c}_{ij})$ is a step function and
3. $F(\boldsymbol{c}_{ij}) = \exp[\theta_1 c_{ij}{}^{(1)}]$ — i.e., $\boldsymbol{\theta}$ is a scalar.

Algorithms for the general case are discussed in Section 5.5.

Each of the methods given below is followed by one or more numerical examples. The purpose of the examples is not to illustrate how the formulæ are used, but rather, to provide a feel for the behavior of the procedures and also to make some practical suggestions. Some of the data used are given in the appendix, so that readers who wish to duplicate our work might do so.

5.3.1 The DSF Procedure

It is easily seen from (5.9), (5.10) and (5.11) that when $\boldsymbol{\theta}$ is known, ML estimates of $\boldsymbol{A}$, $\boldsymbol{B}$ and of the T_{ij}'s can be obtained by solving (5.12). A computational procedure for doing so is given below in Theorem 5.2.

Theorem 5.2 *Let $F_{ij} \geq 0$, $O_i > 0$, $D_j > 0$, $P_i > 0$, and $Q_j > 0$ be any set of numbers for $i \in I$ and $j \in J$, and let*

$$T_{ij}{}^{(0)} = P_i Q_j F_{ij} \quad \textit{for all } i \in I \textit{ and } j \in J. \tag{5.61}$$

Consider the iterations

$$T_{ij}{}^{(2r-1)} = T_{ij}{}^{(2r-2)} O_i / T_{i\oplus}{}^{(2r-2)} \tag{5.62}$$

and

$$T_{ij}{}^{(2r)} = T_{ij}{}^{(2r-1)} D_j / T_{\oplus j}{}^{(2r-1)}, \tag{5.63}$$

where $r = 1, 2, \cdots$. *Under Condition ML3, for each* i *and* j*, the sequence* $T_{ij}{}^{(s)}$ *converges to a unique limit* $\hat{T}_{ij}$ *of the form (5.49) and obeying (5.50).*

Although the procedure given in Theorem 5.2 had been previously given by Kruithof (1937, see also Beardwood and Kirby, 1975, and Macgill, 1977), it was first introduced into the statistical literature by Deming and Stephan (1940), and to the transportation literature by Furness (1965). Consequently, we have called it the Deming-Stephan-Furness Procedure or, simply, the DSF Procedure. In the economics literature it is often called the RAS Method (see Bachrach, 1970) and it has also been called the Iterative Proportional Fitting Procedure (IPFP).

Goodman (1968, see also Macgill, 1977) has given a variation of the DSF Procedure which reduces computer space requirements. It consists of setting $B_j{}^{(0)} = Q_j$, and then iterating using

$$A_i{}^{(2r-1)} = O_i / \sum_{j=1}^{J} B_j{}^{(2r-2)} F_{ij} \tag{5.64}$$

and

$$B_j{}^{(2r)} = D_j / \sum_{i=1}^{I} A_i{}^{(2r-1)} F_{ij}. \tag{5.65}$$

At any step $T_{ij}{}^{(s)}$ can be computed by multiplying together F_{ij} and the most recently obtained $A_i{}^{(s)}$ and $B_j{}^{(s)}$. That the two procedures are equivalent is easy to see from

$$\begin{aligned} T_{ij}{}^{(2r)} &= A_i{}^{(2r-1)} B_j{}^{(2r)} F_{ij} = A_i{}^{(2r-1)} D_j F_{ij} / \sum_i (A_i{}^{(2r-1)} F_{ij}) \\ &= A_i{}^{(2r-1)} B_j{}^{(2r-2)} D_j F_{ij} / \sum_i (A_i{}^{(2r-1)} B_j{}^{(2r-2)} F_{ij}) \\ &= T_{ij}{}^{(2r-1)} D_j / T_{\oplus j}{}^{(2r-1)}, \end{aligned} \tag{5.66}$$

and from a similar derivation using (5.64).

Corollary 5.2 *Under the conditions given in Theorem 5.2, as* $r \rightarrow \infty$, $B_j{}^{(2r)} \rightarrow \hat{B}(j)$ *and* $A_i{}^{(2r-1)} \rightarrow \hat{A}(i)$*, where* $\hat{A}(i)$ *and* $\hat{B}(j)$ *satisfy (5.49) and (5.50). Furthermore, the* $\hat{T}_{ij}$*'s defined by (5.49) are unique.*

PROOF: From the first equality in (5.66) and Theorem 5.2 it follows that

$$B_j^{(2r)}/B_j^{(2r-2)} = T_{ij}^{(2r)}/T_{ij}^{(2r-2)} \to 1.$$

Thus, $B_j^{(2r)} \to \hat{B}(j)$ and, similarly, $A_i^{(2r-1)} \to \hat{A}(i)$. That (5.49) and (5.50) would hold also follows from (5.66) and Theorem 5.2, as does the uniqueness of $\hat{T}_{ij}$'s. □

The DSF procedure has a very wide range of applications even outside of gravity models. Within the gravity model context, the DSF procedure has been used both inside and outside the ML arena. Within the ML context, its use has been to obtain estimates for T_{ij}'s or, alternatively, for $A(i)$'s and $B(j)$'s, when a value of $\boldsymbol{\theta}$ is available. This value of $\boldsymbol{\theta}$ could be the final ML estimate, in which case the DSF procedure would give the final estimates of T_{ij}'s, $A(i)$'s and $B(j)$'s; or it could be a trial value of $\boldsymbol{\theta}$ within an iteration that requires the computation of corresponding T_{ij}'s. The latter type of use places the DSF procedure within the steps of more complex procedures. For this type of use, the following notation is helpful. Note that (when $\boldsymbol{c}_{ij}$'s are known) for any value of $\boldsymbol{\theta}$, $F_{ij} = \exp(\boldsymbol{\theta}^t \boldsymbol{c}_{ij})$ can be computed and hence (5.12) can be solved using the DSF procedure. Call the resultant $\hat{T}_{ij}$'s, $\hat{T}_{ij}(\boldsymbol{\theta}; N_{1\oplus}, \ldots, N_{I\oplus}; N_{\oplus 1}, \ldots, N_{\oplus J})$ or, when no ambiguity is possible, $\hat{T}_{ij}(\boldsymbol{\theta})$. Let

$$\hat{\boldsymbol{T}}(\boldsymbol{\theta}) = (\hat{T}_{11}(\boldsymbol{\theta}), \ldots, \hat{T}_{1J}(\boldsymbol{\theta}), \ldots\ldots, \hat{T}_{I1}(\boldsymbol{\theta}), \ldots, \hat{T}_{IJ}(\boldsymbol{\theta}))^t. \tag{5.67}$$

This notation will be used later in several procedures for the ML estimation of $\boldsymbol{\theta}$.

In order to accommodate non-ML applications of the DSF procedure, Theorem 5.2 has been stated in somewhat greater generality than needed just for ML estimation. We have allowed F_{ij}'s to be arbitrary non-negative numbers, which allows one or more of them to be zero, a situation that would be impossible if we restricted ourselves to $F_{ij} = \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}]$. We have also generalized (5.12) to (5.50), where O_i's and D_j's are arbitrary positive numbers (not necessarily integers), so long as Condition ML3 holds.

One particularly valuable use of the DSF procedure is in making forecasts when forecast period $\hat{T}_{i\oplus} = O_i$, $\hat{T}_{\oplus j} = D_j$ and $F_{ij} = F(c_{ij})$'s are given, and we wish to compute forecast period $\hat{T}_{ij}$'s. Such applications are common in transportation planning where $\hat{T}_{i\oplus} = O_i$ and $\hat{T}_{\oplus j} = D_j$ are obtained from trip generation/attraction models and $F_{ij} = F(\boldsymbol{c}_{ij})$'s are obtained using exogenous forecast period $\boldsymbol{c}_{ij}$'s along with a $\hat{\boldsymbol{\theta}}$ which is estimated from base period data and assumed to remain constant into the forecast period (see Theorem 4.17 for conditions).

Proofs of all or part of Theorem 5.2 have been given by several authors, including Fienberg (1970, who has perhaps given the shortest proof and also provided a history of the procedure), Evans (1970), Haberman (1974), Macgill (1977, 1979), Bishop *et al.* (1975, p. 85), Evans and Kirby (1974) and Andersson (1981). For completeness, a proof, following somewhat the lines of Bishop *et al.*, is presented below. This requires some results which we now state in the form of a lemma.

Lemma 5.5 *Let $p_1, \dots, p_n$ and $q_1, \dots, q_n$ be sequences of positive numbers such that $p_\oplus = q_\oplus$. Then*

$$\sum_{i=1}^{n} p_i \log(p_i/q_i) \geq 0, \tag{5.68}$$

with $\sum_{i=1}^{n} p_i \log(p_i/q_i) = 0$ occurring only when $p_i = q_i$, for all $i = 1, \dots, n$.

Proof: (Taken from Rao, 1973 p. 58). For any $\xi > 0$, we get, using Taylor's Theorem

$$\log(\xi) = (\xi - 1) - (\xi - 1)^2 (2\eta^2)^{-1}, \tag{5.69}$$

where η lies between ξ and 1. Application of this to each term on the left of (5.68) yields

$$\begin{aligned}
\sum_{i=1}^{n} p_i \log(p_i/q_i) &= -\sum_{i=1}^{n} p_i \log(q_i/p_i) \\
&= -\sum_{i=1}^{n} p_i \log\left(1 + [(q_i/p_i) - 1]\right) \\
&= -\sum_{i=1}^{n} p_i[(q_i/p_i) - 1] + \sum_{i=1}^{n} p_i[(q_i/p_i) - 1]^2 (2\tilde{\eta}_i^2)^{-1} \\
&= (p_\oplus - q_\oplus) + \sum_{i=1}^{n} p_i[(q_i/p_i) - 1]^2 (2\tilde{\eta}_i^2)^{-1} \\
&= \sum_{i=1}^{n} p_i (q_i - p_i)^2 (2\eta_i^2)^{-1} \geq 0,
\end{aligned} \tag{5.70}$$

where $\tilde{\eta}_i$'s lie between 1 and q_i/p_i and η_i^2 between p_i^2 and q_i^2. The lemma follows. □

Proof of Theorem 5.2: The existence and uniqueness of $\hat{T}_{ij}$'s of the form (5.49) and obeying (5.50) have already been established in Corollary 5.1. Moreover, when $F_{ij} = 0$, $\hat{T}_{ij} = 0$. For the equations in this proof we use the convention that $0 \log(0) = 0$.

Notice that summing both sides of (5.62) over j and adding both sides of (5.63) over i, we get

$$T_{i\oplus}^{(2r-1)} = O_i = \hat{T}_{i\oplus} \text{ and } T_{\oplus j}^{(2r)} = D_j = \hat{T}_{\oplus j} \tag{5.71}$$

and, hence,

$$T^{(s)}_{\oplus\oplus} = \hat{T}_{\oplus\oplus} \tag{5.72}$$

for all s. Let

$$\mu_s = \sum_{ij} \hat{T}_{ij} \log(\hat{T}_{ij}/T_{ij}{}^{(s)}), \quad \nu_{2r} = \sum_j \hat{T}_{\oplus j} \log(\hat{T}_{\oplus j}/T_{\oplus j}{}^{(2r-1)}) \tag{5.73}$$

and

$$\nu_{2r-1} = \sum_i \hat{T}_{i\oplus} \log(\hat{T}_{i\oplus}/T_{i\oplus}{}^{(2r-2)}). \tag{5.74}$$

From (5.63),

$$\begin{aligned} &\log(T_{ij}{}^{(2r)}) - \log(T_{ij}{}^{(2r-1)}) = \log(D_j/T_{\oplus j}{}^{(2r-1)}) \\ &= \log(\hat{T}_{\oplus j}/T_{\oplus j}{}^{(2r-1)}). \end{aligned} \tag{5.75}$$

First multiplying all expressions in (5.75) by $\hat{T}_{ij}$ and then adding $\hat{T}_{ij} \log[\hat{T}_{ij}]$ to the left side of (5.75), and also subtracting it from the same side, we get

$$\begin{aligned} &-\hat{T}_{ij} \log(\hat{T}_{ij}/T_{ij}{}^{(2r)}) + \hat{T}_{ij} \log(\hat{T}_{ij}/T_{ij}{}^{(2r-1)}) \\ &= \hat{T}_{ij} \log(\hat{T}_{\oplus j}/T_{\oplus j}{}^{(2r-1)}), \end{aligned} \tag{5.76}$$

and it follows, on addition over i and j, that

$$-\mu_{2r} + \mu_{2r-1} = \nu_{2r}. \tag{5.77}$$

Similarly, using (5.62), we get

$$-\mu_{2r-1} + \mu_{2r-2} = \nu_{2r-1}. \tag{5.78}$$

By Lemma 5.5 and (5.72), $\nu_s \geq 0$ and $\mu_s \geq 0$. Consequently

$$\mu_s : s = 1, 2, \ldots$$

is a decreasing sequence of non-negative numbers, and must therefore converge. Consequently, $\nu_s \to 0$, and hence, from (5.72), (5.73), (5.74) and (5.70),

$$T_{\oplus j}{}^{(s)} \to \hat{T}_{\oplus j} = O_j \text{ and } T_{i\oplus}{}^{(s)} \to D_j. \tag{5.79}$$

Thus, since $T_{\oplus j}{}^{(s)}$'s are sums of the positive numbers $T_{ij}{}^{(s)}$'s, it follows that for some S, each $T_{ij}{}^{(s)}$ is bounded when $s > S$, and hence, the $T_{ij}{}^{(s)}$'s have at least one cluster point or point of accumulation. That each $T_{ij}{}^{(s)}$ is of the form (5.49) is immediately apparent from (5.61), (5.62) and (5.63), since we start with a $T_{ij}{}^{(0)}$ of this form and at each step we multiply $T_{ij}{}^{(r)}$ by quantities that depend either on i alone or on j alone. Hence, from (5.79) it follows that any point of accumulation of the $T_{ij}{}^{(s)}$'s is of the form (5.49) and obeys (5.50). By Corollary 5.1, such a point is unique. Therefore, the sequence $T_{ij}{}^{(s)}$ converges to a unique limit $\hat{T}_{ij}$ which is of the form (5.49) and which obeys (5.50). □

Initial Values and Stopping Rules

It has been the experience of the authors that unless the situation is terribly ill-conditioned, i.e., Condition ML3 is close to being violated (see the section below on numerical examples), initial values are not too critical. Obviously, if we actually knew the value of $T_{ij}{}^{(s)}$ for some iteration s, with s large, we would substantially cut down iterations. However, if we do not know such values, the difference between two reasonable guesses is usually no more than one or two iterations. We have found the use of $P_i = Q_j = 1$ quite satisfactory under most circumstances.

Stopping rules are of two types: those that stop iterations when we know that we are sufficiently close to the desired result and those that stop when improvements at each iteration become so small that it would not appear cost effective to continue. An example of the first type of stopping rule is

$$\sum_{i=1}^{I} |T_{i\oplus}{}^{(r)} - O_i| + \sum_{j=1}^{J} |T_{\oplus j} - D_j| < \delta \tag{5.80}$$

where $\delta > 0$ is a suitable preassigned number. Examples of the second kind include $\sum_{ij} |T_{ij}{}^{(s)} - T_{ij}{}^{(s-1)}|$ or $\sum_{i=1}^{I} |A_i{}^{(2r+1)} - A_i{}^{(2r-1)}|$ or $\sum_{j=1}^{J} |B_j{}^{(2r)} - B_j{}^{(2r-2)}|$ being less than some preset number.

An example of a stopping criterion which is attractive because it is simultaneously of both kinds mentioned above is

$$\sum_{ij} \delta_{ij} T_{ij}{}^{(s)} \log \left[\frac{T_{ij}{}^{(s)}}{T_{ij}{}^{(s-1)}} \right] < \delta \tag{5.81}$$

where δ_{ij} is as in the statement of ML3 and $\delta > 0$ is some preassigned number. By (5.70), we see that (5.81) measures the size of improvement from iteration to iteration. Moreover, from (5.62) and (5.63) and depending on the value of s, the left side of (5.81) can be written as either $\sum_j D_j \log \left[D_j / T_{\oplus j}{}^{(s-1)}\right]$ or as $\sum_i O_i \log \left[O_i / T_{i\oplus}{}^{(s-1)}\right]$, both of which measure departure from the desired end product of the DSF procedure. The only difficulty with this procedure is that it depends on logarithms which take a fair amount of time to compute, particularly if these computations are conducted at every iteration and if I and J are large. Therefore, we have tended to use (5.80), especially since $T_{i\oplus}{}^{(s)}$ and $T_{\oplus j}{}^{(s)}$ need to be computed anyway. If (5.64) and (5.65) are used, $T_{i\oplus}{}^{(s)}$ and $T_{\oplus j}{}^{(s)}$ can be obtained easily from the quantities that need to be computed for the iterations.

NUMERICAL EXAMPLES

Consider the following matrix of F_{ij}'s:

$$((F_{ij})) = \begin{pmatrix} 10 & 8 & 1 & 6 & 7 \\ 8 & 10 & 8 & 6 & 5 \\ 1 & 8 & 10 & 5 & 3 \\ 6 & 6 & 5 & 10 & 2 \\ 7 & 5 & 3 & 2 & 10 \end{pmatrix} \tag{5.82}$$

and let the desired vectors of O_i and D_j be

$$\boldsymbol{O}^t = (\ 100 \quad 150 \quad 50 \quad 200 \quad 250\ ) \tag{5.83}$$

$$\text{and} \quad \boldsymbol{D}^t = (\ 200 \quad 200 \quad 100 \quad 150 \quad 100\ ). \tag{5.84}$$

After three full iterations of the DSF Procedure (i.e., for $s = 6$), using as initial values $P_i = Q_j = 1$, we get

$$((T_{ij}{}^{(6)})) = \begin{pmatrix} 35.1 & 28.9 & 2.6 & 20.0 & 13.3 \\ 36.7 & 47.3 & 27.4 & 26.2 & 12.5 \\ 2.2 & 17.9 & 16.2 & 10.3 & 3.5 \\ 45.4 & 46.7 & 28.1 & 71.7 & 8.2 \\ 80.6 & 59.3 & 25.7 & 21.9 & 62.5 \end{pmatrix}. \tag{5.85}$$

Any noticeable difference between $T_{i\oplus}{}^{(6)}$ and O_i or between $T_{\oplus j}{}^{(6)}$ and D_j is only due to round-off.

A situation where Condition ML3 does not hold is illustrated by:

$$((F_{ij})) = \begin{pmatrix} X & 0 & 0 \\ 0 & X & X \\ 0 & X & X \end{pmatrix}, \ \boldsymbol{O} = \begin{pmatrix} 10 \\ X \\ X \end{pmatrix} \ \text{and } \boldsymbol{D} = \begin{pmatrix} 8 \\ X \\ X \end{pmatrix} \tag{5.86}$$

where $X \equiv$'anything' (they do not need to be equal either). Quite obviously, (since T_{11} needs to be a 8 for D_1 to be 8 and 10 for O_1 to be 10) it is impossible to get a set of T_{ij}'s meeting the requirements of Theorem 5.2. In a case like this, after a few iterations, values of $T_{ij}{}^{(s)}$ alternate, with $T_{ij}{}^{(2r-1)}$'s meeting row constraints and $T_{ij}{}^{(2r)}$'s meeting column constraints. (It may be shown that the sequence $T_{ij}{}^{(2r-1)} : r = 1, 2, \cdots$ converges to a unique limit, as does $T_{ij}{}^{(2r)} : r = 1, 2, \cdots$ — see Sinkhorn, 1967, Weber, 1987.)

The following values

$$((F_{ij})) = \begin{pmatrix} 1 & 4 \\ 5 & 0 \end{pmatrix}, \ \boldsymbol{O} = \begin{pmatrix} 10 \\ 5 \end{pmatrix} \ \text{and } \boldsymbol{D} = (\ 5.1 \quad 9.9\ ) \tag{5.87}$$

illustrate a 'borderline' situation where ML3 does hold, but barely so. [Notice that if $\boldsymbol{D} = (5 \ \ 10)$ instead of being $(5.1 \ \ 9.9)$, then no solution is possible. T_{12} would have to be 10, because $D_2 = 10$. But then $T_{11} + T_{12} > 10$ for any $T_{11} > 0$. $\boldsymbol{D} = (4.9 \ \ 10.1)$ creates a clearer violation of ML2]. Here, using (5.64) and (5.65), with $Q_j = 1$, the values of $T_{11}{}^{(s)}$ and $T_{12}{}^{(s)}$ for the first 41 iterations were found to be, respectively:

1, 2, 1.5, 1.3, 1.0, .95, .82, .76, .67, .64, .58, .55, .51, .49, .45, .44, .41, .40, .38, .37, .35, .34, .32, .32, .30, .30, .29, .28, .27, .27, .26, .25, .25, .24, .24, .23, .23, .22, .22, .22, .21, .21,

and

8, 9.9, 8.7, 9.9, 9.0, 9.9, 9.2, 9.9, 9.4, 9.9, 9.4, 9.9, 9.5, 9.9, 9.6, 9.9, 9.6, 9.9, 9.6, 9.9, 9.7, 9.9, 9.7, 9.9, 9.7, 9.9, 9.7, 9.9, 9.7, 9.9, 9.7, 9.9, 9.8, 9.9, 9.8, 9.9, 9.8, 9.9, 9.8, 9.9, 9.8.

Convergence did occur ultimately, but over 100 iterations were needed for reasonable results and oscillations are obvious in the case of ${T_{12}}^{(s)}$.

Typically slow convergence of the DSF Procedure occurs in such 'borderline' situations which are usually linked to a preponderance of zero-valued or near zero-valued F_{ij}'s. Since most ML algorithms discussed in Section 5.3 and 5.5 have the DSF Procedure imbedded within them, these remarks are germane to them.

5.3.2 The Evans-Kirby Procedure

Let V_0 be a set such that for all $i \in I$ and $j \in J$, $\boldsymbol{c}_{ij} \in V_0 \subset R^K$. [The Euclidean space R^K is used here simply because this chapter is devoted to Model (5.2). The procedure can accommodate ordinal and nominal measures as can readily be seen.] Divide V_0 into subsets $V_1, \ldots, V_{K'}$, such that

$$V_0 = \bigcup_{m=1}^{K'} V_m \quad \text{and} \quad V_m \cap V_n = \emptyset, \text{ for all } m, n = 1, \ldots, K'. \tag{5.88}$$

If $\boldsymbol{c}_{ij} \in V_m$, set $F(\boldsymbol{c}_{ij}) = F_m$, where each F_m is a parameter to be estimated. Defined thus, $F(\boldsymbol{c}_{ij})$ is obviously a step function.

If for some m there is no $\boldsymbol{c}_{ij} \in V_m$, no estimate of F_m can be found. Hence, we assume that such V_m's have been deleted from consideration and that for each m, there exists an i and a j such that $\boldsymbol{c}_{ij} \in V_m$. We further assume that for all m, $F_m \geq 0$.

In practical applications with travel time as the only measure of separation, these V_m's are intervals of, say, length 1 minute. A very large number of parameter estimation algorithms for gravity models in transportation planning use step functions and at least in part are similar to the Evans-Kirby procedure, described in this section.

Let

$$\delta_{ij}^{(m)} = \begin{cases} 1 \text{ if } \boldsymbol{c}_{ij} \in V_m \\ 0 \text{ otherwise.} \end{cases}$$

Since for any i and j, $\delta_{ij}^{(m)} = 1$ for one and only one value of m, it follows that

$$\sum_m \delta_{ij}^{(m)} = 1, \quad \sum_m \delta_{ij}^{(m)} T_{ij} = T_{ij}, \quad \sum_m \delta_{ij}^{(m)} N_{ij} = N_{ij} \tag{5.89}$$

for all $i \in I$ and $j \in j$, and

$$\sum_{i,j,m} \delta_{ij}{}^{(m)} T_{ij} = T_{\oplus\oplus}, \quad \sum_{i,j,m} \delta_{ij}{}^{(m)} N_{ij} = N_{\oplus\oplus}. \tag{5.90}$$

From the assumption made above that each V_m contains at least one $\boldsymbol{c}_{ij}$, it follows that

$$\delta_{\oplus\oplus}{}^{(m)} > 0 \quad \text{for all } m \in K. \tag{5.91}$$

Now, let

$$\theta_m = \log(F_m). \tag{5.92}$$

Since, for fixed i and j, $\delta_{ij}{}^{(m)}$ takes the value 1 for one and only one value of m, it follows that

$$\sum_{m=1}^{K'} \theta_m \delta_{ij}{}^{(m)} = \theta_m = \log[F_m]$$

if V_m contains $\boldsymbol{c}_{ij}$. Therefore,

$$F(\boldsymbol{c}_{ij}) = F_m = \exp[\sum_{m=1}^{K'} \theta_m \delta_{ij}{}^{(m)}], \tag{5.93}$$

which is of the form (2.126), the only differences being that $\delta_{ij}{}^{(k)}$'s correspond to the $c_{ij}{}^{(k)}$'s and that the index m (going from 1 to K') corresponds to the index k (going from 1 to K). Therefore, as mentioned in Section 2.3.2, step function $F(\boldsymbol{c}_{ij})$'s can be accommodated within the general exponential form (5.2). Consequently, ML estimates of F_m's, $A(i)$'s and B_j's may be found by solving (5.12) and (5.13). Moreover, under Condition ML2, there exists a unique set of $\hat{T}_{ij}$'s of the requisite form — (5.42).

If, in order to identify more closely with results obtained earlier, we replace the index $m = 1, \ldots, K'$ with $k = 1, \ldots, K$, let $\delta_{ij}{}^{(k)}$'s replace corresponding $c_{ij}{}^{(k)}$'s in the matrix M (of Section 5.2), we see that Condition ML1 does not hold here since from the first equation in (5.89),

$$\sum_{i=1}^{I} \boldsymbol{m}_s = \sum_{j=1}^{J} \boldsymbol{m}_{I+j} = \sum_{k=1}^{K} \boldsymbol{m}_{I+J+k} = (1, \ldots, 1)'. \tag{5.94}$$

However, by choosing the values of one $a(i)$ and one $b(j)$ or one $a(i)$ and one θ_k or one $b(j)$ and one θ_k as arbitrary positive numbers, we may verify that the remaining $a(i)$'s, $b(j)$'s and θ_k's are uniquely determined by the $\hat{T}_{ij}$'s.

Theorem 5.3 *Let $P_i{}^{(0)}$, $Q_j{}^{(0)}$, $R_k{}^{(0)}$ be arbitrary positive numbers for $i \in I$, $j \in J$ and $k \in K$, and let $T_{ij}{}^{(0)} = P_i{}^{(0)} Q_j{}^{(0)} R_k{}^{(0)} \delta_{ij}{}^{(k)}$. Consider the* *iterations,*

$$T_{ij}^{(3r+1)} = T_{ij}^{(3r)} R_k^{(r)}, \tag{5.95}$$

$$T_{ij}^{(3r+2)} = T_{ij}^{(3r+1)} P_i^{(r)}, \tag{5.96}$$

$$T_{ij}^{(3r+3)} = T_{ij}^{(3r+2)} Q_j^{(r)}, \tag{5.97}$$

where $r = 1, 2, \cdots$ *and*

$$R_k^{(r)} = \sum_{ij} N_{ij}\delta_{ij}^{(k)} / \sum_{ij} T_{ij}^{(3r)}\delta_{ij}^{(k)}, \tag{5.98}$$

$$P_i^{(r)} = N_{i\oplus}/T_{i\oplus}^{(3r+1)}, \quad Q_j^{(r)} = N_{\oplus j}/T_{\oplus j}^{(3r+2)}. \tag{5.99}$$

Under Condition ML2, for each i *and* j *the sequence of* $T_{ij}^{(s)}$*'s converges to a unique limit* $\hat{T}_{ij}$ *which is of the form (5.42), where* $\boldsymbol{\theta}$ *is given by (5.92) and* $\delta_{ij}^{(k)}$*'s replace* $c_{ij}^{(k)}$*'s. Moreover,* $\hat{T}_{ij}$ *obeys (5.12) and (5.13).*

Furthermore, values of $A(i)$*,* $B(j)$ *and* θ_k *can be obtained from*

$$A(i) = \prod_{s=0}^{\infty} P_i^{(s)}, \quad B(j) = \prod_{s=0}^{\infty} Q_j^{(s)}, \quad F_k = \prod_{s=0}^{\infty} R_k^{(s)}. \tag{5.100}$$

While these values are not unique, if for any i*,* j *and* k*, any two of the numbers* $A(i) > 0$*,* $B(j) > 0$ *and* $F_k > 0$ *are arbitrarily fixed, the remaining* $A(i)$*'s,* $B(j)$*'s and* F_k*'s are uniquely determined.*

PROOF: The existence and uniqueness of $\hat{T}_{ij}$'s follow from Lemma 5.3. The convergence of $T_{ij}^{(s)}$'s can be shown as for Theorem 5.2 (Note that (5.90) is needed in the proof). The rest of the theorem is trivial. □

A discussion of starting values and stopping rules would be very similar to that for the DSF procedure and is hence omitted.

The procedure presented in Theorem 5.3 was given by Evans and Kirby (1974) after whom we call it the Evans-Kirby Procedure. Evans and Kirby have given a proof of Theorem 5.3. An attractive feature of the procedure is that not much needs to be assumed about $F(c_{ij})$ in order to approximate it with a step function. Indeed, as already mentioned at the beginning of this section, there is no need for separation measures to even be costs; nominal, and ordinal measures will do as well, since no ordering of separation measure values is implied in the treatment given above. A disadvantage is that a good approximation frequently requires a large number of parameters F_m with consequently large standard errors. Attempts to reduce these standard errors can result in huge sample sizes (i.e., in huge $N_{\oplus\oplus}$'s).

Like the DSF Procedure, the Evans-Kirby Procedure has an equivalent which might simplify computational needs. This procedure consists of arbitrarily setting $A(i)^{(0)} > 0$ and $B(j)^{(0)} > 0$ and then iterating using

$$F_k^{(r)} = \sum_{ij} \delta_{ij}^{(k)} N_{ij} / \sum_{ij} \delta_{ij}^{(k)} A(i)^{(r-1)} B(j)^{(r-1)}, \tag{5.101}$$

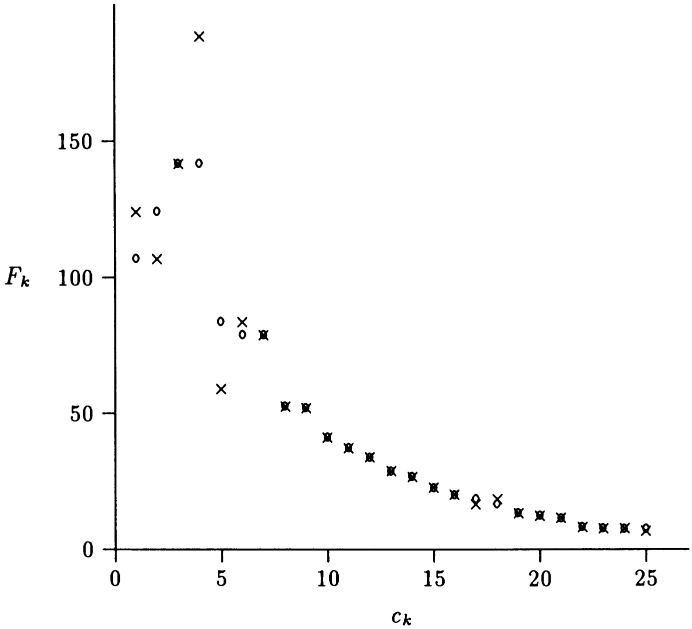

Exhibit 5.1: Original and Smoothed Values of F_k Plotted against c_k.

NOTES: A o denotes a smoothed value and a × an original (pre-smoothing) point. c_k's are in minutes.

$$A(i)^{(r)} = \sum_{jk} \delta_{ij}{}^{(k)} N_{ij} / \sum_{jk} \delta_{ij}{}^{(k)} B(j)^{(r-1)} F_k{}^{(r)}, \tag{5.102}$$

$$B(j)^{(r)} = \sum_{ik} \delta_{ij}{}^{(k)} N_{ij} / \sum_{ik} \delta_{ij}{}^{(k)} A(i)^{(r)} F_k{}^{(r)}, \tag{5.103}$$

for $r = 1, 2, \cdots$. The equivalence of (5.101), (5.102), (5.103) and (5.95), (5.96), (5.97) may be verified as for the DSF procedure.

A Numerical Example

The example given below illustrates how the flexibility of the $F(c_{ij})$ found by the Evans-Kirby procedure can be exploited. Since a large number of parameters are used in the procedure, we need to illustrate it with a reasonably large data set. One such set (which has also been used in other examples in this book) has been made available to us by Professor James Foerster, a colleague of one of the authors. This data set is based on 25247 trips allocated over 42 origins and destinations in the village of Skokie, Illinois. Several of the rows and columns of the N_{ij} matrix were zeros; their deletion resulted in a 40×18 matrix which is presented in the appendix. This data set which includes several different $c_{ij}{}^{(k)}$'s will be referred to as the 'Skokie' data set.

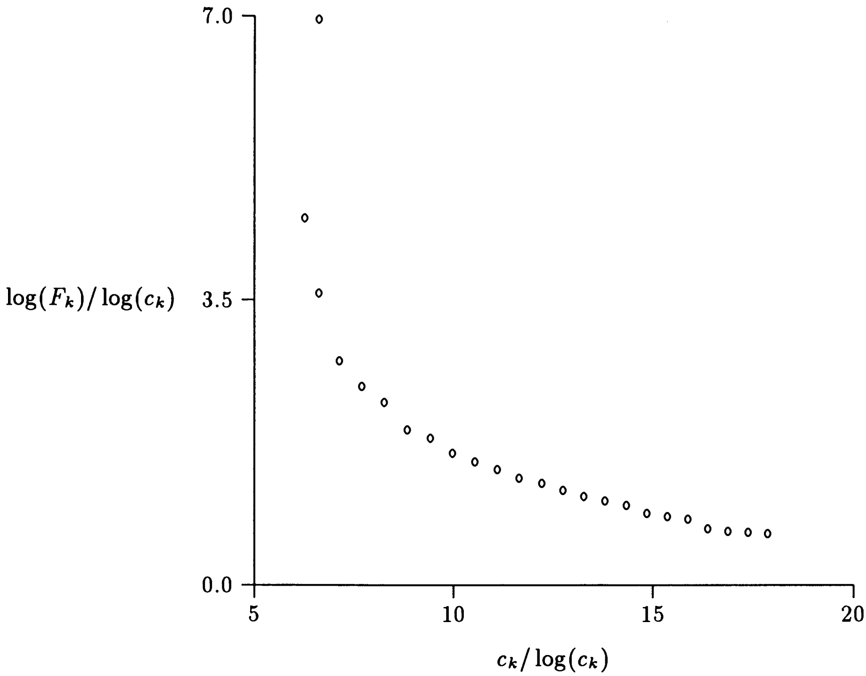

Exhibit 5.2: Plot of $[\log(F_k)/\log(c_k)]$ against $[c_k/\log(c_k)]$.

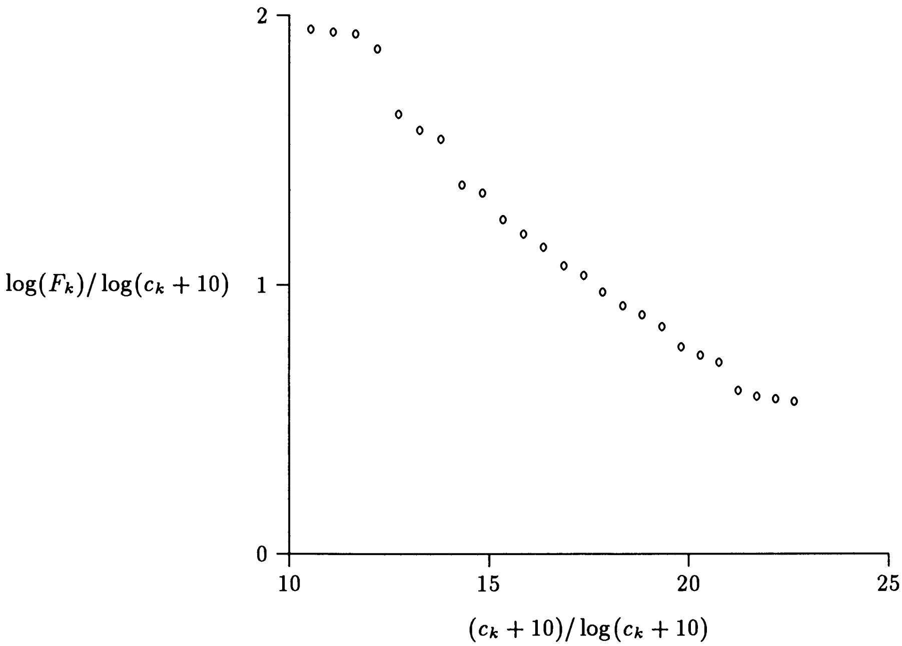

Exhibit 5.3: Plot of $[\log(F_k)/\log(c_k+10)]$ against $[(c_k+10)/\log(c_k+10)]$.

Since in this data set, travel times are given in integer values (minutes), each V_m was set equal to each such value — call them

$$c_1 < c_2 < \ldots < c_{25}. \tag{5.104}$$

After 15 complete iterations (i.e., $r = 15$), starting with the initial values $P_i^{(0)} = Q_j^{(0)} = R_k^{(0)} = 1$, the values of F_k shown in Exhibit 5.1 were obtained.

The plot is somewhat 'ragged' for low values of c_k's. Consequently, we smoothed it using a median smoothing procedure recommended by Tukey (1977). This procedure consists of replacing F_k for each $k = 2, \ldots, 24$ by the median of F_{k-1}, F_k and F_{k+1}. For F_1 an artificial point F_0 is created by linearly extrapolating $F_2^{(S)}$ and $F_3^{(S)}$, where the superscript S indicates that these values have already been smoothed. Then, F_1 is replaced by the median of F_0, F_1 and $F_2^{(S)}$. F_{25} is treated similarly. It should be mentioned that Tukey describes other steps which were not carried out here.

It would appear from Exhibit 5.1 that F_{ij} cannot be of the form $F_{ij} = \exp[\theta_1 c_{ij}]$; consequently, we considered the form

$$F_{ij} = F(c_{ij}) = c_{ij}^{\theta_2} \exp[\theta_1 c_{ij}] = \exp[\theta_2 \log(c_{ij}) + \theta_1 c_{ij}], \tag{5.105}$$

where c_{ij} is travel time in minutes. This form has been widely used in Gravity Model applications in transportation and land-use studies, and probably owes its popularity to a similarity to the Γ-function and the Chi-square density function [see also Section 2.3.2]. It follows from (5.105) that

$$[\log(F_{ij})/\log(c_{ij})] = \theta_2 + \theta_1[c_{ij}/\log(c_{ij})] \tag{5.106}$$

which, in terms of the notation of this section, is

$$[\log(F_k)/\log(c_k)] = \theta_2 + \theta_1[c_k/\log(c_k)]. \tag{5.107}$$

The form of (5.107) implies that the points $(c_k/\log(c_k), \log(F_k)/\log(c_k))$ all lie on a straight line. However, the plot shown in Exhibit 5.2 does not appear to be linear. The situation improves if we replace c_{ij} by $c_{ij} + 10$ in (5.106) — as shown by Exhibit 5.3. The 10 minutes added to c_{ij} has an obvious intuitive explanation if not appeal as representing the general additional penalty associated with any trip [e.g., putting on a coat, locking the door, finding a parking place, etc; The term 'shift' has been applied to the 10 minutes in Section 2.3.2].

Actually, the plot in Exhibit 5.3 is still not quite straight and replacing 10 by an even larger number improves the fit still further, although very slightly. However, the addition of 10 gives an adequate fit, and we shall not pursue this matter further.

5.3.3 The Hyman Procedure

The Hyman procedure may be used when $K = 1$, i.e., when $\boldsymbol{c}_{ij}$ has a single component $c_{ij}{}^{(1)}$. Correspondingly $\boldsymbol{\theta} = (\theta_1)$. The aim is to solve (5.12) and (5.13).

For any value of θ_1, the DSF procedure (with $O_i = N_{i\oplus}$ and $D_j = N_{\oplus j}$) may be used to solve (5.12). Using the notation of Section 5.3.1, let $\hat{T}_{ij}(\theta_1)$'s be the values of $\hat{T}_{ij}$'s obtained after applying the DSF Procedure. Write

$$\sum_{ij} c_{ij}{}^{(1)}\hat{T}_{ij}(\theta_1) = C(\theta_1) \qquad \text{and} \quad \sum_{ij} c_{ij}{}^{(1)}N_{ij} = C. \tag{5.108}$$

Since for any θ_1, because of the application of the DSF procedure, (5.12) is already satisfied, all we need do is solve

$$C(\theta_1) = C. \tag{5.109}$$

One way of solving (5.109) is the Newton-Raphson Method.

Hyman (1969) suggested such a procedure, the steps of which are as follows: Select an initial value $\theta_1{}^{(0)}$ of θ (Hyman suggests using $\theta_1{}^{(0)} = -1.5N_{\oplus\oplus}/C$). Then, using the DSF Procedure, obtain $C(\theta_1{}^{(0)})$ and compute

$$\theta_1{}^{(1)} = \theta_1{}^{(0)}C(\theta_1{}^{(0)})/C. \tag{5.110}$$

After that, $\theta_1{}^{(r)} : r = 2, 3, \cdots$ are computed using the formula

$$\theta_1{}^{(r)} = \frac{[C(\theta_1{}^{(r-1)}) - C]\theta_1{}^{(r-2)} - [C(\theta_1{}^{(r-2)}) - C]\theta_1{}^{(r-1)}}{C(\theta_1{}^{(r-1)}) - C(\theta_1{}^{(r-2)})}. \tag{5.111}$$

In this step we are performing a linear interpolation between the points $(\theta_1{}^{(r-1)}, C(\theta_1{}^{(r-1)}))$ and $(\theta_1{}^{(r-2)}, C(\theta_1{}^{(r-2)}))$ to find a value of $\theta_1{}^{(r)}$ corresponding to C, i.e., the step consists of finding a solution for $\theta_1{}^{(r)}$ from

$$\frac{C(\theta_1{}^{(r-2)}) - C}{C(\theta_1{}^{(r-1)}) - C} = \frac{\theta_1{}^{(r-2)} - \theta_1{}^{(r)}}{\theta_1{}^{(r-1)} - \theta_1{}^{(r)}}. \tag{5.112}$$

Equation (5.112) is equivalent to

$$\begin{aligned}(C(\theta_1{}^{(r-2)}) - C)&(\theta_1{}^{(r-1)} - \theta_1{}^{(r)}) \\ &= (C(\theta_1{}^{(r-1)}) - C)(\theta_1{}^{(r-2)} - \theta_1{}^{(r)})\end{aligned}$$

which, in turn, is equivalent to

$$\begin{aligned}(C(\theta_1{}^{(r-2)}) - C)&\theta_1{}^{(r-1)} - (C(\theta_1{}^{(r-1)}) - C)\theta_1{}^{(r-2)} \\ &= \theta_1{}^{(r)}(C(\theta_1{}^{(r-2)}) - C(\theta_1{}^{(r-1)})),\end{aligned}$$

whence (5.111). Notice that for each r, the computation of $C(\theta_1{}^{(r)})$ requires an application of the DSF procedure.

(a)

s	$\theta_1^{(s)}$
1	-0.5
2	-0.39773
3	-0.13177
4	-0.17274
5	-0.13272
6	-0.13455
7	-0.13448
8	-0.13448

(b)

s	$\theta_1^{(s)}$
1	-1.5
2	-1.02882
3	2.45667
4	0.39974
5	-8.89020
6	-3.67594
7	31.5698
8	*Overflow*

Exhibit 5.4: Values of $\theta_1{}^{(s)}$ in Examples of Hyman's Procedure

Like any Newton-Raphson Procedure, there is no guarantee of convergence; but, in practice, if $\theta_1{}^{(0)}$ is not chosen particularly poorly, convergence will nearly always occur.

Evans (1971) has shown that the function $C(\theta_1)$ is monotonic with respect to θ_1 [see Section 5.2.2 and see Thomas (1977) for an empirical example $C(\theta_1)$]. Other methods for obtaining ML estimates of θ_1 have been given by Evans (1971), Hathaway (1975; see also Southworth, 1979) and Williams (1976, 1977). Williams has also compared a number of methods for the case $K = 1$ and found Hyman's Procedure to be as good as any in terms of computational efficiency.

A Numerical Example

On applying Hyman's procedure to the Skokie data that we used in the last section, we obtained successive values of $\theta_1{}^{(s)}$, shown in Exhibit 5.4(a). Clearly, in this case, adding a 10 to c_{ij} makes no difference, since all that it does is introduce a constant factor into $F(c_{ij})$, which may be absorbed into either the $A(i)$'s or the $B(j)$'s. We did not use Hyman's suggested initial value because it was very close to the final value of θ_1. Exhibit 5.4(b) shows what happens when the initial value chosen is quite bad.

5.4 The LDSF Procedure

In this section we present a procedure which we call the LDSF procedure. Since the last section covered algorithms for obtaining ML estimates in some special cases, and the next section covers algorithms for obtaining ML estimates in the general case, the location of this section might appear odd. The reason for its location is that several methods for computing maximum likelihood estimates can be improved by incorporating within their steps, some steps of the LDSF procedure. The procedure is described below in Section 5.4.1 and is used to construct ML estimation algorithms in Section 5.5.

Since the LDSF procedure can be used also for purposes other than ML estimation algorithms, our treatment is fuller than that required for this application alone. Another application — to short term forecasting — is discussed in Section 5.4.3. Yet another application — to the computation of covariances of future $\hat{T}_{ij}$'s — is discussed in Section 5.7.2.

5.4.1 The Procedure

For all $i \in I$ and $j \in J$, let T_{ij} be a known set of expected flows and let $O_i = T_{i\oplus}$ and $D_j = T_{\oplus j}$. Assume that $\Delta\boldsymbol{O} = (\Delta O_1, \ldots, \Delta O_I)^t$ and $\Delta\boldsymbol{D} = (\Delta D_1, \ldots, \Delta D_J)^t$ are small changes in the values of $\boldsymbol{O} = (O_1, \ldots, O_I)^t$ and $\boldsymbol{D} = (D_1, \ldots, D_J)^t$, and let ΔF_{ij} be a small change in $F_{ij} = F(c_{ij})$ for all $i \in I$ and $j \in J$. The LDSF procedure obtains approximate small changes ΔT_{ij} in each T_{ij} when values of $T_{i\oplus} = O_i$, $T_{\oplus j} = D_j$ and F_{ij} undergo these small changes.

Let

$$\Delta T_{ij}{}^{(0)} = \Delta F_{ij} T_{ij} / F_{ij}, \tag{5.113}$$

$$\Delta T_{ij}{}^{(2r-1)} = \Delta T_{ij}{}^{(2r-2)} + (T_{ij}/O_i)(\Delta O_i - \Delta T_{i\oplus}{}^{(2r-2)}) \tag{5.114}$$

and

$$\Delta T_{ij}{}^{(2r)} = \Delta T_{ij}{}^{(2r-1)} + (T_{ij}/D_j)(\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}), \tag{5.115}$$

for $i \in I$, $j \in J$ and $r = 0, 1, 2, \cdots$. Application of (5.113), and the subsequent iterative use of (5.114) and (5.115) constitute the LDSF procedure.

Equation (5.113) is obtained from the first term in the last line in

$$\begin{aligned}
\Delta T_{ij} &= [T_{ij} + \Delta T_{ij}] - T_{ij} \\
&= [A(i) + \Delta A(i)][B(j) + \Delta B(j)][F_{ij} + \Delta F_{ij}] - A(i)B(j)F_{ij} \\
&= A(i)B(j)\Delta F_{ij} + A(i)\Delta B(j)F_{ij} + \Delta A(i)B(j)F_{ij} \\
&\qquad\qquad + \text{higher order terms} \\
&\approx A(i)B(j)\Delta F_{ij} + A(i)\Delta B(j)F_{ij} + \Delta A(i)B(j)F_{ij} \\
&= (T_{ij}/F_{ij})\Delta F_{ij} + (T_{ij}/B(j))\Delta B(j) + (T_{ij}/A(i))\Delta A(i).
\end{aligned} \tag{5.116}$$

The two iterative steps (5.114) and (5.115) seek to find values of the other two terms in (5.116) so that

$$\Delta T_{i\oplus} = \Delta O_i \quad \text{and} \quad \Delta T_{\oplus j} = \Delta D_j. \tag{5.117}$$

To make this connection explicit, consider the iterations

$$\Delta A^{(2r-1)}(i) - \Delta A^{(2r-3)}(i) = A(i)(\Delta O_i - \Delta T_{i\oplus}^{(2r-2)})/O_i \tag{5.118}$$

and

$$\Delta B^{(2r)}(j) - \Delta B^{(2r-2)}(j) = B(j)(\Delta D_j - \Delta T_{\oplus j}^{(2r-1)})/D_j \tag{5.119}$$

and set $\Delta T_{ij}^{(s)}$ as the function (5.116) with $\Delta A(i)$ and $\Delta B(j)$ being replaced by the most recent versions of these quantities; i.e.,

$$\begin{aligned}\Delta T_{ij}{}^{(2r)} &= (T_{ij}/A(i))\Delta A^{(2r-1)}(i)\\ &\quad + (T_{ij}/B(j))\Delta B^{(2r)}(j) + (T_{ij}/F_{ij})\Delta F_{ij}.\end{aligned} \tag{5.120}$$

and

$$\begin{aligned}\Delta T_{ij}{}^{(2r-1)} &= (T_{ij}/A(i))\Delta A^{(2r-1)}(i)\\ &\quad + (T_{ij}/B(j))\Delta B^{(2r-2)}(j) + (T_{ij}/F_{ij})\Delta F_{ij}.\end{aligned} \tag{5.121}$$

Of course, we can also use these equations to eliminate $\Delta T_{ij}{}^{(s)}$'s altogether from (5.118) and (5.119).

To verify that (5.119) is indeed the same as (5.115), subtract (5.121) from (5.120). We get

$$\Delta T_{ij}{}^{(2r)} - \Delta T_{ij}{}^{(2r-1)} = (T_{ij}/B(j))(\Delta B^{(2r)}(j) - \Delta B^{(2r-2)}(j)) \tag{5.122}$$

and then, using (5.115), we get (5.119). Similarly, we see that (5.114) leads to (5.118). An important observation that can be made here is that in every iteration one or other (but not both) of the quantities $\Delta A(i)$ or $\Delta B(j)$ is incrementally adjusted.

The two steps (5.114) and (5.115) are actually linearized versions of DSF iterations. To see this, notice that from (5.62) we may write

$$\begin{aligned}&T_{ij} + \Delta T_{ij}{}^{(2r-1)}\\ &\quad = (T_{ij} + \Delta T_{ij}{}^{(2r-2)})(O_i + \Delta O_i)/(T_{i\oplus} + \Delta T_{i\oplus}{}^{(2r-2)}),\end{aligned} \tag{5.123}$$

which yields

$$\begin{aligned}&(1 + \Delta T_{ij}{}^{(2r-1)}/T_{ij})\\ &\quad = (1 + \Delta T_{ij}{}^{(2r-2)}/T_{ij})(1 + \Delta O_i/O_i)/(1 + \Delta T_{i\oplus}{}^{(2r-2)}/O_i)\\ &\quad \approx (1 + \Delta T_{ij}{}^{(2r-2)}/T_{ij})(1 + \Delta O_i/O_i)(1 - \Delta T_{i\oplus}{}^{(2r-2)}/O_i)\\ &\quad \approx 1 + \Delta T_{ij}{}^{(2r-2)}/T_{ij} + \Delta O_i/O_i - \Delta T_{i\oplus}{}^{(2r-2)}/O_i,\end{aligned} \tag{5.124}$$

by using a Taylor's series approximation to get the third line of (5.124) and then multiplying together the three expressions within parentheses, in each case retaining only linear terms. (5.114) follows from (5.124). Similarly, (5.115) follows from (5.63).

Equivalently, we could start with, say, (5.65), and write

$$B(j) + \Delta B^{(2r)}(j) = \frac{D_j + \Delta D_j}{\sum_{i=1}^{I}(A(i) + \Delta A^{(2r-1)}(i))(F_{ij} + \Delta F_{ij})}. \tag{5.125}$$

Since $B(j) + \Delta B^{(2r)}(j) = B(j) + \Delta B^{(2r-2)}(j) + \Delta B^{(2r)}(j) - \Delta B^{(2r-2)}(j)$, we get, on dividing both sides of (5.125) by $B(j) + \Delta B^{(2r-2)}(j)$ and using the discussion just below (5.65),

$$1 + \frac{\Delta B^{(2r)}(j) - \Delta B^{(2r-2)}(j)}{B(j) + \Delta B^{(2r-2)}(j)} = \frac{D_j + \Delta D_j}{\sum_{i=1}^{I}[T_{ij} + \Delta T_{ij}{}^{(2r-1)}]}.$$

Ignoring second order terms, the fraction on the left becomes $(\Delta B^{(2r)}(j) - \Delta B^{(2r-2)}(j))/B(j)$. Since $D_j = T_{\oplus j}$, the right side is

$$\frac{D_j[1 + (\Delta D_j/D_j)]}{T_{\oplus j}[1 + (\Delta T_{\oplus j}^{(2r-1)}/T_{\oplus j})]} \approx 1 + \frac{\Delta D_j}{D_j} - \frac{\Delta T_{\oplus j}^{(2r-1)}}{D_j}$$

from which (5.119) follows. The relationship between (5.64) and (5.118) can be seen in much the same way.

In order to prove convergence of $\Delta T_{ij}{}^{(2r-1)}$'s, in the next theorem, we need the following lemma:

Lemma 5.6 *Assume that for all $i \in I$ and $j \in J$, $T_{ij} > 0$ and $F_{ij} > 0$. Then the $\Delta T_{ij}{}^{(s)}$'s, as given by (5.120) or (5.121), are linear functions of the $\Delta T_{i\oplus}^{(s)}$'s and $\Delta T_{\oplus j}^{(s)}$'s and these functions are independent of s.*

PROOF: Let $s = 2r$ ($s = 2r - 1$ can be handled in the same way). We start by introducing some notation which, except for the definition of Φ below, is essentially local to this proof. Let

$$\begin{aligned}
\Delta \boldsymbol{F} &= (\Delta F_{11}, \ldots, \Delta F_{1J}, \ldots\ldots, \Delta F_{I1}, \ldots, \Delta F_{IJ})^t, \\
\Delta \boldsymbol{T}^{(2r)} &= (\Delta T^{(2r)}{}_{11}, \ldots, \Delta T^{(2r)}{}_{1J}, \ldots\ldots, \Delta T^{(2r)}{}_{I1}, \ldots, \Delta T^{(2r)}{}_{IJ})^t, \\
\Delta \boldsymbol{A}^{(2r-1)} &= (\Delta A^{(2r-1)}(1), \ldots, \Delta A^{(2r-1)}(I))^t, \\
\Delta \boldsymbol{B}^{(2r)*} &= (\Delta B^{(2r)}(1), \ldots, \Delta B^{(2r)}(J-1))^t, \\
\Delta \boldsymbol{T}_{i\oplus}{}^{(2r)} &= (\Delta T_{1\oplus}{}^{(2r)}, \ldots, \Delta T_{I\oplus}{}^{(2r)})^t, \\
\Delta \boldsymbol{T}_{\oplus j}{}^{(*2r)} &= (\Delta T_{\oplus j}{}^{(2r)}, \ldots, \Delta T_{\oplus J-1}{}^{(2r)})^t
\end{aligned}$$

and set

$$\boldsymbol{\mu}_{2r}^t = (\Delta \boldsymbol{F}^t, (\Delta \boldsymbol{T}_{i\oplus}{}^{(2r)})^t, (\Delta \boldsymbol{T}_{\oplus j}{}^{(*2r)})^t)$$

and

$$\boldsymbol{\nu}_{2r}^t = ((\Delta \boldsymbol{T}^{(2r)})^t, (\Delta \boldsymbol{A}^{(2r-1)})^t, (\Delta \boldsymbol{B}^{(2r)*})^t).$$

Let

$$\Phi_1 = \begin{pmatrix} \operatorname{diag}(\boldsymbol{\tau}) & M^{(0)} \\ (M^{(0)})^t & 0 \end{pmatrix}, \tag{5.126}$$

where $\boldsymbol{\tau} = (1/T_{11}, \ldots, 1/T_{1J}, \ldots\ldots, 1/T_{I1}, \ldots, 1/T_{IJ})^t$ and $M^{(0)}$ is $M_{(1)}$ with its last column missing ($M_{(1)}$, is defined just before Lemma 5.1 and, by that same lemma, $M^{(0)}$ may be seen to be of full rank) and $\operatorname{diag}(\cdot)$ stands for a diagonal matrix, the diagonal elements of which are given within the parentheses. If the argument is a vector, then the diagonal elements are the components of the vector written in the same order. Let also

$$\Phi_2 = \begin{pmatrix} \operatorname{diag}(\boldsymbol{\phi}) & 0 \\ 0 & I_{I+J-1} \end{pmatrix},$$

where $\boldsymbol{\phi} = (1/F_{11}, \ldots, 1/F_{1J}, \ldots\ldots, 1/F_{I1}, \ldots, 1/F_{IJ})^t$ and an I with a subscript is an identity matrix of dimension given by the subscript, and

$$\Phi_3 = \begin{pmatrix} I_{IJ} & 0 & 0 \\ 0 & (\operatorname{diag}(\boldsymbol{A}))^{-1} & 0 \\ 0 & 0 & (\operatorname{diag}(B_1, \ldots, B_{J-1})^t)^{-1} \end{pmatrix}.$$

The system of equations consisting of

$$\begin{aligned} \Delta F_{ij} &= (F_{ij}/T_{ij})\Delta T_{ij}^{(2r)} \\ &\quad - (F_{ij}/A(i))\Delta A^{(2r-1)}(i) - (F_{ij}/B(j))\Delta B^{(2r)}(j) \end{aligned} \tag{5.127}$$

(which is a rearrangement of (5.120)) and the identities

$$\begin{aligned} \sum_{j=1}^{J} \Delta T_{ij}^{(2r)} &= \Delta T_{i\oplus}^{(2r)} : i = 1, \ldots, I \\ \text{and } \sum_{i=1}^{I} \Delta T_{ij}^{(2r)} &= \Delta T_{\oplus j}^{(2r)} : j = 1, \ldots, J-1 \end{aligned} \tag{5.128}$$

may be written in matrix form as

$$\boldsymbol{\mu}_{2r} = \Phi_2 \Phi_1 \Phi_3 \boldsymbol{\nu}_{2r}. \tag{5.129}$$

[The limitations of the size of this page require us to introduce this rather cumbersome notation; the reader is encouraged to write out (5.129) fully, making all the substitutions implied by the notation introduced earlier in this proof.]

The diagonal matrices Φ_2 and Φ_3 are clearly nonsingular. To show that Φ_1 is non-singular, notice that since $M^{(0)}$ is of full rank, $M^{(0)} \operatorname{diag}(\sqrt{\boldsymbol{T}})^t$ where

$$\sqrt{\boldsymbol{T}} = (\sqrt{T}_{11}, \ldots, \sqrt{T}_{1J}, \ldots\ldots, \sqrt{T}_{I1}, \ldots, \sqrt{T}_{IJ})^t$$

is also of full rank. Therefore,

$$M^{(0)^t} \operatorname{diag}(\boldsymbol{T}) M^{(0)} = M^{(0)^t} \operatorname{diag}(\sqrt{\boldsymbol{T}}) \operatorname{diag}(\sqrt{\boldsymbol{T}}) M^{(0)}$$

is nonsingular. Using the general result (see Rao, 1973, p. 32, Example 2.4):

$$\det \begin{pmatrix} R & V \\ S & U \end{pmatrix} = \det(R) \cdot \det(U - SR^{-1}V),$$

where det($\cdot$) stands for the 'determinant of,' R and U are square and R is non-singular, we see that the determinant det$[\Phi_1]$ of Φ_1 is

$$(\det[\operatorname{diag}(\boldsymbol{T})])^{-1} \cdot \det[-M^{(0)^t} \operatorname{diag}(\boldsymbol{T}) M^{(0)}] > 0. \tag{5.130}$$

Hence Φ_1 is non-singular. [Alternatively, we can show the non-singularity of Φ_1 without use of determinants as follows: For any $\boldsymbol{x}$ and $\boldsymbol{y}$ of suitable dimension, if

$$\begin{pmatrix} \operatorname{diag}(\boldsymbol{\tau}) & M^{(0)} \\ (M^{(0)})^t & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \mathbf{o} \\ \mathbf{o} \end{pmatrix},$$

then

$$\operatorname{diag}(\boldsymbol{\tau})\boldsymbol{x} + M^{(0)}\boldsymbol{y} = \mathbf{o} \text{ and } (M^{(0)})^t\boldsymbol{x} = \mathbf{o}.$$

Thus, $\boldsymbol{x}^t M^{(0)} = \mathbf{o}$ and

$$\begin{aligned} \mathbf{o} &= \boldsymbol{x}^t[\operatorname{diag}(\boldsymbol{\tau})\boldsymbol{x} + M^{(0)}\boldsymbol{y}] \\ &= \boldsymbol{x}^t \operatorname{diag}(\boldsymbol{\tau})\boldsymbol{x} + \boldsymbol{x}^t M^{(0)}\boldsymbol{y} = \boldsymbol{x}^t \operatorname{diag}(\boldsymbol{\tau})\boldsymbol{x}, \end{aligned}$$

i.e., $\boldsymbol{x} = \mathbf{o}$. And since $\boldsymbol{x} = \mathbf{o}$ implies $M^{(0)}\boldsymbol{y} = \mathbf{o}$, it follows that $\boldsymbol{y} = \mathbf{o}$, proving the non-singularity of Φ_1.]

Set $\Phi = \Phi_3^{-1}\Phi_1^{-1}\Phi_2^{-1}$. Since Φ_1, Φ_2 and Φ_3 do not depend on s, nor does Φ. Moreover, $\Phi\boldsymbol{\mu}_{2r} = \boldsymbol{\nu}_{2r}$, and by following very similar steps it may be shown that $\Phi\boldsymbol{\mu}_{2r-1} = \boldsymbol{\nu}_{2r-1}$. [Notice that we would not just be replacing $2r$ by $2r-1$ in the superscripts of ΔA and ΔB; the superscripts will be determined by (5.127). However, Φ would remain unchanged.] The lemma follows on examination of the first IJ rows of $\boldsymbol{\nu}$. □

Notice that since $\sum_{j=1}^J \Delta T_{\oplus j}^{(s)} = \sum_{i=1}^I \Delta T_{i\oplus}^{(s)}$, one of the identities (5.128) is redundant. One $\Delta T_{\oplus j}^{(s)}$ or $\Delta T_{i\oplus}^{(s)}$ can be written as a linear combination of the others and can, therefore, be eliminated from consideration. In the above proof, we eliminated $\Delta T_{\oplus J}$. If we had chosen all $\Delta T_{\oplus j}^{(s)}$'s or $\Delta T_{i\oplus}^{(s)}$'s as arbitrary numbers first, then (5.128) would not necessarily hold.

Also notice that, if all F_{ij}'s, $A(i)$'s, $B(j)$'s, ΔF_{ij}'s and ΔT_{ij}'s were known, we could solve (5.116) for $\Delta A(i)$'s and $\Delta B(j)$'s, but the solution would not be unique. For any solution, we could add a number to all the $\Delta A(i)$'s and subtract the same number from all $\Delta B(j)$'s and the resultant quantities would also be a solution. Or we could set the value of one $\Delta B(j)$ arbitrarily (say, set $\Delta B(J) = 0$), and the remaining $\Delta A(i)$'s and $\Delta B(j)$'s could be obtained (in fact, uniquely) by solving (5.116). This motivated our removal of $\Delta B_J{}^{(2r)}$ in the above proof.

The change from $M_{(1)}$ to $M^{(0)}$ is, of course, the corresponding change.

Corollary 5.3 *For all $i \in I$ and $j \in J$, let F_{ij}'s, $A(i)$'s and $B(j)$'s be given positive numbers. If $T_{ij} = A(i)B(j)F_{ij}$ for all i and j, then for a given set* *of numbers ΔO_i and ΔD_j, with $\Delta O_{\oplus\oplus} = \Delta D_{\oplus\oplus}$, there is a unique set of numbers ΔT_{ij} of the form given by the last line in (5.116) and obeying (5.117).*

PROOF: The proof of Lemma 5.6 can be adapted to prove the corollary: simply strip out all the superscripts — $(2r)$ or $(2r-1)$ and replace $\Delta T_{i\oplus}$'s and $\Delta T_{\oplus j}$'s by ΔO_i's and ΔD_j's, respectively, i.e., if

$$\boldsymbol{\mu}^t = (\Delta \boldsymbol{F}^t, (\Delta \boldsymbol{O}_{i\oplus})^t, (\Delta \boldsymbol{D}_{\oplus j}{}^{(*)})^t)$$

and

$$\boldsymbol{\nu}^t = ((\Delta \boldsymbol{T})^t, (\Delta \boldsymbol{A})^t, (\Delta \boldsymbol{B}^*)^t),$$

where the $*$ still implies deletion of ΔD_J and $\Delta B(J)$, we get, using (5.116) and (5.117), $\Phi\boldsymbol{\mu} = \boldsymbol{\nu}$, where Φ is as in the proof of Lemma 5.6.

Alternatively, we could construct a proof similar to that of Lemma 5.3. Consider the problem of minimizing

$$\sum_{ij} \frac{(\Delta T_{ij})^2}{T_{ij}} - 2\frac{\Delta F_{ij}}{F_{ij}} T_{ij}$$

subject to the constraints:

$$\sum_j \Delta T_{ij} = \Delta O_i \text{ and } \sum_i \Delta T_{ij} = \Delta D_j.$$

[This optimizing problem has been considered in Tobler (1983) in a derivation of a linear spatial interaction model.] The constraints can be written as

$$M_{(1)}{}^t \Delta \boldsymbol{T} = (\Delta \boldsymbol{O}^t, \Delta \boldsymbol{D}^t)^t,$$

where

$$\Delta \boldsymbol{T} = (\Delta T_{11}, \ldots, \Delta T_{1J}, \ldots\ldots, \Delta T_{I1}, \ldots, \Delta T_{IJ})^t.$$

Since $\Delta O_{\oplus\oplus} = \Delta D_{\oplus\oplus}$, one of the O_i's or D_j's (say D_J) can be written as a combination of the others and is redundant. Thus the constraints may be written as

$$M^{(0)^t} \Delta \boldsymbol{T} = \Delta \boldsymbol{\psi}, \tag{5.131}$$

where $\Delta\boldsymbol{\psi}$ is the vector $(\Delta \boldsymbol{O}^t, \Delta \boldsymbol{D}^t)$ with its last component missing. Since $M^{(0)}$ is of full rank, there is a solution to the system of equations (5.131). In fact, it defines a $IJ - I - J + 1$ dimensional subspace. Thus, the constraint space for the optimizing problem is non-empty and a solution exists. And it is unique since the objective function is clearly convex.

Using Lagrange multipliers, the solution is seen to be given by

$$2\Delta T_{ij}/T_{ij} - 2\Delta F_{ij}/F_{ij} - \lambda_i - \nu_j = 0.$$

Writing $\Delta A(i) = \lambda_i A(i)$ and $\Delta B(j) = \nu_j B(j)$, we see that ΔT_{ij} is of the appropriate form. □

Theorem 5.4 *If for all $i \in I$ and $j \in J$, $T_{ij} > 0$, and if ΔF_{ij}, ΔO_i and ΔD_j are arbitrary numbers such that $\Delta O_\oplus = \Delta D_\oplus$, then for each i and j the sequence $\Delta T_{ij}{}^{(s)} : s = 1, 2, 3, \cdots$ given by (5.113), (5.114) and (5.115) converges to ΔT_{ij}, such that (5.117) holds.*

In practice, one seldom encounters a $T_{ij} = 0$. Moreover, since we have assumed that $A(i) > 0$ and $B(j) > 0$ in this chapter, (5.2) cannot yield a zero valued T_{ij}. However, we point out that the condition in Theorem 5.4 that T_{ij}'s be positive can be relaxed, although the proof then becomes somewhat messy.

What makes the LDSF Procedure very attractive is that in most practical situations, convergence is very rapid — often one or two pairs of iterations being adequate. An example is given below (see also Weber and Sen, 1985).

PROOF OF THEOREM 5.4: Since $\sum_i T_{ij}/D_j = 1$, adding over $i = 1, 2, \ldots, I$, we get from (5.115),

$$\Delta T_{\oplus j}{}^{(2r)} = \Delta D_j. \tag{5.132}$$

Similarly, from (5.114), it follows that

$$\Delta T_{i\oplus}{}^{(2r-1)} = \Delta O_i. \tag{5.133}$$

Therefore,

$$\Delta T_{\oplus\oplus}{}^{(s)} = \Delta D_\oplus = \Delta O_\oplus \tag{5.134}$$

for all $s = 1, 2, \cdots$. From (5.133) and (5.115), we get

$$\begin{aligned} \Delta O_i - \Delta T_{i\oplus}{}^{(2r)} &= \Delta T_{i\oplus}{}^{(2r-1)} - \Delta T_{i\oplus}{}^{(2r)} \\ &= -\sum_j (T_{ij}/D_j)(\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}). \end{aligned} \tag{5.135}$$

Therefore, and since by (5.134)

$$\begin{aligned} &\{\min_j (T_{ij}/D_j)\} \sum_j (\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}) \\ &\quad = \{\min_j (T_{ij}/D_j)\}(\Delta D_\oplus - \Delta T_{\oplus\oplus}{}^{(2r-1)}) = 0, \end{aligned}$$

we get

$$\begin{aligned} &|\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}| = |\sum_j (T_{ij}/D_j)(\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)})| \\ &= \Big|\{\min_j (T_{ij}/D_j)\} \sum_j (\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}) \\ &\quad + \sum_j \{(T_{ij}/D_j) - \min_j (T_{ij}/D_j)\}(\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)})\Big| \\ &\le \sum_j \{(T_{ij}/D_j) - \min_j (T_{ij}/D_j)\}|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}|. \end{aligned} \tag{5.136}$$

But

$$\begin{aligned}(T_{ij}/D_j)|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)})| \\ = \{\min_j(T_{ij}/D_j)\}|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}| \\ + \{(T_{ij}/D_j) - \min_j(T_{ij}/D_j)\}|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}|.\end{aligned} \tag{5.137}$$

Hence, summing over j and using (5.136)

$$\begin{aligned}\sum_j(T_{ij}/D_j)|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}| - |\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}| \\ \geq \sum_j\{\min_j(T_{ij}/D_j)\}|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}| \geq 0.\end{aligned} \tag{5.138}$$

In particular, this implies the following result, which could also have been established directly from (5.135):

$$\begin{aligned}\sum_i|\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}| \leq \sum_{ij}(T_{ij}/D_j)|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}| \\ = \sum_j|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}|.\end{aligned} \tag{5.139}$$

A similar argument shows that

$$\sum_j|\Delta D_j - \Delta T_{\oplus j}{}^{(2r+1)}| \leq \sum_i|\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}|. \tag{5.140}$$

Because of (5.139) and (5.140), both $\sum_j|\Delta D_j - \Delta T_{\oplus j}{}^{(2r+1)}|$ and $\sum_i|\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}|$ are non-negative and non-increasing sequences which, therefore, must converge — and to the same limit L (say).

But it also follows from (5.138), on summing over i that

$$\begin{aligned}\sum_j|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}| - \sum_i|\Delta O_i - \Delta T_{i\oplus}{}^{(2r)}| \\ \geq \sum_i\{\min_{ij}(T_{ij}/D_j)\}\sum_j|\Delta D_j - \Delta T_{\oplus j}{}^{(2r-1)}|.\end{aligned}$$

Taking limits as $r \to \infty$, we see that L must be zero (notice that by assumption, $T_{ij} > 0$ for all i and j). Hence,

$$\Delta T_{\oplus j}{}^{(2r-1)} \to \Delta D_j \text{ for all } j \in J \tag{5.141}$$

and

$$\Delta T_{i\oplus}{}^{(2r)} \to \Delta O_i \text{ for all } i \in I. \tag{5.142}$$

That the sequence of $\Delta T_{ij}{}^{(s)}$'s as $s \to \infty$ is convergent then follows from the proofs of Lemma 5.6 and Corollary 5.3 (see the first part of its proof), since $\boldsymbol{\nu}_s - \boldsymbol{\nu} = \Phi[\boldsymbol{\mu}_s - \boldsymbol{\mu}] \to \mathbf{o}$ (where $\mathbf{o}$ is a null vector). That the limit has the requisite properties follows from (5.142), (5.141), (5.132) and (5.133). This proves the theorem. □

A NUMERICAL EXAMPLE

In order to illustrate the LDSF procedure, we need a set of T_{ij}'s. For the purpose of this example, we construct the T_{ij}'s in the following way: Set the matrix of F_{ij}'s to be

$$((F_{ij})) = \begin{pmatrix} 25 & 15 & 9 & 67 \\ 15 & 9 & 6 & 41 \\ 9 & 6 & 3 & 25 \\ 67 & 41 & 25 & 183 \end{pmatrix}, \tag{5.143}$$

set the O_i's to be 340, 862, 206 and 1765, and the D_j's to be 814, 370, 122 and 1867. The corresponding T_{ij}'s — for reasons that will be clear shortly, call them $T_{ij}(1)$'s — can be found using the DSF procedure. They are

$$\begin{pmatrix} 88.50 & 39.51 & 12.68 & 199.31 \\ 220.87 & 98.61 & 35.17 & 507.35 \\ 51.98 & 25.78 & 6.90 & 121.34 \\ 452.65 & 206.10 & 67.24 & 1039.00 \end{pmatrix}. \tag{5.144}$$

The increments

$$\Delta O_1 = 68, \Delta D_2 = 18, \Delta D_3 = 50 \text{ and } \Delta F_{44} = 36.15 \tag{5.145}$$

were then used in an LDSF procedure. After 3 half iterations, i.e., after obtaining $\Delta T_{ij}{}^{(3)}$, subsequent iterations yielded no further changes in the first four decimal places. As mentioned earlier, this is not at all unusual in applications of the DSF procedure. The matrix of $\Delta T_{ij}{}^{(3)}$'s is shown below:

$$\begin{pmatrix} 21.52 & 11.55 & 8.29 & 26.64 \\ 9.29 & 9.00 & 15.91 & -34.20 \\ 2.32 & 2.42 & 3.14 & -7.87 \\ -33.13 & -4.96 & 22.66 & 15.43 \end{pmatrix}.$$

These are the 'adjustments' to the $T_{ij}(1)$'s corresponding to the changes given by (5.145). When these adjustments are added to the $T_{ij}(1)$'s from (5.144) we get the adjusted T_{ij}'s.

The adjustments in the T_{ij}'s could be made in another way as well. We could make the changes (5.145) in the original F_{ij}'s, O_i's and D_j's, and run a DSF procedure. Call the adjusted T_{ij}'s obtained in this way $T_{ij}(2)$'s. Indeed, the whole purpose of the LDSF procedure is to provide an approximation to $T_{ij}(2) - T_{ij}(1)$.

The ratios

$$\frac{T_{ij}(1) + \Delta T_{ij}{}^{(3)}}{T_{ij}(2)}$$

for different ij's are presented below:

$$\begin{pmatrix} 1.00 & 0.99 & 0.94 & 1.01 \\ 1.01 & 1.01 & 1.00 & 0.99 \\ 1.01 & 1.01 & 1.00 & 0.99 \\ 0.99 & 1.00 & 1.02 & 1.00 \end{pmatrix}.$$

This shows that the LDSF procedure provides a remarkably good approximation to the DSF procedure, or more precisely, to the differences between the results obtained from two applications of the DSF procedure. Notice that the changes (5.145) are neither very small nor in any way arranged to favor the LDSF procedure.

A question that might arise at this stage is why have the LDSF procedure at all, when two applications of the DSF procedure gives the same — or better — results. The key advantages of the LDSF procedure are:

1. The $\Delta T_{ij}{}^{(s)}$'s given by the LDSF procedure are linear in the ΔO_1's, ΔD_j's and ΔF_{ij}'s, which makes it of value in computing covariances and the like.

2. It converges very rapidly, so that algebraic expressions can be constructed which combine the first few iterations and these expressions provide closed form approximations for $\Delta T_{ij}{}^{(s)}$'s and, hence, for $T_{ij}(2)$'s. These expressions can then stand in lieu of the DSF procedure within steps of algorithms where repeated applications of the DSF procedure are called for. Section 5.4.2 seeks to find just expressions for use in Section 5.5.

5.4.2 An Approximation Useful for ML Estimation Algorithms

We develop here a special case of the LDSF procedure that we shall use in Section 5.5 to develop procedures for maximum likelihood estimation of parameters of the general model given by (5.1) and (5.2). For this purpose we need, for a small change $\Delta\boldsymbol{\theta}$ in $\boldsymbol{\theta}$ and $\Delta\boldsymbol{O} = \Delta\boldsymbol{D} = \mathbf{o}$, an approximation for the corresponding small change ΔT_{ij} for each T_{ij}.

Let ΔF_{ij} be a small change in $F_{ij} = F(\boldsymbol{c}_{ij}) = \exp(\sum_{k=1}^{K} c_{ij}{}^{(k)}\theta_k)$ due to changing θ_k's to $(\theta_k + \Delta\theta_k)$'s. Then

$$\Delta F_{ij} \approx \sum_{k=1}^{K} \frac{\partial F_{ij}}{\partial \theta_k}\Delta\theta_k = \sum_{k=1}^{K} c_{ij}^{(k)} F_{ij}\Delta\theta_k, \tag{5.146}$$

and, therefore, from (5.113), we have

$$\Delta T_{ij}^{(0)} = \Delta F_{ij} T_{ij} / F_{ij} \approx T_{ij} \sum_{k=1}^{K} (c_{ij}^{(k)} \Delta\theta_k) \quad \text{for all } i \in I \text{ and } j \in J \tag{5.147}$$

and since $\Delta\boldsymbol{O} = \Delta\boldsymbol{D} = \mathbf{o}$, the first two LDSF steps are (from (5.114) and (5.115)),

$$\Delta T_{ij}^{(1)} = \Delta T_{ij}^{(0)} - \Delta T_{i\oplus}^{(0)} \left(\frac{T_{ij}}{O_i}\right) \tag{5.148}$$

and

$$\Delta T_{ij}^{(2)} = \Delta T_{ij}^{(1)} - \Delta T_{\oplus j}^{(1)} \left(\frac{T_{ij}}{D_j}\right). \tag{5.149}$$

Recall that we mentioned earlier that quite often one or two pairs of LDSF iterations get us very close to the limit for the procedure. In fact, two steps are often adequate for a reasonably good approximation. Therefore, combining (5.148) and (5.149), we get

$$\begin{aligned}
\Delta T_{ij} &\approx \Delta T_{ij}^{(2)} \\
&= \Delta T_{ij}^{(0)} - \Delta T_{i\oplus}^{(0)} \left(\frac{T_{ij}}{O_i}\right) \\
&\quad - \left[\Delta T_{\oplus j}^{(0)} - \sum_i (\Delta T_{i\oplus}^{(0)} \left(\frac{T_{ij}}{O_i}\right))\right] \left(\frac{T_{ij}}{D_j}\right) \\
&= \Delta T_{ij}^{(0)} - \Delta T_{i\oplus}^{(0)} \left(\frac{T_{ij}}{O_i}\right) \\
&\quad - \Delta T_{\oplus j}^{(0)} \left(\frac{T_{ij}}{D_j}\right) + \sum_i \left[\Delta T_{i\oplus}^{(0)} \left(\frac{T_{ij}}{O_i}\right)\right] \left(\frac{T_{ij}}{D_j}\right).
\end{aligned} \tag{5.150}$$

Using (5.147), it follows that

$$\begin{aligned}
\Delta T_{ij} &\approx T_{ij} \sum_k (c_{ij}^{(k)} \Delta\theta_k) - \sum_j \left[T_{ij} \sum_k (c_{ij}^{(k)} \Delta\theta_k)\right] \left(\frac{T_{ij}}{O_i}\right) \\
&\quad - \sum_i \left[T_{ij} \sum_k (c_{ij}^{(k)} \Delta\theta_k)\right] \left(\frac{T_{ij}}{D_j}\right) \\
&\quad + \sum_i \left[\sum_j \left[T_{ij} \sum_k (c_{ij}^{(k)} \Delta\theta_k)\right] \left(\frac{T_{ij}}{O_i}\right)\right] \left(\frac{T_{ij}}{D_j}\right).
\end{aligned} \tag{5.151}$$

Therefore, adding together all the coefficients for each $\Delta\theta_k$ on the right side of the above approximation, we get

$$\begin{aligned}\Delta T_{ij} \approx \sum_{k=1}^{K} \Bigg[\Delta\theta_k \Bigg\{ & c_{ij}^{(k)} T_{ij} - \sum_j \left[c_{ij}^{(k)} T_{ij} \right] \left(\frac{T_{ij}}{O_i} \right) \\ & - \sum_i \left[c_{ij}^{(k)} T_{ij} \right] \left(\frac{T_{ij}}{D_j} \right) \\ & + \sum_i \left[\sum_j \left[c_{ij}^{(k)} T_{ij} \right] \left(\frac{T_{ij}}{O_i} \right) \right] \left(\frac{T_{ij}}{D_j} \right) \Bigg\} \Bigg] \\ = \sum_{k=1}^{K} \left[S_{ij}^{(k)} \Delta\theta_k \right], &\end{aligned} \tag{5.152}$$

where

$$\begin{aligned} S_{ij}^{(k)} = & c_{ij}^{(k)} T_{ij} - \sum_j [c_{ij}^{(k)} T_{ij}] \left(\frac{T_{ij}}{O_i} \right) - \sum_i [c_{ij}^{(k)} T_{ij}] \left(\frac{T_{ij}}{D_j} \right) \\ & + \sum_i \left[\sum_j [c_{ij}^{(k)} T_{ij}] \left(\frac{T_{ij}}{O_i} \right) \right] \left(\frac{T_{ij}}{D_j} \right) \end{aligned} \tag{5.153}$$

for all $i \in I$, $j \in J$ and $k \in K$. The $S_{ij}^{(k)}$'s do not include quantities with Δ's in them and $O_i = T_{i\oplus}$ and $D_j = T_{\oplus j}$. Therefore, if T_{ij}'s are known, the only unknowns in the last expression in (5.152) are the $\Delta\theta_k$'s.

We shall use (5.152) and (5.153) repeatedly in Section 5.5.

5.4.3 Application to Short-term Forecasting

The LDSF procedure has several applications outside of developing algorithms for ML estimates. One such application is in making short term forecasts. Let ${N_{ij}}^{(f)}$ be a future observation of the flow from i to j. A future observation is a random variable representing an observation which will occur in the future. If the future is not too far off, ${N_{ij}}^{(f)}$ and N_{ij} will usually be highly (serially) correlated. For example, in the case of commuter trips, both N_{ij} and ${N_{ij}}^{(f)}$ will be counts of mostly the same people going to the same jobs. Then

$$\operatorname{var}({N_{ij}}^{(f)} - N_{ij}) = \operatorname{var}({N_{ij}}^{(f)}) + \operatorname{var}(N_{ij}) - 2\operatorname{Cov}({N_{ij}}^{(f)} N_{ij}), \tag{5.154}$$

will be small and could be much smaller than $\operatorname{var}({N_{ij}}^{(f)})$, where $\operatorname{var}(\cdot)$ stands for 'the variance of' and $\operatorname{Cov}(\cdot)$ for 'the covariance of.' In such cases, especially if N_{ij}'s are known it would be preferable to predict $({N_{ij}}^{(f)} - N_{ij})$'s and add the predictions to the N_{ij}'s. A possible set of predictions for $({N_{ij}}^{(f)} - N_{ij})$'s are $({\hat{T}_{ij}}{}^{(f)} - \hat{T}_{ij})$'s. Thus, if the future is not too far off,

$$N_{ij} + ({\hat{T}_{ij}}{}^{(f)} - \hat{T}_{ij}), \tag{5.155}$$

may yield a better prediction than $\hat{T}_{ij}$, where the superscript f on $\hat{T}_{ij}$, like that on N_{ij}, refers to the future. The computation of $\Delta\hat{T}_{ij} = \hat{T}_{ij}{}^{(f)} - \hat{T}_{ij}$ can be made easily using the LDSF Procedure.

Although this discussion is being presented as describing a possible LDSF procedure application, the key point, that (5.155) might be a better short-term forecast than $\hat{T}_{ij}{}^{(f)}$, is independent of how $(\hat{T}_{ij}{}^{(f)} - \hat{T}_{ij})$'s are calculated. However, the use of the LDSF procedure leads to simpler calculations and would be particularly useful when forecasts need to be made for a number of different sets of changes in O_i's, D_j's and F_{ij}'s.

5.5 General Algorithms for ML Estimates

As mentioned earlier, ML estimates can be obtained in several ways, including

1. Solving (5.12) and (5.13). In Sections 5.5.1 and 5.5.2, we shall pursue this approach.

2. Directly maximizing (5.8), by some 'hill-climbing procedure', i.e., a mathematical programming procedure which seeks to increase the value of (5.8) at every iteration. An example of such a procedure is the gradient search procedure. Several variations of the gradient search procedure will be treated in Sections 5.5.3 and 5.5.4.

3. Using procedures which consist of repeated use of least squares. Examples include procedures of the Gauss-Newton family and the GLIM procedure. The latter is commonly recommended for ML estimation of a wide class of models — a class which includes the gravity model as a special case. Section 5.5.5 is devoted to a brief description of the GLIM procedure.

4. Computing values of the likelihood function, or preferably, its logarithm $\mathcal{L}$, for a large number of $\boldsymbol{\theta}$'s and picking the value of $\boldsymbol{\theta}$ corresponding to the largest value of $\mathcal{L}$. Although other designs are sometimes used, frequently, the values of $\boldsymbol{\theta}$ examined are selected over a uniform grid; i.e., uniformly spaced values of each component of $\boldsymbol{\theta}$ around some previously chosen value of $\boldsymbol{\theta}$ [e.g., one obtained by least squares — see Chapter 5]. This method is usually quite time consuming and, consequently, is not discussed further in this book [Hathaway (1975) has given an example of such an approach].

Although we have made the distinctions we have, there is an inherent similarity between these different approaches — see Judge *et al.* (1985, Appendix B).

In the next section — Section 5.6 — the performance of these different procedures will be compared. For readers who prefer not to read the entire section, we mention that the clear winner of this comparison is a procedure we call the modified scoring procedure, described in Section 5.5.2. In it the LDSF procedure is used to obtain algorithms for the solution of (5.12) and (5.13).

As mentioned in the introduction to this chapter, and for the reasons given there, in this section as well as in the next, we shall focus on the estimation of $\boldsymbol{\theta}$.

5.5.1 Scoring Methods

One method of obtaining ML estimates is by solving (5.12) and (5.13). However, these equations, which are linear in T_{ij}'s are nonlinear in the parameters, which are the $A(i)$'s, $B(j)$'s, and θ_k's. Systems of nonlinear equations frequently arise in ML estimation and a method that has been recommended for solving them is the *method of scoring*. The method is essentially a Newton-Raphson method; its statistical name refers to the fact that the partial derivatives of the log likelihood function $\mathcal{L}$ with respect to the parameters are called efficient scores and the method of scoring makes these scores vanish.

In the method of scoring, initial values of parameters are assumed. Then these initial values are augmented by the addition of increments and then, using a Taylor series approximation, the likelihood equations (e.g., (5.12), (5.13)) are approximated by linear functions of these increments. Then the resultant equations are solved for the increments. The initial choice of parameter values are augmented by the values of the increments thus found, and these sums are the starting point for the next iteration.

However, this process is time consuming. In several practical situations, the unmodified use of this process to solve (5.12) and (5.13) could be difficult to shoe-horn within the space limitations of even the largest computers, since we have $I + J + K$ equations in as many unknowns and I and J can be very large. However, since for a given and fixed $\boldsymbol{\theta}$, we can readily solve (5.12) using the DSF procedure, we can use a procedure akin to Hyman's (Section 5.3.3). Such a procedure is described below. Another approach, which also uses the DSF procedure to solve (5.12), will be presented in Section 5.5.2.

Extension of Hyman Approach

The Hyman Procedure can readily be extended to $K > 1$. For all $k \in K$ define

$$C_k(\boldsymbol{\theta}) = \sum_{ij} c_{ij}{}^{(k)}\hat{T}_{ij}(\boldsymbol{\theta}) \quad \text{and} \quad C_k = \sum_{ij} c_{ij}{}^{(k)}N_{ij}, \tag{5.156}$$

where $\hat{T}_{ij}(\boldsymbol{\theta})$'s are the T_{ij}'s such that (5.12) is solved. Such $\hat{T}_{ij}(\boldsymbol{\theta})$'s may be obtained from the DSF procedure, but for the present, we just treat the $C_k(\boldsymbol{\theta})$'s as functions with the appropriate properties. At every step, $r = 1, 2, \cdots$, assume that the current value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(r)}$ obtained from a previous iteration or, when $r = 0$, as a given initial value. We can express $C_k(\boldsymbol{\theta}^{(r)} + \Delta\boldsymbol{\theta}^{(r)})$ as a power series in $\Delta\boldsymbol{\theta}^{(r)}$, and after deleting all but linear terms from this series we have left

$$C_k(\boldsymbol{\theta}^{(r)} + \Delta\boldsymbol{\theta}^{(r)}) \approx C_k(\boldsymbol{\theta}^{(r)}) + \Delta\boldsymbol{\theta}^{(r)t}\frac{\partial C_k(\boldsymbol{\theta}^{(r)})}{\partial \Delta\boldsymbol{\theta}} = P_k(\boldsymbol{\theta}^{(r)},\, \Delta\boldsymbol{\theta}^{(r)}) \text{ (say).}$$

We can solve the system of K equations typified by

$$P_k(\boldsymbol{\theta}^{(r)}, \Delta\boldsymbol{\theta}^{(r)}) = C_k \tag{5.157}$$

to obtain $\Delta\boldsymbol{\theta}^{(r)}$ and set $\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \Delta\boldsymbol{\theta}^{(r)}$.

The partial derivatives $\partial C_k(\boldsymbol{\theta})/\partial\theta_l : k, l = 1, 2, \ldots, K$, which are needed in the computation of $P_k(\boldsymbol{\theta}^{(r)},\, \Delta\boldsymbol{\theta}^{(r)})$, can be obtained numerically from $\boldsymbol{\theta}^{(r)}, \ldots, \boldsymbol{\theta}^{(r-K)}$ by solving, for each k, the system of approximations

$$\begin{aligned} &C_k(\boldsymbol{\theta}^{(r)}) - C_k(\boldsymbol{\theta}^{(r-s)}) \\ &\qquad \approx (\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-s)})^t \left(\frac{\partial C_k(\boldsymbol{\theta}^{(r)})}{\partial\theta_1}, \ldots, \frac{\partial C_k(\boldsymbol{\theta}^{(r)})}{\partial\theta_k}\right)^t, \end{aligned} \tag{5.158}$$

where $s = 1, \ldots, K$. Notice that, since there are K functions $C_k(\boldsymbol{\theta})$'s and K parameters θ_k's, at every step we would be solving K systems of K equations each — one system for each set $\partial C_k(\boldsymbol{\theta}^{(r)})/\partial\theta_p : k = 1, \ldots, K$. Also notice that for the first iteration, we would need $K + 1$ initial values $\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(-K)}$.

However, unlike Hyman's Procedure, its $K \geq 2$ dimensional analog does not perform very well. The values of $\boldsymbol{\theta}^{(r)}$ often jump around—on occasion even causing 'underflows' or 'overflows' in the computer. This problem with the method of scoring has been noted by Sen (1986) and previously by Batty (1976) who has compared the performance of several procedures. While Batty restricted himself to the special case when $K = 2$, his procedures can be readily generalized to $K > 2$.

A further modification of the method of scoring is described below.

5.5.2 The Modified Scoring Procedure

In the method which we shall call the *modified scoring* procedure [and which will emerge in the next section as the best procedure], we proceed by choosing an initial $\boldsymbol{\theta}$ and then obtaining $T_{ij}(\boldsymbol{\theta})$'s by using the DSF procedure to solve (5.12). We then change θ_k's to $\theta_k + \Delta\theta_k$ and compute, using the LDSF procedure, with $\Delta O = \mathbf{o}$ and $\Delta D = \mathbf{o}$, the corresponding change $\Delta T_{ij}(\boldsymbol{\theta}, \Delta\boldsymbol{\theta})$'s in $T_{ij}(\boldsymbol{\theta})$'s as functions of $\Delta\theta_k$'s. Finally, we insert the $(T_{ij}(\boldsymbol{\theta}) + \Delta T_{ij}(\boldsymbol{\theta}, \Delta\boldsymbol{\theta}))$'s into the left side of (5.13) and then solve the resultant equations for $\Delta\theta_k$'s. For the next iteration, θ_k is set equal to

$\theta_k + \Delta\theta_k$. The use of the LDSF procedure assures us that the $(T_{ij} + \Delta T_{ij})$'s will approximately solve (5.12), while we approach a solution to (5.13).

We now describe in detail a single iteration of the modified scoring procedure. Since we confine our attention to a single iteration for the moment, no confusion need arise if we set $T_{ij} = T_{ij}(\boldsymbol{\theta}^{(r)})$ and $\Delta T_{ij} = \Delta T_{ij}(\boldsymbol{\theta}^{(r)}, \Delta\boldsymbol{\theta}^{(r)})$. Inserting $T_{ij} + \Delta T_{ij}$ in place of T_{ij} in (5.13), we get

$$\sum_{ij} c_{ij}^{(k)}(T_{ij} + \Delta T_{ij}) = \sum_{ij} c_{ij}^{(k)} N_{ij} \quad \text{for all } k \in K. \tag{5.159}$$

If these ΔT_{ij}'s are computed using (5.152) of Section 5.4.2, then (5.12) would remain approximately satisfied, while (5.159) would become, approximately,

$$\begin{aligned}
\sum_{ij} c_{ij}^{(1)} S_{ij}^{(1)} \Delta\theta_1 + \ldots + \sum_{ij} c_{ij}^{(1)} S_{ij}^{(K)} \Delta\theta_K &= \sum_{ij} c_{ij}^{(1)} (N_{ij} - T_{ij}) \\
\sum_{ij} c_{ij}^{(2)} S_{ij}^{(1)} \Delta\theta_1 + \ldots + \sum_{ij} c_{ij}^{(2)} S_{ij}^{(K)} \Delta\theta_K &= \sum_{ij} c_{ij}^{(2)} (N_{ij} - T_{ij}) \\
\cdots\cdots\cdots\cdots\cdots\cdots &= \cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots &= \cdots\cdots \\
\sum_{ij} c_{ij}^{(K)} S_{ij}^{(1)} \Delta\theta_1 + \ldots + \sum_{ij} c_{ij}^{(K)} S_{ij}^{(K)} \Delta\theta_K &= \sum_{ij} c_{ij}^{(K)} (N_{ij} - T_{ij}).
\end{aligned} \tag{5.160}$$

A solution of (5.160) for $\Delta\boldsymbol{\theta}$ would yield $(T_{ij}+\Delta T_{ij})$'s which typically would come closer to solving (5.159) and hence (5.13). With $\Delta\boldsymbol{\theta}$ thus computed, $(\boldsymbol{\theta} + \Delta\boldsymbol{\theta})$ would replace $\boldsymbol{\theta}$ for the next iteration.

Notice that (5.160) gives a system of K linear equations in K unknowns ($\Delta\theta_k$'s) which can be solved by any of the standard solution methods (e.g., Gaussian elimination). Notice also that in practice K is frequently quite small (e.g., 2, 3 or 4).

The steps of the modified scoring procedure may be summarized as follows:

1. Select an initial value $\boldsymbol{\theta}^{(0)}$.

2. At iteration r, using the DSF procedure (with $O_i = N_{i\oplus}$, and $D_j = N_{\oplus j}$), obtain T_{ij} $(= T_{ij}(\boldsymbol{\theta}^{(r)}))$. Compute the corresponding coefficients $S_{ij}^{(k)}$'s (from (5.153)) and solve (5.160) for $\Delta\theta_k$'s.

3. Revise the value of $\boldsymbol{\theta}^{(r)}$ using:

$$\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(r-1)} + \Delta\boldsymbol{\theta}^{(r-1)}. \tag{5.161}$$

4. Iterate steps 2 and 3 until a suitable stopping criterion (see below) is satisfied.

Initial Values and Stopping Criteria

There is no guarantee that this procedure always converges. However, it seems to do so for reasonable choices of initial values. In a very large number of runs of the modified scoring procedure, we only failed to get convergence when we deliberately set out to do so — by using for example large positive coefficients for travel time and distance. (If the user has no idea what initial $\boldsymbol{\theta}$ to use, some procedure, such as the least squares procedure of Chapter 6, could be useful.)

As mentioned earlier, stopping rules are basically of two types: those that stop iterations when we get sufficiently close to the desired result and those that stop when improvements at each iteration become small enough. A criterion of the first type is the following: Since we are trying to solve (5.12) and (5.13), one stopping rule could be to stop iterations when (5.9), (5.10) and (5.11) are all small enough. This is our preferred choice.

Examples of the second type of stopping rule include:

1. Stop when changes in parameter estimates from iteration to iteration are smaller than a predefined criterion.

2. Stop when changes in the scaled deviance (see Section 5.5.5) from iteration to iteration are small enough,

5.5.3 Gradient Search Procedures

Gradient search procedures are among a class of 'hill-climbing' procedures that are used to maximize a function without solving the equations resulting from setting the derivatives of the function equal to zero. Other examples of hill-climbing procedures have been given in Judge *et al.* (1985, Appendix B, see also Batty, 1976, Batty and Mackie (1972), Putman and Ducca (1978a), Sen and Srivastava, 1990, Appendix C). Examples include Fibonacci search and the simplex method.

The gradient of a function $G(x_1, \ldots, x_n)$ is the vector

$$\nabla G = \left(\frac{\partial G}{\partial x_1}, \ldots, \frac{\partial G}{\partial x_n}\right).$$

It is known that the gradient is the direction at $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ along which the rate of increase of G is maximized. Thus, in the gradient search procedure, which is iterative, if $\boldsymbol{x}^{(r)} = (x_1{}^{(r)}, \ldots, x_n{}^{(r)})^t$ is the current estimate of $\boldsymbol{x}$, the value of

$$G(\boldsymbol{x}^{(r)} + \rho^{(r)} \nabla G(\boldsymbol{x}^{(r)})),$$

where $\nabla G(\boldsymbol{x}^{(r)})$ indicates that the derivatives are being taken at $\boldsymbol{x}^{(r)}$, would be higher than $G(\boldsymbol{x}^{(r)})$, if $\rho^{(r)}$ is small enough. Clearly, small movements in directions other than along the gradient will also cause G to increase; in fact, any direction $P\nabla G$ where P is a positive definite matrix will have this property and different choices of P lead to different hill-climbing algorithms.

In many situations, including that involving the log-likelihood function of the gravity model, computation of the gradient vector is relatively easy. However, the computation of the step size $\rho^{(r)}$ presents some difficulties. If it is chosen too small, the procedure converges slowly. If it is too large, $G(\boldsymbol{x}^{(r)} + \rho^{(r)}\nabla G(\boldsymbol{x}^{(r)}))$ may not be much larger than $G(\boldsymbol{x}^{(r)})$ and may even be smaller. One approach is to choose $\rho^{(r)}$ to maximize $G(\boldsymbol{x}^{(r)} + \rho^{(r)}\nabla G(\boldsymbol{x}^{(r)}))$. Frequently, this is done by using a golden section or bisection method. However, it is generally preferable to use an algebraic formula for this purpose, even if it does not quite give a maximum value of G.

In the case of $\mathcal{L}$ given by (5.8), we have $I + J + K$ parameters and, as already mentioned, I and J can each be very large. Therefore, the use of the usual gradient search procedure is not very efficient, as has been verified by a student of one of the authors. Batty (1976) also evaluated this procedure.

Therefore, Sen (1986, see also Sen and Matuszewski, 1991) introduced a version of the gradient search procedure which we refer to as the 'general procedure'. This procedure was claimed by Sen to be the fastest available procedure which could handle large O-D matrices. We shall use it as a benchmark for comparisons in Section 5.6.3 and also modify it using the LDSF procedure in the next subsection. We state the procedure below. After that, in a separate subsection, we provide a motivation for the way the step size was chosen.

The 'General Procedure'

After choosing an initial value $\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$, the method consists of computing $\boldsymbol{\theta}^{(r+1)}$ from $\boldsymbol{\theta}^{(r)}$ as follows:

1. Apply the DSF Procedure with $\boldsymbol{\theta} = \boldsymbol{\theta}^{(r)}$, $O_i = N_{i\oplus}$ and $D_j = N_{\oplus j}$; and, as before, designate by $T_{ij}(\boldsymbol{\theta}^{(r)})$ each T_{ij} that result. Notice that since $T_{ij}(\boldsymbol{\theta}^{(r)})$'s are computed using the DSF Procedure, (5.12) is satisfied (actually only approximately so, since in practice only a finite number of iterations are conducted—but this very slight approximation of language need cause no confusion in the material to follow). Consequently, from (5.9) and (5.10), for all $i \in I$ and $j \in J$.

$$\partial\mathcal{L}/\partial A(i) = \partial\mathcal{L}/\partial B(j) = 0, \tag{5.162}$$

at the point $(\boldsymbol{A}^{(r)}, \boldsymbol{B}^{(r)}, \boldsymbol{\theta}^{(r)})$ where

$$T_{ij}(\boldsymbol{\theta}^{(r)}) = A^{(r)}(i)B^{(r)}(j)\exp[\boldsymbol{\theta}^{(r)^t}\boldsymbol{c}_{ij}]. \tag{5.163}$$

It follows that at this point, $\nabla\mathcal{L} = (0, \ldots, 0, \nabla_{\boldsymbol{\theta}})^t$ where $\mathcal{L}$ is the log-likelihood function given by (5.8), $\nabla_{\boldsymbol{\theta}} = (\nu_1^{(r)} \ldots, \nu_K{}^{(r)})^t$ and (see (5.11))

$$\nu_k^{(r)} = \sum_{ij} c_{ij}{}^{(k)}[N_{ij} - T_{ij}(\boldsymbol{\theta}^{(r)})]. \tag{5.164}$$

2. Compute

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \rho^{(r)}(\nu_1^{(r)}, \ldots, \nu_K^{(r)})^t \tag{5.165}$$

where $\rho^{(r)}$ is a solution for ρ in the equation

$$\begin{aligned}\sum_{k=1}^{K}\Big(&\sum_{ij} c_{ij}{}^{(k)}N_{ij}\\ &- \tau\sum_{ij}[c_{ij}{}^{(k)}T_{ij}(\boldsymbol{\theta}^{(r)})\exp(\sum_{l=1}^{K}\rho c_{ij}{}^{(l)}\nu_l{}^{(r)})]\Big)\nu_k{}^{(r)} = 0\end{aligned} \tag{5.166}$$

with

$$\tau^{-1} = \sum_{ij} T_{ij}(\boldsymbol{\theta}^{(r)})\exp(\sum_{l=1}^{K}\rho c_{ij}{}^{(l)}\nu_l{}^{(r)})/N_{\oplus\oplus}. \tag{5.167}$$

The iterations are continued until a suitable stopping criterion is met (e.g., the $\nu_k{}^{(r)}$ are small enough).

In our experience this procedure has always yielded reasonably good estimates. However, when the vectors

$$(c^{(k)}{}_{11}, \ldots, c^{(k)}{}_{1J}, \ldots\ldots, c^{(k)}{}_{I1}, \ldots, c^{(k)}{}_{IJ})^t : k = 1, \ldots, K \tag{5.168}$$

are close to linear dependency, even after several iterations, the $\boldsymbol{\theta}^{(r)}$'s tend to jump around — but any of these $\boldsymbol{\theta}^{(r)}$'s yield very similar sets of $\hat{T}_{ij}(\boldsymbol{\theta})$'s. A shortcoming this procedure shares with most procedures based on gradient search is that, after moving rapidly in earlier iterations, when values of $\boldsymbol{\theta}^{(r)}$ get close to the limit, the rate of progress sometimes gets to be excruciatingly slow.

Motivation for choice of Step Size

Because of the use of the DSF procedure, $T_{\oplus\oplus}(\boldsymbol{\theta}^{(r)}) = N_{\oplus\oplus}$ and hence, from (5.166) and (5.167), we see that if $\rho = 0$ solves (5.166), then $\tau = 1$ and each $\nu_k{}^{(r)} = 0$ [since (5.166) becomes $\sum_{k=1}^{K}(\nu_k{}^{(r)})^2 = 0$]. Therefore, convergence can only occur when this condition holds (see also (5.165) and (5.164)), i.e., when we reach a maximum of $\mathcal{L}$. Note also that under Conditions ML1 and ML2 of Theorem 5.1, $\mathcal{L}$ is bounded. Therefore, if each iteration resulted in an increase in $\mathcal{L}$, the iterations would converge.

Unfortunately, since it is not clear that $\rho^{(r)}$ always gets chosen in such a way that $\mathcal{L}$ increases when $\boldsymbol{\theta}^{(r)}$ is incremented by $\rho^{(r)}\boldsymbol{\nu}^{(r)}$, it is not known whether the sequence $\boldsymbol{\theta}^{(r)}$: $r = 1, 2, \cdots$ will always converge.

Therefore, the recommendation made by Sen (1986) was based essentially on the fact that the procedure 'works.' Nevertheless, it seems appropriate to describe how (5.166) was chosen. It is the result of 'tweaking' another procedure which can be shown to converge but is computationally inefficient. We describe this procedure below and to avoid confusion with the general procedure described above, we call it Procedure G1.

The only difference between Procedure G1 and the general procedure is in the computation of step size. Instead of (5.165), we have

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \tilde{\rho}^{(r)}\boldsymbol{\nu}^{(r)}, \tag{5.169}$$

where the variables are the same as for (5.165) except for $\tilde{\rho}^{(r)}$, which is a solution for ρ in

$$\begin{aligned}\delta^{(r)}(\rho) = \sum_{k=1}^{K}\Big(&\sum_{ij} c_{ij}{}^{(k)} N_{ij}\\ &- \sum_{ij}[c_{ij}{}^{(k)} T_{ij}(\boldsymbol{\theta}^{(r)}) \exp(\sum_{l=1}^{K} \rho c_{ij}{}^{(l)} \nu_l{}^{(r)})]\Big)\nu_k{}^{(r)} = 0\end{aligned} \tag{5.170}$$

Equation (5.170) is the same as (5.166) with τ set equal to one.

For procedure G1, the equation (5.170) which defines step size was obtained in the following way. Because of (5.162), since $\mathcal{L}$ is a function of the θ_k's, $A(i)$'s and $B(j)$'s

$$d\mathcal{L}/d\rho = \sum_{k=1}^{K}(\partial\mathcal{L}/\partial\theta_k)(d\theta_k/d\rho), \tag{5.171}$$

and at the point $(\boldsymbol{A}^{(r)}, \boldsymbol{B}^{(r)}, \boldsymbol{\theta}^{(r)} + \rho\boldsymbol{\nu}^{(r)})$, this becomes, on using (5.11) and (5.165),

$$\begin{aligned}d\mathcal{L}/d\rho = \sum_{k=1}^{K}[C_k - &\sum_{ij} c_{ij}{}^{(k)} A^{(r)}(i) B^{(r)}(j)\\ &\cdot \exp[\sum_{l=1}^{K} c_{ij}{}^{(l)}(\theta_l{}^{(r)} + \rho\nu_l{}^{(r)})]]\nu_k{}^{(r)} = \delta^{(r)}(\rho),\end{aligned} \tag{5.172}$$

which is the same as (5.170). Thus, a solution to (5.170), yields the maximum value of $\mathcal{L}$ for all $\boldsymbol{\theta}^{(r+1)}$ given by (5.169).

It is easy to see from (5.172) and (5.164) that

$$\delta^{(r)}(0) = (\boldsymbol{\nu}^{(r)})^t(\boldsymbol{\nu}^{(r)}) \geq 0, \tag{5.173}$$

and a little more tedious to verify that

$$
\begin{aligned}
& d\delta^{(r)}(\rho)/d\rho \\
&= -\sum_{k=1}^{K}\sum_{ij} \nu_k{}^{(r)} c_{ij}{}^{(k)} T_{ij}(\boldsymbol{\theta}^{(r)}) \sum_{l=1}^{K} c_{ij}{}^{(l)} \nu_l{}^{(r)} \exp(\sum_{s=1}^{K} \rho c_{ij}{}^{(s)} \nu_s{}^{(r)}) \\
&= -\sum_{ij} (T_{ij}(\boldsymbol{\theta}^{(r)}) \exp(\sum_{s=1}^{K} \rho c_{ij}{}^{(s)} \nu_s{}^{(r)})(\sum_{l=1}^{K} c_{ij}{}^{(l)} \nu_l{}^{(r)})^2) \leq 0.
\end{aligned}
\tag{5.174}
$$

Therefore, $\delta^{(r)}(\rho)$ is a decreasing function of ρ but with a positive value at $\rho = 0$. Hence, there must be a unique value $\tilde{\rho} > 0$ of ρ along the gradient vector for which $\mathcal{L}$ is maximized, and this occurs when $\delta^{(r)}(\rho) = 0$.

Lemma 5.7 *Under the conditions of Theorem 5.1, the successive values* $\boldsymbol{\theta}^{(r)} : r = 1, 2, \cdots$ *converge to the unique ML estimate of* $\boldsymbol{\theta}$.

PROOF: After $\boldsymbol{\theta}^{(r+1)}$ has been obtained, the DSF Procedure is applied to obtain $\boldsymbol{T}(\boldsymbol{\theta}^{(r+1)})$. Since in the DSF Procedure, $\boldsymbol{\theta}$ is kept fixed at $\boldsymbol{\theta}^{(r+1)}$, $\mathcal{L}$ attains its maximum when (5.12) holds; therefore, the application of the DSF Procedure also raises $\mathcal{L}$. Thus $\mathcal{L}$ is monotonically increasing as r increases. Under Conditions ML1 and ML2 of Theorem 5.1, it is bounded above, and consequently, the sequence of values of $\mathcal{L}$ as $r = 1, 2, \cdots$, converges. From (5.172) and (5.173) it can be shown that this can only happen when $\boldsymbol{\nu}^{(r)} \rightarrow 0$. Therefore, $\boldsymbol{\theta}^{(r)}$'s converge to the unique value $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ which is the ML estimate of $\boldsymbol{\theta}$. □

While convergence can be established for this procedure, the unfortunate practical fact is that the values of $\tilde{\rho}^{(r)}$ obtained at each step are far too small, and consequently far too many steps are required. This is because we ignored (5.12) while we searched for a suitable $\tilde{\rho}^{(r)}$ and these conditions were often severely violated for larger values of $\tilde{\rho}^{(r)}$. The solution seemed to lie in 'adjusting' (5.170). Since, at that time, it was difficult to incorporate two sets of constraints into (5.170) [although, this essentially has been done in Section 5.5.4], we proceeded as follows.

Notice that $T_{ij}(\boldsymbol{\theta}^{(r)}) \exp(\sum_{l=1}^{K} \rho c_{ij}{}^{(l)} \nu_l{}^{(r)})$ which appears in both (5.166) and (5.167) is simply the value of T_{ij} for each value of ρ. Call it $T_{ij}{}^{(r)}(\rho)$. τ is simply $N_{\oplus\oplus}/T_{\oplus\oplus}{}^{(r)}(\rho)$. It scales each $T_{ij}{}^{(r)}(\rho)$. Its purpose is to prevent $T_{\oplus\oplus}{}^{(r)}(\rho)$'s from getting too far away from $N_{\oplus\oplus}$ and perhaps keeping $T_{i\oplus}{}^{(r)}(\rho)$ close to $N_{i\oplus}$ and $T_{\oplus j}{}^{(r)}(\rho)$ close to $N_{\oplus j}$ and thus protecting the next use of the DSF procedure from having to make very major changes.

We can combine both Procedure G1 and the general procedure into a single procedure, which would be reasonably fast and would have desirable theoretical properties. Such a procedure is described in the course of the statement of Theorem 5.5 below, the proof of which follows immediately from Lemma 5.7.

Theorem 5.5 *Consider the iterations given by (5.165), where now $\rho^{(r)}$ is either a solution to (5.166) or to (5.170) — whichever makes the value of $\mathcal{L}$ computed for $\boldsymbol{T}(\boldsymbol{\theta}^{(r+1)})$ the larger. Then, under the conditions of Theorem 5.1, $\boldsymbol{\theta}^{(r)} : r = 1, 2, \cdots$ converges to the unique ML estimate of $\boldsymbol{\theta}$.*

In applications, the step sizes obtained from this combined procedure were always exactly the same as those of the general procedure. Therefore, in the computations shown in Section 5.6, computer time was saved by not computing $\tilde{\rho}$ at all and simply using the general procedure.

5.5.4 Modified Gradient Search Procedures

As mentioned in the last section, in gradient search procedures, finding the direction $\nabla\mathcal{L}$ is easy. The difficulties lie in computing how far we should move along this direction before we stop and initiate the next iteration. In (5.171), (5.172) and (5.170), we essentially computed this step size ρ to make $d\mathcal{L}/d\rho = 0$. In computing this derivative, we held the $A(i)$'s and $B(j)$'s constant with respect to ρ.

However, if we wish (5.12) to continue to hold as we moved along the gradient, it may be desirable to allow the $A(i)$'s and $B(j)$'s to change so that $\sum_i \Delta T_{ij}(\boldsymbol{\theta}^{(r)}, \rho) \approx 0$ and $\sum_j \Delta T_{ij}(\boldsymbol{\theta}^{(r)}, \rho) \approx 0$, where $T_{ij}(\boldsymbol{\theta}^{(r)} + \rho\boldsymbol{\nu}^{(r)}) = T_{ij}(\boldsymbol{\theta}^{(r)}) + \Delta T_{ij}(\boldsymbol{\theta}^{(r)}, \rho)$. This is exactly what (5.152) was designed for. Its application would assure that the entire process remains close to the subspace described by (5.12), which is what the original intent of the general procedure was.

Here as in the last section, the iterations are given by

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \Delta\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(r)} + \rho^{(r)}(\nu_1^{(r)}, \ldots, \nu_K^{(r)})^t.$$

The modifications are in the way the $\rho^{(r)}$'s are calculated. We discuss this below and then present a step by step description of the whole procedure.

Since, in the following discussion we consider a single iteration r, as in Section 5.5.2, some simplicity in notation is achieved if we temporarily drop the r from our notation and also write $T_{ij}(\boldsymbol{\theta}^{(r)})$ as simply T_{ij} and $\Delta T_{ij}(\boldsymbol{\theta}^{(r)}, \rho)$ as ΔT_{ij}. Now, along the direction of the gradient

$$\mathcal{L} = \sum_{ij}[-(T_{ij} + \Delta T_{ij}) + N_{ij}\log(T_{ij} + \Delta T_{ij}) - \log(N_{ij}!)], \qquad (5.175)$$

where T_{ij} is independent of ρ and hence

$$\frac{\partial\mathcal{L}}{\partial\Delta T_{ij}} = -1 + N_{ij}(T_{ij} + \Delta T_{ij})^{-1}. \qquad (5.176)$$

Let

$$\delta(\rho) = \frac{d\mathcal{L}}{d\rho} = \sum_{ij}\frac{\partial\mathcal{L}}{\partial\Delta T_{ij}} \cdot \frac{d\Delta T_{ij}}{d\rho}. \qquad (5.177)$$

Since ΔT_{ij}'s are given by (5.152) in which the $S_{ij}{}^{(k)}$'s do not contain anything that changes with $\Delta\boldsymbol{\theta}$ or ρ, and since $\Delta\theta_k = \rho\nu_k$

$$\frac{d\Delta T_{ij}}{d\rho} = \sum_{k=1}^{K}[\nu_k \cdot S_{ij}^{(k)}] = \sum_{k=1}^{K} Q_{ij}^{(k)} \text{ (say)} \qquad \text{for all } i \in I, j \in J \tag{5.178}$$

where ν_k is as in (5.164) and $S_{ij}^{(k)}$'s are given by (5.153). Since $Q_{ij}^{(k)}$ is free of ΔT_{ij}, integration of the two ends of (5.178) yields $\Delta T_{ij} = \rho\sum_{k=1}^{K} Q_{ij}^{(k)}$ and, from (5.176),

$$\frac{\partial\mathcal{L}}{\partial\Delta T_{ij}} = -1 + N_{ij}(T_{ij} + \rho\sum_{k=1}^{K} Q_{ij}^{(k)})^{-1}. \tag{5.179}$$

Consequently, from (5.177), (5.178) and (5.179), it follows that the function $\delta(\rho)$ can be written as

$$\delta(\rho) = \sum_{ij}\left([N_{ij}(T_{ij} + \rho\sum_{k=1}^{K} Q_{ij}^{(k)})^{-1} - 1]\cdot(\sum_{k=1}^{K} Q_{ij}^{(k)})\right). \tag{5.180}$$

Let $\rho^{(r)}$ be a solution to $\delta(\rho) = 0$. Newton-Raphson iterations can be used for this purpose. This $\rho^{(r)}$ then is the step size for this modified gradient search procedure, which we call **Procedure MGS1**.

Alternatively, the equation $\delta(\rho) = 0$ can be linearized using a Taylor series approximation and the resultant linear equation in one unknown can be solved. Such an approximation for (5.180) is

$$\begin{aligned}
\delta(\rho) &= \sum_{ij}\left([N_{ij}(T_{ij} + \rho\sum_{k=1}^{K} Q_{ij}^{(k)})^{-1} - 1]\cdot(\sum_{k=1}^{K} Q_{ij}^{(k)})\right)\\
&= \sum_{ij}\left([N_{ij}(1 + \rho\sum_{k=1}^{K} Q_{ij}^{(k)}/T_{ij})^{-1}\cdot T_{ij}^{-1} - 1]\cdot(\sum_{k=1}^{K} Q_{ij}^{(k)})\right)\\
&= \sum_{ij}\left([N_{ij}(1 - \rho\sum_{k=1}^{K} Q_{ij}^{(k)}/T_{ij})\cdot T_{ij}^{-1} - 1]\cdot(\sum_{k=1}^{K} Q_{ij}^{(k)})\right) \qquad (5.181)\\
&= \sum_{ij}\left(\frac{N_{ij}}{T_{ij}}\cdot\sum_{k=1}^{K} Q_{ij}^{(k)}\right) - \rho\cdot\sum_{ij}\left(\frac{N_{ij}}{T_{ij}^2}\cdot(\sum_{k=1}^{K} Q_{ij}^{(k)})^2\right)\\
&\quad - \sum_{ij}\left(\sum_{k=1}^{K} Q_{ij}^{(k)}\right).
\end{aligned}$$

Notice that this approximation (taking Taylor series and retaining only linear term) would be allowable since $\rho\sum_{k=1}^{K} Q_{ij}^{(k)} = \Delta T_{ij}$ in

$$(1 + \rho\sum_{k=1}^{K} Q_{ij}^{(k)}/T_{ij})^{-1}$$

would usually be much smaller than T_{ij}. Now setting (5.181) equal to 0, we have

$$\rho = \frac{\sum_{ij}\left(N_{ij}T_{ij}^{-1}\sum_{k=1}^{K}Q_{ij}^{(k)}\right) - \sum_{ij}\left(\sum_{k=1}^{K}Q_{ij}^{(k)}\right)}{\sum_{ij}\left(N_{ij}T_{ij}^{-2}(\sum_{k=1}^{K}Q_{ij}^{(k)})^2\right)}. \tag{5.182}$$

The resultant procedure is designated as **Procedure MGS2**. Notice that while a linear approximation will lead to greater computational efficiency in each iteration, this approximation might increase the total number of iterations.

Algorithms for Procedures MGS1 and MGS2

We give below the steps of modified gradient search procedures in some detail

1. Select an initial value $\boldsymbol{\theta}^{(0)}$.

2. Compute $F(\mathbf{c}_{ij}^t; \boldsymbol{\theta}^{(r)})$ and apply the DSF procedure with origin and destination totals given by (5.12), to obtain the $T_{ij}(\boldsymbol{\theta}^{(r)})$'s.

3. Set $\nu_k^{(r)} = \sum_{ij} c_{ij}^{(k)}[N_{ij} - T_{ij}(\boldsymbol{\theta}^{(r)})]$, compute the terms in $\delta(\rho)$ and solve the equation $\delta(\rho) = 0$ for ρ either by

   Procedure MGS1: using an iterative Newton-Raphson algorithm,

   $$\rho_{(n)} = \rho_{(n-1)} - \frac{\delta(\rho_{(n-1)})}{\delta'(\rho_{(n-1)})} \tag{5.183}$$

   where the parenthetic subscripts (n) and $(n-1)$ denote the nth and $n-1$th iterations in the Newton-Raphson method and,

   $$\begin{aligned}\delta'(\rho) &= d\delta(\rho)/d\rho \\ &= -\sum_{ij}\left(\frac{N_{ij}}{(T_{ij}/\sum_{k=1}^{K}Q_{ij}^{(k)} + \rho)^2}\right).\end{aligned} \tag{5.184}$$

   Procedure MGS2: directly solving the linear equation for ρ given by (5.182).

   Either way, call the solution $\rho^{(r)}$.

4. Update $\boldsymbol{\theta}^{(r)}$:

   $$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} + \rho^{(r)}\boldsymbol{\nu}^{(r)}. \tag{5.185}$$

5. Iterate steps 2 through 4 until a suitable stopping criterion is met (e.g., $\boldsymbol{\nu}^{(r)}$ is small enough).

5.5.5 GLIM

The GLIM procedure is a general procedure for obtaining ML estimates of parameters for a wide variety of models in common use in applied statistics. This class of models, called *generalized linear models,* may be described as follows. Let y_p's be a set of P observations with $\mathrm{E}[y_p] = \mu_p$. Assume that corresponding to each y_p, we have a set of Q independent variables x_{pq}. Let

$$\eta_p = \sum_{q=1}^{Q} \beta_q x_{pq} \text{ for } p = 1, \ldots, P, \tag{5.186}$$

where the β_q's are parameters to be estimated. If the η_p's are linked to the μ_p's by a *link function*

$$g(\mu_p) = \eta_p, \tag{5.187}$$

then we have a generalized linear model. A special case of this, when g is the identity function (i.e., $\mu_p = \eta_p$), is the usual linear model. A wide class of distributions for the y_p's can be accommodated. If in the case of the linear model, the y_p's are normally distributed, then the ML estimates are the same as least squares estimates.

Generalized linear models and the GLIM procedure have been described in McCullach and Nelder (1989, see, in particular, p. 27 for a description of the model and pp. 40–41 for an overview of the estimation procedure). In this section, we describe the model and the procedure in the context of gravity models. The reader is referred to the book just mentioned and the various manuals for further details, including discussions of how certain situations (e.g., when the argument of a logarithm turns out to be zero) are to be handled.

In the gravity model context, we see from (5.15), which is repeated below for convenience,

$$t_{ij} = a(i) + b(j) + \sum_{k=1}^{K} \theta_k c_{ij}^{(k)}, \tag{5.188}$$

that t_{ij} is a linear function of the parameters $a(i)$, $b(j)$ and θ_k. We also know that $\mathrm{E}[N_{ij}] = T_{ij}$ and that $\log[T_{ij}] = t_{ij}$. Thus the gravity model may be written as a generalized linear model with the link function $g[T_{ij}] = \log[T_{ij}]$.

Consequently, it might be convenient to bear in mind that the y_p's mentioned above correspond to our N_{ij} and μ_p and η_p to T_{ij} and t_{ij} respectively.

The GLIM procedure is iterative, with each iteration consisting of an application of weighted least squares. At an iteration $r + 1$ suppose the estimate of η_p from the previous iteration is $\eta_p{}^{(r)}$. Also let the corresponding value of μ_p, using (5.187), be $\mu_p{}^{(r)}$. Then the dependent variable for the weighted squares just mentioned is

$$z_p{}^{(r)} = \eta_p{}^{(r)} + (y_p - \mu_p{}^{(r)}) \left(\frac{d\eta}{d\mu}\right)_{\mu_p^{(r)}}, \tag{5.189}$$

where the subscript $\mu_p^{(r)}$ indicates that the derivative is being evaluated at that point. Since the derivative of $\log[x]$ is x^{-1}, in our case, (5.189) becomes

$$z_{ij}{}^{(r)} = t_{ij}{}^{(r)} + \frac{N_{ij} - T_{ij}{}^{(r)}}{T_{ij}{}^{(r)}}. \tag{5.190}$$

The weight $W_p{}^{(r)}$ for the weighted least squares is defined by

$$[W_p{}^{(r)}]^{-1} = \left(\frac{d\eta}{d\mu}\right)^2_{\mu_p^{(r)}} V_{\mu_p^{(r)}} \tag{5.191}$$

where V is the variance of N_{ij} evaluated at $\mu_p^{(r)}$. In our case, since N_{ij} is assumed to be Poisson, and hence the variance of N_{ij} is T_{ij},

$$W_{ij}{}^{(r)} = T_{ij}{}^{(r)}. \tag{5.192}$$

The independent variable values in the gravity model context can easily be seen from (5.15) to be given by the design matrix M, illustrated just below (5.15). Corresponding to N_{ij}, the independent variable values are given by the (ij)th row of M. Thus the model may be written as follows:

$$t_{ij} = \log(T_{ij}) = \sum_{r=1}^{I} \delta_r^{(A)} a_r + \sum_{s=1}^{J} \delta_s^{(B)} b_s + \sum_{k=1}^{K} c_{ij}^{(k)} \theta_k \tag{5.193}$$

where the indicator variables $\delta_r^{(A)}$'s and $\delta_s^{(B)}$'s are defined as:

$$\delta_r^{(A)} = \begin{cases} 1 & \text{when } r = i \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_s^{(B)} = \begin{cases} 1 & \text{when } s = j \\ 0 & \text{otherwise,} \end{cases}$$

and the $\boldsymbol{\beta}$ of (5.186) is given by

$$\boldsymbol{\beta} = (a_1, \ldots, a_I, b_1, \ldots, b_J, \theta_1, \ldots, \theta_K)^t.$$

Thus, the first $I+J$ independent variables are dummy or indicator variables and the next k independent variables are the $c_{ij}{}^{(k)}$'s. Since M is singular, one of the dummy variables needs to be deleted and the corresponding $a(i)$ or $b(j)$ set to a constant.

For the first iteration, McCullach and Nelder (1989) suggest setting $\mu_p{}^{(0)}$ equal to y_p. Thus, for the gravity model, we would use $T_{ij}{}^{(0)} = N_{ij}$, $z_{ij}{}^{(0)} = \log[N_{ij}]$ and $W_{ij}{}^{(0)} = N_{ij}$.

Let the log likelihood for the general linear model, based on a set of independent observations $\boldsymbol{y} = (y_1, \ldots, y_n)^t$, be $\mathcal{L}(\boldsymbol{\mu}; \boldsymbol{y})$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^t$ and let

$$D(\boldsymbol{y}; \boldsymbol{\mu}) = 2\mathcal{L}(\boldsymbol{y}; \boldsymbol{y}) - 2\mathcal{L}(\boldsymbol{\mu}; \boldsymbol{y}). \tag{5.194}$$

D is called the *scaled deviance*. For the gravity model given by (5.1) and (5.2), this becomes

$$D(\boldsymbol{N};\boldsymbol{T}) = 2\sum_{ij}[N_{ij}\log(N_{ij}/T_{ij}) - (N_{ij} - T_{ij})]. \tag{5.195}$$

For the exponential family of distributions, of which the Poisson is a member, $\mathcal{L}(\boldsymbol{y};\boldsymbol{y})$ is the maximum log likelihood achievable for an exact fit in which the fitted values are equal to the observed data. Since $\mathcal{L}(\boldsymbol{y};\boldsymbol{y})$ does not depend on the parameters, maximizing $\mathcal{L}(\boldsymbol{\mu};\boldsymbol{y})$ is equivalent to minimizing $D(\boldsymbol{y};\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$. The iterations of the GLIM procedure are deemed to have converged if the absolute change in scaled deviance between iterations is less than a prespecified criterion number.

The GLIM procedure has been recommended as a method of estimating gravity model parameters — largely in the United Kingdom, perhaps because the GLIM procedure is widely available in computer installations in that country. It might be noted that it is possible, without too much difficulty, to use the NLIN procedure in SAS (SAS, 1985b), which is widely available in the U.S. to construct the GLIM procedure. Note further that since the parameter vector $\boldsymbol{\beta}$ is typically large for the gravity model, a very large weighted least squares problem needs to be solved. This puts a heavy burden on the memory of even supercomputers.

5.6 Performance of General Algorithms

This section reports the results of the application of the procedures developed in the last section to a number of data sets. The attempt here is to assess how the procedures behave as well as to compare them with each other. Since a very large number of runs were made, in order not to get too tedious, our discussion needs to be somewhat abbreviated. A fuller description of the comparative merits, using a wider array of criteria, is presented in Yun (1992). As mentioned in the introduction to the previous section, the ultimate winner in our comparisons, and by a long margin, is the modified scoring procedure.

While several criteria (such as simplicity of implementation, code size, robustness, and memory usage) may be used to judge algorithms, execution time as measured by a real time clock will be our primary means of evaluation.

For all but the GLIM procedure, the IBM 3090–300J at the University of Illinois at Chicago was used with the VM/XA SP5.0 operating system and VS Fortran 2.5. Its cycle time was 14.5 nano seconds. For the GLIM procedure, the CRAY Y-MP/48 at the National Center for Supercomputing Applications (located at the University of Illinois at Urbana-Champaign) with the UNICOS 5.0 operating system and Fortran CF77 5.0 was used.

Its cycle time was 6 nano seconds. These cycle times are important because their ratio allows us to roughly compare procedures running on the two systems. For example, for the GLIM procedure, which was run on the Cray, we multiplied the actual execution times by the ratio to compute standardized times which are comparable to the execution times of procedures running on the IBM 3090–300J. In tables where execution times are presented, standardized times are given as well as the actual execution times (within parentheses after a comma separating them from the number of iterations)

Computer programs in FORTRAN were written for all procedures, except for the GLIM procedure for which the *G02GCF* routine of NAG (Numerical Algorithms Group) Fortran Library Mark 14 (which fits a generalized linear model with Poisson errors) was used.

In order to make the convergence criterion similar we first ran the modified scoring procedure setting the requirement that the efficient scores (the right side of (5.160)) be less than 10^{-10}. The parameter values obtained in this way [call these values $\hat{\boldsymbol{\theta}}_{\text{base}} = (\theta_{1,\text{base}}, \ldots, \theta_{k,\text{base}})^t$] were taken to be the 'correct values.' Whenever, in the course of iterations, each component $\theta_{k,\text{alt}}$ of parameter values $\hat{\boldsymbol{\theta}}_{\text{alt}}$ was such that

$$\left|\frac{\theta_{k,\text{alt}} - \theta_{k,\text{base}}}{\theta_{k,\text{base}}}\right| \leq \text{TOL},$$

for all $k \in K$, the corresponding procedure was deemed to have converged and the time taken to achieve convergence are the ones reported in various tables given below.

The stopping criterion of the GLIM procedure is different from that of the other procedures. Therefore, we manually computed our stopping criterion for each iteration of GLIM. This is the criterion used in most of the tables.

Several data sets were used to evaluate the procedures. These are described below. These data have been used in a number of practical applications and illustrate the construction and use of different $c_{ij}^{(k)}$'s.

5.6.1 The Data

Four data sets were used for the comparisons. Some of these data have also been used for other purposes later in the book (and, in one case, has been used earlier in this chapter). Moreover, the data sets illustrate gravity model applications which are fairly 'standard' in terms of the unusual circumstances and difficulties they present. Consequently, they are described below in some detail.

Since several data sets are used for making comparisons, each exhibit given below identifies the data used in making the runs the table is based on.

Skokie Data

One data set is a modified version of the Skokie data described earlier in this chapter. The $c_{ij}^{(1)}$ was set equal to the Skokie travel times. Distances were also computed, but since these would be highly correlated with travel times, these were permuted over the various origin-destination pairs to essentially remove multicollinearity. These permuted distances constitute the $c_{ij}^{(2)}$. While, after these permutations, the resultant data are neither 'real' nor 'like real,' as, say, simulated data would be, they still serve a useful purpose when purely computational properties of algorithms [e.g., speed] are examined.

In order to check for effects of multicollinearity, we sometimes used the same data as above but with the distances left unpermuted.

Simulated data

The data set for Skokie mentioned before was used as the basis for simulated data that we have used here and in other sections of the book. Several different sets of simulated data were constructed. In one, which was used when a high level of multicollinearity was desired, we added small random numbers to the travel time variable ${c_{ij}}^{(1)}$ to create a new variable ${c_{ij}}^{(2)}$.

For another type of simulated data, used in Section 5.9.2, we set $c_{ij}^{(1)}$ equal to the Skokie travel times and $c_{ij}^{(2)}$ as permuted distance. We also set $\theta_1 = -\cdot 02$ and $\theta_2 = -\cdot 5$. These numbers are rounded versions of parameter estimates for the Skokie data without permuting distances. Inserting these values into (5.2), we obtained $F(\boldsymbol{c}_{ij})$. Applying the DSF procedure with these $F(\boldsymbol{c}_{ij})$'s and with origin and destination totals (i.e., $N_{i\oplus}$'s and $N_{\oplus j}$'s) from the Skokie data, we obtained T_{ij}'s. We applied a Poisson random number generator to these T_{ij}'s to obtain sets of simulated N_{ij}'s.

When we needed particularly small sample size data in Section 5.9.2, we divided the T_{ij}'s obtained above by some number, usually 10, 50 or 100, before applying the Poisson generator. In the absence of a better nomenclature, we shall call these simulated data sets 'T_{ij}', '$T_{ij}/10$', '$T_{ij}/50$', '$T_{ij}/100$' data sets. Averages of histogram entries of the numerical values of N_{ij} for the last two are displayed in Exhibit 5.5, which illustrates how small the individual flows are for them.

In this section, we constructed '$T_{ij}/50$' and '$T_{ij}/100$' data sets by simply setting $T_{ij} = \hat{T}_{ij}$ from the Skokie data with permuted distances and then dividing these T_{ij}'s by 50 or 100 (i.e., we did not set $\theta_1 = -\cdot 02$ and $\theta_2 = -\cdot 5$).

Similar simulated data were also constructed based on some of the other data sets mentioned below.

Sometimes we deliberately tried to create data sets with extra variation (see Section 3.9.2) and data sets with correlated errors. These too are based on the Skokie data and are described in 5.9.2.

N_{ij}	Data '$T_{ij}/50$'	Data '$T_{ij}/100$'
0	67.61	79.27
1	18.50	14.66
2	7.30	4.08
3	3.33	1.31
4	1.59	0.45
5	0.81	0.14
5	0.81	0.14
6	0.42	0.05
7	0.21	0.02
8	0.10	0.01
9	0.06	0.007
10	0.03	0.003
11	0.02	
12	0.01	

Exhibit 5.5: Percentage of N_{ij}'s With Each Integer Value

2000 Origin-Destination Work Trip Data

When we wished to evaluate the performance of the algorithms on a larger data set, we used the data set described in Sööt and Sen (1991). This data set (expanded since the publication of the paper mentioned above) includes a work trip O-D matrix (N_{ij}), obtained from the Census CTPP package, with almost two thousand origin and two thousand destination zones for the Chicago Metropolitan Area. Four $c_{ij}{}^{(k)}$'s (i.e., for $k = 1, \ldots, 4$) were used in the model we constructed.

Three were the travel times by fastest mode, the difference in travel times between the fastest and second fastest modes and the time difference between the fastest and third fastest modes.

Our reasoning for the choice of these variables was as follows: A composite cost method proposed by Williams (1978) is the following: Let c_{ijm} be the cost of travel by mode m from i to j and let T_{ijm} be the corresponding expected number of trips. Then the customary logit mode split model gives

$$T_{ijm}/T_{ij} = \frac{\exp[-\mu c_{ijm}]}{\sum_m \exp[-\mu c_{ijm}]}, \tag{5.196}$$

and if the gravity model is

$$T_{ij} = A_i B_j \exp[-\beta \tilde{c}_{ij}], \tag{5.197}$$

then,

$$T_{ijm} = A_i B_j \exp[-\beta \tilde{c}_{ij}] \frac{\exp[-\mu c_{ijm}]}{\sum_m \exp[-\mu c_{ijm}]}. \tag{5.198}$$

If we assume that

$$T_{ijm} = A_i B_j \exp[-\mu c_{ijm}], \tag{5.199}$$

then a solution to all the above equations is achieved by writing

$$\exp[-\beta \tilde{c}_{ij}] = \sum_m \exp[-\mu c_{ijm}] \tag{5.200}$$

or

$$\tilde{c}_{ij} = -\beta^{-1} \ln(\sum_m \exp[-\mu c_{ijm}]). \tag{5.201}$$

This last expression defines a composite cost $\tilde{c}_{ij}$. By taking Taylor expansions of both sides of (5.200) we get

$$-\beta \tilde{c}_{ij} \approx m - 1 - \sum_m \mu c_{ijm}. \tag{5.202}$$

This makes the gravity model

$$A_i B_j \exp[-\beta \tilde{c}_{ij}] = A_i B_j \exp[(m-1)] \exp[-\mu \sum_m c_{ijm}]. \tag{5.203}$$

The first exp on the right being a constant can be absorbed into the A_i and B_j terms. Hence, we get essentially the form we have used. Use of the fastest mode, etc., was an attempt at introducing the Fisk-Boyce (1984) idea of the prior usage weights, and subtracting the smallest c_{ijm} from the next smallest, etc., was an attempt at getting some measure of orthogonality.

The three modes used were cars, commuter rail and bus/rapid transit. Travel times for the last two were obtained from time-tables after making adjustments for transfer times and access/egress times. Car travel times on expressways were average peak hour times obtained from the Illinois Department of Transportation which, using sensors under the pavement to obtain occupancy rates and volumes, continuously monitors speed on Chicago Area expressways. Arterial speeds were estimated from a regression equation which had as the only independent variable, distance from the center of the city. They were combined using an optimal route choice algorithm (which actually was also used to route public transportation users). In most cases the car provided the fastest travel times.

The fourth measure was an occupational compatibility index. If a residential zone has all blue collar employees, there would not be many people from there holding jobs in a predominantly white collar area. One might construct a separate model for each occupational category, but the necessary data are not available. Therefore, we proceeded as follows. Let ${p_i}^{(1)}, \ldots, {p_i}^{(Q)}$ be the proportion of employees in each of Q employment categories living in i and let ${q_j}^{(1)}, \ldots, {q_j}^{(Q)}$ be the proportion of jobs in each of these categories in employment zone j. Our measure is

$$c_{ij}{}^{(4)} = \log[\sum_{m=1}^{Q} \sqrt{p_i{}^{(m)}} \sqrt{q_j{}^{(m)}}]. \tag{5.204}$$

By the Cauchy-Schwartz inequality,

$$\sum_{m=1}^{Q} \sqrt{p_i{}^{(m)}}\sqrt{q_j{}^{(m)}} \le [\sum_{m=1}^{Q} p_i{}^{(m)}]^{1/2}[\sum_{m=1}^{Q} q_j{}^{(m)}]^{1/2} = 1, \qquad (5.205)$$

with the equality holding if and only if $p_i{}^{(m)} = q_j{}^{(m)}$ for all m. This suits our purpose well, because it makes $\exp[c_{ij}{}^{(4)}] = 1$ when the match is 'perfect' [i.e., when $p_i{}^{(m)} = q_j{}^{(m)}$ for all m], $\exp[c_{ij}{}^{(4)}] \approx 0$ when the match is essentially nonexistent [i.e., when $q_j{}^{(m)} p_i{}^{(m)} \approx 0$ for all m] and $\exp[c_{ij}{}^{(4)}]$ increases with improving matches.

The seven occupational categories used were 1) professional, executive, administrative and managerial; 2) technicians and related support; 3) sales; 4) administrative support (including clerical); 5) service; 6) precision products, craft and repair; and 7) operators, fabricators and laborers.

Hospital Patient Flow Data

Another data set consisted of flows of patients, for a 1 year period, between 250 zip code areas and 92 hospitals in the six-county Chicago metropolitan area. We are indebted to John Lowe of School of Urban Planning, University of Illinois at Chicago for making this data set available to us. Much of the work presented using these data form part of his dissertation (Lowe, 1993). Six chronic care and specialty hospitals are not included in these data. Over half the origin-destination pairs had zero flows (N_{ij}'s). Data on travel times, distances and a payer/hospital match index constituted the $c_{ij}^{(k)}$'s. The last variable is similar to the $c_{ij}^{(4)}$ variable for the work trip data just described, except that $p_i{}^{(m)}$'s and $q_j{}^{(m)}$'s are proportions of patients at the hospital and origin ends belonging to the mth payment category. These categories are Medicare, Medicaid, Commercial Insurance, Other Insurance and Self Pay.

All trials using these Hospital data and reported on in this section were based on data as described above and on simulated data based on them. However, in subsequent work, the list of $c_{ij}^{(k)}$'s have been expanded to include transformations of distance and travel time, and other variables [e.g., travel times to teaching hospitals were treated separately from travel times to other hospitals by the use of suitable indicator variables]. Gravity models have also been constructed for patients belonging only to certain diagnostic categories (see Section 5.8).

5.6.2 Convergence

Exhibits 5.6, 5.7, 5.8, 5.9 and 5.10 illustrate values of $\theta^{(r)}$ for the different procedures when applied to the Skokie data with permuted distances. These plots are rather typical in that they have a similar overall shape

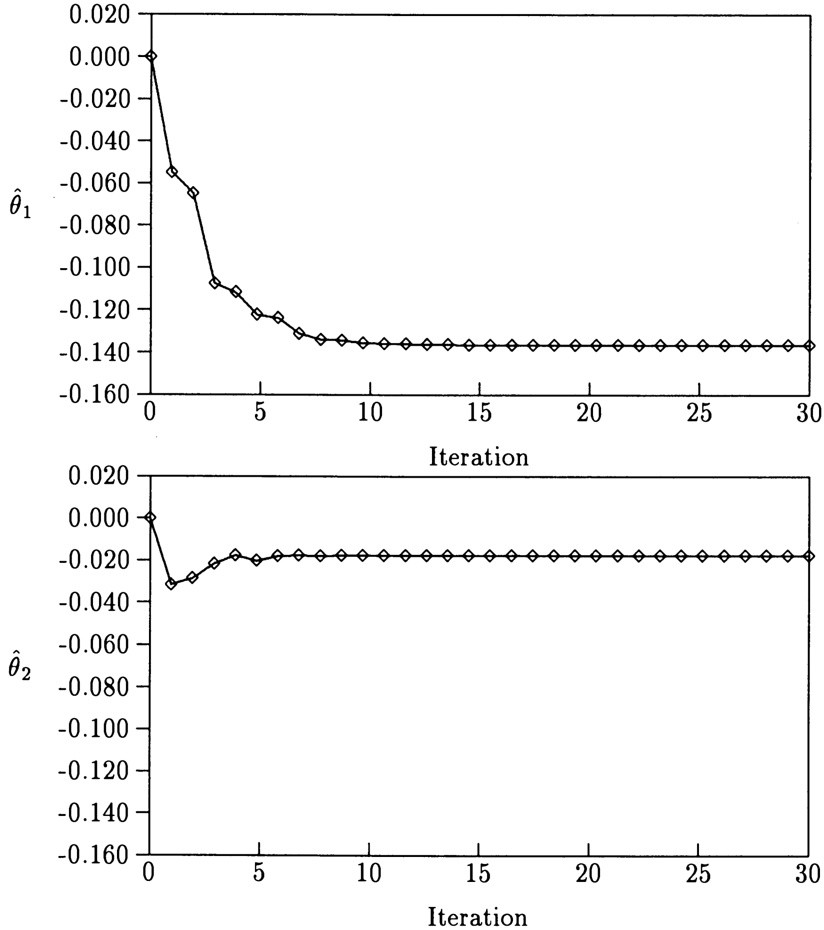

Exhibit 5.6: Rate of Convergence of 'General' Procedure

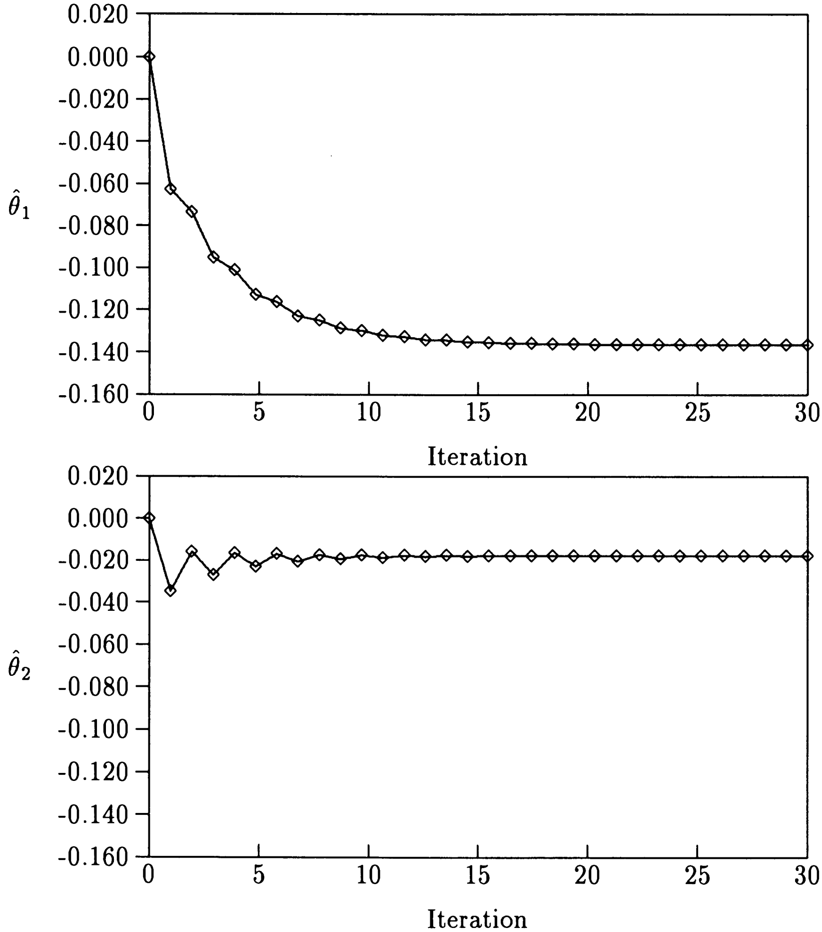

Exhibit 5.7: Rate of Convergence of Procedure MGS1

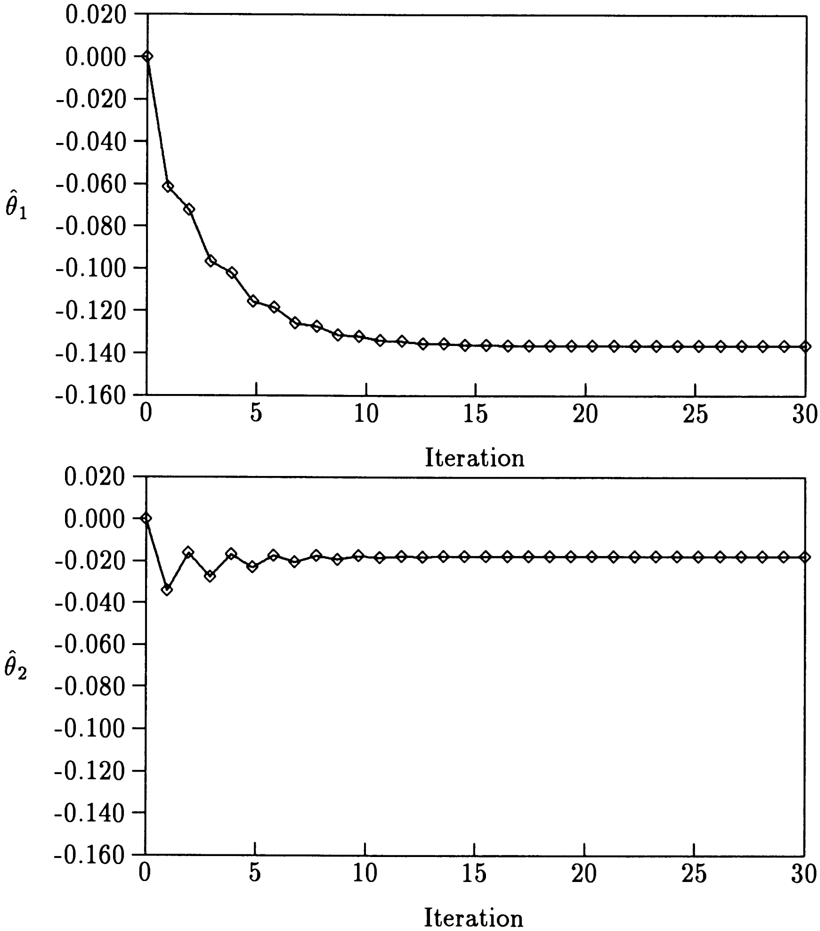

Exhibit 5.8: Rate of Convergence of Procedure MGS2

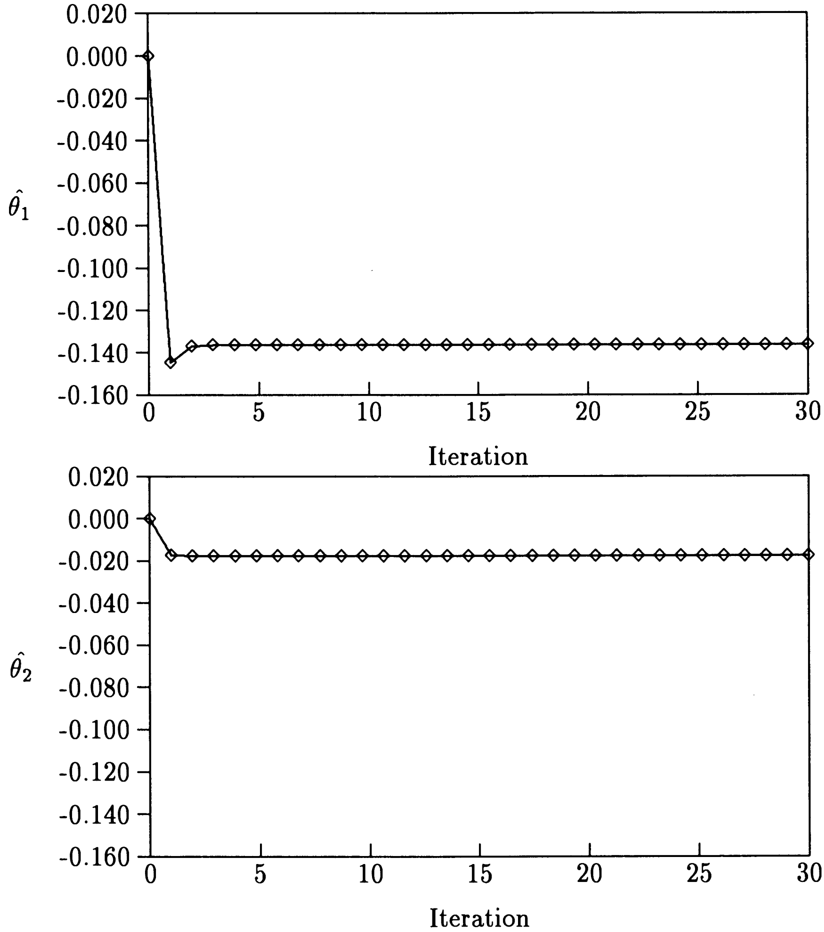

Exhibit 5.9: Rate of Convergence of the Modified Scoring Procedure

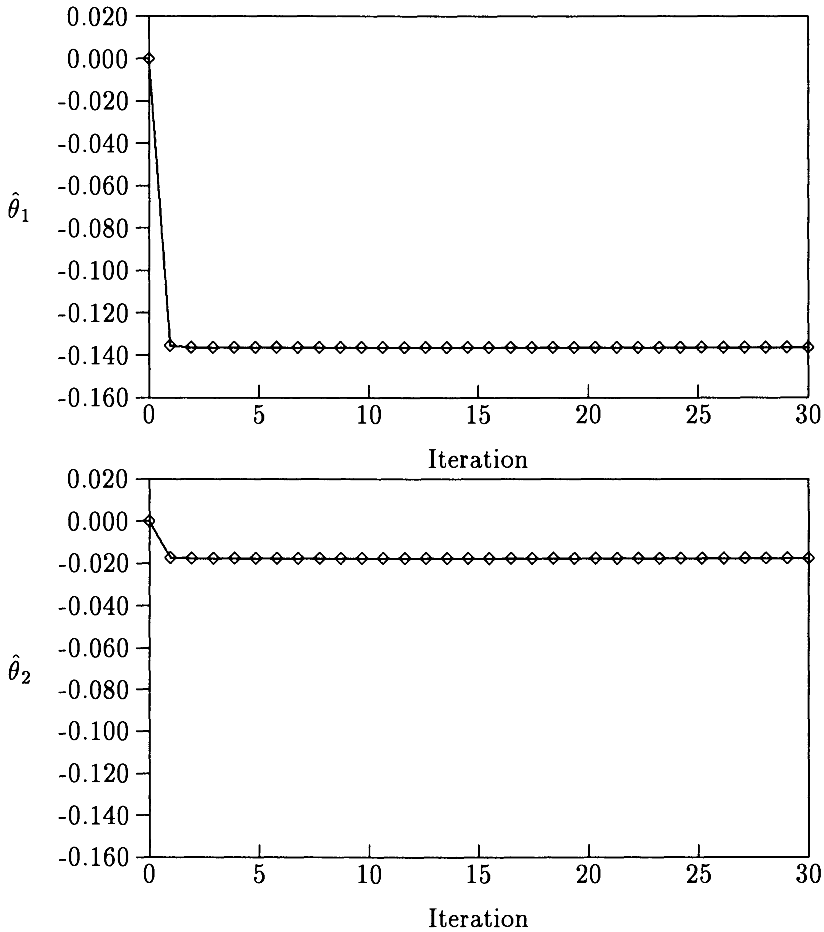

Exhibit 5.10: Rate of Convergence of the GLIM Procedure

to those obtained for the other runs reported in the tables below. While the successive values of the parameters for the modified scoring procedure and the GLIM Procedure appear to move more smoothly than those for the other procedures, for all the procedures, convergence was quite smooth after, perhaps, the first few iterations. However, the number of iterations needed to achieve convergence for various levels of TOL varied considerably as the tables given later will illustrate.

For the general procedure, the issue of convergence has been discussed previously. For the other gradient search procedures, we conjecture that convergence nearly always occurs — albeit sometimes very slowly. Our conjecture is based both on the derivation of the procedures, as well as on the experience with the examples we ran.

Under normal use, the modified scoring procedure always yielded successive iterations which moved smoothly towards the limit. However, we should point out that if the initial values are poorly chosen, methods based on the Newton-Raphson procedure can give overflows or underflows. The modified scoring method is no exception. By giving deliberately poor initial values, we were able to cause overflows. However, with reasonable starting values no problems occurred. If no good initial guess can be made, starting values could perhaps be obtained from running a preliminary least squares procedure (see Chapter 6). In the case of examples we tried, $\boldsymbol{\theta}^{(0)} = \mathbf{o}$ was always satisfactory.

The GLIM procedure ran out of memory even for moderate sized problems. This is not entirely unexpected (see Section 5.5.5). It was also very time consuming. Other than that, it behaved very well.

Except as noted above for the modified scoring procedure, no underflows or overflows in the computer occurred during execution of any of the alternative procedures, and given the smoothness of convergence, no overflows or underflows should be expected in *normal* use. This is in sharp contrast to some competing procedures such as the usual scoring procedure (see Section 5.5.1; also see Batty, 1976 and Sen, 1986) and the method of predetermined step sizes (which is a gradient search procedure with $\rho^{(r)}$'s determined a priori — see Sheffi, 1985, p. 324; this method was investigated but not reported on below), which perform poorly even with reasonable initial values.

5.6.3 Speeds of Procedures

For the Skokie data set with permuted distances, execution times and number of iterations needed to meet different levels of the TOL criterion are shown in Exhibit 5.11. The initial values were set as $\boldsymbol{\theta}^{(0)} = \mathbf{o}$. Exhibit 5.11 also shows speed-up which is defined as:

$$\text{Speed-up} = \frac{\text{CPU time of the general procedure}}{\text{CPU time of the alternative procedure}}.$$

TOL.	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
$\leq \|10^{-1}\|$	.23 (6)	.22 (11)	.10 (7)	.03 (1)	3.84 (5, 1.59)
$\leq \|10^{-2}\|$	.46 (13)	.33 (17)	.17 (13)	.04 (2)	3.84 (5, 1.59)
$\leq \|10^{-3}\|$	.62 (18)	.44 (23)	.25 (19)	.05 (3)	3.84 (5, 1.59)
$\leq \|10^{-4}\|$	32.70 (1000)	.55 (29)	.30 (23)	.05 (3)	3.84 (5, 1.59)
$\leq \|10^{-5}\|$	32.70 (1000)	.62 (33)	.32 (25)	.06 (4)	3.84 (5, 1.59)
Average per iter.	.03	.02	.01	.02	.77

TOL	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
$\leq \|10^{-1}\|$	1.00	1.05	2.30	7.67	.06
$\leq \|10^{-2}\|$	1.00	1.40	2.71	11.50	.12
$\leq \|10^{-3}\|$	1.00	1.41	2.48	12.40	.16
$\leq \|10^{-4}\|$	1.00	59.45	109.00	654.00	8.52
$\leq \|10^{-5}\|$	1.00	52.74	102.19	545.00	8.52

Exhibit 5.11: Execution Times in Secs. (Upper Table) and Speed Up Summary (Lower Table) for Various Procedures and Various Convergence Criteria

NOTES: Numbers of iterations are shown in parentheses. When the convergence criterion was not met before the 1000th iteration, iterations are shown as 1000. The 40×18 Skokie O-D data was used with two $c_{ij}^{(k)}$'s: travel times and permuted distances. Except for the GLIM procedure, (0,0) was used as the initial value. CPU times for GLIM are standardized times — actual times are shown as the second number inside parenthesis.

When the general procedure (base for speed-up computation) does not meet the stopping criteria, the speed-up factor shown in Exhibit 5.11 is *at least* that factor. For instance, the speed-up of the modified scoring procedure is at least 654 for TOL $= 10^{-4}$.

It is easily seen from the tables that the modified scoring procedure is the fastest of the procedures at all tolerance levels. The speed-up of the modified scoring procedure grows rapidly from a factor of 7 to over 600 as the tolerance level gets smaller. It is also noteworthy that, in this case as well as in other exercises we conducted, this procedure required very few iterations. The procedure consists of running the DSF procedure, computing $S_{ij}^{(k)}$'s and solving a set of (usually a small number of) linear equations. Of these the DSF procedure is the most time consuming. Thus, it might be conjectured that the procedure could be speeded up by refining $\Delta\theta_k$'s further between successive applications of the DSF procedure, by repeating step 3 a few times. However, because of the small number of iterations, any savings due to this further modification is not likely to be too large.

The two modified gradient search procedures were the next fastest, with MGS2 being slightly faster of the two. Notice, however, that their relative speeds vis a vis the modified scoring procedure declines slightly with more stringent values of TOL. The GLIM procedure appears to be the slowest for the first three tolerance levels. Because the number of iterations for the GLIM procedure is always 5, we decided to make an additional set of runs in each of which the GLIM procedure was allowed to converge according to different levels of its own convergence criterion. From the parameter estimates thus obtained, we computed values of TOL. These levels were then specified as convergence criteria for the modified scoring procedure (the clear winner from Exhibit 5.11). Exhibit 5.12 was the result.

Notice that the number of iterations used by GLIM was always quite small. This is to be expected since even the first iteration is essentially a weighted least squares procedure. The methods used to speed up least squares that will be discussed in Chapter 6, could perhaps be used here. However, the computation of ${T_{ij}}^{(r)}$'s at every step will require the use of the DSF procedure or some other somewhat time consuming algorithm. Thus it is unlikely that even with the modification, the GLIM procedure will be much faster than the modified scoring procedure.

In the discussion of procedure performance, we have so far concentrated on data in which multicollinearity was deliberately reduced. In practical situations multicollinearity is very frequently present; e.g., if we use travel time and distance as measures of separations, or any ${c_{ij}}^{(k)}$ and its log, there would be fairly high levels of multicollinearity. The presence of multicollinearity usually adversely affects the convergence of most maximum likelihood algorithms, particularly those based on gradient search. Multicollinearity gives rise to likelihood functions which have level curves more or less like elongated ellipses. Gradient search procedures, unless they are

Parameters	TOL	Mod. Scor.	GLIM proc.
$\hat{\theta}_1$ $\hat{\theta}_2$	2.00×10^{-4} 7.72×10^{-3}	.05 (3)	1.93 (2, .80)
$\hat{\theta}_1$ $\hat{\theta}_2$	2.00×10^{-4} 7.72×10^{-3}	.06 (4)	2.44 (3, 1.01)
$\hat{\theta}_1$ $\hat{\theta}_2$	10^{-8} 10^{-8}	.16 (13)	3.84 (5, 1.59)

Exhibit 5.12: Speed Comparison of GLIM and Modified Scoring Procedure

NOTES: CPU times are in seconds and the number of iterations are given in parenthesis. $c_{ij}^{(1)}$ is travel time and $c_{ij}^{(2)}$ is the permuted distance for Skokie. The initial point for the modified scoring procedure was (0,0).

able to compute just the right ρ's can often end up zigzagging in directions which are close to being orthogonal to the major axis of the 'ellipses', rather than getting to the axis and moving along it.

In order to examine the effects of multicollinearity, we used two data sets at this stage. One was the Skokie data with distances not permuted. The other, which we shall call a 'seriously' multicollinear data set, was created by generating pseudo random numbers from a uniform (0,1) distribution and then adding these random numbers to Skokie travel times $c_{ij}^{(1)}$ to get $c_{ij}^{(2)}$. Given that the mean of the $c_{ij}^{(1)}$'s is 12.8 (minutes), one should expect that this would indeed lead to very serious multicollinearity. Exhibit 5.13 displays results from applying the different procedures to these data sets.

It is easily seen that the general procedure performed poorly on multicollinear data. The modified gradient search procedures, surprisingly, did not do nearly as poorly. But most surprising was the fact that the modified scoring procedure and the GLIM procedure seemed hardly affected by multicollinearity. Since the modified scoring procedure at every iteration seeks to solve (5.13), it would be expected that when the rows of the coefficient matrix are close to linear dependency, the quality of solutions would be affected and this in turn would lead to more iterations. However, it would seem from Exhibit 5.13, that even higher levels of multicollinearity would be required before the modified scoring procedure is noticeably affected. However, higher levels of multicollinearity would usually result in the analyst discarding the offending separation measures and, thus would be rare in practice. The GLIM procedure depends at every iteration on a least squares procedure and consequently we expected that its performance would be degraded by multicollinearity. But that too did not happen. Because of the

$c_{ij}^{(2)}$	Init. val.	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
permuted distance	(0,0) LS	.62 (18) 30.6 (*)	.44 (23) .52 (27)	.25 (19) .35 (27)	.05 (3) .04 (2)	3.84 (5, 1.59)
actual distance	(0,0) LS	33.1 (*) 30.7 (*)	.29 (15) .41 (21)	.25 (19) .28 (21)	.04 (2) .03 (1)	4.01 (5, 1.66)
simulated distance	(0,0) LS	30.8 (*) 30.6 (*)	3.2 (172) 1.35 (72)	12.3 (*) 12.3 (*)	.05 (3) .05 (3)	3.87 (5, 1.60)

$c_{ij}^{(2)}$	Init. val.	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
permuted distance	(0,0) LS	1.00 n.a	1.41 1.19	2.48 1.77	12.40 15.50	.16
actual distance	(0,0) LS	1.00 1.08	114.07 80.68	132.32 118.14	827.00 1102.67	8.25
simulated distance	(0,0) LS	1.00 1.01	9.63 22.81	n.a n.a	616.00 616.00	7.96

Exhibit 5.13: Execution Times in Secs. (Upper Table) and Speed Up Summary (Lower Table) for Different Procedures, Data Exhibiting Three Levels of Multicollinearity and Two Initial Values.

NOTES: Numbers of iterations are shown in parentheses. When convergence criterion was not met before the 1000th iteration, iterations are shown as a $*$. A tolerance level of 10^{-3} was used for this table. The 40×18 Skokie O-D data was used with two $c_{ij}^{(k)}$'s. $c_{ij}^{(1)}$ was travel time. When $c_{ij}^{(2)}$ was permuted distance, multicollinearity was low; when it was actual distance, it was quite high; when it was the simulated distance, multicollinearity was severe. LS indicates that initial values were obtained by using a least squares procedure described in Chapter 6.

times per iteration of the GLIM procedure, the modified scoring procedure emerges as the clear winner in this exercise.

We also took the opportunity of these runs to investigate the effects of initial values on the execution time of the procedures. Therefore, for one set of runs we used $\boldsymbol{\theta}^{(0)} = \mathbf{o}$ and for another set we used very good initial values — those given by a least squares procedure (Chapter 6). The use of the latter set of initial values caused the gradient search procedures to slow down in all but one case. This was surprising although not entirely inexplicable, given the general remarks regarding the behavior of gradient search procedures in the presence of multicollinearity. Execution times of

Parameters	TOL	Mod. Scor.	GLIM proc.
$\hat{\theta}_1$ $\hat{\theta}_2$	2.98×10^{-3} 1.28×10^{-3}	.04 (2)	1.79 (2, .74)
$\hat{\theta}_1$ $\hat{\theta}_2$	4.12×10^{-6} 2.22×10^{-6}	.06 (4)	2.42 (3, 1.00)
$\hat{\theta}_1$ $\hat{\theta}_2$	10^{-8} 10^{-8}	.15 (12)	4.01 (5, 1.66)

Exhibit 5.14: Performance of GLIM and Modified Scoring Procedure on Multicollinear Data

NOTES: CPU times are in seconds and the number of iterations are given in parenthesis. $c_{ij}^{(1)}$ is travel time and $c_{ij}^{(2)}$ is the actual distance for Skokie. The initial value for the modified scoring procedure was set as (0,0).

the modified scoring procedure did seem to improve somewhat. Notice that the GLIM procedure does not require user supplied initial values.

We also ran a comparison just for the modified scoring procedure and the GLIM procedure in which we let the GLIM convergence criteria dictate convergence for both procedures (as for Exhibit 5.12). The results are in Exhibit 5.14.

As our next exercise, we decided to increase the number K of separation measures from 2 to 4, by keeping (for the Skokie data) $c_{ij}^{(1)}$ and $c_{ij}^{(2)}$ as travel time and permuted distance, but also using $c_{ij}^{(3)}$ and $c_{ij}^{(4)}$ which were logarithms of travel time and permuted distance. This is a realistic choice of variables given the widespread use of (2.125). However some multicollinearity also gets introduced. Exhibit 5.15 shows results of this effort. Notice that all gradient search procedures performed poorly, perhaps more because of multicollinearity than for any other reason. The GLIM procedure was relatively unaffected and the modified scoring procedure, which was affected adversely, still emerged as the winner. Notice that while there was just one more iteration for the modified scoring procedure (which by itself is not particularly noteworthy), the time per iteration increased, which implies that, while the DSF procedure is very time consuming, the time taken by the rest of the procedure is perhaps not entirely negligible.

We also tried the case $K = 1$ using travel time as the only variable. We already have a well functioning procedure — Hyman's — for this purpose (see Section 5.3.3). However, if the modified scoring procedure were to function well then it could be considered somewhat of an 'all-purpose' procedure. At least in this case it did do well, as Exhibit 5.15 illustrates, converging in two iterations, which was much less than that for Hyman's

K	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
$K = 1$	.26 (8)	.05 (2)	9.43 (*)	.03 (2)	3.31 (5, 1.37)
$K = 2$	.62 (18)	.44 (23)	.25 (19)	.05 (3)	3.84 (5, 1.59)
$K = 4$	40.79 (*)	23.89 (*)	10.37 (*)	.10 (4)	3.92 (5, 1.62)

K	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
$K = 1$	1.00	5.20	n.a	8.67	.08
$K = 2$	1.00	1.41	2.48	12.40	.16
$K = 4$	1.00	n.a	n.a	407.90	10.41

Exhibit 5.15: Execution Times in Secs. (upper table) and Speed Up Summary (Lower Table) for Various Procedures and Three K's

NOTES: Numbers of iterations are shown in parentheses. When convergence criterion was not met before the 1000th iteration, iterations are shown as a $*$. The tolerance level was set at 10^{-3} for this table. CPU times for GLIM are standardized times — actual times are shown as the second number inside parenthesis.

procedure (which took 7 iterations and .08 seconds to reach the same tolerance level and which also uses the DSF procedure at every step). The gradient search procedures, in this case, essentially become efforts to solve the equation $d\mathcal{L}/d\theta_1 = 0$. Their performance as well as that of the GLIM procedure is also shown in Exhibit 5.15.

An issue that we addressed at this stage is how the convergence rates of the procedures would fare for small sizes of T_{ij}. The quality of the parameter estimates obtained in such cases is discussed in Section 5.9.2. We used data simulated as described at the end of the section on simulated data in Section 5.6.1. Since we wanted to see how the procedures behaved with particularly stringent convergence requirements, we set TOL equal to 10^{-6}. For this reason, we have also included in the table (Exhibit 5.16) a row for the 'T_{ij}' data set, where the data was simulated by using a Poisson random number generator using as Poisson parameter the $\hat{T}_{ij}$'s from the Skokie data with permuted distances. It appears that none of the procedures (except for the 'general' procedure) seems to suffer much owing to the smallness of T_{ij}'s. This is noteworthy because the '$T_{ij}/100$' data actually simulates a 1 per cent sample of trips in a small town. Only about 7 per cent of the N_{ij}'s had values of over 1 and the sum $N_{\oplus\oplus}$ of N_{ij}'s was 252. The total number $N_{\oplus\oplus}$ of trips for the '$T_{ij}/50$' data set was 489 and 16 per cent of the N_{ij}'s were greater than 1. The very poor performance of the general procedure was perhaps due to the stringent tolerance level.

Data set	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
'T_{ij}'	32.4 (*)	.59 (22)	.84 (62)	.1 (7)	1.62 (6)
'$T_{ij}/50$'	30.6 (*)	.73 (29)	.13 (9)	.074 (5)	1.62 (6)
'$T_{ij}/100$'	30.7 (*)	.75 (30)	.34 (25)	.086 (6)	1.88 (7)

Exhibit 5.16: Execution Times in Secs. for Different Procedures and Data Sets

NOTES: Numbers of iterations are shown in parentheses. 10^{-6} was set as the tolerance level for this table. When the convergence criterion was not met at or before the 1000th iteration, iterations are shown as a *. Except for the GLIM procedure, (0,0) was used as the initial value. CPU times for GLIM are actual times.

A major problem with algorithms for ML estimation of gravity models is that frequently there are a very large number of parameters. Notice that there are I values of $A(i)$, J values of $B(j)$ and K values of θ_k, and while K is usually not large, I and J often are. Therefore, it is reasonable to ask how the procedures would handle larger data sets. We applied all five procedures to the 250×94 (contrasted with 40×18 Skokie data set) Hospital data set. Execution times and numbers of iteration are shown in Exhibit 5.17.

The GLIM procedure simply could not handle an application of this size. It can be seen from Exhibit 5.17 that for all other procedures, partly because the DSF procedure is imbedded in each iteration, execution time dramatically increases as the size of the application problem get larger. The lack of convergence of the gradient search procedures could be due to multicollinearity since we used both distance and travel time. However, the main point we noticed is that the modified scoring procedure continued to perform well.

Since the modified scoring procedure consistently outperformed its competitors and often by a wide margin, it seemed appropriate at this stage to declare it the winner in our trials. In fact, none of the gradient search procedures handled multicollinearity well (in terms of computing ML estimates of $\boldsymbol{\theta}$ not, as mentioned before, in terms of computing $\hat{T}_{ij}$'s). The GLIM procedure could not handle large data sets. The modified scoring procedure, on the other hand, handled very well a fairly wide selection of actual and simulated data sets, representing a broad range of conditions. A noteworthy feature that the modified scoring procedure shares with the GLIM procedure can only be described as 'even-temperedness.' It always seemed to behave predictably and well.

Consequently, we subjected the modified scoring procedure to further scrutiny. Since it continued to perform well as expected, we have not reported on all the runs made. The only area of concern is that it is possible

Data set	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
Skokie	.62 (18)	.44 (23)	.25 (19)	.05 (3)	3.84 (5, 1.59)
Hospital	1058.94 (*)	556.50 (*)	393.76 (*)	1.43 (3)	n.a

Data set	Gen. proc.	Proc. MGS1	Proc. MGS2	Mod. Scor.	GLIM proc.
Skokie	1.00	1.41	2.48	12.40	.16
Hospital	1.00	n.a	n.a	740.52	n.a

Exhibit 5.17: Execution Times in Secs. (Upper Table) and Speed Up Summary (Lower Table) for Various Procedures and Data Sets with Different I and J

NOTES: Numbers of iterations are shown in parentheses. The tolerance level for this table was 10^{-3}. When the convergence criterion was not met at or before the 1000th iteration, iterations are shown as a *. Except for the GLIM procedure, (0,0) was used as the initial value. CPU times for GLIM are standardized times — actual times are shown as the second number inside parenthesis.

to make the procedure diverge by choosing initial values very poorly. In the case of the hospital data, we succeeded by taking actual parameter estimates and changing signs.

We conclude this section by showing in Exhibit 5.18, the results of applying the modified scoring procedure to three data sets, including the work trip data. Notice that this data set, with about two thousand origins and destinations and four separation measures, is by all standards large, and indeed of a size such that maximum likelihood methods typically would not be considered. Since no comparison of procedures is being made, we could use the stopping criterion which we believe to be the most appropriate for scoring procedures: sizes of the efficient scores (e.s.). Even under very demanding values of e.s., the modified scoring procedure remains an excellent performer.

5.7 Covariance of Estimates

In order to obtain confidence intervals, as well as to carry out certain tests of hypotheses, one needs covariance matrices of ML estimates. As we shall see later, gravity model estimates tend to be asymptotically normal. Then, estimates of parameter values and their covariances give an estimate of

Data Set	e.s. $\leq 10^{-3}$	e.s. $\leq 10^{-5}$	e.s. $\leq 10^{-10}$	Av. time per iter.
Skokie	.09 (7)	.11 (9)	.16 (13)	.01
Hospital	4.54 (12)	5.30 (14)	7.78 (21)	.38
Work Trip	56.34 (14)	64.44 (16)	94.58 (24)	4.00

Exhibit 5.18: Performance of Modified Scoring Procedure for Different Convergence Criteria and Sizes of Data Sets

NOTES: Tolerance levels are in terms of absolute values of efficient scores. Times are in seconds and numbers of iterations are shown in parens.

their distribution. We consider such matrices in this section.

Since the gravity model is a non-linear function of its parameters, obtaining asymptotic covariances is the best we can do. This will be true throughout this section, and hence we shall usually drop the word asymptotic.

In Section 5.7.1 we obtain the covariance matrix of the ML estimate of $\boldsymbol{\theta}$. Gravity models are frequently used for forecasting. In transportation planning, future period $\hat{T}_{ij}$'s are obtained using the DSF procedure with exogenously estimated future values of $T_{i\oplus} = O_i$ and $T_{\oplus j} = D_j$. The factors $F(\boldsymbol{c}_{ij}) = \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}]$ are computed using exogenously given future $\boldsymbol{c}_{ij}$'s and $\boldsymbol{\theta}$ estimated from base period data and assumed to remain the same into the forecast period. Covariance matrices of thus estimated $\hat{T}_{ij}$'s are given in Section 5.7.2. As mentioned in the introduction to the book, other methods might have to be used depending on what assumptions one makes and what estimates are available. These methods of forecasting T_{ij}'s are summarized in Section 5.7.3 as are methods for obtaining their covariance matrices.

5.7.1 Covariance of $\hat{\theta}_k$'s

Since we shall be using Jacobian matrices in this section and since different authors have defined them differently, it is appropriate to define them first. Let $\boldsymbol{g}(\boldsymbol{x})$ be a continuously differentiable vector valued function

$$\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x}))^t$$

of $\boldsymbol{x} = (x_1, \ldots, x_n)^t$. Then the Jacobian matrix $\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})$ of the mapping of $\boldsymbol{x}$ to $\boldsymbol{g}$ is the $m \times n$ dimensional matrix $((\partial g_r / \partial x_s))$. It will also be referred to as the Jacobian matrix of $\boldsymbol{g}$ with respect to $\boldsymbol{x}$.

The property of such matrices that interests us most is that if $\Delta \boldsymbol{x} = (\Delta x_1, \ldots, \Delta x_n)^t$ is a small change in $\boldsymbol{x}$ and $\Delta \boldsymbol{g} = (\Delta g_1, \ldots, \Delta g_m)$ is the corresponding small change in $\boldsymbol{g}$, then

$$\Delta g_r \approx \sum_{s=1}^{n} \frac{\partial g_r}{\partial x_s} \Delta x_s. \tag{5.206}$$

Thus

$$\Delta \boldsymbol{g} \approx \mathcal{J}(\boldsymbol{x} \to \boldsymbol{g}) \Delta \boldsymbol{x}. \tag{5.207}$$

If $\boldsymbol{x}$ and $\boldsymbol{g}$ are random vectors and we let $\Delta \boldsymbol{x} = \boldsymbol{x} - \mathrm{E}[\boldsymbol{x}]$ and $\Delta \boldsymbol{g} = \boldsymbol{g} - \mathrm{E}[\boldsymbol{g}]$, then

$$\begin{aligned}
\mathrm{Cov}\,[\boldsymbol{g}] &= \mathrm{E}[\Delta \boldsymbol{g} \Delta \boldsymbol{g}^t] \approx \mathrm{E}[(\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})\Delta \boldsymbol{x})(\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})\Delta \boldsymbol{x})^t] \\
&= \mathrm{E}[\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})[\Delta \boldsymbol{x}(\Delta \boldsymbol{x})^t]\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})^t] \\
&= \mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})\mathrm{E}[\Delta \boldsymbol{x}(\Delta \boldsymbol{x})^t]\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})^t \\
&\approx \mathcal{J}(\boldsymbol{x} \to \boldsymbol{g}) \ \mathrm{Cov}\,[\boldsymbol{x}]\mathcal{J}(\boldsymbol{x} \to \boldsymbol{g})^t
\end{aligned} \tag{5.208}$$

where, as before, $\mathrm{Cov}\,(\cdot)$ stands for 'the covariance' of. The approximations above become equalities asymptotically. Consequently, the last expression in (5.208) is the asymptotic covariance matrix of $\boldsymbol{g}$. Notice that a special case of (5.208) occurs when $\boldsymbol{g}$ is a linear function $\boldsymbol{g} = M\boldsymbol{x}$ of $\boldsymbol{x}$, where M is a matrix. Then $\mathrm{Cov}\,(M\boldsymbol{x}) = M\ \mathrm{Cov}\,(\boldsymbol{x})M^t$. In this case the relationship is exact (not asymptotic).

As already noted at the beginning of Section 5.2 [see (5.17)], equations (5.12) and (5.13) may be written using matrix notation as $M^t\boldsymbol{T} = M^t\boldsymbol{N}$ and since this relation is obeyed by maximum likelihood estimates, we have,

$$M^t\hat{\boldsymbol{T}} = M^t\boldsymbol{N}, \tag{5.209}$$

where M has been defined in Section 5.2 and, as before,

$$\boldsymbol{T} = (T_{11}, \ldots, T_{1J}, \ldots\ldots, T_{I1}, \ldots, T_{IJ})^t, \tag{5.210}$$

$$\hat{\boldsymbol{T}} = (\hat{T}_{11}, \ldots, \hat{T}_{1J}, \ldots\ldots, \hat{T}_{I1}, \ldots, \hat{T}_{IJ})^t, \tag{5.211}$$

$$\boldsymbol{N} = (N_{11}, \ldots, N_{1J}, \ldots\ldots, N_{I1}, \ldots, N_{IJ})^t. \tag{5.212}$$

The Jacobian matrix of

$$\boldsymbol{t} = (t_{11}, \ldots, t_{1J}, \ldots\ldots, t_{I1}, \ldots, t_{IJ})^t$$

with respect to the concatenation $(\boldsymbol{a}^t, \boldsymbol{b}^t, \boldsymbol{\theta}^t)^t$ of $\boldsymbol{a} = (a(1), \ldots, a(I))^t$, $\boldsymbol{b} = (b(1), \ldots, b(J))^t$ and $\boldsymbol{\theta}$ is, from (5.15), the matrix M. Since

$$\frac{\partial T_{ij}}{\partial A(i)} = \frac{\partial \exp(t_{ij})}{\partial \exp[a(i)]} = T_{ij} \frac{\partial t_{ij}}{\partial a(i)} A(i)^{-1},$$

and since similar results hold for partial derivatives with respect to $B(j)$ and θ_k, it follows that the Jacobian matrix $\mathcal{J}[(\boldsymbol{A}^t, \boldsymbol{B}^t, \boldsymbol{\theta}^t)^t \to \boldsymbol{T}]$ of $\boldsymbol{T}$ with respect to $(\boldsymbol{A}^t, \boldsymbol{B}^t, \boldsymbol{\theta}^t)^t$ is

$$\begin{aligned}
\Psi_1 = {} & \mathrm{diag}\,(\boldsymbol{T}) \cdot M \\
& \cdot \ \mathrm{diag}\,(1/A(1), \ldots, 1/A(I), 1/B(1), \ldots, 1/B(J), 1, \ldots, 1)
\end{aligned} \tag{5.213}$$

where, as before, $\operatorname{diag}(\cdot)$ stands for a diagonal matrix, the diagonal elements of which are given within the parentheses. If the argument is a vector, then the diagonal elements are the components of the vector written in the same order. From (5.213), since M is a constant matrix, the Jacobian matrix of $M^t\boldsymbol{T}$ with respect to $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta})$ is

$$\begin{aligned}\Psi &= M^t \cdot \operatorname{diag}(\boldsymbol{T}) \cdot M \\ &\quad \cdot \operatorname{diag}(1/A(1), \ldots, 1/A(I), 1/B(1), \ldots, 1/B(J), 1, \ldots, 1).\end{aligned} \tag{5.214}$$

Unfortunately, this matrix is singular since M is not of full rank. However, as we have seen before, under Condition ML1, the matrix $M_{(2)}$ obtained by deleting one of the first $I + J$ columns of M is of full rank. Replacing M by $M_{(2)}$ in (5.214) causes no problem since such a replacement in (5.213) implies that one $A(i)$ or one $B(j)$ — say $B(J)$ — has been fixed as an arbitrary positive number, and making the replacement in (5.209) is permissible because one of the equations in (5.12) is redundant (since $\sum_i N_{i\oplus} = \sum_j N_{\oplus j}$). Set

$$\begin{aligned}\Phi &= M_{(2)}^t \cdot \operatorname{diag}(\boldsymbol{T}) \cdot M_{(2)} \\ &\quad \cdot \operatorname{diag}(1/A(1), \ldots, 1/A(I), 1/B(1), \ldots, 1/B(J-1), 1, \ldots, 1).\end{aligned} \tag{5.215}$$

From (5.214), this is the Jacobian matrix of $M^t\boldsymbol{T}$ with respect to

$$(A(1), \ldots, A(I), B(1), \ldots, B(J-1), \theta_1, \ldots, \theta_K)^t.$$

Since

$$\begin{aligned}&\operatorname{diag}(1/A(1), \ldots, 1/A(I), 1/B(1), \ldots, 1/B(J-1), 1, \ldots, 1)^{-1} \\ &\quad = \operatorname{diag}(A(1), \ldots, A(I), B(1), \ldots, B(J-1), 1, \ldots, 1),\end{aligned}$$

the Jacobian matrix of

$$(A(1), \ldots, A(I), B(1), \ldots, B(J-1), \theta_1, \ldots, \theta_K)^t$$

with respect to $M^t\boldsymbol{T}$ is

$$\begin{aligned}\Phi^{-1} = &\operatorname{diag}(A(1), \ldots, A(I), B(1), \ldots, B(J-1), 1, \ldots, 1) \\ &\cdot (M_{(2)}^t \operatorname{diag}(\boldsymbol{T}) M_{(2)})^{-1}.\end{aligned} \tag{5.216}$$

Since the N_{ij}'s have independent Poisson distributions, the covariance matrix $\operatorname{Cov}(\boldsymbol{N})$ of $\boldsymbol{N}$ is $\operatorname{diag}(\boldsymbol{T})$. Hence, the covariance matrix of $M_{(2)}^t\boldsymbol{N}$ is

$$M_{(2)}^t \operatorname{diag}(\boldsymbol{T}) M_{(2)} \tag{5.217}$$

and, because (5.209) holds for maximum likelihood estimates, (5.217) is also the covariance matrix of $M_{(2)}^t\hat{\boldsymbol{T}}$. Therefore, the covariance matrix of

$$(\hat{A}(1), \ldots, \hat{A}(I), \hat{B}(1), \ldots, \hat{B}(J-1), \hat{\theta}_1, \ldots, \hat{\theta}_K)^t$$

is

$$\Phi^{-1}M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)}(\Phi^{-1})^{t}, \tag{5.218}$$

which, using (5.218), (5.216) and

$$\begin{aligned}(M_{(2)}^{t}\ &\mathrm{diag}\,(\boldsymbol{T})M_{(2)})^{-1}(M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)})(M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)})^{-1}\\ &= (M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)})^{-1},\end{aligned}$$

shows that the covariance matrix of

$$(\hat{A}(1),\ldots,\hat{A}(I),\hat{B}(1),\ldots,\hat{B}(J-1),\hat{\theta}_1,\ldots,\hat{\theta}_K)^t$$

may also be written as

$$\begin{aligned}\mathrm{diag}\,&(A(1),\ldots,A(I),B(1),\ldots,B(J-1),1,\ldots,1)\\ &\cdot(M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)})^{-1}\\ &\mathrm{diag}\,(A(1),\ldots,A(I),B(1),\ldots,B(J-1),1,\ldots,1).\end{aligned} \tag{5.219}$$

As mentioned above, matrices (5.218) and (5.219) are actually asymptotic covariance matrices, although, since we do not have exact covariance matrices, we call them covariance matrices.

The matrix

$$((\mathrm{E}[(\partial\mathcal{L}/\partial\xi_s)(\partial\mathcal{L}/\partial\xi_p)])) \tag{5.220}$$

where ξ_s is the sth component of the vector $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{\theta})$ with $B(J)$ deleted, is called an *information matrix*. Write the vector

$$\left(\frac{\partial\mathcal{L}}{\partial\xi_1},\ldots,\frac{\partial\mathcal{L}}{\partial\xi_{I+J+K-1}}\right)^t$$

as $\boldsymbol{\ell}$. Then, (5.220) is $\mathrm{E}[\boldsymbol{\ell}\boldsymbol{\ell}^t]$. From (5.9), (5.10) and (5.11), it follows that

$$\begin{aligned}\boldsymbol{\ell} = \ &\mathrm{diag}\,(1/A(1),\ldots,1/A(I),1/B(1),\ldots,1/B(J-1),1,\ldots,1)\cdot\\ &(N_{1\oplus}-T_{1\oplus},\ldots,N_{I\oplus}-T_{I\oplus},N_{\oplus 1}-T_{\oplus 1},\ldots,N_{\oplus J-1}-T_{\oplus J-1},\\ &\quad\sum_{ij}c_{ij}{}^{(1)}[N_{ij}-T_{ij}],\ldots,\sum_{ij}c_{ij}{}^{(K)}[N_{ij}-T_{ij}])^t\\ = \ &\mathrm{diag}\,(1/A(1),\ldots,1/A(I),1/B(1),\ldots,1/B(J-1),1,\ldots,1)\\ &\cdot M_{(2)}^{t}[\boldsymbol{N}-\boldsymbol{T}].\end{aligned} \tag{5.221}$$

Now, since $\mathrm{E}[\boldsymbol{N}]=\boldsymbol{T}$, it follows that $\mathrm{E}\big([\boldsymbol{N}-\boldsymbol{T}][\boldsymbol{N}-\boldsymbol{T}]^t\big)$ is the covariance matrix of $\boldsymbol{N}$, which is $\mathrm{diag}\,(\boldsymbol{T})$. Thus, $\mathrm{E}[\boldsymbol{\ell}\boldsymbol{\ell}^t]$ becomes

$$\begin{aligned}\mathrm{diag}\,&(1/A(1),\ldots,1/A(I),1/B(1),\ldots,1/B(J-1),1,\ldots,1)\\ &\cdot M_{(2)}^{t}\ \mathrm{diag}\,(\boldsymbol{T})M_{(2)}\\ &\cdot\ \mathrm{diag}\,(1/A(1),\ldots,1/A(I),1/B(1),\ldots,1/B(J-1),1,\ldots,1).\end{aligned} \tag{5.222}$$

Therefore, the inverse of the information matrix $\mathrm{E}[\boldsymbol{\ell}\boldsymbol{\ell}^t]$ is (5.219). It is well known that this inverse is the lower bound of covariance matrices, in the sense that if V is the covariance matrix of any other set of estimates of the parameters $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta})$ with B_J deleted, then $V - \mathrm{E}[\boldsymbol{\ell}\boldsymbol{\ell}^t]$ is non-negative definite (e.g., see Rao, 1973, p. 326). This implies that each diagonal element of V is larger than the corresponding diagonal element of $\mathrm{E}[\boldsymbol{\ell}\boldsymbol{\ell}^t]$, i.e., the ML estimators of gravity model parameters have the least variance of all estimators. This is a standard property of ML estimates under frequently existing conditions. The existence of the conditions is not entirely immediate in the case of the gravity model. At any rate, our treatment above could be seen as an alternative derivation of the expression for the covariance matrix and also provides an independent verification of the applicability of the standard property.

The matrix $M_{(2)}^t \operatorname{diag}(\boldsymbol{T}) M_{(2)}$ has a very simple form:

$$M_{(2)}^t \operatorname{diag}(\boldsymbol{T}) M_{(2)} = \begin{pmatrix} U_1 & U_2 \\ U_2^t & U_3 \end{pmatrix} \tag{5.223}$$

where

$$U_1 = \begin{pmatrix} V_1 & V_2 \\ V_2^t & V_3 \end{pmatrix}, \quad U_2 = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \tag{5.224}$$

$U_3 = ((u_{kl}))$ with $u_{kl} = \sum_{ij} c_{ij}^{(k)} c_{ij}^{(l)} T_{ij}$, $W_1 = ((w_{ik}{}^{(1)}))$ with $w_{ik}{}^{(1)} = \sum_j c_{ij}^{(k)} T_{ij}$, $W_2 = ((w_{jk}{}^{(2)}))$ with $w_{jk}{}^{(2)} = \sum_i c_{ij}^{(k)} T_{ij}$

$$\begin{aligned} V_1 &= \operatorname{diag}(T_{1\oplus}, \ldots, T_{I\oplus}),\ V_2 = ((T_{ij})) \\ \text{and } V_3 &= \operatorname{diag}(T_{\oplus 1}, \ldots, T_{\oplus J-1}). \end{aligned} \tag{5.225}$$

(Notice that the subscript j in each of the matrices above goes only up to $J - 1$.) Computing these matrices algebraically and then inputting them into a computer would usually be better than inputting the component matrices in Φ, since the matrix $M_{(2)}$ would frequently be very large ($M_{(2)}$ has dimension $IJ \times (I+J+K-1)$ whereas $M_{(2)}^t \operatorname{diag}(\boldsymbol{T}) M_{(2)}$ is a square matrix of dimension $I + J + K - 1$). If we replace $A(i)$'s, $B(j)$'s, θ_k's and T_{ij}'s that appear in the right side of (5.219) by their estimates, we get the estimated covariance matrix of $(\hat{A}(1), \ldots, \hat{A}(I), \hat{B}(1), \ldots, \hat{B}(J-1), \hat{\theta}_1, \ldots, \hat{\theta}_K)^t$.

Obviously the bottom right $K \times K$ submatrix of matrix (5.219) is the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$. The bottom right $K \times K$ submatrix of the inverse of (5.223) is (see Rao, 1973, p. 33)

$$(U_3 - U_2^t U_1^{-1} U_2)^{-1}. \tag{5.226}$$

From (5.219), it follows that (5.226) is the covariance matrix of $\hat{\boldsymbol{\theta}}$. Thus, by replacing the T_{ij}'s with their estimates in (5.226) we get the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$.

If, instead of deleting the $(I+J)$-th column, we had deleted some other column, or had made M of full rank in some other way, the full covariance matrix obtained could have been different. However, the covariance matrix of $\boldsymbol{\theta}$ would have been the same, as we show below.

Theorem 5.6 *For a given* $\mathbf{c} = (c_{ij} : ij \in I \times J)$ *and under the conditions of Theorem 5.1,* $\mathrm{Cov}\,(\hat{\boldsymbol{\theta}})$ *is uniquely determined by* $\boldsymbol{T}$.

PROOF: By Lemma 5.4, for any observation of the random vector $\boldsymbol{N}$, the ML estimate $\hat{\boldsymbol{\theta}}$ is unique. Indeed, $\hat{\boldsymbol{\theta}}$ is a measurable function over the sample space of these observations. Since each N_{ij} is independently Poisson with parameter T_{ij}, the distribution of $\boldsymbol{N}$ is uniquely determined by $\boldsymbol{T}$. Consequently, the distribution of $\hat{\boldsymbol{\theta}}$ is uniquely determined by $\boldsymbol{T}$, and, in particular, so is $\mathrm{Cov}\,(\hat{\boldsymbol{\theta}})$. □

Since we are working with matrices, it might be of interest to offer an explanation of Theorem 5.6 in matrix terms. Consider small changes $\Delta\boldsymbol{A}$, $\Delta\boldsymbol{B}$ and $\Delta\boldsymbol{\theta}$ in $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{\theta}$ and let $\boldsymbol{T}$ correspondingly change by $\Delta\boldsymbol{T}$. Then

$$M\Delta\boldsymbol{T} = \Psi(\Delta\boldsymbol{A}^t, \Delta\boldsymbol{B}^t, \Delta\boldsymbol{\theta}^t)^t + o(\Delta\boldsymbol{T}) \tag{5.227}$$

where $o(\Delta\boldsymbol{T})$ stands for an expression such that $o(\Delta\boldsymbol{T})/||\Delta\boldsymbol{T}|| \to 0$ when $\Delta\boldsymbol{T} \to 0$, and $||\cdot||$ denotes the Euclidean norm. Hence,

$$\Psi^- M\Delta\boldsymbol{T} = (\Delta\boldsymbol{A}^t, \Delta\boldsymbol{B}^t, \Delta\boldsymbol{\theta}^t)^t + o(\Delta\boldsymbol{T}), \tag{5.228}$$

where Ψ^- is any member of the family of generalized inverses of Ψ. Letting $\Psi^{(-)}$ be a specific generalized inverse of Ψ, we can write

$$\Psi^- = \Psi^{(-)} + \tilde{\Psi}. \tag{5.229}$$

The last K rows of any such $\tilde{\Psi}$ would have to be zeros, since otherwise, $\Delta\boldsymbol{\theta}$ would not be unique and that would be a violation of Lemma 5.4. The theorem would then follow on inserting $\Psi^- M_{(2)}$ in place of $\Phi^{-1}M_{(2)}$ in (5.218) and then using (5.226).

Notice that when we used $M_{(2)}$ in place of M, we effectively used a particular generalized inverse, since if we augmented Φ^{-1} by a row and a column of zeros in the positions corresponding to where a row and a column got deleted in Ψ to get Φ, the resultant matrix would be a generalized inverse of Ψ (see Rao, 1973, p.27).

Replacing T_{ij}'s by their estimates in (5.226), the covariance matrix of $\boldsymbol{\theta}$ for the Skokie data set (with $c_{ij}{}^{(1)}$ = travel time and $c_{ij}{}^{(2)}$ = distance) was found to be

$$10^{-6} \times \begin{pmatrix} 18.5 & -63.7 \\ -63.7 & 277 \end{pmatrix}. \tag{5.230}$$

The very high correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is readily apparent from (5.230). In the case of the 'T_{ij}', '$T_{ij}/50$' and '$T_{ij}/100$' data sets, the known values of T_{ij} were inserted into (5.226) to obtain, respectively,

$$10^{-7} \times \begin{pmatrix} 33.5 & -5.86 \\ -5.86 & 437 \end{pmatrix}, \quad 10^{-5} \times \begin{pmatrix} 16.7 & -2.93 \\ -2.93 & 219 \end{pmatrix} \tag{5.231}$$

$$\text{and} \qquad 10^{-5} \times \begin{pmatrix} 33.5 & -5.86 \\ -5.86 & 437 \end{pmatrix}. \tag{5.232}$$

That these matrices are multiples of each other is most noticeable. The negative sign of the covariances should not be disconcerting. It shows that, if for some small shift in the observations, $\hat{\theta}_1$ were to increase, $\hat{\theta}_2$ would decrease to 'compensate.' [The situation is akin to that in ordinary least squares, where if we consider the model $\boldsymbol{y} = X\boldsymbol{\beta} = \boldsymbol{\epsilon}$, the covariance of the estimate of β is proportional to the *inverse* of (XX^t).]

Whether $\mathrm{Cov}((\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta}))$ is obtained from (5.218) — since $B(J)$ is set equal to a constant in (5.218), its variance and the covariances involving it are zeros — or whether $\mathrm{Cov}((\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta}))$ is obtained in some other way, it can readily be seen that

$$\mathrm{Cov}(\hat{\boldsymbol{T}}) = \Psi_1 \, \mathrm{Cov}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\theta}) \Psi_1^t. \tag{5.233}$$

5.7.2 Covariance of $\hat{T}_{ij}$

Quite frequently, in applications, $\boldsymbol{T}$ is estimated using O_i's and D_j's which are themselves estimated from exogenous models (e.g., trip generation and trip attraction models in urban transportation planning). In such cases, the computation of $\mathrm{Cov}(\hat{\boldsymbol{T}})$ is considerably simplified by the use of the LDSF Procedure, described in Section 5.4. This section is devoted to obtaining the covariance matrix $\mathrm{Cov}(\hat{\boldsymbol{T}})$ of $\hat{\boldsymbol{T}}$ when exogenously determined O_i's, D_j's and F_{ij}'s are used to estimate $\boldsymbol{T}$.

Since (5.114) and (5.115) are linear they may be expressed in matrix form. One way of writing them is as follows:

$$\boldsymbol{\Delta}^{(2r-1)} = \mathcal{H}_O \boldsymbol{\Delta}^{(2r-2)} \tag{5.234}$$

and

$$\boldsymbol{\Delta}^{(2r)} = \mathcal{H}_D \boldsymbol{\Delta}^{(2r-1)}, \tag{5.235}$$

where $\boldsymbol{\Delta}^{(s)} = (\Delta \boldsymbol{T}^{(s)^t}, \Delta \boldsymbol{O}^t, \Delta \boldsymbol{D}^t)^t$ is an $IJ + I + J$ dimensional vector in which $\Delta \boldsymbol{O}$ and $\Delta \boldsymbol{D}$ are as in Section 5.4.1, and

$$\Delta \boldsymbol{T}^{(s)} = (\Delta T^{(s)}{}_{11}, \ldots, \Delta T^{(s)}{}_{1J}, \ldots\ldots, \Delta T^{(s)}{}_{I1}, \ldots, \Delta T^{(s)}{}_{IJ})^t. \tag{5.236}$$

The augmenting of $\Delta \boldsymbol{T}^{(s)}$ by $\Delta \boldsymbol{O}$ and $\Delta \boldsymbol{D}$ in $\boldsymbol{\Delta}^{(s)}$ is for convenience.

It is quite tedious, but otherwise straightforward to verify that in (5.234)

$$\mathcal{H}_O = \mathcal{I} + \mathcal{D}_O \mathcal{Q}_O, \tag{5.237}$$

where

$$\mathcal{I} = \begin{pmatrix} I_{IJ} & 0 \\ 0 & 0 \end{pmatrix},$$

$\mathcal{D}_O$ is an $IJ + I + J$ dimensional diagonal matrix:

$$\begin{gathered}\mathrm{diag}\,(T_{11}/O_1, T_{12}/O_1, \ldots, T_{1J}/O_1, \\ \ldots\ldots, T_{I1}/O_I, \ldots, T_{IJ}/O_I, 1, \ldots, 1),\end{gathered} \tag{5.238}$$

and

$$\mathcal{Q}_O = \begin{pmatrix} -I_I \otimes \mathbf{1}_J(\mathbf{1}_J)^t & I_I \otimes \mathbf{1}_J & 0 \\ 0 & I_I & 0 \\ 0 & 0 & I_J \end{pmatrix} \tag{5.239}$$

where I_I is the identity matrix of dimension I, $\mathbf{1}_J$ is the vector $(1, \ldots, 1)^t$ of dimension J, 0's are null matrices and $\otimes$ is the usual Kronecker product. [The Kronecker (or tensor or direct) product of two matrices $A = ((a_{pq}))$ of dimension $P \times Q$ and $B = ((b_{rs}))$ of dimension $R \times S$ is defined as follows:

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1Q}B \\ \ldots & \ldots & \ldots \\ a_{P1}B & \ldots & a_{PQ}B \end{pmatrix}$$

where each $a_{pq}B$ is the matrix

$$\begin{pmatrix} a_{pq}b_{11} & \ldots & a_{pq}b_{1S} \\ \ldots & \ldots & \ldots \\ a_{pq}b_{R1} & \ldots & a_{pq}b_{RS} \end{pmatrix}.$$

The matrix $\mathcal{H}_O$ is written out explicitly (without using the Kronecker product) in Weber and Sen, 1985, pp. 322-323.] Throughout this book, the symbol $\mathbf{1}$ represents a vector consisting entirely of 1's.

Similarly,

$$\mathcal{H}_D = \mathcal{I} + \mathcal{D}_D \mathcal{Q}_D \tag{5.240}$$

where $\mathcal{D}_D$ is

$$\begin{gathered}\mathrm{diag}\,(T_{11}/D_1, T_{12}/D_2, \ldots, T_{1J}/D_J, \\ \ldots\ldots, T_{I1}/D_1, \ldots, T_{IJ}/D_J, 1, \ldots, 1)\end{gathered} \tag{5.241}$$

and

$$\mathcal{Q}_D = \begin{pmatrix} -\mathbf{1}_J(\mathbf{1}_J)^t \otimes I_I & 0 & \mathbf{1}_J \otimes I_I \\ 0 & I_I & 0 \\ 0 & 0 & I_J \end{pmatrix}. \tag{5.242}$$

Now that we have expressed the individual iterations of the LDSF procedure in matrix form, we turn our attention to the initial step, i.e., to the computation of $\boldsymbol{\Delta}^{(0)}$. Since, from (5.113) $\Delta \boldsymbol{T}^{(0)} = \phi \Delta \boldsymbol{F}$, where

$$\Delta \boldsymbol{F} = (\Delta F_{11}, \ldots, \Delta F_{1J}, \ldots\ldots, \Delta F_{I1}, \ldots, \Delta F_{IJ})^t$$

and

$$\phi = \mathrm{diag}\,(T_{11}/F_{11}, T_{12}/F_{12}, \ldots, \Delta T_{1J}/F_{1J}, \\ \ldots\ldots, T_{I1}/F_{I1}, \ldots, T_{IJ}/F_{IJ}),$$

it follows that one possible way of writing $\boldsymbol{\Delta}^{(0)}$ is as

$$\boldsymbol{\Delta}^{(0)} = \bar{\Upsilon}(\Delta \boldsymbol{F}^t, \Delta \boldsymbol{O}^t, \Delta \boldsymbol{D}^t)^t$$

where

$$\bar{\Upsilon} = \begin{pmatrix} \phi & 0 \\ 0 & I_{I+J} \end{pmatrix}.$$

However, as we see below, this is not entirely satisfactory and we need to modify $\bar{\Upsilon}$ somewhat.

When, with the superscript r denoting exponentiation, the sequences

$$(\mathcal{H}_D \mathcal{H}_O)^r \Upsilon \text{ and } \mathcal{H}_O (\mathcal{H}_D \mathcal{H}_O)^r \Upsilon, \tag{5.243}$$

multiply $(\Delta \boldsymbol{F}^t, \Delta \boldsymbol{O}^t, \Delta \boldsymbol{D}^t)^t$, they yield $\boldsymbol{\Delta}^{(2r)}$ and $\boldsymbol{\Delta}^{(2r+1)}$ respectively. The first IJ components of $\boldsymbol{\Delta}^{(2r-1)}$ and $\boldsymbol{\Delta}^{(2r)}$ are the left sides of (5.114) or (5.115) and the other components remain constant for all r. Hence, by Theorem 5.4, $\boldsymbol{\Delta}^{(s)}$'s converge, if T_{ij}'s are all positive, $T_{i\oplus} = O_i$, $T_{\oplus j} = D_j$ and $\Delta O_\oplus = \Delta D_\oplus$.

However, if $(\Delta \boldsymbol{F}^t, \Delta O^t, \Delta D^t)^t$ is chosen in such a way that $\Delta O_\oplus \neq \Delta D_\oplus$, the $\boldsymbol{\Delta}^{(s)}$'s need not converge (and, in fact, do not converge — see Weber, 1987). Then, the matrix series (5.243) does not necessarily converge.

Define the matrix Υ as $\bar{\Upsilon}$ with its last row replaced by

$$(\mathbf{o}_{IJ}^t, (\mathbf{1}_I)^t, -(\mathbf{1}_{J-1})^t, 0),$$

where $\mathbf{o}$ is, as before, a null vector of dimension given by the subscript. Then, whatever, ΔO_i's and ΔD_j's we might start with, Υ replaces the ΔD_j with a new $\Delta D_j = \sum_{i=1}^{I} \Delta O_i - \sum_j^{J-1} \Delta D_j$ for which $\Delta O_\oplus = \Delta D_\oplus$, i.e., the last J elements, and the I elements preceding them, in the vector

$$\boldsymbol{\Delta}^{(0)} = \Upsilon(\Delta \boldsymbol{F}^t, \Delta O^t, \Delta D^t)^t, \tag{5.244}$$

have the same sum. If for the original ΔO_i's and ΔD_j's, $\Delta O_\oplus = \Delta D_\oplus$, then ΔD_J remains unchanged.

Now consider the sequence $\mathcal{K}^{(s)} : s = 1, 2, \ldots$ described by

$$\mathcal{K}^{(2r)} = (\mathcal{H}_D \mathcal{H}_O)^r \Upsilon \tag{5.245}$$

and

$$\mathcal{K}^{(2r+1)} = \mathcal{H}_O(\mathcal{H}_D\mathcal{H}_O)^r\Upsilon. \tag{5.246}$$

By Theorem 5.4 and Corollary 5.3, as well as the fact that $\mathcal{H}_O$ and $\mathcal{H}_D$ are matrix representations of the steps of the LDSF procedure given by (5.114) and (5.115), it follows that $\mathcal{K}^{(s)}\boldsymbol{\alpha}$ converges to a unique limit no matter what $\boldsymbol{\alpha}$ may be. Thus $\mathcal{K}^{(s)}$ must itself converge. Let the limit be $\mathcal{K}$ and let its first $I+J$ rows form a matrix $\mathcal{K}^*$. Since when (5.245) and (5.246) multiply $(\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t$ they yield $\Delta\boldsymbol{T}^{(s)}$ obtained at the sth iteration of the LDSF procedure,

$$\Delta\boldsymbol{T} = \mathcal{K}^*(\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t, \tag{5.247}$$

where $\Delta\boldsymbol{T}$ is the unique limit of the LDSF procedure when $\Delta\boldsymbol{F}$, $\Delta\boldsymbol{O}$ and $\Delta\boldsymbol{D}$ are given, as are all the F_{ij}'s and T_{ij}'s (or, alternatively, by virtue of the DSF procedure, the F_{ij}'s, O_i's and D_j's).

It is important to note that both $\mathcal{H}_O$ and $\mathcal{H}_D$ are free of ΔO_i, ΔD_j and ΔF_{ij}, as of course is Υ, and hence, so is $\mathcal{K}$.

Now consider

$$\mathcal{J}((\boldsymbol{F}^t, \boldsymbol{O}^t, \boldsymbol{D}^t)^t \to \boldsymbol{T}) = \Gamma \text{ (say)}. \tag{5.248}$$

Then, we must have

$$\Delta\boldsymbol{T} = \Gamma(\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t. \tag{5.249}$$

Thus Γ and $\mathcal{K}^*$ are essentially the same, in the sense that

$$[\Gamma - \mathcal{K}^*](\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t = \mathbf{o}. \tag{5.250}$$

This equivalence follows from the fact that $\Delta\boldsymbol{T}$ is uniquely and linearly related to $(\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t$. However, because there is a linear relation among the elements of $(\Delta\boldsymbol{F}^t, \Delta\boldsymbol{O}^t, \Delta\boldsymbol{D}^t)^t$ (i.e., the vectors form a subspace of dimension smaller than that of the vectors themselves), Γ is not necessarily exactly the same as $\mathcal{K}^*$. For example, if one of the rows of $\Gamma - \mathcal{K}^*$ were $(\mathbf{o}_{IJ}, (\mathbf{1}_I)^t, -(\mathbf{1}_J)^t)$ and other rows consisted of zeros, (5.250) could hold. Nevertheless, since the covariance of $\Delta\boldsymbol{T}$ is $\mathrm{E}[(\Delta\boldsymbol{T})(\Delta\boldsymbol{T})^t]$, and $\Delta\boldsymbol{T}$ is uniquely determined by either Γ or $\mathcal{K}^*$, treating these two last-mentioned matrices interchangeably presents no problem for us.

Therefore, we may use

$$\begin{aligned} \mathrm{Cov}\,(\hat{\boldsymbol{T}}) &= \Gamma\ \mathrm{Cov}\,(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t\Gamma^t \\ &= (\mathcal{K}^*)\ \mathrm{Cov}\,(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t(\mathcal{K}^*)^t, \end{aligned} \tag{5.251}$$

where $\hat{\boldsymbol{F}} = (\hat{F}_{11}, \ldots, \hat{F}_{1J}, \ldots\ldots, \hat{F}_{I1}, \ldots, \hat{F}_{IJ})^t$, $\hat{\boldsymbol{O}} = (\hat{O}_1, \ldots, \hat{O}_I)^t$ and $\hat{\boldsymbol{D}}_{(J)} = (\hat{D}_1, \ldots, \hat{D}_{J-1})^t$, to compute the covariance matrix of $\hat{\boldsymbol{T}}$. The computation of $\mathrm{Cov}\,[\hat{\boldsymbol{T}}]$ was the primary goal of this section.

We suggest some ways for actually carrying out the computations after the following remark. An alternative way of essentially calculating Γ was actually carried out in Section 5.4.1. An inspection of (5.116), and (5.117) and the steps of the proof of Lemma 5.6 [after stripping out the superscripts $(2r)$ or $(2r-1)$ and replacing $\Delta T_{i\oplus}$'s and $\Delta T_{\oplus j}$'s by ΔO_i's and ΔD_j's] will reveal that $\Phi_2\Phi_1\Phi_3$ is the Jacobian of $(\boldsymbol{F}^t, \boldsymbol{O}^t, (\boldsymbol{D}^*)^t)^t$ with respect to $(\boldsymbol{T}^t, \boldsymbol{A}^t, (\boldsymbol{B}^*)^t)^t$ where $\boldsymbol{D}^* = (D_1, \ldots, D_{J-1})^t$ and $\boldsymbol{B}^* = (B_1, \ldots, B_{J-1})^t$. Consequently, Φ, which is non-singular, is the Jacobian of $(\boldsymbol{T}^t, \boldsymbol{A}^t, (\boldsymbol{B}^*)^t)^t$ with respect to $(\boldsymbol{F}^t, \boldsymbol{O}^t, (\boldsymbol{D}^*)^t)^t$. Thus, the first IJ rows of Φ would play the same role as Γ with its last column deleted. However, this is not too useful for applications. Although Φ_1 is in rather a convenient form for numerical inversion, its size is most daunting. $\mathcal{K}^*$ provides a more promising alternative.

A Computational Suggestion

Fortunately, in the type of situations for which the discussion of this section is intended, certain simplifications may be available. If $\hat{F}_{ij} = F(\hat{\boldsymbol{\theta}}^t \boldsymbol{c}_{ij})$, then $\hat{F}_{ij}$ typically depends on a small number K of parameter estimates $\hat{\theta}_1, \ldots, \hat{\theta}_K$ and fixed values of $\boldsymbol{c}_{ij}$. O_i is often estimated using a model of the form $\hat{O}_i = \boldsymbol{\gamma}^t \boldsymbol{z}_i$ where $\boldsymbol{\gamma}$ consists of estimates of parameters and $\boldsymbol{z}_i$ consists of fixed numbers (e.g., as in the case of trip generation models, including categorical trip generation models — see Thakuriah, *et al.* 1993), and D_j is similarly estimated as a linear function with a small number of parameter estimates $\boldsymbol{\delta}$.

Let X be the Jacobian of $(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t$ with respect to these parameter vectors $(\hat{\boldsymbol{\theta}}^t, \boldsymbol{\gamma}^t, \boldsymbol{\delta}^t)^t$. If $\mathcal{V}$ is the covariance matrix of $(\hat{\boldsymbol{\theta}}^t, \boldsymbol{\gamma}^t, \boldsymbol{\delta}^t)^t$, the (asymptotic) covariance matrix of $(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t$ is $X\mathcal{V}X^t$. Since $\mathcal{V}$ is a covariance matrix and hence symmetric, it may be written as $\mathcal{U}\mathcal{D}\mathcal{U}^t$ where $\mathcal{D}$ is a diagonal matrix with diagonal elements which are the eigenvalues of $\mathcal{V}$; and $\mathcal{U}$ is the orthogonal matrix of the corresponding eigenvectors. Both can be computed using standard packages. The covariance matrix $\mathrm{Cov}\,[(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t]$ of $(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t$ may be written as

$$X\mathcal{U}\mathcal{D}^{1/2}\mathcal{D}^{1/2}\mathcal{U}^t X^t.$$

Let $X^{(1)} = X\mathcal{U}\mathcal{D}^{1/2}$. Then

$$\mathrm{Cov}\,[(\hat{\boldsymbol{F}}^t, \hat{\boldsymbol{O}}^t, \hat{\boldsymbol{D}}_{(J)}{}^t)^t] = X^{(1)}(X^{(1)})^t. \tag{5.252}$$

From (5.251) it follows that $\mathrm{Cov}\,(\hat{\boldsymbol{T}}) = \Gamma X^{(1)}(\Gamma X^{(1)})$. We can apply the LDSF procedure to each column of $X^{(1)}$ treating it as if it were $\boldsymbol{\Delta}^{(0)}$. This would have the effect of multiplying $X^{(1)}$ by the first IJ rows of (5.245) or (5.246), or of approximately multiplying $X^{(1)}$ by Γ. Let the resultant matrix be $X^{(2)} \approx \Gamma X^{(1)}$. The covariance of $\hat{\boldsymbol{T}}$ is

$$\mathrm{Cov}\,(\hat{\boldsymbol{T}}) \approx X^{(2)}(X^{(2)})^t. \tag{5.253}$$

Thus if $\boldsymbol{x}_s{}^{(2)}$ is the sth row of $X^{(2)}$, the variance of $\hat{T}_{ij}$ is $\boldsymbol{x}_s{}^{(2)}(\boldsymbol{x}_s{}^{(2)})^t$ where $s = (i-1)J + j$. The covariance of $\hat{T}_{ij}$ and $\hat{T}_{pq}$ is $\boldsymbol{x}_s{}^{(2)}(\boldsymbol{x}_r{}^{(2)})^t$ where $r = (p-1)J + q$. For particularly large IJ's, the matrix $X^{(2)}$ can be computed and stored and particular variances or covariances can be computed as needed.

Recall that one or two pairs of iterations of the LDSF frequently gets us very close to the limit. Moreover, after computing the requisite sums, each step of the LDSF procedure can be conducted for each element separately. Thus, with some care in programming, no computer storage problems need occur.

Some further economies might occur if (5.118) and (5.119) were used in place of the usual LDSF iterations (5.114) and (5.115). However, we have not explored this possibility.

A Numerical Example

For a numerical example, data were computed in the following way. F_{ij}'s were as described in (5.143), O_i's and D_j's as just under (5.143) and T_{ij}'s as in (5.144). The variances of each F_{ij} was set equal to $.05F_{ij}$ and that of each O_i and D_j were set as $.25O_i$ and $.25D_j$, respectively. All covariances were taken to be zeros, which is a bit unrealistic (since $O_\oplus = D_\oplus$). However, given the procedure described above this does not matter much.

Exhibit 5.19 shows the variances of $\hat{T}_{ij}$'s obtained by using (5.251). The full covariance matrix of the $\hat{T}_{ij}$'s is given in Weber and Sen (1985), but, in order to conserve space, it has not been shown here. Since the matrix of T_{ij}'s was small, matrix methods were used — i.e., the method outlined in the last subsection was not used. Also shown (within parentheses) in Exhibit 5.19 are variances estimated from 10,000 Monte Carlo iterations. Clearly the asymptotic variances computed correspond closely to the Monte Carlo variances. We then computed 90 per cent confidence intervals for each of the estimated variances and covariances from the simulations and found that 232 out of 256 (90.6 per cent) of these confidence intervals included the corresponding computed variances and covariances.

5.7.3 Other Forecasts

Since the primary purpose of gravity models is forecasting, we pay special attention to formulæ for forecasting here, even at the risk of some redundancy. As mentioned in the introduction to this book, the gravity model may be used for forecasting in several different ways, depending on what estimates are available. Let a superscript f denote that the quantity is

	$j=1$	$j=2$	$j=3$	$j=4$
$i=1$	19.00 (18.98)	6.97 (7.10)	1.38 (1.41)	44.50 (43.63)
$i=2$	95.80 (95.30)	38.90 (38.74)	8.93 (9.07)	170.00 (170.03)
$i=3$	14.20 (14.20)	5.95 (6.00)	0.928 (0.931)	30.60 (30.58)
$i=4$	139.00 (138.63)	58.60 (58.46)	15.10 (15.45)	285.00 (288.26)

Exhibit 5.19: Variances of $\hat{T}_{ij}$'s Computed by the Method Described and Variances Estimated by 10000 Monte Carlo Simulations (within parentheses)

an estimate for the forecast period. If ${c_{ij}}^{(k)f}$, $\boldsymbol{\theta}^f$, $A^f(i)$'s and $B^f(j)$'s are available, T_{ij} for the forecast period may be estimated by

$$T_{ij}^f = A^f(i)B^f(j)\exp[(\boldsymbol{\theta}^f)^t \boldsymbol{c}_{ij}^f]. \tag{5.254}$$

Here, typically, $\boldsymbol{c}_{ij}^f$ would be available exogenously, $\boldsymbol{\theta}^f$ would be the estimate $\hat{\boldsymbol{\theta}}$ from the base period which is assumed to remain unchanged into the forecast period. The estimates $A^f(i)$ and $B^f(j)$ could be base period estimates assumed to remain unchanged into the forecast period, or one or both of them could be exogenous. If ${c_{ij}}^{(k)f}$, $\boldsymbol{\theta}^f$, $B^f(j)$'s and ${T_{\oplus j}}^f$'s are available, then we can use

$$T_{ij}^f = {T_{\oplus j}}^f B^f(j)F_{ij}^f / [\sum_{j=1}^{J} B^f(j)F_{ij}^f] \tag{5.255}$$

where $F_{ij}^f = \exp[(\boldsymbol{\theta}^f)^t \boldsymbol{c}_{ij}^f]$. Here, typically, $\boldsymbol{c}_{ij}^f$ and $\boldsymbol{\theta}^f$ would be as described above and $B^f(j)$'s and ${T_{\oplus j}}^f$ could be base period estimates assumed to remain unchanged into the forecast period, or could be obtained exogenously. However, ${T_{\oplus j}}^f$ would usually be obtained exogenously. If ${c_{ij}}^{(k)f}$, $\boldsymbol{\theta}^f$, $A^f(i)$'s and ${T_{i\oplus}}^f$'s are available, then we can use

$$T_{ij}^f = {T_{i\oplus}}^f A^f(i)F_{ij}^f / [\sum_{i=1}^{I} A^f(i)F_{ij}^f] \tag{5.256}$$

and the source of estimates would be similar to those for (5.255). Equations (5.255) and (5.256) follow directly from (5.2), and also from equations (8) and (9) in the introduction to the book. If ${c_{ij}}^{(k)f}$, $\boldsymbol{\theta}^f$, ${T_{\oplus j}}^f$'s and ${T_{i\oplus}}^f$'s are available, then would use the DSF procedure as illustrated in Section 5.7.2. In this case, the ${T_{\oplus j}}^f$'s and ${T_{i\oplus}}^f$ would usually be exogenous while $\boldsymbol{c}_{ij}^f$ and $\boldsymbol{\theta}^f$ would be as before.

The computation of covariance matrices when the DSF procedure is used has already been discussed in Section 5.7.2. In the other cases, (asymptotic) covariance matrices can be obtained as a straightforward, although sometimes tedious, application of (5.208). For example, consider (5.254). Clearly (5.213) is the Jacobian matrix of T_{ij}^f's with respect to the $A^f(i)$'s, $B^f(j)$'s and θ_k^f's. Thus, insertion of this into the last line of (5.208) would yield the necessary covariance matrix. Obviously, we need the covariance matrix of the $A^f(i)$'s, $B^f(j)$'s and θ_k^f's for this purpose. For those estimates that remain unchanged from the base period, the appropriate covariance matrix is the one for the base period. For estimates obtained exogenously, the covariance matrix needs to be supplied exogenously also. Covariances of estimates obtained from different independent data are usually assumed to be zeros. For example, if the $A^f(i)$'s and θ_k^f's are base period estimates, while the $B^f(j)$'s are exogenously given, we would use the base period covariance matrix for the $A^f(i)$'s and θ_k^f's and the covariances between these estimates and the $B^f(j)$'s would be zeros. Covariance matrices for (5.255) require computation of its Jacobian, which is straightforward, followed by the use of (5.208). The covariance matrix for the $B^f(j)$'s, θ_k^f's and $T_{\oplus j}{}^f$'s needs to be developed as just mentioned for (5.254). An application of (5.256) can be treated similarly.

5.8 Goodness of Fit

In this section, we discuss the assessment of goodness of fit both from the standpoint of the entire model (Section 5.8.1) and from that of individual observations or collections of observations (Section 5.8.2).

Notice that the user of (5.2) needs to decide what $c_{ij}^{(k)}$'s should be used. Theoretical arguments are not sufficient guides for this purpose. As in the case of any statistical model building, examination of the data, intuition, and experimentation with an open mind are frequently required. The recommended style of approach is that described in any recent applied statistics text (e.g., Sen and Srivastava, 1990, McCullagh and Nelder, 1989, Bishop, *et al.*, 1975).

The difficulty that this approach presents in the present context is that adequate diagnostic tools are not available. Thus, it might be desirable to make preliminary runs using linear least squares, for which a wide array of diagnostic tools have been constructed and are available in statistical packages. Least squares methods are discussed in the next chapter.

However, overall assessment of models, comparison of models and some outlier analysis is still possible. This section is devoted to these subjects.

5.8.1 Global Measures

There are a number of measures to assess the overall goodness of fit of a model. Perhaps the best known, for Poisson variates is the (Pearson) 'Chi-square' statistic [often written as χ^2].

$$X^2 = \sum \frac{(\text{Observed} - \text{Estimated})^2}{\text{Estimated}} = \sum_{ij} \frac{(N_{ij} - \hat{T}_{ij})^2}{\hat{T}_{ij}}, \tag{5.257}$$

which may be roughly justified by the fact that

$$\mathrm{E}[(N_{ij} - T_{ij})^2] = \mathrm{var}\,[N_{ij}] = T_{ij}.$$

Alternatives to (5.257) include the (Neyman's) modified 'Chi-square' statistic

$$\sum \frac{(\text{Observed} - \text{Estimated})^2}{\text{Observed}} = \sum_{ij} \frac{(N_{ij} - \hat{T}_{ij})^2}{N_{ij}}, \tag{5.258}$$

when $N_{ij} \neq 0$, the simplified Freeman-Tukey statistic

$$\sum_{ij} 4[\sqrt{N_{ij}} - \sqrt{\hat{T}_{ij}}]^2 \tag{5.259}$$

and the scaled deviance (5.195) which, when comparing different ML estimated models can be written as (since $\hat{T}_{\oplus\oplus} = N_{\oplus\oplus}$)

$$2\sum_{ij}[N_{ij}\log(N_{ij}/\hat{T}_{ij})]. \tag{5.260}$$

(See Rao, 1973, p. 352, for additional possibilities.) We use the word 'simplified' when referring to the Freeman-Tukey statistic because the original form is

$$\sqrt{N_{ij}} + \sqrt{N_{ij}+1} - \sqrt{4\hat{T}_{ij}+1}.$$

Since $[N_{ij} - \hat{T}_{ij}]^2/\hat{T}_{ij} = [(N_{ij}\hat{T}_{ij}^{-1/2} - \hat{T}_{ij}^{1/2}]^2$, when each N_{ij} is approximately $\hat{T}_{ij}$, at an intuitive level, the equivalence of (5.259), (5.258) and (5.257) is easily seen, and that of (5.260) and (5.257) follows from (5.70). A formal proof that all four statistics, (5.257), (5.258), (5.259) and (5.260), have asymptotically equivalent distributions is given in Bishop, *et al.* (1975, pp. 513–516). Moreover, when $\hat{T}_{ij}$ is obtained using maximum likelihood, (5.257) has a Chi-square distribution with, under ML1, df$= IJ - I - J - K + 1$ degrees of freedom (see Bishop, *et al.*, 1975, pp. 516–18, Rao, 1973, pp. 391–392).

Each of the goodness of fit statistics given above has relative advantages. For example, if two models yield estimates $\hat{T}_{ij}{}^{(1)}$ and $\hat{T}_{ij}{}^{(2)}$, (5.260) yields by subtraction the statistic $2\sum_{ij} N_{ij}\log[\hat{T}_{ij}{}^{(1)}/\hat{T}_{ij}{}^{(2)}]$. Similarly, it allows easy formation of likelihood ratio test statistics for testing whether individual θ_k's are zero. However, since the Chi-square statistic (5.257) has an easy intuitive meaning and is most widely used, the discussion of the rest of this section is confined to it.

As an aside, it might be mentioned that parameter estimation can occur by minimizing any of the goodness of fit statistics. Since, in this chapter, we consider maximum likelihood, we are discussing methods based on the minimization of the scaled deviance. In the next chapter we shall briefly discuss the minimization of (5.257) and (5.258), as well as $[\log(N_{ij}+\frac{1}{2})-\hat{t}_{ij}]^2$. Clearly each criterion favors results of methods based on it. Consequently, one needs to be very careful when using a goodness of fit statistic to compare different methods of estimation (e.g., least squares vs. maximum likelihood). While such a statement is trite, the literature contains examples where estimates using two procedures are compared using a criterion favoring a specific procedure.

Distribution of the Chi-Square Statistic

As is well known, when T_{ij}'s are large, X^2 has the chi-square distribution (for more on the subject of how large, see Madansky, 1988, p. 43). The difficulty in using this test or one of its asymptotic equivalents in a gravity model context is that several of the T_{ij}'s can be small. In urban transportation studies, most T_{ij}'s are very small. This is usually not because the total $T_{\oplus\oplus}$ is small; rather, the trips are spread out over a large number of i,j-pairs. Since the amount of trip data collected is large (typically in the thousands), it would appear that a reasonable asymptotic distribution should be available even under such circumstances, by exploiting the size of I and J via the law of large numbers and the central limit theorem. This turns out to be the case as we now discuss.

If enough data are available, often $\hat{T}_{ij} \approx T_{ij}$ even if several individual T_{ij}'s are small. In the case of Poisson gravity models given by (5.1) and (5.2), this is because of the existence of sufficient statistics $N_{i\oplus}$, $N_{\oplus j}$ and $\sum_{ij} c_{ij}{}^{(k)} N_{ij}$, which can be, (and very often are) large, even though individual N_{ij}'s may not be so. When $\hat{T}_{ij} \approx T_{ij}$, $X^2 \approx Z^2$ where

$$Z^2 = \sum_{ij} \frac{(N_{ij}-T_{ij})^2}{T_{ij}}. \tag{5.261}$$

Since $\mathrm{E}[N_{ij}] = T_{ij}$ and because N_{ij}'s have independent Poisson distributions

$$\operatorname{var}[N_{ij}] = \mathrm{E}(N_{ij}-T_{ij})^2 = T_{ij}. \tag{5.262}$$

Hence,

$$\mathrm{E}[Z^2] = IJ. \tag{5.263}$$

Thus, the so called *Chi-square ratio*, X^2/df, has an expectation which is asymptotically 1.

From (5.262) and (5.263), it follows that the variance $\operatorname{var}[Z^2]$ of Z^2 is

$$\mathrm{E}\left[\sum_{ij}\left(\frac{(N_{ij}-T_{ij})^2}{T_{ij}}-1\right)\right]^2 = \mathrm{E}\left[\sum_{ij}\left(\frac{(N_{ij}-T_{ij})^2}{T_{ij}}-1\right)^2\right], \quad (5.264)$$

the right side following from the fact that the covariance of two independent random variables is zero. From (5.262), $\mathrm{E}[(N_{ij}-T_{ij})^2/T_{ij}] = 1$ and hence, from (5.264), it follows that

$$\begin{aligned}\operatorname{var}[Z^2] &= \mathrm{E}\left[\sum_{ij}\left(\frac{(N_{ij}-T_{ij})^4}{T_{ij}^2}-1\right)\right] \\ &= \sum_{ij}\left[\frac{T_{ij}+3T_{ij}^2}{T_{ij}^2}-1\right] \approx \sum_{ij}[T_{ij}^{-1}+2],\end{aligned} \quad (5.265)$$

using the usual expression for the fourth moment of a Poisson distribution (see (6.19) in Chapter 6). Thus,

$$\operatorname{var}[Z^2/(IJ)] \approx \sum_{ij}[(T_{ij}I^2J^2)^{-1} + 2(IJ)^{-2}]. \quad (5.266)$$

Hence, if T_{ij}'s are bounded away from zero, the variance of $Z^2/IJ \longrightarrow 0$ as $IJ \longrightarrow \infty$. It follows that when $\hat{T}_{ij} \longrightarrow T_{ij}$ and T_{ij}'s are bounded away from zero, the variance of $X^2/\mathrm{df} \longrightarrow 0$. When T_{ij}'s are fairly large, this convergence would occur fairly rapidly. In practical applications, if I and J are reasonably large and no T_{ij} too small, we would consider the variance to be negligible.

However, when T_{ij}'s are very small and I and J are not very large, the variance may not be negligible. By using the estimates $\hat{T}_{ij}$ in lieu of T_{ij}, we can estimate the variance of the Chi-square ratio. In the next subsection we illustrate such computations for several data sets.

It may be noted that for large IJ, Z^2 is approximately normal. This can be shown by the use of the Liapunov version of the central limit theorem (Rao, 1973, p. 127), which requires that $[\sum_{ij}\tau_{ij}]^{1/3}/[\sum_{ij}\sigma_{ij}^2]^{1/2} \longrightarrow 0$ as $I \longrightarrow \infty$ and $J \longrightarrow \infty$, where τ_{ij} is the third absolute moment $\mathrm{E}[|x_{ij}-1|^3]$ of each term $x_{ij} = [N_{ij}-T_{ij}]^2/T_{ij}$ comprising Z^2 [note that $\mathrm{E}[x_{ij}] = 1$] and σ_{ij}^2 is the corresponding variance $\mathrm{E}[x_{ij}-1]^2$. It is easy, though slightly tedious, to verify that $\tau_{ij} \approx 8 + 22T_{ij}^{-1} + T_{ij}^{-2}$ which like $\sigma^2 \approx 2 + T_{ij}^{-1}$ is bounded by a positive number if T_{ij} is bounded away from zero. Thus,

$$[\sum_{ij}\tau_{ij}]^{1/3}/[\sum_{ij}\sigma_{ij}^2]^{1/2}$$

can be seen to be of order $[IJ]^{-1/6} \longrightarrow 0$.

In assessing goodness of fit in several gravity model studies one of the authors has used the fact, established above, that under extremely mild conditions, the Chi-square ratio has mean approximately 1 and a readily computable variance, which is frequently close to zero. Fortunately, we have found the test to be quite sensitive to even small departures from a good fit, so that even a rough rule of thumb has been adequate. In practical situations, since the Poisson assumption seldom holds perfectly, one can use a rough rule of thumb to assess a gravity model as fitting well if, say, the Chi-square ratio is less than 2.

EXAMPLES

In this section we apply the formulæ presented above to several different data sets, including simulated data based on the Skokie data set and the hospital patient flow data set (see Section 5.6.1). Average simulated N_{ij}'s for the hospital 'T_{ij}' and Skokie 'T_{ij}' are about 22 and 35 trips, respectively, which, of course, makes the averages for the '$T_{ij}/100$' data sets about .22 and .35.

We obtained both the Chi-square ratio and the ratio $\mathrm{E}[Z^2]/IJ$, a principal difference between the two being that in the latter case, we used T_{ij}'s which we knew and had used previously to generate the N_{ij}'s and in the former case, we used $\hat{T}_{ij}$'s which were obtained by estimating gravity model parameters using N_{ij}'s as observations. In each case, the Chi-square ratio was computed along with its standard deviation using (5.266). The results are shown in Table 5.20 and 5.21.

Data Type	Hospital Data	Skokie Data
'T_{ij}'	0.98 (0.10)	1.17 (0.02)
'$T_{ij}/10$'	1.04 (0.33)	1.10 (0.06)
'$T_{ij}/100$'	0.80 (1.07)	1.04 (0.21)

Exhibit 5.20: Z^2/IJ Ratios and Standard Errors (in Parens.)

Data Type	Hospital Data	Skokie Data
'T_{ij}'	0.98 (0.08)	1.04 (0.01)
'$T_{ij}/10$'	0.98 (0.22)	0.91 (0.07)
'$T_{ij}/100$'	0.74 (0.65)	0.87 (0.17)

Exhibit 5.21: Chi-square Ratios and Standard Errors (in Parens.)

The values of the Chi-square ratio in all but one case for both tables lie within the 95 per cent confidence interval. When sample sizes are very small (e.g., as in the '$T_{ij}/100$' data set), the standard errors are not negligible.

An additional data set, which has not been previously introduced was also used. These data, which we call the sampled shopping trip data was used in Kim, Sööt and Sen (1993, as well as in Kim, 1993, and in Kim *et al.*, 1994). This data set consists of a number of different origin destination matrices $((N_{ij}{}^{(s)}))$ where each $N_{ij}{}^{(s)}$ is the number of shopping trips between i and j in DuPage County, which is near Chicago. The subscript s represents the order of the trip within its own circuit, a circuit being a sequence of trips taken by the same person, starting from home and ending at home. For example, $N_{ij}{}^{(1)}$ is the number of trips that start at home, which is at i, and end at a shop in j; $N_{ij}{}^{(2)}$ is the number of trips that start at an origin in i which was reached from home via one trip and end at a shop in j. The largest of the $N_{\oplus\oplus}{}^{(s)}$'s was 1,145 shopping trips involving 240 origins zones and 167 shopping destination zones. The smallest was 40 shopping trips involving 39 origin zones and 31 destination zones. The value of s is called a stage and the number of stages in a circuit determine its circuit size.

Table 5.22 shows Chi-square ratios and their standard errors for separate gravity models fitted to observations for each circuit size and stage in the shopping data. Since these data are actual observations (not simulated), we had expected that the Chi-square ratios would be larger than 1.

Except in two cases, a formal test would indicate that, at a five per cent level, there is no reason to reject the hypothesis that the model constructed is not appropriate. However, the nearly uniformly lower than 1 values of the ratio is noticeable and possibly results from the fact that IJ is not big enough to compensate for the large number of zero-valued N_{ij}'s (also see next subsection). This phenomenon for N_{ij} matrices containing mainly zeros has also been noticed by Boyle and Flowerdew (1993) [and a suggestion of adjustment by altering the degrees of freedom has been made by Paul Metaxotos, a student of one of the authors]. Nevertheless, a rule of thumb like the use of 2 as a cutoff for the Chi-square ratio, still seems reasonable in most cases one might encounter in practice.

Further Examples

The values of the Chi-square statistic for the Skokie data using different $\hat{\boldsymbol{\theta}}^t \boldsymbol{c}_{ij}$'s in different procedures [and, hence, in this case, different models] are given in Exhibit 5.23 along with the ratio X^2/df of the Chi-square value to the appropriate degrees of freedom.

As mentioned earlier, under the Poisson assumption and with T_{ij}'s large enough, $E(X^2/\mathrm{df}) \approx 1$. Thus if the model (5.2) fit perfectly and $\boldsymbol{c}_{ij}$ was chosen perfectly, the ratio of Chi-square to degrees of freedom would be close to 1, especially given the large number of degrees of freedom. However, we know that trips are not independent. About the time these data were collected, there was nationwide, an average of 1.4 person-trips per vehicle-trip. This alone would raise the expectation of the ratio to 1.4. Be-

	cs= 2	cs= 3	cs= 4	cs= 5
Stage 1	0.68 (0.18)	0.71 (0.28)	0.67 (0.31)	1.46 (0.84)
Stage 2		0.60 (0.11)	0.65 (0.24)	0.64 (0.38)
Stage 3			0.58 (0.14)	0.72 (0.25)
Stage 4				0.97 (0.29)

Exhibit 5.22: Chi-square Ratios and their Standard Errors (in parentheses) for Different Circuit Sizes (cs) in Shopping Data Set

Procedure	X^2	X^2/df
Evans-Kirby (trv. time)	2143	3.24
Hyman Proc. (${c_{ij}}^{(1)}$ = trv. time)	2553	3.86
Gen. Proc. (${c_{ij}}^{(1)}$ =trv. time)		
(${c_{ij}}^{(2)} = \log[{c_{ij}}^{(1)}]$)	2551	3.86
(${c_{ij}}^{(2)}$ = distance)	1650	2.50

Exhibit 5.23: Goodness of Fit for Different ${c_{ij}}^{(k)}$'s for Skokie Data.

NOTES: $c_{ij}^{(k)}$'s are shown in parentheses after the name of the procedure used. The Chi-square values for the Evans-Kirby and the Hyman procedure are from the examples in Section 5.3

sides, neighbors find out about job and shopping opportunities from other neighbors, and families frequently travel together for various recreational activities. Consequently, the values in Exhibit 5.23 for the case where travel time and distance are both included are not too bad. However, based on other work we have done with person-trip data, we conjecture that there are still other ${c_{ij}}^{(k)}$'s that should have been included in $F(c_{ij})$. Notice also that travel time alone, regardless of how it is used, gives higher ratios of Chi-square to degrees of freedom (X^2/df).

As another example, consider Exhibit 5.24, where X^2/df ratios are presented for the hospital data and several possible separation measures. It is easily seen that both the square root of travel time and the log of travel time perform better than time itself, and a similar situation holds for distance.

The intuitive rationale for the square root function is its ability to incorporate 'diminishing effects' of longer travel times. While the log function also exhibits such effects, it is much more extreme in nature. This can be illustrated in the present context in the following way. Suppose that the travel times from an origin i to destinations j and h are given respectively

$c_{ij}{}^{(1)}$	Parameter Estimate	X^2/df
Distance	-0.000544	9591.66
Log Distance	-1.928516	23.24
Square Root of Distance	-0.742918	18.28
Travel Time	-0.134393	127.16
Log Travel Time	-3.196524	15.76
Square Root of Travel Time	-1.372817	15.62

Exhibit 5.24: Goodness of Fit for Different Measures of Physical Distance

NOTES: Parameters estimates are ML. No other $c_{ij}^{(k)}$ used in the models for which results are shown.

by $d_{ij} = 10$ minutes and $d_{ih} = 15$ minutes. If the travel time to j is increased to $d'_{ij} = 120$ minutes, then one may ask what increase in travel time, d'_{ih}, to h would offset this in the sense that the expected trip share, T_{ij}/T_{ih}, stay the same. For any exponential gravity model, with travel time as the only separation measure, it may be readily verified that

$$\frac{T_{ij}}{T_{ih}} = \frac{T'_{ij}}{T'_{ih}} \Leftrightarrow d_{ih} - d_{ij} = d'_{ih} - d'_{ij} \tag{5.267}$$

and hence that $d'_{ih} = 125$. Thus, pure travel times fail to incorporate any diminishing effect; i.e., any five minute difference in travel times has the same effect regardless of total trip durations. On the other hand, if travel time is replaced by the log of travel time in (5.267), then the travel time which now yields the same trip share [i.e., which satisfies $\log[10] - \log[5] = \log[d'_{ih}] - \log[120]$] is given by $d'_{ih} = 240$. Hence, while diminishing effects are now present, they are too extreme in that most travelers would consider doubling of a 5-minute trip to 10 minutes to be less objectionable than the doubling of a 2-hour trip to 4 hours. However, if travel time is replaced by the square root of travel time in (5.267), then the new value of travel time to h [i.e., satisfying $\sqrt{10} - \sqrt{5} = \sqrt{d'_{ih}} - \sqrt{120}$] is given by $d'_{ih} = 150$. Here diminishing effects are seen to be of a more reasonable order of magnitude; namely, that doubling a 5-minute trip to 10 minutes is roughly comparable to adding half an hour to a 2-hour trip.

In our experience, transformations are frequently effective. It should be borne in mind that the model attempts to describe how people view and respond to certain factors, and if people respond in a non-linear way, we need to model accordingly.

The Chi-square ratio is still quite high in Exhibit 5.24. Part of the reason

Service Category	Parameter Estimate[a]	X^2/df
Pediatrics	-1.588499	4.98
Obstetrics	-1.398835	6.46
Med/Surgical	-1.393375	12.68
Other	-1.275631	1.72
Oncology	-1.252428	3.09
Psychiatry	-1.161458	2.57
Total Trips	-1.372817	15.62

Exhibit 5.25: Parameter Estimates and Chi-Square Ratios for different Models
NOTES: The estimates are ML. In all runs represented above, the only $c_{ij}^{(k)}$ used was the square root of distance.

is suggested by Exhibit 5.25 which shows results from several gravity models each constructed from a different class of patients, the classification being based on service categories. The X^2/df ratio is considerably lower now for all categories except for the general medical-surgical group, which possibly should be disaggregated further. While the Chi-square values for the other categories are reasonably satisfactory, they can perhaps be lowered still (and have been — in fact, an additional $c_{ij}^{(k)} = c_{ij}^{(4)}$ used for the purpose is described below). Nevertheless, the results are fairly convincing that patients do indeed seem to behave according to the gravity model.

Before concluding this section, we draw attention to the θ_1 value for psychiatry, which is relatively low indicating a lower sensitivity to travel time, and for pediatrics which seems to have a higher sensitivity.

5.8.2 Residuals

As in any model building exercise, examination of residuals is strongly recommended when gravity models are being constructed. They can point out $c_{ij}^{(k)}$'s that could otherwise have been missed, the need to disaggregate the data and construct several gravity models, errors in data collection, etc. However, in order to conduct an analysis of residuals two difficulties must first be addressed.

The first difficulty is that of unequal variances of residuals, a consequence of the Poisson distribution. This problem is easily handled, at least approximately, by considering, instead of the plain residuals $N_{ij} - \hat{T}_{ij}$, the components of any of the measures of fit (5.257) through (5.260). For example, we could use as Cochran suggests (see Rao, 1973, p. 393)

$$\frac{N_{ij} - \hat{T}_{ij}}{\sqrt{\hat{T}_{ij}}}, \tag{5.268}$$

or

$$\sqrt{N_{ij}} - \sqrt{\hat{T}_{ij}}, \tag{5.269}$$

their squares, or

$$N_{ij} \log[N_{ij}/\hat{T}_{ij}]. \tag{5.270}$$

The other problem stems from the fact that there are usually a very large number of residuals. While no set rule can be given, we have found the following two procedures to be helpful:

1. Draw normal plots (sometimes called rankit plots) of residuals. Such plots are described in Sen and Srivastava (1990, pp. 101–103, 165, 168, 171), and may be briefly summarized as follows: If e_s are the chosen residuals ordered from the smallest to largest, then the plot consists of the points (e_s, γ_s) where

   $$\gamma_s = \Phi^{-1}[(e_s - 3/8)(n + 1/4)],$$

   Φ is the distribution function (cdf) of the standard normal distribution and $n = IJ$ is the total number of e_s's. If the e_s's are normal such a plot would lie approximately on a straight line. Points that deviate sharply, near the ends may be tagged as outliers and subjected to further scrutiny. Unusual jumps in the plots or other strongly non-linear shapes could signal the need for transformations of the $c_{ij}{}^{(k)}$'s or for additional $c_{ij}{}^{(k)}$'s.

2. The residuals can be combined over intuitively meaningful sets (e.g., (5.268) or (5.269) could be squared and added over such sets). All origins for a common destination and all destinations for a common origin could be such sets. In fact, it was such an analysis that suggested a very important variable for the hospital study. Destinations which were large teaching hospitals had much larger sum of squares of e_s's than other destinations. This was understandable since such hospitals attract more patients from large distances, and suggested the used of the variable $c_{ij}^{(4)} = \delta c_{ij}^{(1)}$, where $c_{ij}^{(1)}$ is the square root of travel time and δ is 1 for teaching hospitals and 0 otherwise. In effect, we would be seeking a different parameter estimate for the square root of distance for large teaching hospitals.

There are clearly other ways of examining residuals. The potential efficacy of such examinations in the gravity model case should not be seen as less than that for linear models where residual analysis is routinely carried out by all competent analysts.

5.9 Other Properties of ML Estimates

ML estimates are used so frequently because they often possess very desirable large sample properties. These properties are described in Section 5.9.1 where it is shown in particular that they are exhibited by the ML estimates $\hat{\boldsymbol{\theta}}$. In Section 5.9.2 Monte Carlo methods are used to show that ML estimates for the gravity model have pleasant properties even when sample sizes are not very large.

5.9.1 Asymptotic Properties

We have, in fact, already proved one very important property of ML estimates of gravity model parameters — efficiency. There are several definitions of efficiency. Some authors say that an estimate is efficient if its asymptotic variance is lower than that of any other estimate, while others call an estimate efficient if its asymptotic variance equals a theoretical lower bound, which, in our case, is the inverse of the information matrix. The second definition is obviously stronger than the first and, as shown in Section 5.7.1, the expression (5.219) for the asymptotic variance is precisely the inverse of the information matrix. Therefore, ML estimates of all gravity model parameters (including, trivially, the one given an arbitrary constant value) are efficient by either criterion. In particular, ML estimates of the θ_k's are efficient.

Two other highly desirable properties are often associated with ML estimates:

- (Strong) consistency, which implies that as the sample size grows, parameter estimates converge almost surely to their 'true values.'
- Asymptotic normality, which is obviously desirable both for testing and for obtaining confidence intervals.

A third highly desirable property, which we shall address at the end of this section is that of asymptotic robustness.

Large samples occur when $N_{\oplus\oplus}$ is large. As implicit in our discussion of Section 5.8.1, this happens if each T_{ij} is large or if IJ is large and T_{ij}'s are not too small. The former case is the one usually seen in the statistical literature. The latter is often more appropriate for gravity model applications.

It might be noted that the asymptotic variance in the definition of efficiency is achieved when the differences between estimates and true values are small, i.e., when estimates converge to their true values, no matter under what conditions the convergence occurred. Thus, whether an estimate is efficient or not is not affected by the type of asymptotics, as long as estimates converge to their true values.

In order to formally state the next theorem, we need an additional condition:

ML4. For all values of the T_{ij}'s as $T_{\oplus\oplus}$ gets large, each $P_{ij} = T_{ij}/T_{\oplus\oplus}$ remains unchanged.

If P_{ij}'s were to change that would imply a fundamental change in the model. Thus the condition is reasonable. Moreover, one would expect that the growth in $T_{\oplus\oplus}$ would occur by increasing the number of identical time periods over which the data are gathered or by increasing the sampling rate. However, it needs to be stressed that in practical applications $T_{\oplus\oplus}$ never grows; the result below is used to judge what happens when $T_{\oplus\oplus}$ is large — for a single matrix of T_{ij}'s. Notice that Condition ML4 assures that for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$, each $P_{ij} > 0$.

Theorem 5.7 *Under Condition ML4 and Conditions ML1 and ML2 of Theorem 5.1, as each T_{ij} gets large, the ML estimates of θ_k's are strongly consistent, efficient and asymptotically normal.*

A general version of this theorem (i.e., not for the gravity model alone) has been proved, along with other results, in Rao (1973). Rao's conditions can be verified for the gravity model, although such an effort is non-trivial. However, in order to be more self contained, we have given a direct proof, which actually can be seen as an adaptation of Rao's proofs. Treatments paralleling his have been given in Birch (1964), Haberman (1974, Chapter 4), Bishop, *et al.* (1975, p.509 et seq.) and others.

Proof of Theorem 5.7: Efficiency was established in the discussion above. To establish the properties of strong consistency and asymptotic normality it is convenient, partly because of ML4, to work with relative frequencies rather than the frequencies themselves. As in Section 5.2.4, let $P_{ij} = T_{ij}/T_{\oplus\oplus}$, $P^*_{ij} = \hat{T}_{ij}/T_{\oplus\oplus}$ and $\hat{P}_{ij} = N_{ij}/T_{\oplus\oplus}$. Then

$$E(\hat{P}_{ij}) = P_{ij} = A(i)B(j)\exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] : i \in I, j \in J$$

where, as before, we assume that $B(J) = 1$ and the $A(i)$'s above are obtained by dividing the $A(i)$'s in (5.2) by $T_{\oplus\oplus}$. Set

$$\begin{aligned}\boldsymbol{P} &= (P_{11}, \ldots, P_{1J}, \ldots\ldots, P_{I1}, \ldots, P_{IJ})^t\\ \boldsymbol{P}^* &= (P^*_{11}, \ldots, P^*_{1J}, \ldots\ldots, P^*_{I1}, \ldots, P^*_{IJ})^t\\ \hat{\boldsymbol{P}} &= (\hat{P}_{11}, \ldots, \hat{P}_{1J}, \ldots\ldots, \hat{P}_{I1}, \ldots, \hat{P}_{IJ})^t.\end{aligned}$$

Define Ω to be the Jacobian matrix of $M_{(2)}{}^t\boldsymbol{P}$ with respect to

$$\boldsymbol{\xi} = (a(1), \ldots, a(I), b(1), \ldots, b(J-1), \theta_1, \ldots, \theta_K)^t$$

where, as before, small case letters denote logarithms of corresponding capital letters. Since, it may be verified, using methods similar to those in Section 5.7.1, that the Jacobian matrix of $\boldsymbol{P}$ with respect to $\boldsymbol{\xi}$ is $\operatorname{diag}\boldsymbol{P} \cdot M_{(2)}$, it follows that $\Omega = M_{(2)}{}^t \operatorname{diag}\boldsymbol{P} \cdot M_{(2)}$. Since, because of Condition ML4,

$P_{ij} > 0$ for all $i \in I$ and $j \in J$, Ω^{-1} exists and is the Jacobian matrix of $\boldsymbol{\xi}$ with respect to $M_{(2)}{}^t\boldsymbol{P}$.

Let

$$\hat{\boldsymbol{\xi}} = (\hat{a}(1), \ldots, \hat{a}(I), \hat{b}(1), \ldots, \hat{b}(J-1), \hat{\theta}_1, \ldots, \hat{\theta}_K)^t,$$

where $\hat{a}(i)$, $\hat{b}(j)$ and $\hat{\theta}_k$ are the ML estimates corresponding to $\hat{P}_{ij}$'s. Since $\boldsymbol{\xi}$ is the vector of true values and $\hat{\boldsymbol{\xi}}$ consists of estimates, a proof of strong consistency requires that $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} \overset{\text{a.s.}}{\to} \mathbf{o}$ as all $T_{ij} \to \infty$, where '$\overset{\text{a.s.}}{\to}$' denotes almost sure convergence.

It is well known (and can be proved as in Rao, 1973, p. 355–359) that as T_{ij}'s grow large, $\hat{\boldsymbol{P}} \to \boldsymbol{P}$, almost surely, and, hence $M_{(2)}{}^t\hat{\boldsymbol{P}} \overset{\text{a.s.}}{\to} M_{(2)}{}^t\boldsymbol{P}$. Since the elements of $\boldsymbol{P}^*$ are ML estimates, we get from (5.209),

$$M_{(2)}^t\boldsymbol{P}^* = M_{(2)}^t\hat{\boldsymbol{P}} \tag{5.271}$$

and it follows that $M_{(2)}{}^t\boldsymbol{P}^* \overset{\text{a.s.}}{\to} M_{(2)}{}^t\boldsymbol{P}$. It is then easily seen that

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} \overset{\text{a.s.}}{\to} \Omega^{-1}M_{(2)}{}^t\boldsymbol{P}^* - \Omega^{-1}M_{(2)}{}^t\boldsymbol{P} \overset{\text{a.s.}}{\to} \mathbf{o}.$$

This establishes strong consistency of $\hat{\xi}$, and, in particular of $\hat{\boldsymbol{\theta}}$.

A proof of asymptotic normality is almost identical. As each T_{ij} gets large, $\sqrt{T_{\oplus\oplus}}\hat{P}_{ij} \overset{\text{dist}}{\to} z_{ij}$, where z_{ij} is normally distributed (with, as may be verified, mean $\sqrt{T_{\oplus\oplus}}P_{ij}$ and variance P_{ij}) and '$\overset{\text{dist}}{\to}$' denotes convergence in distribution. Hence

$$\begin{aligned}[\sqrt{T_{\oplus\oplus}}\hat{\boldsymbol{\xi}} - \Omega^{-1}M_{(2)}{}^t\boldsymbol{z}] &\overset{\text{dist}}{\to} \Omega^{-1}[\sqrt{T_{\oplus\oplus}}M_{(2)}{}^t\boldsymbol{P}^* - M_{(2)}{}^t\boldsymbol{z}] \\ &\overset{\text{dist}}{\to} \Omega^{-1}[\sqrt{T_{\oplus\oplus}}M_{(2)}{}^t\hat{\boldsymbol{P}} - M_{(2)}{}^t\boldsymbol{z}] \overset{\text{dist}}{\to} 0.\end{aligned}$$

This proves the theorem. □

An examination of the proof of Theorem 5.7 will show that all we needed for the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}$ was the convergence and the asymptotic normality of $M_{(2)}{}^t\hat{\boldsymbol{P}}$. Thus we have the corollary:

Corollary 5.4 *The ML estimate of $\hat{\boldsymbol{\theta}}$ is strongly consistent, efficient and asymptotically normal, if $M_{(2)}{}^t\hat{\boldsymbol{P}}$ is strongly consistent and asymptotically normal provided that Conditions ML1, ML2 and ML4 hold.*

As mentioned earlier, in actual practice, one encounters a single matrix of N_{ij}'s with a corresponding underlying matrix of T_{ij}'s. The theorem above, and particularly its proof shows that when T_{ij}'s are all large, $\hat{P}_{ij}$ would be 'close to' P_{ij} and thus the estimates will be close to their true values. Moreover, the estimates would be asymptotically normal.

A reasonable practical question one might then ask is what happens if all T_{ij}'s do not appear to be large [i.e., we observe that some N_{ij}'s are small, implying the high likelihood that corresponding T_{ij}'s are also small]. Then, if $T_{i\oplus}$'s, $T_{\oplus j}$'s are all large, implying that several T_{ij}'s are large, then $\hat{P}_{i\oplus}$'s and $\hat{P}_{\oplus j}$'s would converge to their expected values almost surely as would the $\sum_{ij} c_{ij}{}^{(k)}\hat{P}_{ij}$'s [since convergence would occur for large values T_{ij}'s and small valued ones would not make much of a contribution to the total]. The question then is what happens to Ω. If Ω has eigenvalues well away from zero, i.e., it is always well removed from singularity, then we get our estimates of $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}$. Then the question that arises is when does Ω have this property.

Notice that relatively small valued T_{ij}'s imply near zero valued P_{ij}'s. If we set the rows of $M_{(2)}$ corresponding to the near zero valued P_{ij}'s to zero, then the $M_{(1)}$ part of this matrix still has rank $I + J - 1$ provided that there is at least one P_{ij} in each row and at least one P_{ij} in each column of the matrix $((P))$ which is not zero [as can be shown as in the proof of Lemma 5.1]. The existence of such non-zero P_{ij}'s is assured by the largeness of $T_{i\oplus}$'s and $T_{\oplus j}$'s. However, whether the remaining columns of $M_{(2)}$ [i.e., those not contained in $M_{(1)}$] cause the modified $M_{(2)}$ to have rank less than $I + J + K - 1$ is not clear. But if it did have rank $I + J + K - 1$, then $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ would be small. Since we expect that in practical situations, the modified $M_{(2)}$ more often than not would be of full rank [of course, this can be verified in a given case], we would expect that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ would be small as long as the $T_{i\oplus}$'s and $T_{\oplus j}$'s are large enough [for fixed I and J].

A similar statement may be made for asymptotic normality, and the arguments in favor of it would be similar.

Asymptotic Robustness

While we have assumed through most of this chapter that N_{ij}'s are Poisson, in practice this is often just an approximation, as described in Section 5.8.1. Therefore, a reasonable question to ask is to what extent do the ML estimates $\hat{\boldsymbol{\theta}}$ depend on the Poisson assumption. Fortunately, the large sample answer to the question is 'not at all' as far as consistency is concerned, so long as the true distribution obeys some very mild conditions. That is, ML estimates obtained under the Poisson assumption remain consistent even if the actual distribution is not Poisson. Another way of putting it is to say that the estimates are asymptotically robust.

The only place we used the distribution of N_{ij}'s in the proof of Theorem 5.7 was in the use of the likelihood equations (5.271). Thus, for any estimates for which (5.271) hold, strong consistency and asymptotic normality would continue to hold as long as individual P_{ij}'s converged almost surely and were asymptotically normal and this is almost universally true. [see Rao, 1973, p. 355–359, see also Corollary 5.4 for greater generality of the above discussion]. In particular, any procedure [e.g., the modified scoring procedure] which solves these equations, regardless of the underlying distributions of the N_{ij}'s would yield results possessing these properties.

Actually,

$$M_{(2)}^t \boldsymbol{P}^* - M_{(2)}^t \hat{\boldsymbol{P}} \xrightarrow{\text{a.s.}} \mathbf{o} \tag{5.272}$$

would have sufficed for the proof, and (5.272) would follow from

$$[\hat{P}_{ij} - P_{ij}^*] \xrightarrow{\text{a.s.}} 0 \text{ for all } i \text{ and } j, \tag{5.273}$$

a property possessed by estimates based on a very wide class of distributions.

Another relevant treatment of robustness is given in Huber (1967, see also Davies and Guy, 1987). Huber has also shown that the property of asymptotic normality is also retained if the distribution is not Poisson (as can also be seen from examining the proof of Theorem 5.7), and both he and Davies and Guy have given expressions for the asymptotic variance.

Not only does (5.273) hold for a wide class of distributions, it also holds for a wide class of estimation procedures. Consequently, the properties mentioned in Theorem 5.7 are not unique to ML estimates but shared by estimates from a number of different procedures. A property called second order efficiency identifies ML estimation as superior to others (see Rao, 1973, pp. 352–353).

Case where I and J are large

Now, turning our attention to the case when I and J are large, we conjecture that similar results hold for this type of large samples. A discussion of why we make this conjecture is given below and practical implications are described at the end of this subsection.

Now since the number of N_{ij}'s is not fixed, we need to impose additional conditions on the behavior of T_{ij}'s and $c_{ij}{}^{(k)}$'s. Therefore, let

$$T^{\#} \geq T_{ij} \geq T_{\#} > 0 \text{ and } c^{\#} \geq c_{ij}{}^{(k)} \geq c_{\#} > 0, \tag{5.274}$$

for all i and j and all $n = IJ$. The former condition assures that $T_{i\oplus}$, $T_{\oplus j}$ and therefore $T_{\oplus\oplus}$ will be large for large I and J. Further assume that $I/J \to \kappa$ as $I, J \to \infty$, where $0 < \kappa < 1$. This last condition assures us that for each i, there are a large number of N_{ij}'s, and for every j also the number of N_{ij}'s is large. Since each T_{ij} is not large, we now must depend on an application of the law of large numbers and of the central limit theorem. We will need an additional condition mentioned later.

Let $Q = ((q_{rs}))$ be the diagonal matrix defined by

$$q_{ss} = \begin{cases} J^{-1} & \text{when } s = 1, \ldots, I \\ I^{-1} & \text{when } s = I+1, \ldots, I+J-1 \\ [IJ]^{-1} & \text{when } s = I+J, \ldots, I+J+K. \end{cases}$$

Then the first I elements of the vector $QM_{(2)}{}^t \boldsymbol{N}$ may be shown to be $J^{-1}\sum_{j=1}^{J} N_{ij}$, the next $J-1$ elements to be $I^{-1}\sum_{i=1}^{I} N_{ij}$ and the last K

elements to be $[IJ]^{-1}\sum_{ij} c_{ij}{}^{(k)}N_{ij}$. Recall that we obtained matrix $M_{(2)}$ from M by removing the $(I+J)$th column.

Under the conditions on T_{ij}'s and $c_{ij}{}^{(k)}$'s given by (5.274), each variance $\text{var}\,[N_{ij}] = T_{ij}$ and each variance $\text{var}\,[c_{ij}{}^{(k)}N_{ij}] = [c_{ij}{}^{(k)}]^2T_{ij}$ is bounded. Hence by the strong law of large numbers (Rao, 1973 p. 114; note the inconsequential 'typo' in the statement in Rao), each of the sequences that constitute the elements of $QM_{(2)}{}^t\boldsymbol{N}$ converge almost surely to their expected values, i.e,

$$QM_{(2)}{}^t\boldsymbol{N} \overset{\text{a.s.}}{\to} QM_{(2)}{}^t\boldsymbol{T}. \tag{5.275}$$

Since every ML estimate $\hat{\boldsymbol{T}}$ must obey the likelihood equations, (5.209), it follows that

$$QM_{(2)}{}^t\hat{\boldsymbol{T}} \overset{\text{a.s.}}{\to} QM_{(2)}{}^t\boldsymbol{T}. \tag{5.276}$$

For every $n = IJ$, define $\Omega_n^* = QM_{(2)}{}^t \,\text{diag}\,\boldsymbol{T}M_{(2)}$. It may be shown, as for Ω in the proof of Theorem 5.7, that Ω_n^* is the Jacobian matrix of $QM_{(2)}{}^t\boldsymbol{T}$ with respect to $\boldsymbol{\xi}$. Also, since Q is of full rank, as is $M_{(2)}{}^t \,\text{diag}\,\boldsymbol{T}\cdot M_{(2)}$, it follows that $[\Omega_n^*]^{-1}$ exists. Since, for small values of $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$, the matrix $[\Omega_n^*]^{-1}$ transforms $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$ into $\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}$, which includes as components $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$, we need a condition such that for all n, $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ would be small whenever $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$ is small.

A possible sufficient condition in order to achieve this is that the matrix $[(\Omega_n^*)^{-1}]^t[(\Omega_n^*)^{-1}]$ have eigenvalues that are bounded for all I and J [see Rao, 1973, p. 62]. This boundedness of eigenvalues of inverse matrices is a natural extension of the condition of non-singularity when the matrix itself varies, and implies that it never gets close to singularity. If we compute the products that constitute Ω_n^*, we will see that the elements of Ω_n^* are always bounded under conditions (5.274), and that for the $(I+J-1)\times(I+J-1)$ submatrix on the top-left, the diagonal terms overwhelm other terms. In fact, these non-diagonal terms $\to 0$. Thus, it is extremely likely that the boundedness of eigenvalues depends on the $c_{ij}{}^{(k)}$'s only and in most practical situations, it holds true.

However, we are uncertain whether this condition is meaningful and what the effects of increasing numbers of components of $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$ might be. For example, we need to be open to the possibility that a large part of the sum of the components of the difference $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$ might be concentrated on the θ_k's, so that $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ could be not too small simply by virtue of the number of components of $QM_{(2)}{}^t[\hat{\boldsymbol{T}}-\boldsymbol{T}]$. However, we conjecture based on an examination of the matrix Ω_n^*, that this does not happen. Notice also that as additional origin or destination zones are added, rows are appended to the matrix $M_{(2)}$, but the dimension of $\hat{\boldsymbol{\theta}}$ does not change. It would indeed be surprising if with the additional information contained in $M_{(2)}$ and the corresponding N_{ij}'s, $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ would not decrease simply by virtue of the size of $M_{(2)}$. Thus we arrive at our conjecture regarding strong consistency.

A very similar reasoning can be presented in support of $\hat{\boldsymbol{\theta}}$ being asymptotically normal as $I, J \to \infty$. By the use of the central limit theorem (e.g., Liapunov's version; see Rao, 1973, p. 127) instead of the law of large numbers, but otherwise following steps similar to those above, we can show that $GM_{(2)}\hat{\boldsymbol{T}}$ is asymptotically normal. But as before translating this into asymptotic normality of $\hat{\boldsymbol{\theta}}$ remains a problem. Thus asymptotic normality of $\hat{\boldsymbol{\theta}}$ as $I, J \to \infty$ must also be left as a conjecture for the moment.

At an intuitive level, the conjectures are based on the fact that since the estimation of $\hat{\boldsymbol{\theta}}$ depends on the sufficient statistics, and when I and J are large, these sufficient statistics could be based on large samples, even though individual N_{ij}'s may be small. Then $\hat{\boldsymbol{\theta}}$ should approach $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ should be asymptotically normal. This intuitive statement has also been made elsewhere in this chapter. Unfortunately, we are not sure of the asymptotic behavior in all cases of the relationship between the parameters and $\boldsymbol{T}$. Of course, the eigenvalues of Ω_n^* may be computed in a given situation, but for large I and J, the computations could be substantial.

5.9.2 Small Sample Properties

In this section we investigate small sample properties. It is unrealistic to expect any estimation procedure to perform well when $T_{\oplus\oplus}$ is very small. Therefore, we concentrate here on situations where T_{ij}'s are small and I and J are of moderate size.

The discussion of this section is based on Monte Carlo simulations. We shall illustrate most of our statements using the 'T_{ij}', '$T_{ij}/50$' and '$T_{ij}/100$' data sets based on Skokie data as described in Section 5.6.1. However, our conclusions are based on work involving several other data sets.

Bias and Variances

Recall that, while Equations (5.12) and (5.13) are linear in T_{ij}'s, they are non-linear in $A(i)$'s, $B(j)$'s and θ_k's. When ML Estimates are obtained by solving non-linear equations, bias usually results; and our situation is no exception (as can be verified applying Jensen's inequality — see Rao, 1973, p. 56 — to the right side of (5.2)). Moreover, the smaller the T_{ij}'s, the worse the bias becomes. However, fortunately, for most practical situations where gravity models are applied, this bias is negligible, as we illustrate below.

Note that $(-\cdot 02, -\cdot 5)$ was the input value of $\boldsymbol{\theta}$ used in constructing the N_{ij}'s used in the simulations (see Section 5.6.1). Hence, $(-\cdot 02, -\cdot 5)$ may be treated as the 'true value' of $\boldsymbol{\theta}$. Consequently, we see from Exhibit 5.26 that bias is very small, even for the two 'small sample' data sets (the '$T_{ij}/50$' and '$T_{ij}/100$' data sets). Indeed, if we divide the standard deviations shown in the table by $\sqrt{n}$ to obtain standard errors of the means, we see that, with one minor exception, the means are well within the 95 per cent interval

Simulated Data Set	Parameter	Sample Mean	Std. Dev. from Sample Variance	Std. Dev. from (5.230), (5.231) and (5.232)
'T_{ij}'	$\hat{\theta}_1$	-.0191	.0020	.00183
	$\hat{\theta}_2$	-.4941	.0068	.0066
'$T_{ij}/50$'	$\hat{\theta}_1$	-.0218	.0134	.0129
	$\hat{\theta}_2$	-.4986	.0447	.0468
'$T_{ij}/100$'	$\hat{\theta}_1$	-.0207	.0181	.0183
	$\hat{\theta}_2$	-.5051	.0619	.0661

Exhibit 5.26: Mean and Variances from Simulations and Theoretically Obtained Variances.

NOTES: $n = 10$ simulations were for 'T_{ij}', $n = 100$ simulations were used for '$T_{ij}/50$' and '$T_{ij}/100$.'

of the true value. Whatever the bias is exactly, it appears that in most practical applications it would have virtually no effect.

Thus, the results above show that even for small T_{ij}'s scattered over a modest sized matrix (see Section 5.6.1, especially Exhibit 5.5), there is no noticeable bias. This is not entirely surprising since ML estimates depend on the N_{ij}'s only through the right hand sides of equations (5.12) and (5.13). These quantities are usually much larger than individual N_{ij}'s. To some extent, these statements are the moderate sized IJ versions of the statements made in Corollary 5.4.

The variances and covariances obtained in Section 5.7.1 are asymptotic. Consequently, it is appropriate to check whether, when T_{ij}'s are small, the variances given by those formulæ match the sample variances of simulated $\hat{\theta}_k$'s. The last two columns of Exhibit 5.26 provide such a comparison. It appears that it is quite safe to use (5.226) even when T_{ij}'s are small, especially since, in practice, variance estimates can be 'rougher' than parameter estimates.

Normality of Estimates

Exhibits 5.27 and 5.28 show a set of normal plots (see Section 5.8.2) for the estimated θ_k's. Recall that if a random variable is approximately normal, its normal plot is approximately a straight line. The plots show the estimates of θ_k's appear indeed to be approximately normal even for data sets containing mainly very small values of $\mathrm{E}[T_{ij}]$.

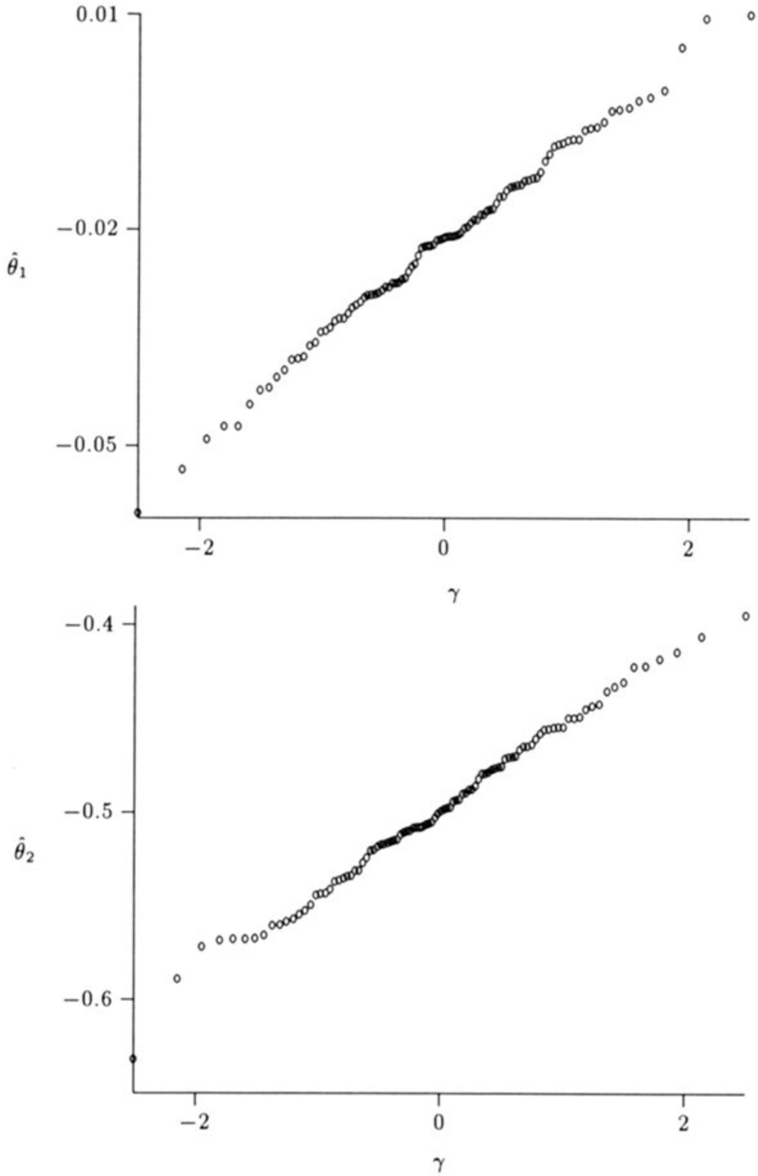

Exhibit 5.27: Normal Plots for $\hat{\theta}_1$ and $\hat{\theta}_2$: '$T_{ij}/50$' Data.

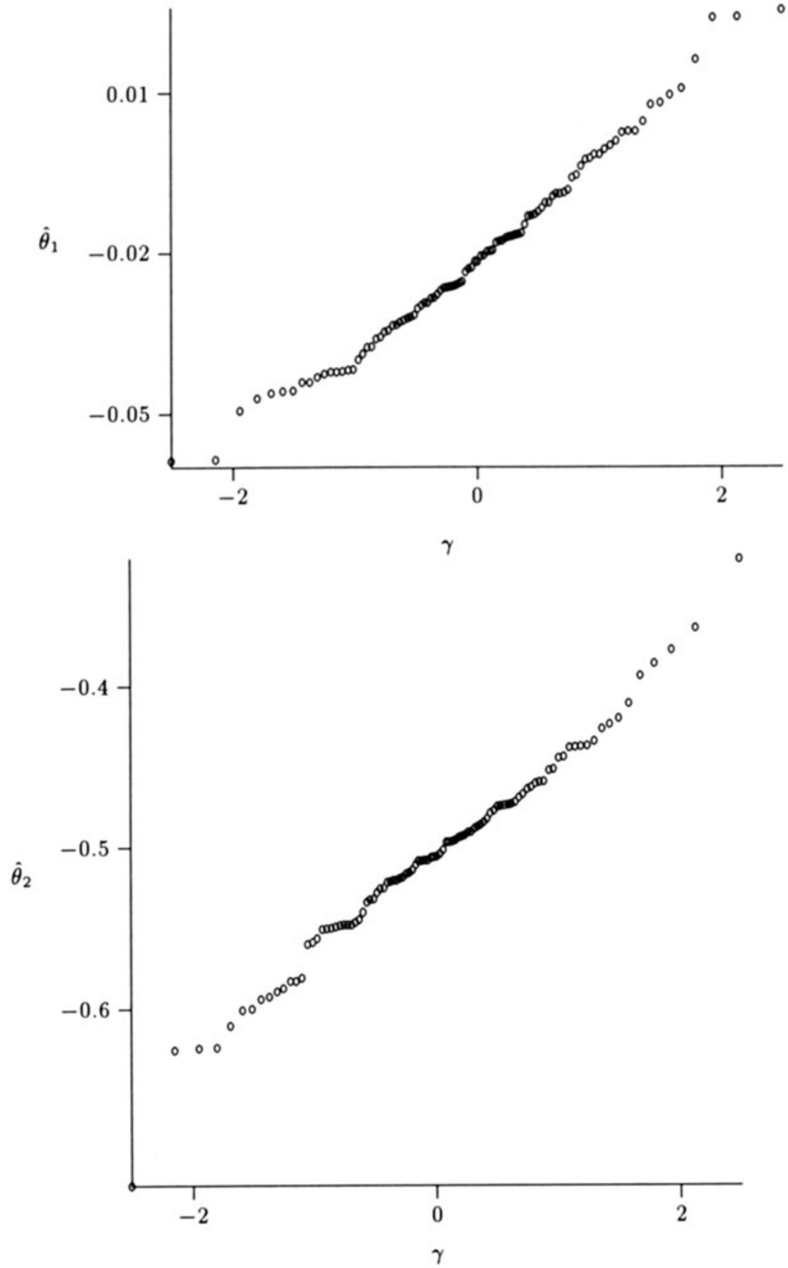

Exhibit 5.28: Normal Plots for $\hat{\theta}_1$ and $\hat{\theta}_2$: '$T_{ij}/100$' Data.

In fact, in no case we tried did a comparison of the empirical distribution of $\hat{\theta}_k$'s with the appropriate (with respect to mean and variance) normal distribution using a Kolmogorov's D statistic (which were computed using SAS Proc Univariate — see SAS, 1985a, p. 1187) even come very close to being significant at a 5 percent level. For the two plots in Exhibit 5.27, the p-values are .67 and .85. The corresponding values for Exhibit 5.28 are .08 and .83. Thus, only for the upper plot of Exhibit 5.28 did we get significance at a 10 per cent level. For all 9 times that we tested for normality of ML estimates of θ_k's, including cases not displayed here, this was the only case of significance at a 10 per cent level.

Therefore, all the usual tests and procedures based on the normal distribution can be safely carried out, even when sample sizes are small. This would, of course, include procedures for determining sample sizes.

ROBUSTNESS

To simulate different realistic departures from the Poisson assumption, we carried out several exercises. In each case we attempted to simulate the situation that trip makers affect the behavior of other trip makers (because, as mentioned in Section 5.8.1, people travel together, suggest shopping and work destinations to friends, etc.) and in that way violate the Poisson assumption.

1. In one exercise we computed two data sets ${N_{ij}}^{(1)}$ and ${N_{ij}}^{(2)}$ and obtained $N_{ij} = {N_{ij}}^{(1)} + {N_{ij}}^{(2)}$. We simulated ${N_{ij}}^{(1)}$'s by applying a Poisson random number generator to $T_{ij}/2$'s. This would represent the trips that were being made independently. In order to have the rest of the trips dependent on these (e.g., when people travel together or choose their destination because of some other person's choice), we then applied a Poisson random number generator to the ${N_{ij}}^{(1)}$'s to obtain ${N_{ij}}^{(2)}$'s.

2. In another exercise, after generating ${N_{ij}}^{(1)}$'s we simply doubled each of them. Here we made half the trips totally dependent on the other half.

3. Clearly, zonal boundaries do not act as barriers to trips influencing each other. To simulate this situation, after obtaining each ${N_{ps}}^{(1)}$'s as above, we added to it the sum of $({N_{ij}}^{(1)} - T_{ij})$'s for i's close to p and j's close to s; i.e., for origin-destination pairs i, j which are close to p, s in terms of location on an origin-destination matrix. We then applied a Poisson generator to the resultant quantities to obtain ${N_{ij}}^{(2)}$'s and added these to the original ${N_{ij}}^{(1)}$'s. This case, which is an example of spatial correlation, is a very severe test of the procedure.

Ten runs were made for each case and the results of these exercises are shown in Exhibit 5.9.2. They indicate that estimates have not been affected

Exercise	Parameter	Mean	Std.Dev.
1	$\hat{\theta}_1$	-.0189	.0027
	$\hat{\theta}_2$	-.4894	.0099
	X^2	1595.4	
2	$\hat{\theta}_1$	-.0201	.0019
	$\hat{\theta}_2$	-.4844	.0092
	X^2	1317.8	
3	$\hat{\theta}_1$	-.0200	.0020
	$\hat{\theta}_2$	-.5010	.0076
	X^2	1345.0	

Exhibit 5.29: Results of Applying ML Procedure to Non-Poisson Data.

much by our tampering. The 'true' values of θ_1 and θ_2 are still $-.02$ and $-.5$. The relatively high Chi-square values (if the N_{ij}'s were Poisson, given that the data are simulated, Chi-square values would have been around 661) indicate that the 'data' are no longer close to being Poisson. Although only these cases are reported above, several other exercises were tried; the results were essentially the same.

Perhaps the greatest testament to robustness of ML estimates is the GLIM procedure. An examination of it (see Section 5.5.5) reveals that save for the assumption that variance be proportional to the mean, the iterations themselves use nothing from a Poisson assumption. Still, the estimates are ML estimates from a Poisson model. Indeed our own work in Section 5.6 shows that the estimates are very close to those from the method of scoring which provides accurate solutions to (5.12) and (5.13).

5.9.3 ML Estimates from Factored Data

We conclude our treatment of maximum likelihood with a discussion of the use of factored data. In many practical situations, the N_{ij}'s used in analysis are computed from a relatively small sample of the total number of interactions. Needless to say, then the model estimated would of course be for these 'sampled' N_{ij}. Let $T_{ij} = \mathrm{E}[N_{ij}]$ be the expectations of the sampled N_{ij}'s and let $T_{ij}{}^{(w)} = \mathrm{E}[N_{ij}{}^{(w)}]$ be the expected total number of trips from i to j. If $T_{ij}/T_{ij}{}^{(w)} = W$, then the sampled N_{ij}'s are often scaled up by an estimate of W^{-1}; the resultant 'data' WN_{ij} are called factored data, and the W are called factors. Equations (5.12) and (5.13) indicate that there is no difference in estimates of $T_{ij}{}^{(w)}$, whether one scales N_{ij}'s first by W^{-1}, or one uses the observed N_{ij}'s and then scales up the estimates of T_{ij}'s obtained from them. [However note that WN_{ij}'s do not have a Poisson distribution unless $W = 1$].

Since

$$T_{ij}{}^{(w)} = W^{-1}A(i)B(j)\exp[\boldsymbol{\theta}^t \mathbf{c}_{ij}] : i \in I, j \in J,$$

if $T_{ij} = A(i)B(j)\exp[\boldsymbol{\theta}^t \mathbf{c}_{ij}]$, the W^{-1} can be absorbed into the $A(i)$'s and $B(j)$'s, and by the uniqueness of $\boldsymbol{\theta}$ [Section 5.2], it follows that the estimate of $\boldsymbol{\theta}$ is not affected by whether one uses the raw data N_{ij} or the factored data $W^{-1}N_{ij}$. In fact, even if W were not the same for all i and j, but were of the form $T_{ij}/T_{ij}{}^{(w)} = W_i W_j$, the W_i's and W_j's could be absorbed into the $A(i)$'s and $B(j)$'s. However, for these and other complex factoring schemes, the log-likelihood function $\mathcal{L}$ would need to be changed and, if it is not or if we use (5.12) and (5.13) which depend on the form of $\mathcal{L}$, the estimates would not necessarily be maximum likelihood estimates. Thus, it is usually safer to estimate parameters using the original data, and scale estimates of T_{ij} later as needed.

5.10 Notes and Concluding Remarks

As in previous chapters, it is appropriate to relate the concepts and results of the chapter to the literature and to emphasize the principal findings. However, this chapter, being concerned with more statistical (as opposed to probability theoretical) aspects of gravity models needs to be treated somewhat differently.

Unlike more theoretical work, attempts at empirical contributions can be made by many. Some not only make such efforts, they also publish their results or incorporate them in software. When a class of publications of this type emerges, it cannot be ignored. We do mention such work below, although mostly without citations.

We first discuss the concepts and results of each section. Since, the immediate ancestry of the results in the chapter have been mentioned in the course of presentation of the results, only broader issues of general setting are addressed. We then conclude with an overview of the findings presented and a very strong recommendation for the use of ML methods for gravity model parameter estimation.

Section 5.1

In developing estimation methods, without the advantage of the understanding of a specific data set, one can only work with a template. That is, a class of models are identified and studied, with the anticipation that they would be adequate for the situations one would wish to model. A justification of the class chosen — exponential Poisson gravity models — has been provided in the earlier chapters and its rather general nature alluded to in Section 5.1.1. However, the suitability of this template model in a specific situation must be determined by the person who knows the situation and the data better. We hope that the results in this book will be helpful to such analysts and we anticipate that they would frequently use the exponential Poisson form. The form itself appears to have been first presented in Sen and Sööt (1981).

While Bayesian methods constitute an important part of the statistician's arsenal, they have not been applied much to gravity models. An exception is a recent paper by West (1994), in which he has considered a generalization of the model considered in this chapter — one which allows for extra variation (see Section 3.9.2).

SECTION 5.2

This section is a revised and updated version of Sen (1986). Results on existence and uniqueness of general ML estimates are widely available, even in textbooks, particularly for distributions of the exponential family (a form in which our distributions can be placed). However, the fact that while M is not of full rank, the ML estimate of $\boldsymbol{\theta}$ is still unique calls for a separate treatment.

The method of proof, used in showing uniqueness by posing a related optimizing problem, borrows from the work of Evans and Kirby (1974) — an approach also used by Evans elsewhere. A treatment which is very relevant to our problem is given in Haberman (1974).

SECTION 5.3

The three methods presented have been found to be very useful by one of the authors who has used the Evans-Kirby procedure and Hyman's procedure in many different contexts. The DSF procedure has been used within just about every other procedure presented in the chapter — except for the Evans-Kirby procedure which is a generalization of the DSF procedure. A significant literature exists around these methods and is referred to in the section. The reader who is particularly interested in the DSF procedure or the Evans-Kirby procedure might aslo wish to consult the relevant literature on doubly stochastic matrices [e.g., see Bapat and Raghavan (1980, 1989) and Parlett and Landis (1982)].

A great number of estimation methods are included in the gravity model literature and related software. A fair proportion of them exist only as algorithms, with little or poor justification as to efficacy. Many of these usually perform poorly and in certain situations can even give 'absurd' results.

A justification given some of these methods is that they are frugal users of computer time and space. None that we have seen are as economical as the linear least squares methods given in the next chapter and few, if any, could compete with the modified scoring procedure.

Section 5.4

The LDSF procedure was introduced in the literature by Weber and Sen (1983, 1985). While the form of presentation and the numerical examples are based on these papers, most of the theoretical results in Section 5.4 are new. This procedure has also been used within other procedures.

Sections 5.5 and 5.6

The results of these sections are largely taken from Yun (1992) and Yun and Sen (1994).

Section 5.7

The discussion of the covariance of $\hat{\boldsymbol{\theta}}$ is based on Sen and Matuszewski (1991) and the discussion of the covariance of $\hat{\boldsymbol{T}}$ is based on Weber and Sen (1983). However, Section 5.7 carries the discussion in these papers considerably further.

Section 5.8

The examination of goodness of fit using the Chi-square statistic is not new. The derivation of the asymptotic distribution of the Chi-square statistic when T_{ij}'s are small but IJ is large appears to have been presented for the first time in a dissertation written under the supervision of one of the authors [Kim, 1994, see also Kim and Sen, 1993].

Three points made in the text bear repeating here. First, goodness of fit criteria are often minimized to yield estimation procedures. Therefore, empirical comparisons of estimation procedures using a fixed goodness of fit criterion can become meaningless. Second, goodness of fit criteria need to be based on the assumed distribution of N_{ij}'s. There is some literature on goodness of fit which ignores this. Third, good gravity model fitting requires the same kind of checking of underlying assumptions that any other model fitting requires. For example, as stated earlier, we have found that travel time as a measure of separation frequently needs to be transformed. This is rarely done in the literature.

Baxter has addressed the issue of testing in a different way [see e.g., Baxter, 1985, 1987].

Section 5.9

As mentioned in the text, large samples can occur if all T_{ij}'s are large or if I and J are large. That ML estimates possess certain very desirable asymptotic properties in the first type of asymptotics is well known; however, the fact that $\hat{\boldsymbol{\theta}}$ possesses them needed to be verified because M is not of full rank. A discussion of asymptotic normality, efficiency and consistency of estimates for the second type of large samples appears to be presented here for the first time.

The treatment of small sample properties borrows heavily from Sen and Matuszewski (1991).

5.10.1 Conclusion

Maximum likelihood estimates of $\boldsymbol{\theta}$ exist and are unique under very mild conditions. Algorithms to compute them in special cases are well behaved and converge relatively quickly. In the general case one method, which we have called the modified scoring procedure, stands out. It is very well behaved and converges rapidly, handling many different situations with ease. It is also very fast, dealing with a 2000×2000 matrix of flows, four $c_{ij}^{(k)}$'s, and an extremely stringent convergence criterion on a regular mainframe computer in a little over a minute and a half.

The estimates themselves have excellent asymptotic properties. Even for very small samples of trips (e.g., for a one per cent sample of trips in a small town divided into over 40 zones) no noticeable bias occurred. The covariance matrix of the $\hat{\theta}_k$'s is easy to compute, and since the estimates are effectively normally distributed, statistical inference is readily possible. ML estimates are robust.

A shortcoming of ML procedures, as it applies to gravity models, is that the variety of diagnostic tools for model building that is available in linear regression, is not available for ML procedures. Thus, linear least squares approaches continue to be of value. Moreover, least squares procedures, as described in the next chapter, are faster — although not by much anymore (in fact, linear least squares, as has often been practiced in past gravity model work, using the M or $M_{(2)}$ as the design matrix, is usually slower than the modified scoring procedure).

All in all, therefore, ML estimation, using the modified scoring procedure, is highly recommended.

CHAPTER 6

Least Squares

6.1 Introduction

Maximum likelihood procedures are very much in favor among those interested in parameter estimation, and as mentioned in Chapter 5 for good reason. However, ML methods are not the only estimation methods. Very frequently used alternatives to maximum likelihood are various methods using linear and non-linear least squares. In this chapter we examine the use of these procedures for estimation of gravity model parameters.

We saw in Chapter 5 that ML methods perform very well. Because the form of the likelihood equations (5.12) and (5.13) are such that ML estimation is based on numbers $N_{i\oplus}$, $N_{\oplus j}$, $\sum_{ij} c_{ij}{}^{(k)} N_{ij}$ which are typically quite large even when individual N_{ij}'s are not, we surmise that the quality of ML estimates will be nearly impossible to surpass. In the past, linear least squares (LS) procedures were used in part because they were better known and more accessible computationally, but also because they were more frugal consumers of computer time and memory. This is no longer the case, entirely. In fact, for moderate to large numbers of origins and destinations, some procedures based on linear least squares have larger appetites for computer resources than does the modified scoring procedure recommended in the last chapter.

However, some linear least squares based procedures, e.g., those given in Sen and Sööt (1981), Sen and Pruthi (1983) and Gray and Sen (1983), are simpler. In fact, they are simple enough that one of the authors has been able to apply one of them to a 25×25 table of N_{ij}'s in about 2 hours using only a hand calculator (see end of Section 6.3.5). Other advantages also exist for these procedures. As mentioned in the last chapter, more diagnostic tools exist for least squares analysis (see, for example, Sen and Srivastava, 1990). Therefore, standard packages can be used to refine model formulation (e.g., in choosing the $c_{ij}{}^{(k)}$'s). Under certain conditions, the odds ratio procedure given in Gray and Sen (1983) can provide a 'non-parametric' estimate of the function $F(c_{ij})$, which also may be useful in model formulation. Moreover, linear least squares procedures, unlike ML procedures, do not require initial values of parameters, and can, in fact, be used to supply such initial estimates for ML procedures.

On the other hand, most LS procedures do not work well in the presence of small T_{ij}'s and virtually none of them perform well when a large proportion of T_{ij}'s are very small. Adjustments proposed in Sen and Sööt (1981), Sen and Pruthi (1983) and Gray and Sen (1983), and described in

Section 6.2.3, allow the use of fairly small T_{ij}'s. Some suggestions which could allow even smaller T_{ij}'s have been made in this chapter. Nevertheless, we do not recommend the use of LS procedures when most T_{ij}'s are very small. We shall present some Monte Carlo exercises which address the issue of 'how small' in Section 6.5. We hasten to add that the above discussion excludes the GLIM procedure of McCullach and Nelder (1989) — see also Section 5.5.5 of Chapter 5 — which, because it yields ML estimates, we see as an ML procedure and have treated it as such in Chapter 5, although it consists of iterations based on least squares. We remind the reader that the GLIM procedure was found to be much more demanding of both computer time and space than the best ML procedure.

Available procedures based on non-linear least squares (NLS) do not demand less computer resources than the modified scoring procedure. We also conjecture that the quality of estimates would be no better. Consequently, we have given only a very brief discussion of such procedures in this chapter.

As in Chapter 5, here too, we adopt the same symbol N_{ij} for a random variable and its observed value and assume that all these observations are collected under the same configuration $c \in C$ so that no explicit mention of a configuration need be made. Furthermore, we largely confine our attention to the model:

$$\boldsymbol{N} \in < \text{POISSON} > \tag{6.1}$$

with

$$E(N_{ij}) = T_{ij} = A(i)B(j)F(\boldsymbol{c}_{ij}) \tag{6.2}$$

where

$$F(\boldsymbol{c}_{ij}) = \exp[\boldsymbol{\theta}^t \boldsymbol{c}_{ij}] = \exp[\sum_{k=1}^{K} \theta_k c_{ij}{}^{(k)}]. \tag{6.3}$$

In the above model definition and elsewhere in the chapter, the notation is exactly the same as that in Chapter 5 but differs slightly from that in earlier chapters in ways described in Section 5.1.1.

In (6.2) and (6.3), the $c_{ij}{}^{(k)}$'s are assumed known and the $A(i)$'s, $B(j)$'s and θ_k's are the parameters to be estimated. Because $A(i)$'s and $B(j)$'s can be obtained easily using the DSF procedure (Section 5.3.1) and because in many applications an estimate of $\boldsymbol{\theta}$ is all that is required, we shall focus, as we did in Chapter 5, mainly on estimating the θ_k's. We remind the reader that, as mentioned in Section 5.1.1, the form (6.3) is much more general than might appear on first sight and could be used to at least approximate most F_{ij}'s that one could reasonably wish for.

6.1.1 A Preview of this Chapter

Since a key reason for using LS procedures is computational simplicity, we begin in Section 6.2 by exploring ways of reducing the number of parameters. This, of course, need not have much effect on the diagnostic procedures mentioned above as the other key reason for using least squares.

It turns out that, whether one uses the above mentioned parameter reduction methods or not, unless suitable adjustments are made, Gauss-Markov Conditions are usually violated. These conditions, described briefly in Section 6.2.2, and in more detail in books on least squares (e.g., Sen and Srivastava, 1990, Ch. 2), are sufficient conditions for LS procedures to have desirable properties. In Section 6.2, adjustments are presented which reduce the extent of these violations. The various methods of parameter reduction and of dealing with Gauss-Markov violations result in different LS procedures. One of these is selected for further examination.

In Section 6.3, large sample properties of the selected linear least squares procedures are presented. In Theorem 6.1, the proof of which occupies most of the section, we address both types of large samples: the one where individual T_{ij}'s are large and that where IJ is large. We show that the estimate of $\boldsymbol{\theta}$ is consistent and, more importantly, the bias is smaller than the standard error for large samples. Corollary 6.1 is a slight generalization of the theorem. The methods mentioned in Section 6.2 involve weighting using a random variable related to the dependent variable as weight. This can lead to *substantial* bias and the failure of consistency. The fact that this does not happen for the procedure we focus on and some related procedures adds importance to Theorem 6.1 and Corollary 6.1.

Section 6.3.5, the final section of Section 6.3, covers some practical hints for the use of LS procedures. In particular, it points out that if a standard statistical package is used, standard errors of estimates provided by such procedures should not be used. The section provides a method for computing standard errors.

Some alternatives to the procedures proposed in Section 6.2 are explored in Section 6.4. Section 6.5 presents Monte Carlo results on small sample properties of the procedures. We see that the procedure selected for study in Section 6.3 performs reasonably well as long as we do not have too many zero-valued N_{ij}'s. Some ad hoc adjustments to accommodate this last mentioned type of situation are also examined and are found to perform reasonably well. However, such ad hoc adjustments are just that (ad hoc) and also increase computer time needs. For these reasons, and also because the modified scoring procedure described in the last chapter performs so well, we do not particularly recommend the adjusted procedures.

The penultimate section, Section 6.6, briefly examines nonlinear least squares (NLS) procedures. Since, it appears unlikely that these procedures would be found preferable to ML procedures, our presentation is brief. Nevertheless, we prove that the parameter estimates exist and are unique. Computational algorithms can be constructed by a slight modification of the modified scoring algorithm, that we saw in the last chapter.

It is appropriate to point out that we have paid little attention to forecasting in this chapter. This is appropriate since LS methods are intended for parameter estimation. When these parameters are used for forecasting there is no particular reason to remember that they are LS estimates (see Section 5.7.3).

6.2 LS Procedures

On substituting (6.3) into (6.2) and taking logarithms of both sides of the resultant model, we get

$$t_{ij} = a(i) + b(j) + \sum_{k=1}^{K} \theta_k c_{ij}^{(k)} \tag{6.4}$$

[which is the same as (5.15); recall also that small case letters denote logarithms of quantities denoted by corresponding capital letters]. It can be readily seen that (6.4) is linear in the parameters $a(i)$, $b(j)$ and θ_k and can, therefore, form the basis for an LS procedure. However, three difficulties need to be addressed in order to construct a procedure that is computationally substantially simpler than ML procedures and that which also yields reasonably good estimates in the sense of having acceptably low bias and variances. These are:

- Notice that (6.4) contains $I+J+K$ parameters [$a(i)$'s, $b(j)$'s and θ_k's]. If I and J are large [in transportation studies in major cities each of them can be larger than 1000], an LS procedure need not remain computationally simple. In fact, if nothing is done to reduce the number of parameters, LS procedures can take substantially more time than the modified scoring procedure. Demands on computer memory can also become prohibitive. This has already been illustrated in Section 5.6, since each iteration of the GLIM procedure is essentially a single application of a least squares procedure. Reducing the number of parameters in order to enhance computational efficiency is discussed below in Section 6.2.1.

- In (6.4), t_{ij} is not known. If we replace it with $\log[N_{ij}]$, substantial bias can result when N_{ij}'s are not very large. This issue is discussed below in Section 6.2.3.

- Logarithms of Poisson variables do not have constant variance. Thus, some form of weighting would usually be called for. This is discussed in Section 6.2.4.

Following these discussions a procedure is identified in Section 6.2.5. Minor variations of this procedure are presented in Section 6.2.5 and Section 6.4.

6.2.1 Reduction of Parameters

As mentioned earlier, if $I+J$ is large, there is little computational simplicity in applying LS procedures to (6.4). However, substantial savings in computational time can occur if the parameters $a(i)$ and $b(j)$ are algebraically eliminated before applying least squares. Sen and Sööt (1981) proposed the use of the model

$$\begin{aligned} t_{ij} - t_{i\bullet} - t_{\bullet j} + t_{\bullet\bullet} &= f_{ij} - f_{i\bullet} - f_{\bullet j} + f_{\bullet\bullet} \\ &= \sum_{k=1}^{K} \theta_k [c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}], \end{aligned} \tag{6.5}$$

where $f_{ij} = f(c_{ij}) = \log[F(c_{ij})]$ and the $\bullet$ indicates that a mean has been taken with respect to the subscript it replaces; e.g., $t_{i\bullet} = J^{-1}\sum_{j=1}^{J} t_{ij}$, $t_{\bullet\bullet} = I^{-1}J^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}$. This model, which may be verified to be a straightforward consequence of (6.4) [see also Cesario (1973, 1974a, 1974b)] has another important advantage over (6.4), which will be described in Section 6.4. Our primary focus will be on procedures based on (6.5).

There are other methods of removing the $a(i)$'s and $b(j)$'s besides (6.5). For example, in Gray and Sen (1983) the use of

$$\begin{aligned} t_{ij} + t_{ji} - t_{ii} - t_{jj} &= f_{ij} + f_{ji} - f_{ii} - f_{jj} \\ &\sum_{k=1}^{K} \theta_k [c_{ij}{}^{(k)} + c_{ji}{}^{(k)} - c_{ii}{}^{(k)} - c_{jj}{}^{(k)}] \end{aligned} \tag{6.6}$$

was discussed. Equation (6.6) also follows from (6.4) via straightforward algebraic manipulations. Since the left side of (6.6) is the logarithm of $[T_{ij}T_{ji}/T_{ii}T_{jj}]$ which has been called the odds ratio (see Bishop, *et al.*, 1975), procedures based on (6.6) may be called odds ratio procedures [the reader might notice that this is reminiscent of elements in Chapters 2 and 4 — e.g., see (2.20)]. Further details can be found in Gray and Sen (1983), which also presents numerical applications (including one for the Skokie data) showing that estimates obtained using (6.6) fit data very well. A generalization of (6.6) is

$$t_{ij} + t_{pq} - t_{iq} - t_{pj} = f_{ij} + f_{pq} - f_{iq} - f_{pj}, \tag{6.7}$$

which could be valuable if certain of N_{ij}'s were not available.

In Sen and Pruthi (1983), the use of

$$t_{ij} - t_{i\bullet} - t_{\bullet j} + t_{\bullet\bullet} - (I-1)^{-1} t_{ji} = f_{ij} - f_{i\bullet} - f_{\bullet j} + f_{\bullet\bullet} - (I-1)^{-1} f_{ji} \tag{6.8}$$

was suggested for cases when $I = J$ and the observations N_{ii} are missing (as would happen in, for example, air travel; the paper itself presents an application on freight movements in India). In (6.8), the computation of means,

$t_{i\bullet}$, $t_{\bullet j}$, etc., would exclude t_{ii}'s and f_{ii}'s, e.g., $t_{i\bullet} = (I-1)^{-1}\sum_{j\neq i} t_{ij}$ [i.e., the notation in (6.5) and in (6.8) are not the same]. Equation (6.7) can also be readily derived from (6.4).

Although we shall not devote significant attention to these models in this book, most of the theoretical results presented in this chapter apply to (6.8) to about the same extent as they do to (6.5). Although we have not examined whether Theorem 6.1 or Corollary 6.1 holds for (6.6), we see no reason why that would not be the case.

Non-Parametric Estimation of F

The following assumptions may be appropriate in several situations:

1. $F_{11} = F_{22} = \cdots = F_{II}$, i.e., within zone movement for one zone is as onerous as it is for any other zone. Then, without loss of generality, we can set $F_{ii} = 1$, for all i.

2. For all i and j, $F_{ij} = F_{ji}$. This would be true, for example, if costs were symmetric, i.e., if $c_{ij}{}^{(k)} = c_{ji}{}^{(k)}$ for all k.

Then, it follows from the first line of (6.6) that

$$[T_{ij}T_{ji}/T_{ii}T_{jj}]^{1/2} = F_{ij}. \tag{6.9}$$

This equation gives a way to explicitly estimate each $F_{ij} = F[\mathbf{c}_{ij}]$ without having to identify $c_{ij}{}^{(k)}$'s, estimate $\boldsymbol{\theta}$ or even to specify an algebraic form for F. F_{ij}'s estimated in this way can be used in many different ways. For example, they could be valuable in exploratory studies to identify the algebraic form of $F[\mathbf{c}_{ij}]$ in a specific application. They could also be used for forecasting or for tracking changes over time of F_{ij}'s without having to get involved in identifying $\mathbf{c}_{ij}$'s. Further discussion and numerical illustrations of the use of (6.9) are presented in Gray and Sen (1983).

6.2.2 Gauss-Markov Conditions

In the last section, we addressed methods by which the number of parameters can be reduced substantially before applying an LS procedure to estimate $\boldsymbol{\theta}$. Regardless of whether or not such a parameter reduction is carried out, we must still address the assumptions underlying the use of LS procedures.

Bias and relatively large variances of LS estimators occur when certain well-known conditions, called Gauss-Markov Conditions, are violated. Moreover, these conditions underlie some of the formulæ used in least squares analysis (e.g., that for the covariance matrix) and would change if the conditions are not met. Therefore, it is appropriate to start this discussion with a description of these conditions.

Consider the familiar linear regression model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e} \tag{6.10}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^t$ is a vector of observations of the dependent variable, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^t$ is a vector of the parameters to be estimated, $\boldsymbol{e} = (e_1, \ldots, e_n)^t$ is a vector of the errors and $X = ((x_{ps}))$ is the $n \times K$ matrix of independent variable values, x_{ps}. According to the Gauss-Markov Theorem (e.g., see Sen and Srivastava, 1990, p. 41), least squares yields a good (i.e., unbiased and among all unbiased linear estimates the one with the smallest variance — the so called BLU estimate) estimate of $\boldsymbol{\beta}$ if the following three conditions, collectively called the Gauss-Markov Conditions, are met:

$$\mathrm{E}(e_i) = 0 \quad \text{for all } i = 1, \ldots, n \tag{6.11}$$

$$\mathrm{E}(e_i^2) = \sigma^2 \quad \text{for all } i = 1, \ldots, n \tag{6.12}$$

$$\mathrm{E}(e_i e_j) = 0 \quad \text{for all } i, j = 1, \ldots, n. \tag{6.13}$$

As an aside it might be pointed out that (6.14) and the proof of Part 3 of Lemma 6.3 actually constitute a proof that these conditions yield the pleasant properties that we have claimed above. The failure of the first of these conditions, i.e., (6.11), usually results in the estimate of $\boldsymbol{\beta}$ becoming biased. But the condition is not a necessary condition: If $\boldsymbol{b}$ is the LS estimate of $\boldsymbol{\beta}$ in the model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$, then it is well known that (see, for example, Sen and Srivastava, 1990, Ch. 2) $\boldsymbol{b} = (X^t X)^{-1} X^t \boldsymbol{y}$. It then follows from (6.10), that

$$\boldsymbol{b} = (X^t X)^{-1}[X^t X\boldsymbol{\beta} + X^t \boldsymbol{e}] = \boldsymbol{\beta} + (X^t X)^{-1} X^t \boldsymbol{e}, \tag{6.14}$$

and $\boldsymbol{b}$ can be unbiased (i.e., $\mathrm{E}[\boldsymbol{b}] = \boldsymbol{\beta}$), if, say, $X^t\mathrm{E}[\boldsymbol{e}] = 0$, even if $\mathrm{E}[\boldsymbol{e}] \neq 0$. We mention this here because an analogous situation occurs later in this chapter.

The failure of either of the other two Gauss-Markov Conditions, (6.12) and (6.13) does not cause bias; however, the estimate then may not have the minimum variance property. Although this latter property may not be too vital because of the large size of $N_{\oplus\oplus}$ that one typically encounters, serious violations of (6.12) and (6.13) should nonetheless be avoided — see also Section 6.4.

6.2.3 Bias

Jensen's inequality (see Rao, 1973, p.58) states that for convex functions $g(x)$ of x, $\mathrm{E}[g(x)] \geq g[\mathrm{E}(x)]$, with the equality holding if and only if the distribution of x is degenerate with all the probability concentrated on one value. It follows that

$$\mathrm{E}[\log(N_{ij})] \neq \log[\mathrm{E}(N_{ij})] = t_{ij}. \tag{6.15}$$

Therefore, the replacement of t_{ij} by $\log(N_{ij})$ in (6.5) (or, for that matter in (6.4)) would usually cause (6.11) to be violated and yield a biased model, since then, because of (6.5),

$$E[z^*_{ij} - z_{i\bullet}{}^* - z_{\bullet j}{}^* + z^*_{\bullet\bullet}] - \sum_{k=1}^{K} \theta_k [c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}] \neq 0$$

where $z^*_{ij} = \log[N_{ij}]$. However, when N_{ij} is Poisson and T_{ij} is not too small, this bias is considerably reduced by the simple expedient of adding a half to N_{ij} before taking logarithms, that is, by replacing t_{ij} by $z_{ij} = \log[N_{ij} + \frac{1}{2}]$ instead of $z^*_{ij} = \log(N_{ij})$. This is shown below.

But before doing this, let us digress to discuss the $\mathcal{O}$ notation which we use throughout this chapter and, in particular, in Lemma 6.1. For any function $g(n)$ of a positive integer n and any number α, the statement '$g(n) = \mathcal{O}(n^{-\alpha})$' or the statement '$g(n)$ is of the order of $n^{-\alpha}$' will mean that, for large enough n, the quotient $g(n)/n^{-\alpha}$ is bounded, i.e., there exist numbers $\kappa > 0$ and $n(\kappa)$, such that $n > n(\kappa) \Rightarrow |g(n)/n^{-\alpha}| \leq \kappa$. Similarly, for a continuous variable x,

$$g(x) = \mathcal{O}(g_1(x)) \Leftrightarrow \limsup_{x\to\infty} |g(x)/g_1(x)| \leq \kappa. \tag{6.16}$$

No confusion need arise and notation will be simplified if we also use $\mathcal{O}(A)$ for a function of order A. Thus we may make statements like $g = \mathcal{O}(A) + \mathcal{O}(B)$ in place of $g = g_1 + g_2$ where $g_1 = \mathcal{O}(A)$ and $g_2 = \mathcal{O}(B)$. The following easily verified relations will also be freely used in the rest of the chapter:

$$\mathcal{O}(A)\mathcal{O}(B) = \mathcal{O}(AB),$$
$$\mathcal{O}(n^{-\alpha_1}) + \mathcal{O}(n^{-\alpha_2}) = \mathcal{O}(n^{-\alpha_1}),$$

if $0 \leq \alpha_1 \leq \alpha_2$. Also the term $g_1 = \mathcal{O}(n^{-\alpha_1})$ will be called of higher order than, or asymptotically smaller than, $g_2 = \mathcal{O}(n^{-\alpha_2})$.

Lemma 6.1 $E[\psi_{ij}] = -1/(24T^2_{ij}) + 29/(12T^3_{ij}) + \cdots = \mathcal{O}(T_{ij}^{-2})$ *where* $\psi_{ij} = z_{ij} - t_{ij}$ *and* $z_{ij} = \log[N_{ij} + \frac{1}{2}]$.

PROOF: Let $\epsilon_{ij} = N_{ij} - T_{ij}$. Then

$$\begin{aligned}\psi_{ij} + t_{ij} = z_{ij} &= \log[N_{ij} + \frac{1}{2}] = \log[T_{ij} + \epsilon_{ij} + \frac{1}{2}] \\ &= \log[T_{ij}] + \log[1 + (\epsilon_{ij} + \frac{1}{2})/T_{ij}] \\ &= t_{ij} + (\epsilon_{ij} + \frac{1}{2})/T_{ij} - (\epsilon_{ij} + \frac{1}{2})^2/2T^2_{ij} \\ &\quad + (\epsilon_{ij} + \frac{1}{2})^3/3T^3_{ij} - (\epsilon_{ij} + \frac{1}{2})^4/4T^4_{ij} \\ &\quad + (\epsilon_{ij} + \frac{1}{2})^5/5T^5_{ij} - (\epsilon_{ij} + \frac{1}{2})^6/6T^6_{ij} + \cdots .\end{aligned} \tag{6.17}$$

Since N_{ij} has a Poisson distribution with expectation T_{ij}, its characteristic function is $\exp[-T_{ij}\{1-\exp(\imath\zeta)\}]$, where $\imath = \sqrt{-1}$ and ζ is the argument of the characteristic function (see Rao, 1973, p.99 *et. seq.*). Therefore, the second characteristic or the cumulant-generating function of N_{ij}, which is the logarithm of the characteristic function, is

$$-T_{ij} + T_{ij}\exp(\imath\zeta). \tag{6.18}$$

The coefficients of $(\imath\zeta)^r/r!$ in a power series expansion of the second characteristic are the cumulants; in this case it can easily be seen that all cumulants after the first are T_{ij}. Using the relationships between cumulants and moments (see Kendall and Stuart, 1963, p.84) we get after some straightforward, but tedious, algebra:

$$\begin{aligned}
&\mathrm{E}[\epsilon_{ij}^2] = T_{ij},\ \mathrm{E}[\epsilon_{ij}^3] = T_{ij},\ \mathrm{E}[\epsilon_{ij}^4] = 3T_{ij}^2 + T_{ij},\\
&\mathrm{E}[\epsilon_{ij}^5] = T_{ij} + 10T_{ij}^2,\ \mathrm{E}[\epsilon_{ij}^6] = T_{ij} + 25T_{ij}^2 + 15T_{ij}^3,\\
&\mathrm{E}[\epsilon_{ij}^7] = T_{ij} + 56T_{ij}^2 + 105T_{ij}^3,\\
&\mathrm{E}[\epsilon_{ij}^8] = T_{ij} + 119T_{ij}^2 + 490T_{ij}^3 + 105T_{ij}^4,\\
&\mathrm{E}[\epsilon_{ij}^9] = T_{ij} + 246T_{ij}^2 + 1918T_{ij}^3 + 1260T_{ij}^4.
\end{aligned} \tag{6.19}$$

The lemma follows on taking expectations of all expressions in (6.17) and using (6.19). □

The proof of the lemma is more detailed than immediately necessary, since moments up to order 4, which are all that are strictly necessary for the lemma, are available from standard sources. However, the additional moments shorten the presentation of some future computations.

Note that if the half were omitted, computations very similar to those in the proof of Lemma 6.1 [essentially leaving the half out of every step in (6.17)] yield

$$\mathrm{E}[\log(N_{ij})] = \log(T_{ij}) - 1/(2T_{ij}), \tag{6.20}$$

resulting, when T_{ij}'s are not too small, in a much larger bias. Corrections like the addition of half are sometimes called Anscombe's corrections (Anscombe, 1948, Rao, 1973, p.426).

Exhibit 6.1 shows values of $E[z_{ij}] - t_{ij}$ for selected values of T_{ij}. Bias appears to be essentially eliminated when $T_{ij} \geq 3$. For smaller values, the addition of $\frac{1}{2}$ is not enough. Thus we see that if T_{ij}'s are large (≥ 3), the use of $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$ as the dependent variable provides an essentially bias-free procedure based on (6.5). The regression model then becomes

$$\begin{aligned}
&z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}\\
&= \sum_{k=1}^{K} \theta_k [c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}] + \psi_{ij} - \psi_{i\bullet} - \psi_{\bullet j} + \psi_{\bullet\bullet},
\end{aligned} \tag{6.21}$$

T_{ij}	$E[z_{ij}] - t_{ij}$
1	.1689
2	.0259
3	.0019
4	-.0020
5	-.0021
6	-.0016
7	-.0012
15	-.0002

Exhibit 6.1: Values of $E[z_{ij}] - t_{ij}$ for selected values of T_{ij}

for $i \in I$ and $j \in J$. It is easy to see that the expectation $\mathrm{E}[e_{ij}]$ of the error

$$e_{ij} = \psi_{ij} - \psi_{i\bullet} - \psi_{\bullet j} + \psi_{\bullet\bullet} \tag{6.22}$$

is very close to zero when T_{ij}'s are moderately large.

We shall see in Section 6.2.4 that when N_{ij}'s are small, they receive a very small weight. Therefore, an occasional small T_{ij} or N_{ij} need not concern us unduly. The Monte Carlo exercises of Section 6.5 gives some practical 'feel' for what might be acceptable.

While any correction based on Exhibit 6.1 should properly be made on the basis of T_{ij}, replacing z_{ij} by v_{ij} where

$$v_{ij} = \begin{cases} z_{ij} - .1689 & \text{when } N_{ij} = 1 \\ z_{ij} - .0259 & \text{when } N_{ij} = 2 \\ z_{ij} & \text{otherwise} \end{cases} \tag{6.23}$$

might yield a useful procedure. When $T_{ij} \geq 1$, we could consider using $v_{ij} - v_{i\bullet} - v_{\bullet j} + v_{\bullet\bullet}$ as the dependent variable instead of $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$. Such a procedure will be explored in Section 6.5.

Bias Corrections for Other Procedures

The bias in LS procedures based on (6.6) or (6.8) can be reduced in exactly the same way. We would add a half to N_{ij} before taking logs and would insert the resultant quantity z_{ij} in place of t_{ij} in the appropriate expression.

Simply replacing T_{ij} by N_{ij} in (6.9) also leads to a biased model. Instead, based on arguments very similar to those above and as mentioned in Gray and Sen (1983), we recommend the use of

$$\left(\frac{(N_{ij} + 1/4)(N_{ji} + 1/4)}{(N_{ii} + 3/4)(N_{jj} + 3/4)} \right)^{1/2}$$

as an estimate of F_{ij} under the two conditions mentioned just before (6.9).

From the proof of Lemma 6.1, it follows that the reduction of bias from the addition of half is determined by the fact that N_{ij} is Poisson. If N_{ij} is not Poisson but if the variance $\mathrm{var}\,[N_{ij}]$ of N_{ij} has the form

$$\mathrm{var}\,[N_{ij}] = \alpha \mathrm{E}[N_{ij}] \tag{6.24}$$

then the addition of $\alpha/2$ reduces bias. Since $\mathrm{var}\,[\alpha^{-1}N_{ij}] = \alpha^{-2}\,\mathrm{var}\,[N_{ij}]$, we can write (6.24) equivalently as

$$\mathrm{var}\,[\alpha^{-1}N_{ij}] = \mathrm{E}[\alpha^{-1}N_{ij}].$$

This is particularly useful in studies of freight movements where α could be seen as some natural unit. Then all one would need to do is divide flows by α tons (say) and then add half, or, equivalently, add $\alpha/2$, as mentioned before. Sen and Pruthi (1983) have applied the procedure implied by (6.8) to food-grain and coal shipments between Indian states. Since these commodities move as train-loads, one train-load was seen to be a 'natural' unit. Thus, total volumes of food-grains were divided by the average size $[\alpha]$ of a train-load of food-grain before the addition of half. Coal was treated similarly. It might be mentioned that the estimates obtained in this way were found to be numerically very close to ML estimates.

Another use of (6.24) is with factored data [Section 5.9.3 p. 468], where factored observations are $N^*_{ij} = \alpha N_{ij}$. Then

$$\mathrm{var}\,[N^*_{ij}] = \alpha^2\,\mathrm{var}\,[N_{ij}] = \alpha^2\mathrm{E}[N_{ij}] = \alpha\mathrm{E}[N^*_{ij}]$$

if N_{ij}'s are Poisson. Here again, we would either divide by α before adding half or simply add $\alpha/2$.

Such factoring using constant factors α creates no problems in any of the procedures we consider in this section. However, more complex factoring schemes, involving factors that vary with i and j, usually do create problems, since they violate the distributional assumptions and produce weights which are inappropriate. Thus, as in Chapter 5, it is safer to use data which either have not been factored or in which all observations have been factored to the same extent. This is true for both linear least squares and non-linear least squares. [See also Kirby and Leese, 1978].

6.2.4 Weighting

For large enough T_{ij}, by substituting a power series for $\log(1 + \{(\epsilon_{ij} + \frac{1}{2})/T_{ij}\})$ — see (6.17) — into

$$\begin{aligned}\mathrm{var}\,[\log(N_{ij} + \frac{1}{2})] &\approx \mathrm{E}[\log(N_{ij} + \frac{1}{2}) - \log(T_{ij})]^2 \\ &= \mathrm{E}[\log\{(N_{ij} + \frac{1}{2})/T_{ij}\}]^2 = \mathrm{E}[\log(1 + \{(\epsilon_{ij} + \frac{1}{2})/T_{ij}\})]^2\end{aligned} \tag{6.25}$$

it may be shown that

$$\breve{T}_{ij}^{-1} = \text{var}\,[\log(N_{ij} + \frac{1}{2})] = T_{ij}^{-1} + \mathcal{O}((T_{ij}^{-2})). \tag{6.26}$$

where the first equality defines $\breve{T}_{ij}$. It follows from (6.26) that for large values of I and J, the variance of $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$ is

$$\begin{aligned} \text{var}\,[z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}] &= [1 - I^{-1} - J^{-1} + I^{-1}J^{-1}]^2\, \text{var}\,[z_{ij}] \\ + I^{-2}[1 - J^{-1}]^{-1} \sum_{\substack{s=1\\ s\neq i}}^{I} \text{var}\,[z_{sj}] &+ J^{-2}[1 - I^{-1}]^{-1} \sum_{\substack{p=1\\ p\neq j}}^{J} \text{var}\,[z_{ip}] \\ + (IJ)^{-2} \sum_{\substack{s=1\\ s\neq i}}^{I} \sum_{\substack{p=1\\ p\neq j}}^{J} \text{var}\,[z_{sp}]. \end{aligned} \tag{6.27}$$

Notice that the computation of the term on the right side of the first line took account of the fact that z_{ij} occurs in $z_{i\bullet}$, $z_{\bullet j}$ and $z_{\bullet\bullet}$, the next term takes into account the fact that $z_{\bullet j}$ includes terms which are also in $z_{\bullet\bullet}$, etc. If $\max \breve{T}_{ij}^{-1} = \breve{T}^{-1}$, the last two lines of (6.27) can be seen to be less than

$$[I^{-1}[1 - J^{-1}]^{-1} + J^{-1}[1 - I^{-1}]^{-1} + (IJ)^{-1}]\breve{T}^{-1}$$

which $\to 0$ if $I \to \infty$ and $J \to \infty$ while $\breve{T}^{-1}$ remains bounded. Hence for large I and J and $\breve{T}_{ij}$'s bounded away from zero, we get

$$\begin{aligned} &\text{var}\,[z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}] \\ &\quad \approx [1 - I^{-1} - J^{-1} + I^{-1}J^{-1}]^2/\breve{T}_{ij} = \gamma^2/\breve{T}_{ij} \text{ (say)} \to T_{ij}^{-1} \end{aligned} \tag{6.28}$$

from (6.26).

The last line of (6.28) would usually be far from a constant as is required by condition (6.12). This problem can be alleviated by weighting by the reciprocal of a quantity roughly proportional to the variance. Consequently, for large values of I, J and T_{ij}, a reasonable set of weights would be T_{ij}, since by equation (6.26),

$$T_{ij} \approx \breve{T}_{ij}. \tag{6.29}$$

However, T_{ij}'s are not available. Therefore, we recommend the use of N_{ij}'s as weights.

This choice of weight carries with it somewhat mixed blessings. It is much simpler than the alternative of using an iteratively reweighted least squares in which we would first run ordinary least squares (i.e., LS 'without weighting' or, correctly stated, LS with equal weights for all observations), obtain estimates $\hat{T}_{ij}$'s of T_{ij}, use $\hat{T}_{ij}^{-1}$ as weights for a second exercise of LS, weight with $\hat{T}_{ij}^{-1}$ obtained from this second exercise, and continue iterating until a suitable convergence criterion is met (see Sen and Srivastava, 1990, Ch 6). Also, weighting by N_{ij}'s removes from the regression all observations z_{ij} for which $N_{ij} = 0$. This latter advantage is particularly important since bias is especially difficult to remove for very small values of T_{ij}, mainly because, since $\log(x) \to -\infty$ as $x \to 0$, the amount of bias is very unstable for such values.

However, using random variables for weighting is dangerous business. It can induce bias on its own. However, as we show in the next section, if certain conditions hold, the bias is negligible.

This weighting does not quite cause Condition (6.12) to be met, as we now show. The weights have the effect of multiplying the dependent variable by $N_{ij}^{1/2}$. If T_{ij} is large,

$$\begin{aligned} \operatorname{var}[N_{ij}^{\frac{1}{2}} \log(N_{ij} + \frac{1}{2})] &\approx \left[\frac{d[T_{ij}^{\frac{1}{2}} \log(T_{ij} + \frac{1}{2})]}{dT_{ij}}\right]^2 \operatorname{var}(N_{ij}) \\ &= \left(T^{-\frac{1}{2}}\left[\frac{T_{ij}}{T_{ij} + \frac{1}{2}} + \frac{1}{2}\log[T_{ij} + \frac{1}{2}]\right]\right)^2 T_{ij} \approx \left(1 + \frac{1}{2}\log[T_{ij}]\right)^2, \end{aligned} \tag{6.30}$$

where the quantity in square bracket on the first line is a derivative [see Sen and Srivastava (1990, p. 115) — the material just preceding equation (6.4) in that book — for a proof of the first line]. While the right side of the last of (6.30) is not exactly a constant, it varies considerably less than does T_{ij}^{-1}. For example, if $T_{11} = 10000$ and $T_{12} = 100$, from (6.26),

$$\frac{\operatorname{var}[y_{12}]}{\operatorname{var}[y_{11}]} \approx \frac{\operatorname{var}[\log(N_{12} + \frac{1}{2})]}{\operatorname{var}[\log(N_{11} + \frac{1}{2})]} \approx \frac{T_{11}}{T_{12}} = 100,$$

where $y_{ij} = z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$; while, using (6.30), we get

$$\frac{\operatorname{var}[N_{11}^{\frac{1}{2}} y_{11}]}{\operatorname{var}[N_{12}^{\frac{1}{2}} y_{12}]} \approx \frac{\operatorname{var}[N_{11}^{\frac{1}{2}} \log(N_{11} + \frac{1}{2})]}{\operatorname{var}[N_{12}^{\frac{1}{2}} \log(N_{12} + \frac{1}{2})]} \approx \frac{31.4}{10.9} < 2.9.$$

The effect of weighting by N_{ij} on the variance of estimates will also be discussed in Section 6.3.4.

6.2.5 Procedures

We are now ready to formulate a recommended procedure based on (6.5). As we saw in the last two sections [6.2.3 and 6.2.4], a desirable procedure would consist of replacing t_{ij} by $z_{ij} = \log[N_{ij} + \frac{1}{2}]$ in (6.5) and using weighted least squares with weights N_{ij}. Thus, the procedure is:

1. Set $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$ as the dependent variable,

2. Set $c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}$ for $k = 1, \ldots, K$ as independent variables,

3. Run a weighted LS procedure with weights N_{ij}'s and with no intercept.

The estimated coefficients of the independent variables will be the estimates of the θ_k's. We call this **Procedure 1.** This procedure is the one we shall examine in detail in Section 6.3.

The Sen and Pruthi (1983) and Gray and Sen (1983) LS procedures can also be summarized using statements made about them in Sections 6.2.3 and 6.2.4. For the Sen and Pruthi procedure, the dependent variable would be $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet} - (I-1)^{-1} z_{ji}$, the independent variables would be $c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)} - (I-1)^{-1} c_{ji}{}^{(k)}$ and the weight would be N_{ij}. For the Gray and Sen procedure, $z_{ij} + z_{ji} - z_{ii} - z_{jj}$ would be the dependent variable, $c_{ij}{}^{(k)} + c_{ji}{}^{(k)} - c_{ii}{}^{(k)} - c_{jj}{}^{(k)}$ the independent variables and the weight would be $[N_{ij}^{-1} + N_{ji}^{-1} - N_{ii}^{-1} - N_{jj}^{-1}]^{-1}$. In neither case would there be an intercept term. We shall not explicitly deal with either procedure in the following sections.

Designate as **Procedure 1A** that where $v_{ij} - v_{i\bullet} - v_{\bullet j} + v_{\bullet\bullet}$ is the dependent variable and everything else is the same as in Procedure 1. Practical details on the use of Procedures 1 and 1A are given in Section 6.3.5.

Obviously, one can construct other LS procedures for the estimation of gravity model parameters. For example, the ordinary least squares version of Procedure 1 and a weighted or ordinary least squares version of a procedure based on (6.4) without any algebraic manipulation are such possibilities. Several such procedures are described and commented on in Section 6.4.

6.3 Large Sample Theory

Unfortunately only asymptotic properties are available for Procedure 1. These are discussed in this section. Small sample properties may be investigated by Monte Carlo methods. The results of such an investigation are presented in Section 6.5.

We begin this section with the description of some notation after which some conditions are stated. These enable us then, in Section 6.3.2, to state Theorem 6.1 and Corollary 6.1, which are the principal results of this section. The remainder of the section is devoted to proving them.

6.3.1 Preliminaries

As in Chapter 5, let

$$\boldsymbol{N} = (N_{11}, \ldots, N_{1J}, \ldots\ldots, N_{I1}, \ldots, N_{IJ})^t, \tag{6.31}$$

$$\boldsymbol{T} \;=\; (T_{11}, \ldots, T_{1J}, \ldots\ldots, T_{I1}, \ldots, T_{IJ})^t \tag{6.32}$$

and $\boldsymbol{\epsilon} = \boldsymbol{N} - \boldsymbol{T}$. Also let

$$\begin{aligned} \breve{\boldsymbol{T}} &= (\breve{T}_{11}, \ldots, \breve{T}_{1J}, \ldots\ldots, \breve{T}_{I1}, \ldots, \breve{T}_{IJ})^t, & (6.33)\\ \boldsymbol{\psi} &= (\psi_{11}, \ldots, \psi_{1J}, \ldots\ldots, \psi_{I1}, \ldots, \psi_{IJ})^t, & (6.34)\\ \boldsymbol{e} &= (e_{11}, \ldots, e_{1J}, \ldots\ldots, e_{I1}, \ldots, e_{IJ})^t, & (6.35)\\ \text{and } \boldsymbol{y} &= (y_{11}, \ldots, y_{1J}, \ldots\ldots, y_{I1}, \ldots, y_{IJ})^t, & (6.36) \end{aligned}$$

where $\breve{T}_{ij}^{-1} = \operatorname{var}[z_{ij}]$, $y_{ij} = z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$, $\psi_{ij} = z_{ij} - t_{ij}$ and $e_{ij} = \psi_{ij} - \psi_{i\bullet} - \psi_{\bullet j} + \psi_{\bullet\bullet}$ have been previously defined in sections 6.2.4 and 6.2.3.

Sometimes, it will be convenient to replace the double subscript by a single subscript, e.g., write $\boldsymbol{N}$ as $(N_1, \ldots, N_n)^t$ with typical component N_s. In that case, if $N_{ij} = N_s$, one could view s as equaling $(i-1)J + j$. Let

$$x_{sk} = c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}$$

and let X be the matrix $X = ((x_{sk}))$ with typical element x_{sk}. Obviously,

$$n = IJ. \tag{6.37}$$

Then the model (6.21) may be written in matrix form as

$$\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{e}. \tag{6.38}$$

We shall assume X is of full rank, i.e., the rank of X is K.

Since we shall frequently be using diagonal matrices constructed from the vectors $\boldsymbol{N}$, $\boldsymbol{e}$, $\boldsymbol{T}$, etc, for this section and the next, we make the convention that such diagonal matrices will be represented by the corresponding unsubscripted non-bold letter; e.g., $N = \operatorname{diag}(\boldsymbol{N})$, $T = \operatorname{diag}(\boldsymbol{T})$, $\breve{T} = \operatorname{diag}(\breve{\boldsymbol{T}})$, $e = \operatorname{diag}(\boldsymbol{e})$, $\epsilon = \operatorname{diag}(\boldsymbol{\epsilon})$, etc., where, as before, diag indicates a diagonal matrix with non-zero entries equal to the elements of the quantity within parentheses. Using this notation, we see that weighting by N_{ij}'s is equivalent to multiplying each term on both sides of (6.38) by the diagonal matrix $N^{1/2}$ on the left. The weighted least squares estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= [X^tNX]^{-1}X^tN\boldsymbol{y} \\ &= [X^tNX]^{-1}X^tN(X\boldsymbol{\theta} + \boldsymbol{e}) \\ &= \boldsymbol{\theta} + [X^tNX]^{-1}X^tN\boldsymbol{e}. \end{aligned} \tag{6.39}$$

Let $U = X^tTX$ and $V = X^t\epsilon X$. Then, since $\boldsymbol{N} = \boldsymbol{T} + \boldsymbol{\epsilon}$,

$$\begin{aligned} [X^tNX]^{-1} &= [U+V]^{-1} = [U(I_K + U^{-1}V)]^{-1} \\ &= (I_K + U^{-1}V)^{-1}U^{-1}, \end{aligned} \tag{6.40}$$

where I_K is the K-dimensional identity matrix, and if we let

$$I_K + W = (I_K + U^{-1}V)^{-1}, \ \boldsymbol{w}_1 = U^{-1}X^tT\boldsymbol{e} \text{ and } \boldsymbol{w}_2 = U^{-1}X^t\epsilon\boldsymbol{e}, \quad (6.41)$$

we see from (6.39) that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (I_K + W)(\boldsymbol{w}_1 + \boldsymbol{w}_2). \quad (6.42)$$

Further, let $\boldsymbol{w} = (X^tTX)^{-1}X^tT(\boldsymbol{e} - \mathrm{E}[\boldsymbol{e}]) = \boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1]$.

We shall also make the following conventions. A matrix (or vector) $H = ((h_{ps}))$ will be said to be less than (not greater than) $G = ((g_{ps}))$, if H and G have the same dimension and, for all p and s, $|h_{ps}| < |g_{ps}|$ ($|h_{ps}| \leq |g_{ps}|$). Note that we are comparing absolute values. If each diagonal element of a matrix $H = ((h_{ps}))$ is less than (less than or equal to) the corresponding diagonal element of the matrix $G = ((g_{ps}))$ of the same dimension, we shall denote this phenomenon by $H \prec G$ ($H \preceq G$). A matrix H will be said to be of order $\mathcal{O}(\rho)$ ($H = \mathcal{O}(\rho)$) if each of its elements is of order ρ. For example, if all x_{ps} are bounded, a matrix with all its elements proportional to $\sum_{p=1}^n x_{ps}^2$ is of order $\mathcal{O}(n)$; and, if all its elements were proportional to $[\sum_{p=1}^n x_{ps}^2]^{-1}$, the matrix would be of order $\mathcal{O}(n^{-1})$.

Since we are interested in asymptotic properties of least squares estimators, we are interested in a sequence of models:

$$\boldsymbol{y}^{(m)} = X^{(m)}\boldsymbol{\theta} + \boldsymbol{e}^{(m)} \quad (6.43)$$

and the corresponding least squares estimators $\hat{\boldsymbol{\theta}}^{(m)}$, where $m = 1, 2, 3, \ldots$. The results that concern us involve the limiting value of $\hat{\boldsymbol{\theta}}^{(m)}$ as $m \to \infty$ under appropriate conditions on $\boldsymbol{y}^{(m)}$ and $X^{(m)}$. However, some simplicity of notation occurs, and nothing much is lost, if where asymptotic results are considered, m is not explicitly mentioned but is simply implied. Thus, we will simply talk about the limiting behavior of $\hat{\boldsymbol{\theta}}$ when we mean $\hat{\boldsymbol{\theta}}^{(m)}$ from the model (6.43).

As mentioned before, in the gravity model context, large samples can occur in two ways:

- $n = IJ$ could be large, or
- each T_{ij} could be large.

Consequently, the key results of this section (Theorem 6.1 and Corollary 6.1) are for either kind of 'large,' as long as all T_{ij}'s are not so small that just taking a log transformation (after adding a half) creates substantial bias (see Section 6.2.3). In order to accommodate both types of 'large' within the same development, we make the following notation: The statement $g(x, n) = \mathcal{O}(n^{-\alpha}g_1(x))$ will mean that there exist numbers κ, $n(\kappa)$ and $x(\kappa)$ such that

$$n > n(\kappa) \text{ and } x > x(\kappa) \Rightarrow |g(x, n) \leq \kappa[n^{-\alpha}g_1(x)]. \quad (6.44)$$

Usually, x would be $T_{\#}$, defined below, and $\lim_{x\to\infty} g_1(x) = 0$ and $\alpha > 0$, so that if $x = T_{\#}$ and n are both large enough, $g(x,n) = \mathcal{O}(n^{-\alpha}g_1(x))$ implies that $g(x,n) \to 0$ if *either* $x \to \infty$ or $n \to \infty$.

For any $\boldsymbol{T}$ let $\max_{ij} T_{ij} = T^{\#}$ and $\min_{ij} T_{ij} = T_{\#}$. [Notice that a superscript (m) should have been placed on each T_{ij}, $T^{\#}$ and $T_{\#}$ but has been omitted by the convention mentioned above.] The following conditions on the relative sizes of T_{ij}'s and I and J are used in Theorem 6.1 below:

LS1. $T_{\#}$ is bounded below and $T^{\#}/T_{\#} = L^{(m)} \to L$.

LS2. $I = \mathcal{O}(n^{\frac{1}{2}})$ and $J = \mathcal{O}(n^{\frac{1}{2}})$.

The latter condition obviously implies that the relative sizes of I and J remain more or less the same as they grow larger. The first part of Condition LS1 prevents us from increasing the number of zones without increasing $T_{\oplus\oplus} = \sum_{ij} T_{ij}$, i.e., trying to get large sample results without necessarily a large number of observations. The second part of LS1 prevents some T_{ij}'s from dominating the computations.

These conditions do not appear to be too restrictive for most applications. However, LS1 can be weakened as follows:

LS1A. $T_{\#}$ is bounded below and $T^{\#}/T_{\#} = \mathcal{O}(T_{\#}^{\alpha})$ where $0 \le \alpha < 1$.

This condition is used in Corollary 6.1 which is a generalization of Theorem 6.1. Also, $T^{\#} = T^{(m)\#}$ and $T_{\#} = T^{(m)}{}_{\#}$ in Conditions LS1 and LS1A can be replaced by $\limsup_{ij} T_{ij}{}^{(m)}$ and $\liminf_{ij} T_{ij}{}^{(m)}$, although the corresponding treatment, which is a straightforward generalization, would still add to the already lengthy treatment below.

Another condition that we shall use is:

LS3. Each matrix $X = X^{(m)}$ is bounded, i.e., for every m and every i, j and k, $|x_{sk}{}^{(m)}| < \mu$ for some $\mu > 0$. [Recall that we can write $s = (i-1)J + j$.]

Although this condition might appear to be unduly stringent, it should usually be expected to hold in the gravity model context. When each $T_{ij} \to \infty$, the separation measures $c_{ij}{}^{(k)}$'s would remain about the same. When $IJ \to \infty$ the increase in IJ would typically occur by subdividing the origin and destination zones, which also would not have any effect on the boundedness of the elements of X.

A final condition we need is:

LS4. $X^t N X$ is non-singular and $(X^t X)^{-1} = \mathcal{O}(n^{-1})$.

The first part of the condition is necessitated by the requirement that we have a unique least squares estimator. It is very unlikely to be violated in practice for moderate to large values of $T_{\#}$. Nevertheless, since it is theoretically possible for a violation to occur (e.g., when $N = 0$) we make this explicit statement.

By the convention mentioned above, the second part of the condition is a simplified restatement of $(X^{(m)^t}X^{(m)})^{-1} = \mathcal{O}(n^{-1})$. Conditions of the type 'as $n \to \infty$, $(X^tX)^{-1} = \mathcal{O}(n^{-1})$' are quite common in establishing asymptotic results related to regression (e.g., see Judge *et al.*, 1985, p. 145). They essentially protect against multicollinearity. However, the following discussion might be helpful.

Let Corr(X) be the sample correlation matrix of the columns of X, i.e., a matrix containing 1's along the diagonal and sample correlations between pairs of columns of X at the off-diagonal positions. Since for every q, the mean of $x_{1q}, \ldots, x_{nq}$ can be verified to be zero, it follows that the correlation between the columns $(x_{1q}, \ldots, x_{nq})^t$ and $(x_{1s}, \ldots, x_{ns})^t$ of X is $\sum_{p=1}^n x_{pq}x_{ps}/[\sum_{p=1}^n x_{pq}^2 \sum_{p=1}^n x_{ps}^2]^{1/2}$. Therefore,

$$\begin{aligned} X^tX = {} & \operatorname{diag}[(\sum_{p=1}^n x_{p1}^2)^{1/2}, \ldots, (\sum_{p=1}^n x_{pk}^2)^{1/2}] \\ & \cdot \operatorname{Corr}(X) \operatorname{diag}[(\sum_{p=1}^n x_{p1}^2)^{1/2}, \ldots, (\sum_{p=1}^n x_{pk}^2)^{1/2}]. \end{aligned} \tag{6.45}$$

If there are of the order of n non-zero x_{pq}'s in every column of X and the eigenvalues of Corr(X) are bounded away from zero, it is easy to see that the inverse of (6.45) is of order n^{-1}. Fewer than $\mathcal{O}(n)$ non-zero valued entries in each column of X implies that a vast majority of ${c_{ij}}^{(k)}$'s are constant — a situation that most users of the gravity model would not use to estimate $\boldsymbol{\theta}$!

6.3.2 The Main Theorem

After stating the conditions we did in the last section, we now are in a position to state our main asymptotic results:

Theorem 6.1 *Under Conditions LS1, LS2, LS3 and LS4,*

1. $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$, *where* $\boldsymbol{\beta}_1 = \mathcal{O}(n^{-1}T_{\#}^{-1})$ *and* $\boldsymbol{\beta}_2 = \mathcal{O}(T_{\#}^{-2})$.

2. $\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t] = \mathcal{O}(n^{-1}T_{\#}^{-1})$.

3. *For fixed n and large enough $T_{\#}$, the vector of standard errors of $\hat{\boldsymbol{\theta}}$ is larger than* $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$.

4. *For $T_{\#} \to \infty$ and fixed n, each component of the random vector $\hat{\boldsymbol{\theta}}$ is asymptotically normal. For large enough $T_{\#}$, each component of $\hat{\boldsymbol{\theta}}$ is asymptotically normal even when only $n \to \infty$.*

The proof of this theorem will be given in a series of lemmas in the next two subsections. It might be noted that somewhat similar types of results are available in Judge, *et al.* (1985, p. 175–177, especially, equations (5.5.24) and (5.5.25)) and elsewhere. However, these results do not extend very easily to the present situation with the consequence that we make no palpable use of them in our treatment.

Notice that, according to Theorem 6.1, the bias $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$ of $\hat{\boldsymbol{\theta}}$ is the sum of two terms. One is of order $n^{-1}T_{\#}^{-1}$ which (by Part 2) is of the same order of magnitude as $\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t]$, and (by Part 3) is much less than the vector of standard errors of $\hat{\theta}_k$'s for large enough $T_{\#}$. Hence, for such T_{ij}'s, the bias from this source will be essentially imperceptible.

The other term is of order $T_{\#}^{-2}$. As we shall see from the proof of Lemma 6.6, its source, for moderate to large $T_{\#}$, is essentially the one considered in Section 6.2.3. At the end of this section, we shall assess its effect and show that if $T_{\#}$ is 15, the effect of this term on the estimate is at most .008 *per cent.* For $T_{\#} = 100$ the effect is at most .00009 per cent. Therefore, for even moderately large $T_{\#}$, there would appear to be little reason to pay much attention to it.

It follows from Theorem 6.1 that, under the conditions given, the LS estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is consistent for largeness defined by $T_{ij} \longrightarrow \infty$ for all i and j. It can also be verified, using methods essentially paralleling those used in the various lemmas that constitute the proof of Theorem 6.1, that, for $T_{ij} \longrightarrow \infty$ for all i and j, the estimate of $\boldsymbol{\theta}$ is asymptotically robust over the class of distributions with the property that $\mathrm{E}[N_{ij}] = T_{ij}$ and $\mathrm{var}\,[N_{ij}] = \Xi T_{ij}$ where Ξ is a constant. On intuitive grounds, it appears extremely unlikely that N_{ij}'s would have a distribution far outside of this class.

Some greater generality (in comparison with Theorem 6.1) is provided by the following corollary:

Corollary 6.1 *Assume that Conditions LS1A, LS2, LS3 and LS4 hold. Then*

1. $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$, *where* $\boldsymbol{\beta}_1 = \mathcal{O}(n^{-1}T_{\#}^{\alpha/2-1})$ *and* $\boldsymbol{\beta}_2 = \mathcal{O}(T_{\#}^{-2})$.

2. $\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t] = \mathcal{O}(n^{-1}T_{\#}^{\alpha-1})$.

3. *For fixed n and large enough $T_{\#}$, the vector of standard errors of $\hat{\boldsymbol{\theta}}$ is larger than* $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$.

4. *For $T_{\#} \longrightarrow \infty$ and fixed n, each component of the random vector $\hat{\boldsymbol{\theta}}$ is asymptotically normal. For large enough $T_{\#}$, each component of $\hat{\boldsymbol{\theta}}$ is asymptotically normal even when only $n \longrightarrow \infty$.*

6.3.3 A Projection Matrix

Notice that the transformation $\boldsymbol{\psi} \to \boldsymbol{e}$ [recall that $e_{ij} = \psi_{ij} - \psi_{i\bullet} - \psi_{\bullet j} + \psi_{\bullet\bullet}$] is linear. Thus we can write $\boldsymbol{e} = \mathcal{P}\boldsymbol{\psi}$ where $\mathcal{P}$ is a matrix. In fact, it is the same matrix that converts the separation configuration matrix c [with columns

$$\boldsymbol{c}^{(k)} = (c_{11}{}^{(k)}, \ldots, c_{1J}{}^{(k)}, \ldots\ldots, c_{I1}{}^{(k)}, \ldots c_{IJ}{}^{(k)})^t$$

for $k = 1, \ldots, K$, i.e., $c = ((c_{sk}))$ where $c_{sk} = c_s{}^{(k)} = c_{ij}{}^{(k)}$] into the matrix X; i.e., $X = \mathcal{P}c$. It is simple though tedious to verify that $\mathcal{P}$ can be explicitly described as follows:

$$\mathcal{P} = I_{IJ} - J^{-1} I_I \otimes \mathbf{1}_J \mathbf{1}_J^t - I^{-1} \mathbf{1}_I \mathbf{1}_I^t \otimes I_J + (IJ)^{-1} \mathbf{1}_{IJ} \mathbf{1}_{IJ}^t, \qquad (6.46)$$

where, as before, a subscripted I stands for an identity matrix, $\mathbf{1}$ is vector consisting only of 1's, the subscript gives the size and $\otimes$ is the Kronecker product (as defined in Section 5.7.2). Since $\mathcal{P}$ will play a major role in this chapter, we advise the careful reader to examine it closely and, perhaps, write out the individual components for some I and J.

An attempt at a verbal description of $\mathcal{P}$ follows. Let the four components on the right side of (6.46) be called Q_1, Q_2, Q_3 and Q_4. The second matrix, Q_2, is a block diagonal matrix, with each of the non-zero partitions of Q_2 consisting entirely of the same elements — $1/J$. The $c^{(k)}{}_{i\bullet}$'s are created by the multiplication of $\boldsymbol{c}^{(k)}$ by Q_2. Matrix Q_3 consists of identical partitions; each partition is a diagonal matrix with non-zero elements $1/I$. It creates the $c^{(k)}{}_{\bullet j}$'s. Q_4 consists of identical elements $1/[IJ]$ and creates the $c^{(k)}{}_{\bullet\bullet}$'s.

It follows from the above discussion that another way of writing $\mathcal{P}$ is as follows. Let the elements of $\mathcal{P} = ((\mathcal{P}_{rs}))$ be $\mathcal{P}_{rs}$ and let $s = (i-1)J + j$ and $r = (i'-1) + j'$. Then,

$$\mathcal{P}_{ss} = \gamma = 1 - I^{-1} - J^{-1} + (IJ)^{-1}, \qquad (6.47)$$

and if $r \neq s$,

$$\mathcal{P}_{rs} = \begin{cases} -J^{-1} + (IJ)^{-1} & \text{if } i = i' \\ -I^{-1} + (IJ)^{-1} & \text{if } j = j' \\ (IJ)^{-1} & \text{otherwise.} \end{cases} \qquad (6.48)$$

We present some properties of $\mathcal{P}$ in the following lemma:

Lemma 6.2 *Let*

$$\boldsymbol{g} = (g_{11}, \ldots, g_{1J}, \ldots\ldots, g_{I1}, \ldots, g_{IJ})^t \qquad (6.49)$$

$$\textit{and} \quad \boldsymbol{D} = (D_{11}, \ldots, D_{1J}, \ldots\ldots, D_{I1}, \ldots, D_{IJ})^t \qquad (6.50)$$

be any vectors and let $D = \operatorname{diag}(\boldsymbol{D})$. *Then*

1. *$\mathcal{P}$ is idempotent (i.e., $\mathcal{P}^2 = \mathcal{P}$) and symmetric.*

2. *The rank of $\mathcal{P}$ is $(I-1)(J-1)$ and that of $I_n - \mathcal{P}$ is $I+J-1$.*

3. $g^t[I_n - \mathcal{P}]g \geq 0$.

4. $\mathcal{P}(I_n - \mathcal{P}) = 0$.

5. $g^t\mathcal{P}D\mathcal{P}g = \sum_{i=1}^{I}\sum_{j=1}^{J}(g_{ij} - g_{i\bullet} - g_{\bullet j} + g_{\bullet\bullet})^2 D_{ij}$.

6. *If $\mathcal{P} = ((\mathcal{P}_{rs}))$, then*

$$\sum_{s=1}^{n} \mathcal{P}_{rs}^2 = \gamma = 1 - J^{-1} - I^{-1} + I^{-1}J^{-1}.$$

7. $$\begin{aligned}\mathcal{P} &= [I_{IJ} - J^{-1}I_I \otimes \mathbf{1}_J\mathbf{1}_J^t][I_{IJ} - I^{-1}\mathbf{1}_I\mathbf{1}_I^t \otimes I_J] \\ &= [I_{IJ} - I^{-1}\mathbf{1}_I\mathbf{1}_I^t \otimes I_J][I_{IJ} - J^{-1}I_I \otimes \mathbf{1}_J\mathbf{1}_J^t].\end{aligned}$$

PROOF: As before, call the four components of $\mathcal{P}$ in (6.46) Q_1, Q_2, Q_3 and Q_4, i.e.,

$$\mathcal{P} = Q_1 - Q_2 - Q_3 + Q_4. \tag{6.51}$$

It can be verified by matrix multiplication that

$$\begin{aligned} Q_1Q_2 = Q_2Q_2 = Q_2,\ \ Q_1Q_3 = Q_3Q_3 = Q_3, \\ \text{and } Q_1Q_4 = Q_2Q_3 = Q_2Q_4 = Q_3Q_4 = Q_4Q_4 = Q_4.\end{aligned} \tag{6.52}$$

Notice also that each of the products in (6.52) could have been written in the reverse order (e.g., $Q_2Q_1 = Q_2$). It is easily verified from (6.52) that

$$\begin{aligned} Q_2\mathcal{P} &= Q_2[Q_1 - Q_2 - Q_3 + Q_4] \\ &= Q_2Q_1 - Q_2Q_2 - Q_2Q_3 + Q_2Q_4 = Q_2 - Q_2 - Q_4 + Q_4 = 0.\end{aligned} \tag{6.53}$$

Similarly it may be seen that

$$Q_3\mathcal{P} = Q_4\mathcal{P} = 0. \tag{6.54}$$

Thus it follows that $\mathcal{P}^2 = \mathcal{P}Q_1 = \mathcal{P}$. Hence, $\mathcal{P}$ is idempotent. Symmetry follows from the fact that each of the matrices Q_p is symmetric. Part 4 is a trivial consequence of Part 1.

The characteristic roots of an idempotent matrix are only zeros and ones and, therefore, the number of ones is equal to the rank of the matrix (for an overview of matrix algebra see Sen and Srivastava, 1990, Appendix A). Also, the rank of an idempotent matrix is its trace (which, we remind the reader, is the sum of the diagonal elements of a square matrix). It can be verified by inspection that

$$\operatorname{tr} Q_1 = IJ,\ \ \operatorname{tr} Q_2 = IJ/J,\ \ \operatorname{tr} Q_3 = IJ/I \text{ and } \operatorname{tr} Q_4 = IJ/IJ \tag{6.55}$$

where tr stands for the 'trace of'. Therefore, $\operatorname{tr}\mathcal{P} = IJ - I - J + 1 = (I-1)(J-1)$. Since $(I_n - \mathcal{P})(I_n - \mathcal{P}) = I_n - \mathcal{P} - \mathcal{P} + \mathcal{P} = I_n - \mathcal{P}$, $I_n - \mathcal{P}$ is idempotent. Its trace is $IJ - \operatorname{tr}\mathcal{P} = I + J - 1$. This proves Part 2.

Since the characteristic roots of $I_n - \mathcal{P}$ are all non-negative, $I_n - \mathcal{P}$ is positive semi-definite. Consequently, Part 3 follows. Part 5 follows from the fact that $\mathcal{P}\boldsymbol{g}$ is a vector with typical component $g_{ij} - g_{i\bullet} - g_{\bullet j} + g_{\bullet\bullet}$, and that $\boldsymbol{g}^t\mathcal{P} = (\mathcal{P}\boldsymbol{g})^t$.

To prove Part 6, notice that if $\boldsymbol{p}_r$ is the rth column of $\mathcal{P}$, we get, because $\mathcal{P}$ is symmetric,

$$\sum_{s=1}^{n} \mathcal{P}_{rs}^2 = \boldsymbol{p}_r'\boldsymbol{p}_r$$

which is the (r,r)th element of $\mathcal{P}^2 = \mathcal{P}$, i.e., it is $\mathcal{P}_{rr} = \gamma$.

Part 7 is a straightforward consequence of the fact that $Q_2Q_3 = Q_4$. □

Lemma 6.2 shows that $\mathcal{P}$ is a projection operator (Rao, 1973, p. 46), projecting any vector into the orthogonal complement of the subspace containing the columns of Q_2 and Q_3. Since $\mathcal{P}c = X$, it follows that $X = \mathcal{P}^2c = \mathcal{P}X$. Since it may be verified that

$$\mathcal{P}\mathbf{1} = [Q_1 - Q_2 - Q_3 + Q_4]\mathbf{1} = \mathbf{1} - \mathbf{1} - \mathbf{1} + \mathbf{1} = \mathbf{o},$$

where $\mathbf{1}$ is the vector with all components 1 and $\mathbf{o}$ is a vector of zeros, the following corollary follows from (6.53) and (6.54):

Corollary 6.2 $X^t\mathbf{1} = \mathbf{o}$ *and* $X^tQ_2 = X^tQ_3 = X^tQ_4 = 0$.

Since ψ_s and ψ_p are independent for $s \neq p$, we get from Lemma 6.1,

$$\mathrm{E}[\psi_s\psi_p] = \mathrm{E}[\psi_s]\mathrm{E}[\psi_p] = \mathcal{O}(T_{\#}^{-4}).$$

From (6.26) and Lemma 6.1,

$$\breve{T}_{ij}^{-1} = \operatorname{var}[\psi_{ij}] = \mathrm{E}[\psi_{ij}^2] - (\mathrm{E}[\psi_{ij}])^2 = \mathrm{E}[\psi_{ij}^2] + \mathcal{O}(T_{\#}^{-4}).$$

Hence $\mathrm{E}[\psi_s^2] = \breve{T}_s^{-1} + \mathcal{O}(T_{\#}^{-4})$, so that $\mathrm{E}[\boldsymbol{\psi}\boldsymbol{\psi}^t] = \breve{T}^{-1} + \mathcal{O}(T_{\#}^{-4})$. Notice that $\mathcal{P}\mathcal{O}(T_{\#}^{-4})$, [i.e., the product of $\mathcal{P}$ with a matrix all of the elements of which are $\mathcal{O}(T_{\#}^{-4})$] is itself of order $\mathcal{O}(T_{\#}^{-4})$. This follows from the fact that each element of $\mathcal{P}\mathcal{O}(T_{\#}^{-4})$ consists of one term [containing a diagonal element of $\mathcal{P}$] of order $\mathcal{O}(T_{\#}^{-4})$ and the remaining $n-1$ terms are of order $\mathcal{O}(n^{-1})$ times $\mathcal{O}(T_{\#}^{-4})$. [Computations like this will be carried out throughout the remainder of this section.] Therefore, $\mathcal{P}\mathcal{O}(T_{\#}^{-4})\mathcal{P} = \mathcal{O}(T_{\#}^{-4})$. Thus, using the fact that $\boldsymbol{e} = \mathcal{P}\boldsymbol{\psi}$, it follows that, for large values of $\breve{T}_{ij}$,

$$\mathrm{E}[\boldsymbol{e}\boldsymbol{e}^t] = \mathcal{P}\breve{T}^{-1}\mathcal{P} + \mathcal{O}(T_{\#}^{-4}) \approx \mathcal{P}T^{-1}\mathcal{P}. \tag{6.56}$$

From (6.56) and (6.51), it can be shown that when $I \to \infty$, $J \to \infty$ and $T_{\#} \to \infty$, then $\mathrm{E}[\boldsymbol{e}\boldsymbol{e}^t]$ approaches a diagonal matrix. For example, the matrix $Q_2\breve{T}^{-1}Q_2$ can be seen to be a block diagonal matrix, the ith block of which is $J^{-1}[J^{-1}\sum_{j=1}^{J}\breve{T}_{ij}^{-1}]\mathbf{1}_J\mathbf{1}_J^t$. Clearly this goes to zero as $J \to \infty$.

However, even when $I \to \infty$ and $J \to \infty$, these non-diagonal elements have an effect on variances of estimates. Computations of variances require multiplying the matrix (6.56) from both sides by a vector and the large number of off-diagonal elements can, in a linear combination of n terms, yield a non-negligible quantity, even though each non-diagonal element is small. That this indeed happens may be shown as follows. Considering only the diagonal elements of (6.56), $\boldsymbol{g}^t\mathcal{P}\breve{T}^{-1}\mathcal{P}\boldsymbol{g}$ would be $\sum_{n=1}^{N} g_n^2\breve{T}_n^{-1}$; however, from Part 5 of Lemma 6.2 we get quite a different expression.

Let $\breve{\boldsymbol{y}} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{e}]$. Clearly, from (6.38) $\mathrm{E}[\breve{\boldsymbol{y}}] = X\boldsymbol{\theta}$, i.e., if we had observations on $\breve{y}$, the linear least squares [weighted by fixed positive numbers] estimate of $\boldsymbol{\theta}$ would have been unbiased. Consider a linear function $\boldsymbol{\ell}^t\boldsymbol{\theta}$. Suppose we estimate this function using two different unbiased linear estimates of $\boldsymbol{\theta}$ based on $\breve{\boldsymbol{y}}$:

- the first one being the usual least squares estimate, weighted by T_{ij}'s:

$$\boldsymbol{h}^t\breve{\boldsymbol{y}} = \boldsymbol{\ell}^t(X^tTX)^{-1}X^tT\breve{\boldsymbol{y}} \tag{6.57}$$

- and the second one being any linear unbiased estimate $\boldsymbol{g}^t\breve{\boldsymbol{y}}$.

Then,

$$\mathrm{E}[\boldsymbol{g}^t\breve{\boldsymbol{y}}] = \boldsymbol{g}^tX\boldsymbol{\theta} = \boldsymbol{\ell}^t\boldsymbol{\theta}. \tag{6.58}$$

Since this is true for all $\boldsymbol{\theta}$, we get $\boldsymbol{\ell}^t = \boldsymbol{g}^tX$. Therefore,

$$\boldsymbol{h}^t = \boldsymbol{g}^tX(X^tTX)^{-1}X^tT. \tag{6.59}$$

It is trivially verified that

$$\mathrm{var}\,[\boldsymbol{h}^t\breve{\boldsymbol{y}}] = \mathrm{var}\,[\boldsymbol{h}^t\boldsymbol{y}] \text{ and } \mathrm{var}\,[\boldsymbol{g}^t\breve{\boldsymbol{y}}] = \mathrm{var}\,[\boldsymbol{g}^t\boldsymbol{y}]. \tag{6.60}$$

Lemma 6.3 *For $\boldsymbol{g}$ and $\boldsymbol{h}$ defined as above,*

1. $\mathrm{var}\,[\boldsymbol{g}^t\boldsymbol{y}] = \boldsymbol{g}^t\mathcal{P}\breve{T}^{-1}\mathcal{P}\boldsymbol{g} = \sum_{i=1}^{I}\sum_{j=1}^{J}[g_{ij} - g_{i\bullet} - g_{\bullet j} + g_{\bullet\bullet}]^2/\breve{T}_{ij}$.

2. $\mathrm{var}\,[\boldsymbol{h}^t\boldsymbol{y}] = \boldsymbol{h}^t\mathcal{P}\breve{T}^{-1}\mathcal{P}\boldsymbol{h} = \sum_{i=1}^{I}\sum_{j=1}^{J}[h_{ij} - h_{i\bullet} - h_{\bullet j} + h_{\bullet\bullet}]^2/\breve{T}_{ij}$.

3. $\boldsymbol{h}^tT^{-1}\boldsymbol{h} \le \boldsymbol{g}^tT^{-1}\boldsymbol{g}$.

4. $\sum_{i=1}^{I}\sum_{j=1}^{J}[h_{ij} - h_{i\bullet} - h_{\bullet j} + h_{\bullet\bullet}]^2/T_{ij} \ge \boldsymbol{h}^tT^{-1}\boldsymbol{h}$.

PROOF: Since

- $\boldsymbol{y} - \mathrm{E}[\boldsymbol{y}] = X\boldsymbol{\theta} + \boldsymbol{e} - X\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{e}] = \boldsymbol{e} - \mathrm{E}[\boldsymbol{e}]$,

- the components ψ_s of ψ are independent and
- the variance of ψ_{ij} is $\breve{T}_{ij}^{-1}$,

we get from Part 5 of Lemma 6.2,

$$\begin{aligned}\text{var}\,[\boldsymbol{g}^t\boldsymbol{y}] &= \text{var}\,[\boldsymbol{g}^t\boldsymbol{e}] = \boldsymbol{g}^t\text{E}[(\boldsymbol{e}-\text{E}(\boldsymbol{e}))(\boldsymbol{e}-\text{E}(\boldsymbol{e}))^t]\boldsymbol{g} \\ &= \boldsymbol{g}^t\text{E}[\mathcal{P}(\boldsymbol{\psi}-\text{E}(\boldsymbol{\psi}))(\boldsymbol{\psi}-\text{E}(\boldsymbol{\psi}))^t\mathcal{P}^t]\boldsymbol{g} = \boldsymbol{g}^t\mathcal{P}\breve{T}^{-1}\mathcal{P}\boldsymbol{g} \\ &= (\mathcal{P}\boldsymbol{g})^t\breve{T}^{-1}\mathcal{P}\boldsymbol{g} = \sum_{i=1}^{I}\sum_{j=1}^{J}[g_{ij}-g_{i\bullet}-g_{\bullet j}+g_{\bullet\bullet}]^2/\breve{T}_{ij}.\end{aligned} \tag{6.61}$$

Part 2 is a special case of Part 1.

Let $\boldsymbol{g}_1 = T^{-1/2}\boldsymbol{g}$ and $Z = T^{1/2}X$. Then, from (6.59),

$$\begin{aligned}\boldsymbol{h}^tT^{-1}\boldsymbol{h} &= \boldsymbol{g}^tX(X^tTX)^{-1}X^tTT^{-1}TX(X^tTX)^{-1}X^t\boldsymbol{g} \\ &= \boldsymbol{g}^tX(X^tTX)^{-1}X^t\boldsymbol{g} = \boldsymbol{g}_1^tZ(Z^tZ)^{-1}Z^t\boldsymbol{g}_1.\end{aligned} \tag{6.62}$$

Also, $\boldsymbol{g}^tT^{-1}\boldsymbol{g} = \boldsymbol{g}_1^t\boldsymbol{g}_1$. Therefore,

$$\boldsymbol{g}^tT^{-1}\boldsymbol{g} - \boldsymbol{h}^tT^{-1}\boldsymbol{h} = \boldsymbol{g}_1^t[I_n - Z(Z^tZ)^{-1}Z^t]\boldsymbol{g}_1 \geq 0, \tag{6.63}$$

since, because

$$Z(Z^tZ)^{-1}Z^tZ(Z^tZ)^{-1}Z^t = Z(Z^tZ)^{-1}Z^t,$$

the matrix $Z(Z^tZ)^{-1}Z^t$ is idempotent and, therefore, as in the proof of Lemma 6.2, $I_n - Z(Z^tZ)^{-1}Z^t$ is positive semi-definite. This proves Part 3. Part 4 follows as a consequence if we let $\boldsymbol{g} = \mathcal{P}\boldsymbol{h}$. □

The main result in Lemma 6.3 is its Part 4, which we shall use shortly. However, it might be mentioned in passing that for the 'T_{ij}' data set (see page 415), which is very much like the Skokie data set, the left side of Part 4 (which, by Part 2, equals $\text{var}\,[\boldsymbol{h}^t\boldsymbol{y}]$) is about four times the right side.

6.3.4 Proof of Theorem 6.1

The proof of Theorem 6.1, being rather long, is split into a number of lemmas. Lemma 6.4, which examines some implications of Condition LS1, and Lemma 6.5 establish some preliminary results. Subsequent lemmas and corollaries, then address different parts of the statement of Theorem 6.1.

Lemma 6.4 *Under Condition LS1,*

$$\begin{aligned}\text{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}) &\Leftrightarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}) \\ &\Leftrightarrow (X^tX)^{-1} = \mathcal{O}(n^{-1})\end{aligned}$$

where (as defined in Section 6.3.1) $\boldsymbol{w} = (X^tTX)^{-1}X^tT(\boldsymbol{e}-\text{E}[\boldsymbol{e}])$.

PROOF: Notice that all three of the matrices in the statement of the lemma are covariance matrices (for some random vector). Because the matrices have this form, it is sufficient to prove the lemma for diagonal elements, since, by application of the Cauchy-Schwartz inequality, it can be seen that the other elements will follow suit. Let $\boldsymbol{u}^{(\rho)}$ be a vector, the ρth component of which is 1 and all other components zeros, and let $\boldsymbol{h}^t = \boldsymbol{u}^{(\rho)^t}(X^tTX)^{-1}X^tT$, i.e., $\boldsymbol{\ell} = \boldsymbol{u}^{(\rho)}$ in (6.57). Recall that $\boldsymbol{e} = \mathcal{P}\boldsymbol{\psi}$ and the fact that the components ψ_s of $\boldsymbol{\psi}$ are independent with $\operatorname{var}[\psi_s] = \breve{T}_s$. Therefore, the ρth diagonal element of $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]$ is easily seen to be

$$(\boldsymbol{u}^{(\rho)})^t\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]\boldsymbol{u}^{(\rho)} = \boldsymbol{h}^t \operatorname{Cov}[\boldsymbol{e}]\,\boldsymbol{h} = \boldsymbol{h}^t\mathcal{P}\breve{T}^{-1}\mathcal{P}\boldsymbol{h}, \tag{6.64}$$

which, by Parts 2 and 4 of Lemma 6.3, is not less than

$$\begin{aligned}\beta\boldsymbol{h}^tT^{-1}\boldsymbol{h} &= \beta\boldsymbol{u}^{(\rho)^t}(X^tTX)^{-1}X^tTT^{-1}TX(X^tTX)^{-1}\boldsymbol{u}^{(\rho)}\\ &= \beta\boldsymbol{u}^{(\rho)^t}(X^tTX)^{-1}\boldsymbol{u}^{(\rho)},\end{aligned} \tag{6.65}$$

where $\beta = \max_s[T_s/\breve{T}_s]$. Since the last line in (6.65) equals β times the ρth diagonal element of $(X^tTX)^{-1}$, it follows that

$$\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}) \Rightarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}). \tag{6.66}$$

On the other hand, if $\min_s \breve{T}_s = \breve{T}_{\#}$,

$$\begin{aligned}\operatorname{var}[\boldsymbol{h}^t\boldsymbol{y}] &= \sum_{i=1}^{I}\sum_{j=1}^{J}[h_{ij} - h_{i\bullet} - h_{\bullet j} + h_{\bullet\bullet}]^2/\breve{T}_{ij}\\ &\leq \breve{T}_{\#}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}[h_{ij} - h_{i\bullet} - h_{\bullet j} + h_{\bullet\bullet}]^2\\ &= \breve{T}_{\#}^{-1}\boldsymbol{h}^t\mathcal{P}\mathcal{P}\boldsymbol{h} = \breve{T}_{\#}^{-1}\boldsymbol{h}^t\mathcal{P}\boldsymbol{h} \leq \breve{T}_{\#}^{-1}\boldsymbol{h}^t\boldsymbol{h},\end{aligned} \tag{6.67}$$

the first line of (6.67) following from Part 1 of Lemma 6.3 and the last inequality from Part 3 of Lemma 6.2. But

$$\boldsymbol{h}^t\boldsymbol{h} = \sum_{i=1}^{I}\sum_{j=1}^{J}h_{ij}^2 \leq T^{\#}\sum_{i=1}^{I}\sum_{j=1}^{J}h_{ij}^2T_{ij}^{-1} = T^{\#}\boldsymbol{h}^tT^{-1}\boldsymbol{h}, \tag{6.68}$$

which, as in (6.65), is $T^{\#}$ times the ρth diagonal element of $(X^tTX)^{-1}$.

From (6.67) and (6.68) it follows that

$$\operatorname{var}[\boldsymbol{h}^t\boldsymbol{y}] \leq [T^{\#}/\breve{T}_{\#}]\boldsymbol{h}^tT^{-1}\boldsymbol{h}. \tag{6.69}$$

From Part 2 of Lemma 6.3 and (6.64), we see that the left side of (6.69) is the ρth diagonal element of $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]$. By (6.65), (6.26) and Condition LS1 we see that the right side of (6.69) is less than a constant times a diagonal element of $(X^tTX)^{-1}$. Hence,

$$\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}) \Leftarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}). \tag{6.70}$$

If the columns of X are linearly independent, it is easily seen that the columns of $T^{1/2}X$ will also be linearly independent; and if the columns of $T^{1/2}X$ are linearly independent the same would be true of X. Therefore, the existence of the inverse of X^tX is guaranteed by the existence of $(X^tTX)^{-1}$, and *vice versa*. Since $(X^tX)^{-1}$ and $(X^tTX)^{-1}$ are covariance matrices, their diagonal elements and their eigenvalues are non-negative. Since their trace is equal to the sum of their eigenvalues, it may be seen that, for either matrix, each eigenvalue is numerically smaller than the sum of diagonal elements and each diagonal element is numerically smaller than the sum of eigenvalues. Notice also that both matrices have dimension K which is a fixed number, that is usually small and not affected by n. Thus, for diagonal elements to be of any given order it is necessary and sufficient that the eigenvalues be of the same order.

Let $\lambda_\tau(G)$ be the τth (in descending order) eigenvalue of G. Then, as a special case of a result in Anderson and Dasgupta (1963, reproduced in Rao, 1973, p.68), we have for any two symmetric matrices of dimension K

$$\lambda_\tau(G+H) \leq \lambda_1(G) + \lambda_\tau(H) \tag{6.71}$$

for $\tau = 1, \ldots, K$. Let $T_s^\circ = T_s/T_{\#} \geq 1$ and let T° be the diagonal matrix with T_s°'s as diagonal elements. Clearly, $X^tT^\circ X = [(T^\circ)^{1/2}X]^t[(T^\circ)^{1/2}X]$ is positive definite and $-X^t(I_K - T^\circ)X$ is non-negative definite. That is, the eigenvalues of $X^tT^\circ X$ are all positive and those of $-X^t(I_K - T^\circ)X$ are non-negative. It then follows from (6.71) that

$$\lambda_\tau(X^tX) \leq \lambda_\tau(X^tT^\circ X) + \lambda_1(X^t(I_K - T^\circ)X) \leq \lambda_\tau(X^tT^\circ X). \tag{6.72}$$

Hence, each eigenvalue of $X^tT^\circ X$ is greater than or equal to the corresponding one of X^tX. The eigenvalues of $(X^tT^\circ X)^{-1}$, are reciprocals of those of $X^tT^\circ X$. Hence the sum of eigenvalues of $(X^tT^\circ X)^{-1}$ is less than or equal to the sum of eigenvalues of $(X^tX)^{-1}$. Hence,

$$\begin{aligned}&(X^tX)^{-1} = \mathcal{O}(n^{-1}) \Rightarrow \\ &\qquad (X^tTX)^{-1} = T_{\#}^{-1}(X^tT^\circ X)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}).\end{aligned} \tag{6.73}$$

Similarly, by comparing $T^{\#}X^t \operatorname{diag}(T_1/T^{\#}, \ldots, T_n/T^{\#})X$ with X^tX we can establish

$$(X^tX)^{-1} = \mathcal{O}(n^{-1}) \Leftarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}). \tag{6.74}$$

The results (6.66), (6.70), (6.73) and (6.74) together prove the lemma. □

Corollary 6.3

1. $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}) \Rightarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1})$,

2. $(X^tX)^{-1} = \mathcal{O}(n^{-1}) \Rightarrow (X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1})$.

3. $X^tTX = \mathcal{O}(n) \Rightarrow X^tX = \mathcal{O}(n)$.

PROOF: The proofs of Parts 1 and 2 follow from the proofs of the corresponding parts of Lemma 6.4 on noticing that Condition LS1 on the upper bound of $T^{\#}/T_{\#}$'s was never used in establishing the results. Part 3 follows from the first sentence following (6.72). □

Corollary 6.4 *Under Condition LS1A,*

$$(X^tTX)^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1}) \Rightarrow \mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{\alpha-1}).$$

PROOF: Consider (6.69). Apply Part 2 of Lemma 6.3 and (6.64) to its left side. Substitute for $\boldsymbol{h}^tT^{-1}\boldsymbol{h}$ on the right side from (6.65) and for $[T^{\#}/\breve{T}_{\#}]$ using Condition LS1A. Note that LS1A makes sense as a generalization of LS1 only when $T_{\#} \to \infty$ and then $\breve{T}_{\#} \to T_{\#}$. □

Lemma 6.5, below, uses the results that $\mathrm{E}[\epsilon\boldsymbol{e}] \approx \mathbf{1}$ and $X^t\mathbf{1} = \mathbf{o}$ to show that $\boldsymbol{w}_2 = U^{-1}X^t\epsilon\boldsymbol{e}$ — see (6.41) — is small.

Lemma 6.5 *Assume that LS1, LS2, LS3 and LS4 hold and let* $\eta_s = \mathrm{E}[\epsilon_s e_s]$ *and* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^t$.

1. $\mathrm{E}[X^t\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon X] - X^t\boldsymbol{\eta}\boldsymbol{\eta}^tX$ *is a matrix of the form* $X^tSX + \mathcal{O}(T_{\#}^{-2}n)$, *where* S *is a diagonal matrix.*

2. $S = [2 + \mathcal{O}(T_{\#}^{-1}) + \mathcal{O}(n^{-1/2})]I_n$ *and* $X^t\boldsymbol{\eta} = \mathcal{O}(nT_{\#}^{-2})$.

3. $\mathrm{E}[\boldsymbol{w}_2\boldsymbol{w}_2^t] = \mathcal{O}(n^{-1}T_{\#}^{-2}) + \mathcal{O}(T_{\#}^{-6})$.

PROOF: We may write the matrix $\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon$ as

$$(\epsilon_1 e_1, \ldots, \epsilon_n e_n)(\epsilon_1 e_1, \ldots, \epsilon_n e_n)^t.$$

Hence, any off-diagonal term in its expectation is of the form $\mathrm{E}[\epsilon_e e_s \epsilon_p e_p]$ with $p \neq s$. Since $\boldsymbol{e} = \mathcal{P}\boldsymbol{\psi}$,

$$e_p = \sum_{q=1}^{n} \mathcal{P}_{pq}\psi_q, \tag{6.75}$$

where $\mathcal{P} = ((\mathcal{P}_{pq}))$. On substituting this expression for e_p and a similar one for e_s, with $s \neq p$, into $\epsilon_s e_s \epsilon_p e_p$, it may be seen that of the resultant n^2 terms, the only ones that can have non-zero expectations are the two terms containing $\epsilon_s\psi_s\epsilon_p\psi_p$ (since the ϵ_s's and the ψ_s's are independent unless they have common subscripts). One of these terms has coefficient

$$\mathcal{P}_{pp}\mathcal{P}_{ss} = [1 - I^{-1} - J^{-1} + I^{-1}J^{-1}]^2 = \gamma^2, \tag{6.76}$$

by (6.47). Since, as can be easily verified, $\mathrm{E}[\epsilon_s e_s] = \mathcal{P}_{ss}\mathrm{E}[\epsilon_s\psi_s]$, this term can be written as

$$\mathrm{E}[\epsilon_s e_s]\mathrm{E}[\epsilon_p e_p] = \eta_s\eta_p. \tag{6.77}$$

The other term has coefficient $\mathcal{P}_{sp}\mathcal{P}_{ps} = \mathcal{P}_{sp}^2 = \phi_{sp}$(say), which for $s \neq p$ equals [see (6.48)]

$$\phi_{sp} = \begin{cases} [J^{-1}(1 - I^{-1})]^2 & \text{if } \psi_s,\ \psi_p \text{ in same row of } ((\psi_{ij})) \\ [I^{-1}(1 - J^{-1})]^2 & \text{if } \psi_s,\ \psi_p \text{ in same column of } ((\psi_{ij})) \\ [IJ]^{-2} & \text{otherwise.} \end{cases} \tag{6.78}$$

Define the matrix $\breve{\Phi} = ((\breve{\phi}_{sp}))$, with $\breve{\phi}_{sp} = \mathrm{E}[\phi_{sp}\epsilon_s\psi_s\epsilon_p\psi_p]$ for $p \neq s$ and $\breve{\phi}_{ss} = \phi_{ss}$ for all s, where $\phi_{ss} = [J^{-1}(1 - I^{-1})]^2 + [I^{-1}(1 - J^{-1})]^2 + [IJ]^{-2}$. This choice of the diagonal elements ϕ_{ss} is merely for convenience at this stage; we will subtract it out later.

Notice that it follows from the above arguments that the off-diagonal elements of $\mathrm{E}[\epsilon \boldsymbol{e}\boldsymbol{e}^t\epsilon]$ are exactly the off-diagonal elements of the sum of $\breve{\Phi}$ and $\boldsymbol{\eta}\boldsymbol{\eta}^t$ [(see (6.77)].

Set $\Phi = ((\phi_{sp}))$. Its diagonal elements are the ϕ_{ss}'s defined above and its off-diagonal elements are as in (6.78). It may be verified that

$$\Phi = J^{-1}(1 - 2I^{-1})Q_2 + I^{-1}(1 - 2J^{-1})Q_3 + I^{-1}J^{-1}Q_4. \tag{6.79}$$

The Q_r's are as defined just before (6.48) and the material there may be helpful in reading the current material. In (6.79), the first term on the left comes from expanding $J^{-2}[1 - I^{-1}]^2$ from the first line of (6.78), contributing the term $I^{-2}J^{-2}$ to some of the elements not covered by the last line of (6.78) and noting that Q_2 already includes a J^{-1}. It then follows from Corollary 6.2 that

$$X^t\Phi = X^t[J^{-1}(1 - 2I^{-1})Q_2 + I^{-1}(1 - 2J^{-1})Q_3 + I^{-1}J^{-1}Q_4] = 0. \tag{6.80}$$

By using a power series for $\log[(N_s + \frac{1}{2})/T_s] = \log[(\epsilon_s + \frac{1}{2})/T_s + 1]$, it may be verified, in a manner similar to the proof of Lemma 6.1, that

$$\begin{aligned} \mathrm{E}(\epsilon_s\psi_s) &= \mathrm{E}[\epsilon_s\{\log(N_s + \frac{1}{2}) - \log(T_s)\}] \\ &= 1 + \mathcal{O}(T_s^{-2}). \end{aligned} \tag{6.81}$$

Thus,

$$\begin{aligned} \mathrm{E}[\phi_{sp}\epsilon_s\psi_s\epsilon_p\psi_p] &= \phi_{sp}\mathrm{E}[\epsilon_s\psi_s]\mathrm{E}[\epsilon_p\psi_p] = \phi_{sp}[1 + \mathcal{O}(T_{\#}^{-2})] \\ &= \phi_{sp} + \mathcal{O}(n^{-1}T_{\#}^{-2}) \end{aligned} \tag{6.82}$$

since, using Condition LS2 and (6.78), it may be seen that $\phi_{sp} = \mathcal{O}(n^{-1})$. It follows that

$$\breve{\Phi} = \Phi + \mathcal{O}(n^{-1}T_{\#}^{-2}). \tag{6.83}$$

Then, using (6.80), we get

$$X^t\breve{\Phi}X = X^t\Phi X + X^t((\mathcal{O}(n^{-1}T_{\#}^{-2})))X = \mathcal{O}(T_{\#}^{-2}n). \tag{6.84}$$

In the above arguments, we have been mainly concerned with off-diagonal terms of the matrix $\mathrm{E}[\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon]$, each of which is the sum of the corresponding off-diagonal elements of $\boldsymbol{\eta}\boldsymbol{\eta}^t$ (from (6.77)) and $\breve{\Phi}$. But we have introduced incorrect, though convenient, diagonal elements as place holders. We now need to subtract these out and replace them with the correct diagonal elements which are $\mathrm{E}[\epsilon_s^2 e_s^2]$. Let G be the diagonal matrix of the diagonal elements of $\boldsymbol{\eta}\boldsymbol{\eta}^t$ and $D = \operatorname{diag}(\phi_{11}, \ldots, \phi_{nn})$ and let

$$E = \operatorname{diag}(\mathrm{E}[\epsilon_1^2 e_1^2], \ldots, \mathrm{E}[\epsilon_n^2 e_n^2])$$

be the matrix of the diagonal terms of $\mathrm{E}[\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon]$. Then, since, as mentioned earlier, the off-diagonal elements of $\mathrm{E}[\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon]$ are exactly the off-diagonal elements of the sum of $\breve{\Phi}$ and $\boldsymbol{\eta}\boldsymbol{\eta}^t$

$$\mathrm{E}[\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon] = \boldsymbol{\eta}\boldsymbol{\eta}^t + \breve{\Phi} + E - G - D.$$

Hence, using (6.83) and (6.84), we get

$$\begin{aligned}\mathrm{E}[X^t\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon X] &= X^t[E + \boldsymbol{\eta}\boldsymbol{\eta}^t - G + \Phi - D]X + \mathcal{O}(T_{\#}^{-2}n)\\ &= X^t\boldsymbol{\eta}\boldsymbol{\eta}^t X + X^t[E - G - D]X + \mathcal{O}(T_{\#}^{-2}n).\end{aligned}$$

This proves Part 1 of the lemma on setting $S = E - G - D$.

From Part 6 of Lemma 6.2, we have

$$\sum_{\substack{p=1\\ p\neq s}}^{n} \mathcal{P}_{sp}^2 = \mathcal{O}(n^{-1/2}).$$

Consequently, we have for the elements $\mathrm{E}[\epsilon_s^2 e_s^2]$ in E

$$\begin{aligned}\mathrm{E}[\epsilon_s^2(\sum_{p=1}^{n}\mathcal{P}_{sp}\psi_p)^2] &= \mathcal{P}_{ss}^2\mathrm{E}[\psi_s^2\epsilon_s^2] + \mathrm{E}[\epsilon_s^2]\sum_{\substack{p=1\\ p\neq s}}^{n}\mathcal{P}_{sp}^2\mathrm{E}[\psi_p^2]\\ &\leq \gamma^2\mathrm{E}[\psi_s^2\epsilon_s^2] + T^{\#}/\breve{T}_{\#}\sum_{\substack{p=1\\ p\neq s}}^{n}\mathcal{P}_{sp}^2\\ &= [1 + \mathcal{O}(n^{-1/2})]\mathrm{E}[\psi_s^2\epsilon_s^2] + \mathcal{O}(n^{-1/2}).\end{aligned} \tag{6.85}$$

As in the case of (6.81), it may be shown using (6.19) that

$$\mathrm{E}[\psi_s^2\epsilon_s^2] = 3 + \mathcal{O}(T_s^{-1}). \tag{6.86}$$

Therefore, from (6.85), we get

$$E = (3 + \mathcal{O}(T_{\#}^{-1}) + \mathcal{O}(n^{-1/2}))I_n. \tag{6.87}$$

By (6.81),

$$\begin{aligned}\eta_s &= \mathrm{E}[\epsilon_s \sum_{p=1}^{n} \mathcal{P}_{sp}\psi_p] = \mathcal{P}_{ss}\mathrm{E}[\epsilon_s\psi_s] \\ &= \gamma[1 + \mathcal{O}(T_s^{-2})] = \gamma + \mathcal{O}(T_s^{-2}) \\ &= 1 + \mathcal{O}(n^{-1/2}) + \mathcal{O}(T_s^{-2}).\end{aligned} \tag{6.88}$$

Thus

$$G = (1 + \mathcal{O}(T_{\#}^{-2}) + \mathcal{O}(n^{-1/2}))I_n. \tag{6.89}$$

Clearly, $D = \mathcal{O}(n^{-1})$. Thus, from (6.87) and (6.89),

$$S = E - G - D = (2 + \mathcal{O}(T_{\#}^{-1}) + \mathcal{O}(n^{-1/2}))I_n. \tag{6.90}$$

This proves the first portion of Part 2 of the lemma. To prove the rest of Part 2, notice that by Corollary 6.2, $X^t\gamma\mathbf{1} = 0$ and, therefore, from the next to last line of (6.88),

$$X^t\boldsymbol{\eta} = X^t\mathcal{O}(T_{\#}^{-2}) = \mathcal{O}(T_{\#}^{-2}n), \tag{6.91}$$

since each row of X^t has n terms. Thus, Part 2 is proved.

From (6.90), since S is a diagonal matrix, it follows from LS3 that $X^tSX = \mathcal{O}(n)$ [any element in the product X^tSX contains n terms, each a product two elements of X and one element from the diagonal of S]. Therefore, since from Lemma 6.4, $U^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1})$, we get

$$\begin{aligned}\mathrm{E}[\boldsymbol{w}_2\boldsymbol{w}_2^t] &= U^{-1}[X^t\mathrm{E}[\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon]X]U^{-1} \\ &= U^{-1}X^tSXU^{-1} + U^{-1}[X^t\boldsymbol{\eta}\boldsymbol{\eta}^tX]U^{-1} + U^{-2}\mathcal{O}(T_{\#}^{-2}n) \\ &= \mathcal{O}(n^{-1}T_{\#}^{-2}) + \mathcal{O}(T_{\#}^{-6}).\end{aligned} \tag{6.92}$$

[Notice that U is a $K \times K$ matrix.] This proves Part 3. □

In order not to break the continuity of the proof of the next few lemmas we present some matrix results which will be used in them. Let D and D' be diagonal matrices with $D \preceq D'$ and let d_s and d'_s be their typical diagonal elements. Consider any matrix $H = ((h_{rs}))$ of suitable dimension. Then the rth diagonal element of HDH^t is $\sum_s h_{rs}^2 d_s \leq \sum_s h_{rs}^2 d'_s$ which is the rth diagonal element of $HD'H^t$. Thus

$$HDH^t \preceq HD'H^t. \tag{6.93}$$

Since for any random variables g_k and h_k, the Cauchy-Schwartz inequality yields

$$|\mathrm{E}[g_k h_k]| \leq (\mathrm{E}[g_k^2]\mathrm{E}[h_k^2])^{1/2}$$

and

$$|\sum_{k=1}^{K} g_k h_k| \leq [\sum_{k=1}^{K} g_k^2]^{1/2} [\sum_{k=1}^{K} h_k^2]^{1/2},$$

it follows that

$$\begin{aligned} |\sum_{k=1}^{K} \mathrm{E}[g_k h_k]| &\leq |\sum_{k=1}^{K} \mathrm{E}[g_k^2]^{1/2} \mathrm{E}[h_k^2]^{1/2}| \\ &\leq [\sum_{k=1}^{K} \mathrm{E}(g_k^2)]^{1/2} [\sum_{k=1}^{K} \mathrm{E}(h_k^2)]^{1/2}. \end{aligned} \tag{6.94}$$

It follows from (6.94) that for any random square matrix $G = ((g_{pq}))$ and any random vector $\boldsymbol{g} = (g_1, \ldots, g_K)^t$ of common dimension K, the numerical value of a component of the expectation of their product $\boldsymbol{h} = (h_1, \ldots, h_K)^t$ can be written as

$$|h_p| = |\mathrm{E}[\sum_{q=1}^{K} g_{pq} g_q]| \leq \left(\mathrm{E}[\sum_{q=1}^{K} g_{pq}^2]\right)^{1/2} \left(\mathrm{E}[\sum_{q=1}^{K} g_q^2]\right)^{1/2}. \tag{6.95}$$

Notice that $\sum_{q=1}^{K} g_{pq}^2$ is a diagonal element of GG^t and $\sum_{q=1}^{K} g_q^2$ is $\boldsymbol{g}^t \boldsymbol{g}$.

It also follows from (6.94) that if the diagonal elements of the matrix $\mathrm{E}[GG^t]$, of fixed dimension K, are all of order o_G and if the diagonal elements of the matrix $\mathrm{E}[HH^t]$, also of fixed dimension K are all of order o_H, then

$$\mathrm{E}[GH] = \mathcal{O}([o_G o_H]^{1/2}). \tag{6.96}$$

If any matrix $G = ((g_{pq}))$ is of the form of a covariance matrix, then $q_{pq} \leq g_{pp}^{1/2} g_{qq}^{1/2}$ by the Cauchy Schwartz inequality, since the off-diagonal entries of such a matrix are the covariances and the diagonal entries are variances. Since for any numbers r and s, $r^2 + s^2 - 2rs = (r - s)^2 \geq 0$ and therefore, $r^2 + s^2 \geq 2rs$, it follows that $q_{pq} \leq g_{pp}^{1/2} g_{qq}^{1/2} \leq g_{pp} + g_{qq}$ and hence, because the diagonal elements of G are all positive, the trace of G is larger than any element. Thus, if the trace is of a certain order of magnitude, all elements will be of the same order of magnitude.

In a lemma to follow, we shall encounter a situation where we need to determine the order of magnitude of a matrix like $\mathcal{G} = \mathrm{E}([(G + H)(G + H)^t]^2)$ given the order (say) o_G of $\mathrm{E}([GG^t]^2)$ and that of $\mathrm{E}([HH^t]^2)$ which is o_H. If $o_G \geq o_H$, then

$$\mathrm{E}([(G + H)(G + H)^t]^2 = \mathcal{O}(o_G). \tag{6.97}$$

as we now show. If we carry out the multiplication and the squaring implied, we see that $\mathcal{G}$ is the sum of sixteen expectations. One of these is $\mathrm{E}([GG^t]^2)$ and another is $\mathrm{E}([HH^t]^2)$. Yet another is

$$\mathrm{E}(GG^t HH^t) \leq [\mathrm{E}([GG^t]^2)\mathrm{E}([HH^t]^2)]^{1/2}.$$

But others are slightly more complicated to work with. Consider the term $\mathrm{E}(GH^tHG^t)$. Its order of magnitude is the same as that of its trace

$$\operatorname{tr}\mathrm{E}(GH^tHG^t) = \mathrm{E}[\operatorname{tr}(GH^tHG^t)] = \mathrm{E}[\operatorname{tr}(G^tGH^tH)].$$

Since G^tG is symmetric, it has the same diagonal elements as GG^t. Therefore $\mathrm{E}(GH^tHG^t) = \mathcal{O}(o_G^{1/2}o_H^{1/2})$. Going through all the terms in like manner, and using the above discussion involving the Cauchy-Schwartz inequality, it can be seen that (6.97) holds.

Lemma 6.6 *Under LS1, LS3 and LS4,* $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$, *where* $\boldsymbol{\beta}_1 = \mathcal{O}(n^{-1}T_{\#}^{-1})$ *and* $\boldsymbol{\beta}_2 = \mathcal{O}(T_{\#}^{-2})$.

PROOF: From (6.42) we get

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (I + W)(\boldsymbol{w}_1 + \boldsymbol{w}_2) = \boldsymbol{w}_1 + \boldsymbol{w}_2 + W\boldsymbol{w}_1 + W\boldsymbol{w}_2. \qquad (6.98)$$

We shall consider below each of the terms on the right side, starting with $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$.

From Lemma 6.4 we see that, because of Condition LS4, $(X^tTX)^{-1} = U^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-1})$. From Lemma 6.1 invoking Condition LS1, we get $\mathrm{E}[\psi_{ij}] = \mathcal{O}(T_{\#}^{-2})$ for all $i \in I$ and $j \in J$. Since $\mathrm{E}[\psi_{i\bullet}]$, $\mathrm{E}[\psi_{\bullet j}]$ and $\mathrm{E}[\psi_{\bullet\bullet}]$ are of the same order, it follows that $\mathrm{E}[\boldsymbol{e}] = \mathcal{O}(T_{\#}^{-2})$. Therefore,

$$\mathrm{E}[\boldsymbol{w}_1] = U^{-1}X^tT\mathrm{E}[\boldsymbol{e}] = U^{-1}X^tT\mathcal{O}(T_{\#}^{-2}) = \mathcal{O}(T_{\#}^{-2}), \qquad (6.99)$$

since, because each element of $X^tT\mathcal{O}(T_{\#}^{-2})$ contains n terms each including a T_s, a x_{ks} and a term of order $T_{\#}^{-2}$, we get $X^tT\mathcal{O}(T_{\#}^{-2}) = \mathcal{O}(nT_{\#}^{-1})$ from LS1 and LS3.

From Part 2 of Lemma 6.5, we see that

$$\mathrm{E}[\boldsymbol{w}_2] = U^{-1}X^t\mathrm{E}[(\epsilon_s e_s, \ldots, \epsilon_n e_n)^t] = U^{-1}X^t\boldsymbol{\eta} = \mathcal{O}(T_{\#}^{-3}). \qquad (6.100)$$

In order to discuss the last two terms in (6.98), we first examine $\mathrm{E}[\boldsymbol{w}_1^t\boldsymbol{w}_1]$ and $\mathrm{E}[\boldsymbol{w}_2^t\boldsymbol{w}_2]$. Since

$$\boldsymbol{w} = U^{-1}X^tT(\boldsymbol{e} - \mathrm{E}[\boldsymbol{e}]) = \boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1], \qquad (6.101)$$

it follows that

$$\mathrm{E}[\boldsymbol{w}_1\boldsymbol{w}_1^t] = \mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] + \mathrm{E}[\boldsymbol{w}_1]\mathrm{E}[\boldsymbol{w}_1^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}) + \mathcal{O}(T_{\#}^{-4}) \qquad (6.102)$$

by Lemma 6.4 and (6.99). Notice that $\boldsymbol{w}_1^t\boldsymbol{w}_1$ is a scalar. Therefore, $\boldsymbol{w}_1^t\boldsymbol{w}_1 = \operatorname{tr}[\boldsymbol{w}_1^t\boldsymbol{w}_1]$ which is the same as $\operatorname{tr}[\boldsymbol{w}_1\boldsymbol{w}_1^t]$ (since $\operatorname{tr}[GH] = \operatorname{tr}[HG]$). Consequently,

$$\mathrm{E}[\boldsymbol{w}_1^t\boldsymbol{w}_1] = \mathcal{O}(n^{-1}T_{\#}^{-1}) + \mathcal{O}(T_{\#}^{-4}). \qquad (6.103)$$

Similarly, from Lemma 6.5,

$$\mathrm{E}[\boldsymbol{w}_2^t\boldsymbol{w}_2] = \mathcal{O}(n^{-1}T_{\#}^{-2}) + \mathcal{O}(T_{\#}^{-6}). \tag{6.104}$$

We can write $I_K + W = (I_K + U^{-1}V)^{-1}$ as a power series:

$$I_K - U^{-1}V + [U^{-1}V]^2 - [U^{-1}V]^3 + \cdots \tag{6.105}$$

in powers of $U^{-1}V$ by repeated use of the identity:

$$(I_K + GH)^{-1} = I_K - G(I_K + HG)^{-1}H$$

where G and H are arbitrary matrices of dimension K for which the inverse exists. Since $\mathrm{E}[\epsilon XX^t\epsilon]$ is an n-dimensional diagonal matrix with diagonal elements $\sum_{s=1}^K x_{ps}^2\mathrm{E}(\epsilon_p^2) \leq K\mu^2T_p$ (because $\mathrm{E}(\epsilon_p^2) = T_p$ and $|x_{ps}| \leq \mu$ by Condition LS3), it follows from (6.93) that

$$\begin{aligned} \mathrm{E}[U^{-1}V(U^{-1}V)^t] &= U^{-1}X^t\mathrm{E}[\epsilon XX^t\epsilon]XU^{-1} \\ &\preceq K\mu^2U^{-1}X^tTXU^{-1} = K\mu^2\,U^{-1}, \end{aligned} \tag{6.106}$$

which is of order $\mathcal{O}(n^{-1}T_{\#}^{-1})$.

We also need to gauge the order of matrices $\mathrm{E}([U^{-1}V(U^{-1}V)^t]^m)$ for $m > 1$ to show that are asymptotically much smaller than (6.106). Unfortunately, this requires methods that are less tidy than the one we used for when $m = 1$. Consider $m = 2$, i.e., consider

$$U^{-1}X^t\epsilon XX^t\epsilon XU^{-1}U^{-1}X^t\epsilon XX^t\epsilon XU^{-1}.$$

Notice first that the elements of $X^t\epsilon X$ are of the form

$$\sum_{s=1}^n x_s{}^{(k)}x_s{}^{(\ell)}\epsilon_s, \tag{6.107}$$

where $x_s{}^{(k)}$ is the (s,k)th element of X. Since $x_p{}^{(k)}$ is the (p,k)th element of X, it should properly have been called x_{pk}, as has been done before, but the notation just introduced is more convenient here. If this $X^t\epsilon X$ is multiplied by another $X^t\epsilon X$ or its transpose, each element of the resultant product would consist of a linear combination of k terms each of which is a product of two linear combinations like (6.107). Following this type of argument further, and then multiplying out the terms and regrouping them, we see that $[U^{-1}V(U^{-1}V)^t]^2$ consists of a linear combination of terms each of which is the product of four ϵ_s's. When the expectation of this entire linear combination is taken, we see that the only non-zero terms are the n^2 terms which include $\epsilon_s^2\epsilon_p^2$. When $s \neq p$ the expectation of $\epsilon_s^2\epsilon_p^2$ is T_sT_p, and when $s = p$ it is of the order of T_s^2. Either way, $\epsilon_s^2\epsilon_p^2 = \mathcal{O}(T_{\#}^2)$ by LS1. Now consider the coefficients of these terms. Each of them contain four $x_s{}^{(k)}$'s from the four X's. The product of four $x_s{}^{(k)}$'s is less than μ^4 by LS3. The four U^{-1}'s contribute $\mathcal{O}(n^{-4}T_{\#}^{-4})$ to the coefficient. While every time K dimensional square matrices are multiplied, sums of k terms are created, but since K is fixed, the coefficient is of the order of $n^{-4}T_{\#}^{-4}$. As already mentioned, there are n^2 such non-zero terms [which are multiplied by these coefficients], each of order $T_{\#}^2$. Hence it follows that $\mathrm{E}([U^{-1}V(U^{-1}V)^t]^2) = \mathcal{O}(n^{-2}T_{\#}^{-2})$. The argument above can be extended to show that

$$\mathrm{E}([U^{-1}V(U^{-1}V)^t]^m) = \mathcal{O}((nT_{\#})^{-m}). \tag{6.108}$$

Applying (6.96) to $[U^{-1}V]^r$ and $\boldsymbol{w}_1$, and using (6.108), (6.103) and the fact that for any $g_1 > 0$ and $g_2 > 0$, $g_1^{1/2} + g_2^{1/2} > [g_1 + g_2]^{1/2}$, we see that, for $r \geq 1$, $\mathrm{E}[(U^{-1}V)^r \boldsymbol{w}_1] = \mathcal{O}((T_{\#}n)^{-\frac{1}{2}(r+1)}) + \mathcal{O}(n^{-\frac{1}{2}r}T_{\#}^{-2-\frac{1}{2}r})$. Therefore,

$$\mathrm{E}[W\boldsymbol{w}_1] = \mathcal{O}(n^{-1}T_{\#}^{-1}) + \mathcal{O}(n^{-1/2}T_{\#}^{-5/2}). \tag{6.109}$$

Similarly, from (6.104),

$$\mathrm{E}[W\boldsymbol{w}_2] = \mathcal{O}(n^{-1}T_{\#}^{-3/2}) + \mathcal{O}(n^{-1/2}T_{\#}^{-7/2}). \tag{6.110}$$

The lemma follows from (6.98), (6.99), (6.100), (6.109) and (6.110) on setting $\boldsymbol{\beta}_1$ as the sum of the first terms on the right side of (6.109) and (6.110); and $\boldsymbol{\beta}_2$ as the sum of (6.99), (6.100) and the second terms on the right in (6.109) and in (6.110). □

Corollary 6.5 *Under LS1A, LS3 and LS4,* $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$, *where* $\boldsymbol{\beta}_1 = \mathcal{O}(n^{-1}T_{\#}^{\alpha/2-1})$ *and* $\boldsymbol{\beta}_2 = \mathcal{O}(T_{\#}^{-2})$.

PROOF: The proof is essentially the same as for Lemma 6.6. The key difference is the insertion of the result in Corollary 6.4 (note Part 2 of Corollary 6.3) into (6.102) and the results so obtained into (6.109). □

Notice that from (6.105), (6.106) and (6.108) we get, under the same conditions as in the above lemma,

$$\mathrm{E}[WW^t] = \mathcal{O}(U^{-1}) = \mathcal{O}(n^{-1}T_{\#}^{-1}). \tag{6.111}$$

Lemma 6.7 *Under LS1, LS2, LS3 and LS4,* $\mathrm{Cov}[\hat{\boldsymbol{\theta}}] = \mathcal{O}(n^{-1}T_{\#}^{-1})$.

PROOF: Notice that

$$\begin{aligned}\mathrm{Cov}[\hat{\boldsymbol{\theta}}] &= \mathrm{E}[(\hat{\boldsymbol{\theta}} - \mathrm{E}(\hat{\boldsymbol{\theta}}))(\hat{\boldsymbol{\theta}} - \mathrm{E}(\hat{\boldsymbol{\theta}}))^t] \\ &= \mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \mathrm{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \mathrm{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))^t].\end{aligned} \tag{6.112}$$

We get from (6.98)

$$\begin{aligned}
&\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \mathrm{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\
&\quad = \boldsymbol{w}_1 + \boldsymbol{w}_2 + W\boldsymbol{w}_1 + W\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2] \\
&\qquad - \mathrm{E}[W\boldsymbol{w}_1 + W\boldsymbol{w}_2) \\
&\quad = \boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_3 + \boldsymbol{h}_4
\end{aligned} \tag{6.113}$$

where

$$\begin{aligned}
\boldsymbol{h}_1 &= \boldsymbol{w}_1 - \mathrm{E}(\boldsymbol{w}_1) + \boldsymbol{w}_2 - \mathrm{E}(\boldsymbol{w}_2) \\
\boldsymbol{h}_2 &= W[\boldsymbol{w}_1 - \mathrm{E}(\boldsymbol{w}_1)] + W[\boldsymbol{w}_2 - \mathrm{E}(\boldsymbol{w}_2)] \\
\boldsymbol{h}_3 &= W[\mathrm{E}(\boldsymbol{w}_1) + \mathrm{E}(\boldsymbol{w}_2)] \\
\boldsymbol{h}_4 &= -\mathrm{E}[W\boldsymbol{w}_1 + W\boldsymbol{w}_2].
\end{aligned}$$

Notice that since for any $\boldsymbol{\xi}$, $\mathrm{E}[\boldsymbol{\xi} - \mathrm{E}(\boldsymbol{\xi})] = 0$, it follows from (6.113) that $\mathrm{E}[\sum_{\ell=1}^{4} \boldsymbol{h}_\ell] = 0$. Also $\boldsymbol{h}_4$ is a constant. Therefore,

$$\begin{aligned}
\mathrm{Cov}\,[\hat{\boldsymbol{\theta}}] &= \mathrm{E}[(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell)^t] \\
&= \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell)^t] + \mathrm{E}[\boldsymbol{h}_4(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell)^t] \\
&= \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell)^t] = \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)^t] + \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)\boldsymbol{h}_4^t] \\
&= \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)^t] + \mathrm{E}[(\sum_{\ell=1}^{4} \boldsymbol{h}_\ell - \boldsymbol{h}_4)\boldsymbol{h}_4^t] \\
&= \mathrm{E}[(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)(\sum_{\ell=1}^{3} \boldsymbol{h}_\ell)^t] - \boldsymbol{h}_4\boldsymbol{h}_4^t.
\end{aligned} \tag{6.114}$$

Under LS1 and LS4, we have, by (6.101) and Lemma 6.4,

$$\mathrm{E}[(\boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1])(\boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1])^t] = \mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] = \mathcal{O}(n^{-1}T_{\#}^{-1}). \tag{6.115}$$

Under the conditions of the lemma, we get, by (6.41) and Lemma 6.5

$$\begin{aligned}
&\mathrm{E}[(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])^t] = \mathrm{E}[\boldsymbol{w}_2\boldsymbol{w}_2^t] - \mathrm{E}[\boldsymbol{w}_2]\mathrm{E}[\boldsymbol{w}_2^t] \\
&\quad = U^{-1}X^t\epsilon\boldsymbol{e}\boldsymbol{e}^t\epsilon XU^{-1} - U^{-1}X^t\mathrm{E}[\epsilon\boldsymbol{e}]\mathrm{E}[\epsilon\boldsymbol{e}]^tXU^{-1} \\
&\quad = U^{-1}[X^tSX + X^t\boldsymbol{\eta}\boldsymbol{\eta}^tX + \mathcal{O}(T_{\#}^{-2}n)]U^{-1} \\
&\qquad - U^{-1}X^t\boldsymbol{\eta}\boldsymbol{\eta}^tXU^{-1} \\
&\quad = U^{-1}[X^tSX + \mathcal{O}(T_{\#}^{-2}n)]U^{-1} = \mathcal{O}(n^{-1}T_{\#}^{-2}).
\end{aligned} \tag{6.116}$$

By applying the Cauchy-Schwartz inequality (see (6.96)), we see that

$$\mathrm{E}[(\boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1])(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])^t] = \mathcal{O}(n^{-1}T_{\#}^{-3/2}) \tag{6.117}$$

and

$$\mathrm{E}[(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_1 - \mathrm{E}[\boldsymbol{w}_1])^t] = \mathcal{O}(n^{-1}T_{\#}^{-3/2}). \tag{6.118}$$

Therefore,

$$\begin{aligned} &\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_1^t] \\ &= \mathrm{E}[(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])^t] \\ &= \mathcal{O}(n^{-1}T_{\#}^{-1}). \end{aligned} \tag{6.119}$$

Notice that when $T_{\#}$ is not very small, (6.115) dominates $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_1^t]$, since all other terms in $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_1^t]$ are of higher order. It will be apparent in the course of the remainder of this proof that $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_1^t]$ and, therefore, (6.115) dominate the other terms in $\mathrm{Cov}\,[\hat{\boldsymbol{\theta}}]$. This fact is of some importance in numerical approximations.

Now consider the term involving $\boldsymbol{h}_2$. Following arguments very similar to those used to establish (6.108)) in Lemma 6.6, it might be shown that $\mathrm{E}([U^{-1}V]^p[(U^{-1}V)^t]^q) = \mathcal{O}((nT_{\#})^{-p-q})$. Thus, writing W as a power series using (6.105):

$$W = -U^{-1}V + [U^{-1}V]^2 - \ldots,$$

it may be shown that $\mathrm{E}[(WW^t)^2] = \mathcal{O}(n^{-2}T_{\#}^{-2})$.

A similar argument may be made to compute the order of $\mathrm{E}[(\boldsymbol{w}\boldsymbol{w}^t)^2]$. We outline it below. It is easily verified from (6.22) that each element in $\boldsymbol{w}$ is a linear combination of $(\psi_s - \mathrm{E}[\psi_s])$'s where, of course, $s = 1, \ldots, n$. Each element in $\boldsymbol{w}\boldsymbol{w}^t$, therefore, consists of the product of two such linear combinations, and each element in $(\boldsymbol{w}\boldsymbol{w}^t)^2$ is the sum of a finite number (K) of products of four such linear combinations. Each of these last-mentioned products, if multiplied out in full, would yield n^4 terms each involving products of four $(\psi_s - \mathrm{E}[\psi_s])$'s. However, most of these terms would have an expected value of zero. The terms with non-zero expectations are those including $\mathrm{E}[(\psi_s - \mathrm{E}[\psi_s])^2(\psi_p - \mathrm{E}[\psi_p])^2]$ and $\mathrm{E}[(\psi_s - \mathrm{E}[\psi_s])^4]$. There would be, at most, n^2 expressions of the first kind and n expression of the second kind. The coefficients of the corresponding terms consist of the product of a quantity of order $(n^{-1}T_{\#}^{-1})^4$ (from the four U^{-1}'s — see the definition of $\boldsymbol{w}$) with one which is less than $\mu^4T_{\#}^4$ (from the X and T). From (6.26), we see that $\mathrm{E}[(\psi_s - \mathrm{E}[\psi_s])^2(\psi_p - \mathrm{E}[\psi_p])^2] = \mathcal{O}(T_{\#}^{-2})$ plus terms of higher order in $T_{\#}$. Moreover, from (6.17), it may be shown that $\mathrm{E}[(\psi_s - \mathrm{E}[\psi_s])^4 = \mathcal{O}(T_{\#}^{-2})$. Thus, $\mathrm{E}[(\boldsymbol{w}\boldsymbol{w}^t)^2] = \mathcal{O}(n^{-2}T_{\#}^{-2})$.

That $\mathrm{E}([(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_2])^t]^2) = \mathcal{O}(n^{-2}T_{\#}^{-4})$ can be established in much the same way. On using the Cauchy-Schwartz inequality [see (6.96) and (6.97)] on the above results, it may be shown that

$$\begin{aligned} \mathrm{E}[\boldsymbol{h}_2\boldsymbol{h}_2^t] = \mathrm{E}[W\{&(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2]) \\ &\cdot(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])^t\}W^t] = \mathcal{O}(n^{-2}T_{\#}^{-2}). \end{aligned} \tag{6.120}$$

Now applying (6.96) to (6.119) and (6.120) it follows that both

$$\mathrm{E}[\boldsymbol{h}_2\boldsymbol{h}_1^t] = \mathrm{E}[W\{(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])^t\}^t]$$

and

$$\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_2^t] = \mathrm{E}[\{(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])(\boldsymbol{w}_1 + \boldsymbol{w}_2 - \mathrm{E}[\boldsymbol{w}_1 + \boldsymbol{w}_2])^t\}W^t]$$

are of lesser order than $\mathcal{O}(n^{-1}T_{\#}^{-1})$. Notice that $\mathrm{tr}\,[\boldsymbol{h}_3\boldsymbol{h}_3^t] = \mathrm{tr}\,[(\mathrm{E}(\boldsymbol{w}_1) + \mathrm{E}(\boldsymbol{w}_2))WW^t(\mathrm{E}(\boldsymbol{w}_1) + \mathrm{E}(\boldsymbol{w}_2))^t]$ since for any two matrices G and H of appropriate dimension, $\mathrm{tr}\,[GH] = \mathrm{tr}\,[HG]$. Therefore, from (6.99), (6.100) and (6.111) and the fact that any element of such a matrix is smaller than its trace, we see that $\mathrm{E}[\boldsymbol{h}_3\boldsymbol{h}_3^t]$ is of order $n^{-1}T_{\#}^{-5}$. Therefore, using (6.119) and (6.120) we may verify that $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_3^t]$, $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_2^t]$, $\mathrm{E}[\boldsymbol{h}_1^t\boldsymbol{h}_3]$ and $\mathrm{E}[\boldsymbol{h}_1^t\boldsymbol{h}_3]$ are all of lesser order of magnitude than $\mathcal{O}(n^{-1}T_{\#}^{-1})$.

From (6.109) and (6.110), we get

$$\boldsymbol{h}_4 = \mathcal{O}(n^{-1}T_{\#}^{-1}) + \mathcal{O}(n^{-1/2}T_{\#}^{-5/2}).$$

Thus, $\boldsymbol{h}_4\boldsymbol{h}_4^t$ only includes terms of order higher than $n^{-1}T_{\#}^{-2}$.

The lemma follows on combining together (6.114), (6.119), (6.120) and the results mentioned in the last three paragraphs, □

From the above proof we see that for large T_{ij}'s, the covariance matrix of $\boldsymbol{w}_1$ dominates all other terms in the covariance of $\hat{\boldsymbol{\theta}}$. Even for the '$T_{ij}$' data set (see page 415), for which T_{ij}'s are hardly large, with travel times as the only cost, the variance of $\boldsymbol{w}_1$ is about 94 per cent of the variance of $\hat{\theta}_1$. Therefore, at least as far as the covariance matrix is concerned, weighting by N_{ij}'s as opposed to T_{ij}'s makes very little difference when n and T_{ij}'s are moderately large (see the definitions of $\boldsymbol{w}_1$, $\boldsymbol{w}$, etc. in (6.41) and the material following it).

Corollary 6.6 *Under LS1A, LS2, LS3 and LS4,* $\mathrm{Cov}\,[\hat{\boldsymbol{\theta}}] = \mathcal{O}(n^{-1}T_{\#}^{\alpha-1})$

PROOF: The proof is similar to that for Lemma 6.7. Under LS1A and LS4, we need to use Corollary 6.3 and Corollary 6.4 to appropriately change the right end of (6.115) and then make corresponding changes in the rest of the proof. □

Since the covariance matrix of $\hat{\boldsymbol{\theta}}$ is roughly of the same small order as $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$, it is natural to conjecture that for large enough $T_{\#}$ and n the elements of $\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$ are much smaller then their respective standard errors (which are the square roots of the of the diagonal elements of the covariance matrix of $\hat{\boldsymbol{\theta}}$). This would be a valuable result, since it implies that the bias is so much smaller than the standard error that it would not be noticeable. A proof is given below.

Corollary 6.7 *Under LS1, LS2, LS3 and LS4, as* $T_{\#} \to \infty$,

$$\mathrm{E}[\hat{\theta}_k - \theta_k]/\mathrm{s.e.}[\theta_k] = [\mathcal{O}(n^{-1/2}T_{\#}^{-1/2})],$$

where n *is fixed,* s.e.$[\theta_k]$ *is the square root of the* k*th diagonal element of* Cov $[\hat{\boldsymbol{\theta}}]$ *(i.e., it is the standard error of* $\hat{\theta}_k$*).*

PROOF: From (6.98), (6.99) and (6.100) it follows that

$$\mathrm{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \mathrm{E}[W(\boldsymbol{w}_1 + \boldsymbol{w}_2)] + \mathcal{O}(T_{\#}^{-2}). \tag{6.121}$$

From the proof of Lemma 6.7 (notice, in particular, (6.115) and (6.116)) and the use of (6.96), we can see that for large $T_{\#}$,

$$\mathrm{E}([\boldsymbol{w}_1 + \boldsymbol{w}_2][\boldsymbol{w}_1 + \boldsymbol{w}_2]^t) = \mathcal{O}(n^{-1}T_{\#}^{-1}).$$

Applying (6.95), we get

$$\begin{aligned}
&\mathrm{E}[W(\boldsymbol{w}_1 + \boldsymbol{w}_2)] \\
&\quad \leq [\mathrm{E}(\,\mathrm{diag}\,[WW^t])]^{1/2}[\mathrm{E}([\boldsymbol{w}_1 + \boldsymbol{w}_2]^t[\boldsymbol{w}_1 + \boldsymbol{w}_2])]^{1/2} \\
&\quad \leq [\mathrm{E}(\,\mathrm{diag}\,[WW^t])]^{1/2}\Xi
\end{aligned} \tag{6.122}$$

where $\Xi = [\mathcal{O}(n^{-1/2}T_{\#}^{-1/2})]$ and, as before, diag(A) stands for the diagonal matrix whose non-zero elements are the diagonal elements of A.

From (6.111), the diagonal elements of $\mathrm{E}[WW^t]$ are less than a constant times U^{-1}. By a combination of (6.64) and (6.65), the diagonal elements of U^{-1} are less than those of $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]$. From the proof of Lemma 6.7, we can see that as $T_{\#} \to \infty$, $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] \to \mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t]$. Thus, it follows, from (6.122) and (6.121) that

$$\mathrm{E}[\hat{\theta}_k - \theta_k] \leq \mathrm{s.e.}[\theta_k][\mathcal{O}(n^{-1/2}T_{\#}^{-1/2})].$$

This proves the corollary. □

In fact, we conjecture that, if $T_{\#}$ is not too small, as $n \to \infty$, $\boldsymbol{\beta}_1 = \mathrm{diag}(\,\mathrm{Cov}\,[\hat{\boldsymbol{\theta}}])^{1/2}\Xi$. An examination of the proof of the above corollary reveals that, except for terms of the order of $T_{\#}$, the only point where $T_{\#} \to \infty$ is essential is in showing $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] \to \mathrm{Cov}\,[\hat{\boldsymbol{\theta}}]$. In fact, it would have sufficed to show that

$$\mathrm{diag}\,(\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]) < \mathrm{diag}\,(\,\mathrm{Cov}\,[\hat{\boldsymbol{\theta}}]). \tag{6.123}$$

The diagonal elements of the left side are essentially variances of the estimates of components of $\boldsymbol{\theta}$ in a weighted least squares exercise with weights T_{ij}, while the diagonal elements on the right side are the variances of the estimates of components of $\boldsymbol{\theta}$ in a regression using the random variables N_{ij} as weights. Since $\mathrm{E}[N_{ij}] = T_{ij}$, it is reasonable to believe that (6.74) holds, although a proof has been elusive.

Corollary 6.8 *Under LS1A, LS2, LS3 and LS4, as $T_{\#} \to \infty$, but n remains fixed,*

$$\mathrm{E}[\hat{\theta}_k - \theta_k]/\text{s.e.}[\theta_k] = [\mathcal{O}(n^{-1/2}T_{\#}^{(\alpha-1)/2})]$$

PROOF: The proof is very similar to that for Corollary 6.7. A key difference is the use of Corollary 6.3 and Corollary 6.4 to determine the order of $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t]$. Note also that under LS1A, as under LS1, $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] \to \mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t]$. The rest of the proof is identical. □

Lemma 6.8 *Under Conditions LS1 through LS4, For $T_{\#} \to \infty$ and fixed n, each component of the random vector $\hat{\boldsymbol{\theta}}$ is asymptotically normal. For large enough $T_{\#}$, each component of $\hat{\boldsymbol{\theta}}$ is asymptotically normal even when only $n \to \infty$.*

PROOF: As already mentioned, the proof of Lemma 6.7 shows that for large T_{ij}'s the covariance matrix of $\boldsymbol{w}_1$ dominates all other terms in the covariance of $\hat{\boldsymbol{\theta}}$. Therefore, the distribution of $\boldsymbol{w}_1$ approaches that of $\hat{\boldsymbol{\theta}} - \mathrm{E}[\hat{\boldsymbol{\theta}}]$ as $T_{\#} \to \infty$. Since the N_{ij}'s are Poisson and consequently are asymptotically normal as $T_{\#} \to \infty$, it follows that the ϵ_{ij}'s are asymptotically normal. The variance of ϵ_{ij} is T_{ij} and hence the variance of v_{ij} is 1, where $\epsilon_{ij} = v_{ij}\sqrt{T_{ij}}$. Making this last mentioned substitution for ϵ_{ij} in (6.17) it is seen that for large T_{ij}, ψ_{ij} is asymptotically $v_{ij}/\sqrt{T_{ij}}$ with the other terms in (6.17) much smaller than it. Therefore, the ψ_i's are also asymptotically normal. Since each component of $\boldsymbol{w}_1$ is a linear function with constant coefficients of a finite number of e_i's and, therefore, of a finite number of ψ_i's, it follows that each component of $\boldsymbol{w}_1$ is asymptotically normal for large $T_{\#}$.

Now consider the situation when $n \to \infty$. It follows from (6.40), (6.41), (6.105), (6.106) and (6.108) that

$$[X^tNX]^{-1} = (I + W)U^{-1} \to U^{-1} = [X^tTX]^{-1}. \tag{6.124}$$

Therefore, from (6.41) and (6.42) when $n \to \infty$, the distribution of $\hat{\boldsymbol{\theta}}$ is equivalent to that of $\boldsymbol{w}_1 + \boldsymbol{w}_2$.

First consider $\boldsymbol{w}_1$. As before, let the elements of the kth column of X be $x_p{}^{(k)} = x_{ij}{}^{(k)}$ under the convention described in Section 6.3.1, i.e., $x_{ij}{}^{(k)} = c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}$. The elements of the kth row of X^tT are $x_{ij}{}^{(k)}T_{ij}$ and the elements in the kth row of $X^tT\mathcal{P}$ are

$$z_{qr}{}^{(k)} = x_{qr}{}^{(k)}T_{qr} - I^{-1}\sum_{i=1}^{I} x_{ir}{}^{(k)}T_{ir} - J^{-1}\sum_{j=1}^{J} x_{qj}{}^{(k)}T_{qj}$$
$$+ I^{-1}J^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}{}^{(k)}T_{ij}.$$

Write, as we have done before, $z_s{}^{(k)} = z_{qr}{}^{(k)}$, where $s = 1, \ldots, n$ and $s = (q-1)J + r$. Since

$$X^t T e = X^t T \mathcal{P} \psi$$

the kth component of $X^t T e$ is a linear combination of ψ_s's in which the coefficient of ψ_s is $z_s{}^{(k)}$. Therefore, each component of

$$\sqrt{n} w = \sqrt{n} U^{-1} X^t T[e - \mathrm{E}(e)]$$

is a linear combination of $(\psi_s - \mathrm{E}[\psi_s])$'s where $s = 1, \ldots, n$ and the coefficient of $(\psi_s - \mathrm{E}[\psi_s])\check{T}^{1/2}$ for the pth component of w is

$$\sum_{k=1}^{K} \sqrt{n} u_{pk} z_s{}^{(k)} / \check{T}^{1/2},$$

where $U^{-1} = (\!(u_{pk})\!)$. By condition LS3 we know that each $x_s{}^{(k)}$ is bounded. Therefore, each $z_s{}^{(k)} = \mathcal{O}(T^{\#})$ and since $u_{pk} = \mathcal{O}(n^{-1} T_{\#}^{-1})$ for all p and k, because of (both parts of) LS1, the coefficient of each $(\psi_s - \mathrm{E}[\psi_s])\check{T}^{1/2}$ is $\mathcal{O}(n^{-1/2})$. From (6.26) and the definition of ψ [see statement of Lemma 6.1], we see that the $[(\psi_s - \mathrm{E}[\psi_s])\check{T}^{1/2}]$'s are independent with identical means and variances. Their coefficients go to zero as $n \to \infty$ and the sum of squares of the coefficients is a finite number. Hence, by a version of the central limit theorem reproduced from Gnedenko and Kolmogorov (1954) as Theorem 5.2 in Sen and Srivastava (1990) Chapter 5, it follows that any component of $\sqrt{w}$ is normal. Alternatively, notice that the second and third absolute moments of $\psi_s - \mathrm{E}[\psi_s]\check{T}^{1/2}$ are independent of n and bounded above and below. Hence, it is easily shown that the conditions of Liapunov's central limit theorem (see Rao, 1972, p. 127) are met.

Now consider w_2. It is easily seen that an element of $w_2 = X^t \epsilon e$ is of the form $\sum_{s=1}^{n} \sum_{p=1}^{n} x_s{}^{(k)} \mathcal{P}_{sp} \epsilon_s \psi_p$. An inspection of the proof of Lemma 6.5 will show that, when $n \to \infty$, all terms including $\epsilon_s \psi_p$ with $s \neq p$ contribute relatively little to the variance of a component of w_2. Thus, we need only consider $\sum_{k=1}^{K} u_{pk} \sum_{s=1}^{n} x_s{}^{(k)} \mathcal{P}_{ss} \epsilon_s \psi_s$. The asymptotic normality of this can be established in much the same way as we did for the components of w_1.

This proves the lemma. □

Corollary 6.9 *Under Conditions LS1A through LS4, For $T_{\#} \to \infty$ and fixed n, each component of the random vector $\hat{\theta}$ is asymptotically normal. For large enough $T_{\#}$, each component of $\hat{\theta}$ is asymptotically normal even when only $n \to \infty$.*

The proof is very similar to that for Lemma 6.8. Given these results, we now present a proof of Theorem 6.1.

Proof of Theorem 6.1: The proof of the theorem consists of combining Lemma 6.4, Lemma 6.6, Lemma 6.7, Lemma 6.8 and Corollary 6.7. □

The effect of $\boldsymbol{\beta}_2$ in Theorem 6.1 can be assessed somewhat by comparison with

$$\boldsymbol{\theta} = (X^t T X)^{-1} X^t T X \boldsymbol{\theta} = U^{-1} X^t T \mathcal{P} \boldsymbol{t}, \tag{6.125}$$

where $\boldsymbol{t} = (t_{11}, \ldots, t_{1J}, \ldots\ldots, t_{I1}, \ldots, t_{IJ})^t$. The bias term under consideration is approximately

$$U^{-1} X^t T \mathcal{P} \boldsymbol{v}, \tag{6.126}$$

where $\boldsymbol{v} = (v_{11}, \ldots, v_{1J}, \ldots\ldots, v_{I1}, \ldots, v_{IJ})^t$ and $v_{ij} = -1/[24T_{ij}^2]^{-1}$. A comparison of (6.125) and (6.126) will show that if $T_{\#}$ is 15 the effect of the bias on the estimate is at most .008 *per cent*. For $T_{\#} = 100$ the effect is at most .00009 per cent. Therefore, for large $T_{\#}$, there would appear to be little reason to pay much attention to it.

Proof of Corollary 6.1: The proof follows from Corollary 6.5, Corollary 6.6, Corollary 6.8 and Corollary 6.9. □

6.3.5 Some Practical Details

When n and the T_{ij}'s are large, we recommend Procedure 1. While it is biased, the bias is imperceptible when T_{ij}'s are large and some mild conditions are met. While this statement is asymptotic, the procedure can be safely used if only a few out of many T_{ij}'s are quite small. Even if the T_{ij}'s are not large, if they are not too small, under the mild conditions mentioned above, the bias and standard errors go to zero as $n \to \infty$.

It is important to note that if a standard statistical package is used for obtaining the least squares estimates, the estimates of the standard errors would be wrong. This is because, the package would assume that the e_s's are independent and would estimate the covariance matrix of $\hat{\boldsymbol{\theta}}$ to be $(X^t N X)^{-1}$, which would be incorrect as can be seen from an inspection of the proof of Lemma 6.7. Thus, the standard errors need to be estimated separately.

The proof of Lemma 6.7 gives us a way to estimate an approximate set of standard errors. If $T_{\#}$ is very large, since then the covariance of $\boldsymbol{w}$ is the dominant component in the covariance of $\hat{\boldsymbol{\theta}}$, we could use the diagonal elements of

$$\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^t] \approx U^{-1} X^t T \mathcal{P} T^{-1} \mathcal{P} T X U^{-1}, \tag{6.127}$$

noting that $\mathrm{E}[(\boldsymbol{e} - \mathrm{E}(\boldsymbol{e}))(\boldsymbol{e} - \mathrm{E}(\boldsymbol{e}))^t] = \mathcal{P}\mathrm{E}[(\boldsymbol{\psi} - \mathrm{E}(\boldsymbol{\psi}))(\boldsymbol{\psi} - \mathrm{E}(\boldsymbol{\psi}))^t]\mathcal{P}^t = \mathcal{P}\breve{T}^{-1}\mathcal{P}^t \approx \mathcal{P}T^{-1}\mathcal{P}$, using (6.26). Of course, in practice, we would have to replace T by its estimate $\hat{T}$.

When n is large and $T_{\#}$ not too small, a better approximation can be obtained by using $\mathrm{E}[\boldsymbol{h}_1\boldsymbol{h}_1^t]$, i.e.,

$$\begin{aligned}\mathrm{E}[(\boldsymbol{w}_1+\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_1+\boldsymbol{w}_2])(\boldsymbol{w}_1+\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_1+\boldsymbol{w}_2])^t]\\ =\mathrm{E}[(\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1])(\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1])^t]\\ +\mathrm{E}[(\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_2])^t]\\ +\mathrm{E}[(\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1])(\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_2])^t]\\ +\mathrm{E}[(\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1])^t].\end{aligned}\tag{6.128}$$

Since $\boldsymbol{w}=\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1]$, (6.127) is an approximation for the term on the second line in (6.128). An approximation to the term on the third line is given by (6.116). Since $U^{-1}\mathcal{O}(T_{\#}^{-2}n)U^{-1}=\mathcal{O}(n^{-1}T_{\#}^{-4})$, it is $U^{-1}X^tSXU^{-1}$ where S is given in Lemma 6.5 and is approximately $2I_n$.

In order to compute approximations for the terms on the last two lines of (6.128), let $\tilde{\psi}_s=\psi_s-\mathrm{E}[\psi_s]$. Consider the (s,r)th element of the matrix $\epsilon[\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]][\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]]^tT$, which is of the form

$$\epsilon_s[e_s-\mathrm{E}[e_s]][e_r-\mathrm{E}[e_r]]T_r=\epsilon_s\sum_{p=1}^{n}\mathcal{P}_{sp}\tilde{\psi}_p\sum_{q=1}^{n}\mathcal{P}_{rq}\tilde{\psi}_qT_r,\tag{6.129}$$

on using the relation $e_s=\sum_{p=1}^n\mathcal{P}_{sp}\psi_p$. Taking the expectations of the terms on the right side of (6.129), we see that the only term with non-zero expectation is $\epsilon_s\mathcal{P}_{ss}\tilde{\psi}_s\mathcal{P}_{rs}\tilde{\psi}_sT_r$, whose expectation is $\gamma\mathcal{P}_{rs}T_r\mathrm{E}[\epsilon_s\tilde{\psi}_s\tilde{\psi}_s]\approx\gamma\mathcal{P}_{rs}T_rT_s^{-1}$ from (6.17) and (6.19) in Lemma 6.1. Thus

$$\epsilon[\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]][\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]]^tT\approx\gamma T^{-1}\mathcal{P}T.\tag{6.130}$$

Therefore, using (6.99) and (6.100), we get,

$$\begin{aligned}\mathrm{E}[(\boldsymbol{w}_2-\mathrm{E}[\boldsymbol{w}_2])(\boldsymbol{w}_1-\mathrm{E}[\boldsymbol{w}_1])^t]\\ =U^{-1}X^t\epsilon[\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]][\boldsymbol{e}-\mathrm{E}[\boldsymbol{e}]]^tTXU^{-1}\\ \approx\gamma U^{-1}X^tT^{-1}\mathcal{P}TXU^{-1}.\end{aligned}\tag{6.131}$$

This is, therefore, an approximation for the term in the last line of (6.128). The second from last line is approximated by its transpose.

Thus the standard errors of the components $\hat{\theta}_k$ of $\hat{\boldsymbol{\theta}}$ may be approximated by the square roots of the diagonal elements of

$$U^{-1}X^t[T\mathcal{P}T^{-1}\mathcal{P}T+S+\gamma T^{-1}\mathcal{P}T+\gamma T\mathcal{P}T^{-1}]XU^{-1},\tag{6.132}$$

after replacing T by its estimate $\hat{T}$.

However, it should be noted that these expressions are valid for T_{ij}'s which are not too small, since we have used the first few terms in power series in deriving all these terms and when T_{ij}'s are small these approximations cannot possibly hold. In the presence of a large number of sufficiently small T_{ij}'s, even negative estimates of variances can occur! Still, if most T_{ij}'s are moderately large, while a few are fairly small, these formulæ should provide reasonable estimates. Obviously, since T_{ij}'s are unknown, in practice their estimates will need to be used to compute estimates of covariance matrices.

Notice that in the least squares exercise in Procedures 1 and 1A, no 'intercept' term is used. If one is inadvertently left in, Corollary 6.2 will be false and substantial bias will result. This bias alone can, and often does, make the 'intercept' term 'significant.'

We conclude this section with the following description of the hand-calculator version of Procedure 1, referred to in the introduction to this chapter. It works fairly well when I and J are not too large and when K is either 1 or 2.

Step 1 Using N_{ij}'s, compute z_{ij}'s and arrange them in the form of a matrix or spreadsheet with each row representing an i and each column representing a j.

Step 2 Compute the mean of all z_{1j}'s in row 1; i.e., compute $z_{1\bullet}$. Subtract this mean from each z_{1j} to get $z_{1j} - z_{1\bullet}$ and replace each z_{1j} with the corresponding $z_{1j} - z_{1\bullet}$. Do the same to all rows. Now we have a matrix with elements $z_{ij} - z_{i\bullet}$.

Step 3 For each column of the matrix just created, compute means. For each column j, the mean is easily seen to be $z_{\bullet j} - z_{\bullet\bullet}$. Now subtract this mean from each element of the corresponding column. The result will be a matrix with elements $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$.

Step 4 For each value of k, an application of Steps 2 and 3 above to the matrix of $c_{ij}{}^{(k)}$'s yields the matrix with elements $c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}$.

Step 5 Multiply each $(c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)})$ and $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$ by the corresponding $N_{ij}^{1/2}$ and apply the usual formulæ for calculation of ordinary least squares estimates for models with no intercept [for $K = 1$, the estimate is $\hat{\theta}_1 = \sum_{s=1}^{n} \tilde{y}_s \tilde{x}_{1s} / \sum_{s=1}^{n} \tilde{x}_{1s}^2$, where $\tilde{y}_s$'s are the products of $N_{ij}^{1/2}$ with $(z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet})$'s and $\tilde{x}_{1s}$'s are the products of $N_{ij}^{1/2}$ with $c_{ij}{}^{(1)} - c_{i\bullet}{}^{(1)} - c_{\bullet j}{}^{(1)} + c_{\bullet\bullet}{}^{(1)}$'s — see Sen and Srivastava, 1990, p. 11].

Alternatively, we can proceed graphically. If there is only one separation measure — i.e., $K = 1$ — plot $(z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet})$'s against $c_{ij}{}^{(1)} - c_{i\bullet}{}^{(1)} - c_{\bullet j}{}^{(1)} + c_{\bullet\bullet}{}^{(1)}$'s, fit a line by eye through the origin and read off its slope. If there are two separation measures, we may use the fact that $y = \gamma_1 x_1 + \gamma_2 x_2$ is equivalent to $y/x_1 = \gamma_1 + \gamma_2 x_2/x_1$. Then we can plot $[z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}]/[c_{ij}{}^{(1)} - c_{i\bullet}{}^{(1)} - c_{\bullet j}{}^{(1)} + c_{\bullet\bullet}{}^{(1)}]$ against $[c_{ij}{}^{(2)} - c_{i\bullet}{}^{(2)} - c_{\bullet j}{}^{(2)} + c_{\bullet\bullet}{}^{(2)}]/[c_{ij}{}^{(1)} - c_{i\bullet}{}^{(1)} - c_{\bullet j}{}^{(1)} + c_{\bullet\bullet}{}^{(1)}]$

and the intercept and the slope would be the estimates of the two parameters. We have found plots obtained by eye to yield estimates surprisingly close to *appropriately* weighted least squares estimates.

6.4 Alternative Methods

In Section 6.2 we saw some alternatives to Procedure 1 — the procedure we examined in detail in the last section. In this section we discuss some other alternatives, mainly modifications of Procedure 1.

6.4.1 Use of Iterative Reweighting in Procedure 1

Instead of weighting simply by the random variables N_{ij}, one might consider doing a few iterative reweighting steps. Here, we would use as weights, for every step, $s+1$, after the first step, values of estimates $\hat{T}_{ij}$ of T_{ij} from the previous step, s. Such values could be obtained using the DSF Procedure with $N_{i\oplus}$'s and $N_{\oplus j}$'s for marginal totals and F_{ij}'s based on estimates of $\boldsymbol{\theta}$ from step s. For future reference we call such a procedure **Procedure 2**. Alternatively, $\hat{T}_{ij}$ could be obtained from

$$\sum_{k=1}^{K} \hat{\theta}_k [c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}] + z_{i\bullet} + z_{\bullet j} - z_{\bullet\bullet} \tag{6.133}$$

which is a rough estimate of $\hat{t}_{ij}$, since

$$t_{ij} - t_{i\bullet} - t_{\bullet j} + t_{\bullet\bullet} = \sum_{k=1}^{K} \hat{\theta}_k [c_{ij}{}^{(k)} - c_{i\bullet}{}^{(k)} - c_{\bullet j}{}^{(k)} + c_{\bullet\bullet}{}^{(k)}].$$

Notice that the use of the word 'reweighting' here does not imply using GLIM (Section 5.5.5), a procedure with which the word has become associated; the use here is just in the sense of weighting over or weighting repeatedly, as in Sen and Srivastava (1990, Chapter 6).

We have already seen that when T_{ij}'s are large, the covariance of $\boldsymbol{w}_1$ dominates the expression for the covariance of $\hat{\boldsymbol{\theta}}$. The covariance of $\boldsymbol{w}_1$ would have been the covariance of $\hat{\boldsymbol{\theta}}$ had we weighted by T_{ij}'s instead of N_{ij}'s. Therefore, an iterative weighting scheme, the purpose of which is to find better estimates (than N_{ij}'s) of T_{ij}'s to use as weights, is unlikely to improve covariances much. As far as bias is concerned, the expression of the order $T_{\#}^{-2}$ is mainly due to taking the logarithm of $N_{ij} + 1/2$ and cannot be improved by weighting by other estimates of the T_{ij}'s (see the proof of Lemma 6.6). Therefore, the only place where we might be able to see any significant improvement with Procedure 2 is with the bias term that is of order $n^{-1}T_{\#}^{-1}$, in particular that associated with $\mathrm{E}[W\boldsymbol{w}_1]$.

However, we *conjecture* that even here we will not see much improvement, and, indeed the bias could get larger. One reason that $\mathrm{E}[W\boldsymbol{w}_1] = \mathcal{O}(n^{-1}T_{\#}^{-1})$ is the following:

$$\mathrm{E}[W\boldsymbol{w}_1] \approx \mathrm{E}[U^{-1}V\boldsymbol{w}_1] = \mathrm{E}[U^{-1}X^t\epsilon XU^{-1}X^tT\mathcal{P}\boldsymbol{\psi}]$$

is the expectation of a matrix, each element of which is of the form $\epsilon_u\psi_v$ times a constant and $\mathrm{E}[\epsilon_u\psi_v] = 0$ for $u \neq v$. If N_{ij}'s are replaced as weights by some other estimate of T_{ij} — say $T_{ij} + \breve{\epsilon}_{ij}$ — then $\breve{\epsilon}_{ij}$ would be a complex function of several ψ_s's at least in the later iterations. Consequently, the matrix $\mathrm{E}[\breve{\boldsymbol{\epsilon}}\boldsymbol{\psi}^t]$ would not usually be too sparse. Thus, it would appear that each element of the matrix would have to be of order n^{-2} for this part of the bias to remain at the same order of magnitude.

Moreover, it appears unlikely that Part 2 of Lemma 6.6 would hold if ϵ_{ij} were to be replaced by $\breve{\epsilon}_{ij}$. As can be seen by examining the proof of Lemma 6.6, this can cause the bias terms involving $\boldsymbol{w}_2$ to go up, perhaps substantially.

Thus, it appears that Procedure 2, which is more difficult to use and which requires substantially more computer time, might not necessarily yield better results and could yield poorer estimates than Procedure 1.

6.4.2 Not Reducing Parameters

Another procedure could be constructed based directly on (6.4). It would have as dependent variable $\log[N_{ij} + \frac{1}{2}]$ and as independent variables the columns of the matrix M (as defined at the beginning of Section 5.2). However, since M is singular, one of its first $I + J$ columns could be deleted. (As an alternative, Cesario, in several papers, e.g., 1974a, used a set of constraints and obtained a procedure somewhat similar to Procedure 4 below.) As in (5.215), call the resultant matrix $M_{(2)}$. Designate as **Procedure 3** that of applying weighted (by N_{ij}) least squares to such a model.

The difficulty with this procedure is that Corollary 6.2 no longer holds. Hence, Lemma 6.5 is no longer available. Therefore, $\boldsymbol{w}_2$ provides a term

$$(M_{(2)}^tTM_{(2)})^{-1}M_{(2)}^t\boldsymbol{1}\boldsymbol{1}^tM_{(2)}(M_{(2)}^tTM_{(2)})^{-1}, \tag{6.134}$$

in the variance of $\hat{\boldsymbol{\theta}}$ (see proof of Lemma 6.5), which does not go to zero with increasing n. The fact that Corollary 6.2 does not hold also adds a term $(M_{(2)}^tTM_{(2)})^{-1}M_{(2)}^t\boldsymbol{1}$ to the bias (see (6.100)) and this term too does not go to zero. Therefore, such a Procedure 3 will usually give estimates which, for large n, could be inferior to those from Procedure 1. [This is the usual type of problem one encounters when weighting with a random variable. The availability of Corollary 6.2 was able to save Procedure 1 from this fate.]

In the case of Procedure 3, iterative weighting could make an enormous improvement. Here, since the dependent variable is $\log[N_{ij} + \frac{1}{2}]$, the 'predicted' values $\hat{t}_{ij}$ are estimates of $\log[T_{ij}]$ and, consequently, $\exp[\hat{t}_{ij}]$ can be used in iterations after the first. However, we doubt that the estimates would be much better than those from Procedure 1. Moreover, Procedure 3 even by itself (without iterations) takes much more computer time than Procedure 1, since we need to invert a $I+J+K$-dimensional matrix. Worse still, it needs a fair amount of computer space if I and J are large. It might be mentioned in passing that Procedure 3 has a slight formal resemblance to the maximum likelihood procedure. In maximum likelihood, we minimize the log-likelihood function (5.8), which is equivalent to minimizing

$$\sum_{ij} N_{ij}[\log(N_{ij}+\frac{1}{2}) - \log(T_{ij})] + T_{ij}, \tag{6.135}$$

whereas in Procedure 3 we minimize

$$\sum_{ij} N_{ij}[\log(N_{ij}+\frac{1}{2}) - \log(T_{ij})]^2. \tag{6.136}$$

In fact, with iterative reweighting and some updating of the dependent variable at each iteration, it has some resemblance to the GLIM procedure described Section 5.5.5 and the computer time needed might be similar.

6.4.3 Use of OLS

Two other LS procedures merit attention. Both are similar to Procedure 1. In one, which for the sake of consistency we call **Procedure 4**, we would simply use ordinary least squares (OLS), instead of weighted least squares, on the model (6.38). The estimate of $\boldsymbol{\theta}$ is $(X^tX)^{-1}X^t\boldsymbol{y}$. Then the only bias will be due to that described by Lemma 6.1 and will be $\mathcal{O}(T_{\#}^{-2})$, and, if we notice that $\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{e}$, $\boldsymbol{e} = \mathcal{P}\psi$ and $\text{var}[\psi_s] = \breve{T}_s$, we see that the covariance matrix of the estimate of $\boldsymbol{\theta}$ would be

$$(X^tX)^{-1}X^t\mathcal{P}\breve{T}\mathcal{P}X(X^tX)^{-1} = (X^tX)^{-1}X^t\breve{T}X(X^tX)^{-1}, \tag{6.137}$$

since $X^t\mathcal{P} = X^t$ by Corollary 6.2. [Note that here too the estimates of variances of $\hat{\theta}_k$'s would not be the diagonal elements of $s^2(X^tX)^{-1}$, where s is the root mean square of the residuals, as a least squares package program would give us]. This procedure is about as easy to use as Procedure 1. However, the variances are quite different. Under LS4, (6.137) is $\mathcal{O}(n^{-1}T_{\#})$ while the variance of estimates given by Procedure 1 was shown in Lemma 6.7 to be $\mathcal{O}(n^{-1}T_{\#}^{-1})$. Even in some small sample situations (the Skokie data and the 'T/50' data) for which we have computed such variances, variances obtained from (6.137) have been four to six times larger than corresponding variances for Procedure 1 estimates. Besides, this procedure gives undue importance to observations with smaller T_{ij}'s.

It might be noted some of the variance for this procedure and for Procedure 1 come to occur in somewhat different ways. While for Procedure 1, much of it is due to the presence of off-diagonal terms in the covariance matrix $\mathcal{P}\check{T}^{-1}\mathcal{P}$ of $\boldsymbol{y}$, these off-diagonal terms do not have any untoward effect in (6.137). The size of (6.137) is largely due to the unequal sizes of the T_{ij}'s.

If Procedure 3 were to be modified so that we used OLS, some of the difficulties we mentioned earlier would go away but the variances, under a condition akin to LS4, would still be of the order of $\mathcal{O}(n^{-1}T_{\#})$ and undue emphasis would be placed on observations with small T_{ij}'s.

6.4.4 Use of Generalized Inverses

Another procedure would involve using a generalized LS procedure incorporating the generalized inverse $(\mathcal{P}\check{T}^{-1}\mathcal{P})^{-}$ of $\mathrm{Cov}[e] = \mathcal{P}\,\mathrm{Cov}[\psi]\mathcal{P} = \mathcal{P}\check{T}^{-1}\mathcal{P}$. Notice that in Procedure 1, the weights N_{ij} do not take into account the fact that the least squares model (6.38) has correlated errors. The use of the generalized inverse of $\mathcal{P}T^{-1}\mathcal{P}$ would do this. The estimate of $\boldsymbol{\theta}$ would be then

$$[X^t(\mathcal{P}\check{T}^{-1}\mathcal{P})^{-}X]^{-1}X^t(\mathcal{P}\check{T}^{-1}\mathcal{P})^{-}\boldsymbol{y} \tag{6.138}$$

and its covariance matrix would be $[X^t(\mathcal{P}\check{T}^{-1}\mathcal{P})^{-}X]^{-1}$. We call this **Procedure 5**. Clearly, $\check{T} \approx T$ would need to be estimated iteratively, as in the case of Procedure 2.

Because it is not very convenient to compute generalized inverses for very large matrices, we estimated $\boldsymbol{\theta}$ and computed covariances using portions of the 'T_{ij}' data set. This generalized inverse procedure yields variances about a third the size of those from procedure 1. However, it appears to be far from being easy to use.

Procedures we did not explore include modifications of the GLIM procedure incorporating some form of parameter reduction (Section 6.2.1).

Among approaches that we did examine, Procedure 1 appears to be an excellent choice for large n and T_{ij}'s. It is tied with Procedure 4 as the two least computationally demanding procedures, and, except for the difficult-to-use Procedure 5, it appears to provide estimates which are at least as good as those from any other procedure described in this section.

6.5 Small Sample Properties

While large sample properties of LS procedures can be examined analytically, this appears to be impossible when T_{ij}'s are small, particularly when several N_{ij}'s are zeros. Therefore, in this section, Monte Carlo methods are used to investigate small sample properties. By small, in this section, we mean the case when T_{ij}'s are small and n not too large. This situation was left uncovered by the asymptotic discussion of the last section, but is one sometimes encountered in transportation planning.

6.5.1 The Procedures

We shall investigate Procedures 1 and 1A described in Section 6.2.5. Notice that, although when $N_{ij} = 0$, weighting eliminates from consideration the corresponding $z_{ij} - z_{i\bullet} - z_{\bullet j} + z_{\bullet\bullet}$ (or $v_{ij} - v_{i\bullet} - v_{\bullet j} + v_{\bullet\bullet}$), zero-valued N_{ij}'s do enter the analysis through the means $z_{i\bullet}$, $z_{\bullet j}$ etc. When there are only a few zero valued N_{ij}'s, this role is insignificant, but when the vast majority of N_{ij}'s are zeros, the effect can be noticeable. Since in Procedure 3, zero valued N_{ij}'s play no role whatsoever, we decided to examine its performance as well as that of another procedure described below.

We propose as **Procedure 1B,** a two step procedure. The first step is exactly Procedure 1A. After θ_k's have been estimated using this method, we obtain estimated T_{ij}'s — call them $\hat{T}_{ij}$'s — using the DSF procedure (as in Procedure 2 of Section 6.4). We then define

$$z_{ij}^* = \begin{cases} \log(\hat{T}_{ij}) & \text{when } N_{ij} = 0 \\ z_{ij} & \text{otherwise.} \end{cases} \tag{6.139}$$

We use $z_{ij}^* - z_{i\bullet}{}^* - z_{\bullet j}{}^* + z_{\bullet\bullet}^*$ as the dependent variable in the second step with the independent variables and weights remaining as in the last two methods.

Replacing 'troublesome' observations with estimates is not unusual. Indeed, one might be tempted to iterate the second step, using in each iteration, the $\hat{T}_{ij}$'s from the previous one. Exhibit 6.2 shows estimates of θ_1 obtained in each iteration for five sets of simulated data sets with a large number of zero values (Data '$T_{ij}/100$' described in the next section). The 'true' value of θ_1 was .15. Notice that estimates in the second row are the ones provided by Procedure 1B. Thus, at least in this example, additional iterations of the second step makes the estimates worse.

A plausible explanation for this might be the following. Use of Jensen's inequality yields $\mathrm{E}[\log(\hat{T}_{ij})] < \log(T_{ij})$ when $E[\hat{T}_{ij}] = T_{ij}$, and consequently it might be surmised that most $\log[\hat{T}_{ij}]$'s will be underestimates. Using low estimates of T_{ij} will typically cause the estimate of $\hat{T}_{ij}$ used by (6.139) to be even lower in the next step. Another underlying factor is that when Procedure 1B would differ from Procedure 1, $z_{\oplus\oplus}{}^* > z_{\oplus\oplus}$.

We stress here that our primary interest lies with Procedure 1. The performance of the other procedures were examined partly for reasons already mentioned — but also because the procedures appeared promising when the simulations were conducted.

Iter. No.	Data No. 1	Data No. 2	Data No. 3	Data No.4	Data No. 5
1	-.12	-.13	-.11	-.12	-.12
2	-.16	-.14	-.14	-.15	-.15
3	-.20	-.15	-.17	-.18	-.16
4	-.23	-.16	-.18	-.19	-.17

Exhibit 6.2: Illustration of Consequence of Adding Further Iterations to Procedure 1B

6.5.2 The Simulations

Since, the LS procedures are essentially the same for any value of K, in order to save computer time we mainly worked with one independent variable. This also avoided any multicollinearity problems and yet demonstrated the performance of the procedures. The simulated data sets based on Skokie data, described in Section 5.6.1 and used in many of the simulations of the last chapter, were also used here. We set $c_{ij}^{(1)}$ equal to the Skokie travel times. Rounding estimates obtained from the Skokie data, we set $\theta_1 = -\cdot 15$. From these we obtained $F(\boldsymbol{c}_{ij}) = \exp[\theta_1 c_{ij}{}^{(1)}]$. Applying the DSF procedure with these $F(\boldsymbol{c}_{ij})$'s and origin and destination totals (i.e., $N_{i\oplus}$'s and $N_{\oplus j}$'s) from the Skokie data, we obtained T_{ij}'s. We applied a Poisson random number generator to these T_{ij}'s to obtain sets of simulated N_{ij}'s.

As in Chapter 5, when we needed particularly small sample size data, we divided the T_{ij}'s obtained above by some number, usually 50 or 100, before applying the Poisson generator. Also as in Chapter 5, we call these simulated data sets 'T_{ij}', '$T_{ij}/50$', '$T_{ij}/100$' data sets. The distribution of sizes of N_{ij}'s are just about the same as in Exhibit 5.5.

To investigate some special situations, other types of simulated data were generated. One is where some observations are small and others large. Such situations occur when transportation costs are relatively high (e.g., in the case of freight movements involving very bulky commodities). This case also strains the condition on the boundedness of $T^{\#}/T_{\#}$ that we have imposed in some of our derivations in Section 6.3. To generate such data, the Skokie $c_{ij}^{(1)}$'s were altered as follows:

$$c_{ij}^{**} = \begin{cases} c_{ij}^{(1)}/3.5 & \text{when } c_{ij}^{(1)} < 10 \\ 5.6(c_{ij}^{(1)} - 10) + 3 & \text{when } 10 \le c_{ij}^{(1)} \le 15 \\ 2c_{ij}^{(1)} & \text{when } c_{ij}^{(1)} > 15. \end{cases} \tag{6.140}$$

This yielded a data set for which roughly a third of T_{ij}'s were less than 5 and a third greater than 30. There were several zero valued N_{ij}'s and a few N_{ij}'s were over a thousand.

Procedure	Data 'T_{ij}'	Data '$T_{ij}/50$'	Data $T_{ij}/100$'
3	-.143	-.062	-.039
1	-.154	-.121	-.115
1A	-.151	-.123	-.112
1B	-.148	-.146	-.159

Exhibit 6.3: Means of estimates of θ_1 from simulated data sets.

Usually 500 iterations were used. As the standard errors of θ_1 in Exhibit 6.4 show, this number of iterations makes the standard errors of mean (taken over the different simulations) of θ_1 (not shown, but which are approximately 1/22 times the standard deviations) quite small.

6.5.3 Results from Simulations

Bias

Exhibit 6.3 presents a summary of our investigations of bias. For each method and each type of data, the mean of $\hat{\theta}_1$ taken over all simulations is presented.

A quick comparison with standard errors will show that the hypothesis of 'no bias' can be rejected for all entries in Exhibit 6.3. Even the entry .151 which is close to .15 is about 8 standard errors away. However, for most practical purposes, all the methods, with the possible exception of Procedure 3, do fairly well for the 'T_{ij}' data sets. Therefore, for typical urban travel data using 100 per cent samples the last three of the procedures in the exhibit should yield estimates with acceptably low bias levels. However, when we consider smaller than 100 per cent samples, differences in performance are more obvious. Procedure 3 does extremely poorly, while Procedure 1B does very well indeed, although even with this procedure the level of bias starts to deteriorate by the time we reach the 1 per cent sample (but recall, from Section 5.6.1, that in this case about 80 per cent of the origin-destination pairs have no trips and almost 95 per cent had one or less trips). Methods 1 and 1A perform similarly. The relatively poor performance of Procedure 3 can perhaps be attributed to factors that were discussed in Section 6.4.

In the case where some observations are very large and others small the mean of the estimate of θ_1 using Procedure 1 over 50 simulations, was found to be -.150 and using Procedures 1A and 1B, it was -.149 and -.146, respectively. As in the last set of runs, the 'true' value of θ_1 was .15.

Procedure	Data'T_{ij}'	Data '$T_{ij}/50$'	Data $T_{ij}/100$'
3	.0020 (.0020)	.0129 (..0093)	.0140 (.0113)
1	.0027 (.0018)	.0092 (.0119)	.0163 (.0164)
1A	.0027 (.0018)	.0098 (.0121)	.0166 (.0162)
1B	.0026 (.0017)	.0146 (.0126)	.0212 (.0209)
Asymp. s.e.'s	.0029	.0205	.0290
ML s.e.'s	.0021	.0151	.0213

Exhibit 6.4: Standard Errors of $\hat{\theta}_1$'s

VARIANCES

Exhibit 6.4 shows standard errors obtained for the four methods and the three data sets. The unparenthetic numbers given in the first four rows are 'sample' standard deviations computed using as 'observations' the simulated $\hat{\theta}_1$'s. The numbers within parentheses are the means of the corresponding standard errors computed by the least squares package. The fact that these numbers are somewhat close when Procedure 3 is applied to the 'T_{ij}' data set should come as no surprise. It perhaps suggests that our earlier assertion that simply replacing T_{ij} by N_{ij} as weights has little effect on covariances when T_{ij}'s are large also extends to smaller samples. The fact that the match is poor for the other methods should also have been expected, since they mainly illustrate what happens to variances when we have 'correlated errors' in a least squares exercise (in fact, they illustrate what happens even when the correlations are very small). The good match for the $T_{ij}/100$ data set is noteworthy but the reasons (if any) underlying the phenomenon are not clear — but could have to do with the sparsity of the data.

The second last line in the table gives estimates of standard errors using formula (6.127) presented in Section 6.3.5. As we had stated there, these asymptotic standard errors are, in fact, too asymptotic. Only in the case of the 'T_{ij}' data set does the estimate of variance come even close to being a reasonable approximation. We should point out that these standard error estimates relate only to procedures 1 and 1A. We also should mention that (6.128) gave rather poor estimates here, possibly because of the large number of small T_{ij}'s.

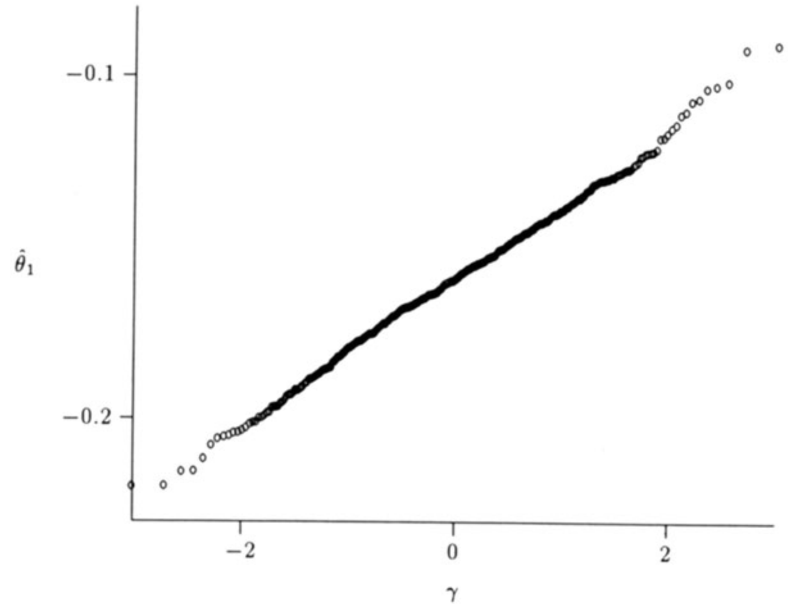

Exhibit 6.5: Normal Plot of Estimated θ_1's for $T_{ij}/50$ Data

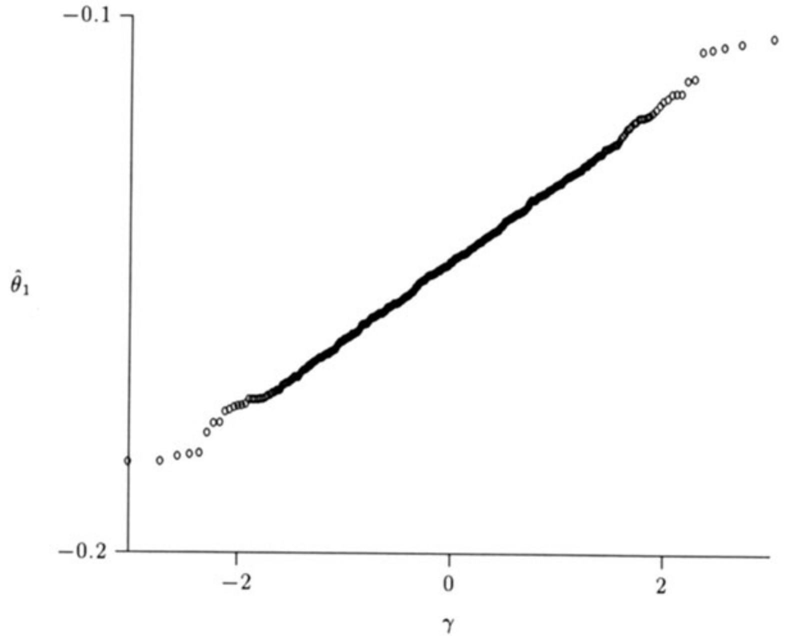

Exhibit 6.6: Normal Plot of Estimated θ_1's for $T_{ij}/100$ Data

Data	Mean of θ_1	s.e. of θ_1
(1)	-.151	.0038
(2)	-.152	.0024

Exhibit 6.7: Means and standard errors of θ_1 for non-Poisson data.

The last line of the table gives the standard errors for maximum likelihood estimates computed as discussed in Section 5.226. For reasons not entirely clear to us, these are very close to the simulated standard errors, particularly for the '$T_{ij}/50$' and the '$T_{ij}/100$' data sets. Note that these ML standard errors are easier to compute than the corresponding maximum likelihood estimates. Although there is perhaps not enough evidence to make a recommendation, it would appear that for data sets comparable to the 'T_{ij}' data set, the asymptotic standard error may be used as a crude estimate of the standard error, and for data sets with smaller T_{ij}'s and procedure 1B the maximum likelihood standard errors could be investigated further, at least for Procedure 1B. Recall that, for most practical purposes, fairly rough estimates of standard errors usually suffice.

For significance testing, it is probably almost as valuable to know that the distributions of estimated θ_1's are close to the normal distribution. This was verified using the usual normal plots (see page 456 for a brief description of such plots), two of which are shown in Exhibit 6.5 and 6.6. All plots were very straight, showing that the estimates have a distribution very close to the normal.

ROBUSTNESS

For the same reasons mentioned in Section 5.9.2 we need to see what happens under realistic departures from the Poisson distribution. For LS procedures we considered only two types of departures — types 1 and 2 of Section 5.9.2. Exhibit 6.7 shows the results based on 10 simulations for each of the two types of data described above, using in each case Procedure 1B. The means of θ_1 from Procedure 1 using 25 simulations were .148 and .149. Therefore, both methods appear to be quite robust.

6.5.4 CONCLUSIONS

One conclusion that we draw from the simulations is that for data sets like the Skokie ('T_{ij}') data set, Procedure 1 (or 1A) work reasonably well. Bias is small and estimates of standard errors using formulæ recommended are not too poor. In addition, estimates appear to be normally distributed and robust. Thus, it would appear that, from a highly practical viewpoint, the asymptotic results established in Theorem 6.1 extend even to moderate-sized data sets.

Beyond that, it is somewhat unclear what conclusions one might draw from the Monte Carlo exercises. It appears that for a data set like the Skokie data set, Procedures 1, 1A or 1B give reasonably good results and the asymptotic variance formula gives acceptable estimates. For much smaller sample sizes (i.e., smaller values of T_{ij}'s) the results are too mixed for a recommendation. However, LS estimates from smaller sample sizes can be useful as inputs into an ML procedure.

6.6 Non-linear Least Squares

For the sake of completeness, we need to at least briefly consider nonlinear least squares (NLS). In the gravity model context, such an approach to estimation has been taken by several authors, including Cesario (e.g., Cesario, 1973, 1974a, 1975), Kirby (1974), Kirby and Leese (1978) and Openshaw (1976, 1979). Most statistical packages (e.g., SAS, SYSTAT) include NLS programs (see, also, Sen and Srivastava, 1990, Appendix C). However, because of the large number of parameters that are frequently involved, their use for gravity model estimation is not very convenient, and, as in the case of ML procedures, gravity model specific methods are desirable. As noted below, it would appear that such methods can be constructed fairly easily by modifying ML methods.

In this section, we shall use the same gravity model that we have been using in the rest of this chapter and in Chapter 5, i.e., we assume that the flows N_{ij} are Poisson with expectations given by (6.2) and (6.3). In NLS, estimates of $A(i)$'s, $B(j)$'s and θ_k's are obtained by minimizing the weighted sum of squares

$$\sum_{ij} W_{ij}(N_{ij} - T_{ij})^2$$

where W_{ij} are weights and $T_{ij} = A(i)B(j)\exp[\sum_{k=1}^{K} \theta_k c_{ij}{}^{(k)}]$. In most gravity model uses of NLS, W_{ij} has been set equal to one. However, since N_{ij}'s are assumed to have a Poisson distribution and therefore, $\mathrm{E}[(N_{ij} - T_{ij})^2] = T_{ij}$, it is preferable and appropriate to set $W_{ij} = T_{ij}^{-1}$ and minimize

$$\chi^2 = \sum_{ij}[N_{ij} - T_{ij}]^2/T_{ij}. \tag{6.141}$$

For obvious reasons, such an approach has been called a minimum chi-square (MCS) procedure. An alternative to (6.141) consists of setting $W_{ij} = N_{ij}^{-1}$ and using

$$\chi^2 = \sum_{ij}[N_{ij} - T_{ij}]^2/N_{ij}. \tag{6.142}$$

The resultant approach which has been called a modified minimum chi-square (MMCS) procedure is only usable when $N_{ij} \neq 0$ for all i and j. As a near trivial application of Lemma 5.5, it may be shown that either method is asymptotically equivalent to the maximum likelihood procedure (see also Amemiya, 1976, Rao, 1973, p. 352, Kendall and Stuart, 1967, p. 93, Bishop, *et al.*, 1975, p. 514) and, indeed, MCS estimates are frequently numerically nearly identical to ML estimates.

Let $F_{ij} = \exp[\sum_{k=1}^{K} \theta_k c_{ij}{}^{(k)}]$. Then

$$
\begin{aligned}
\frac{\partial \chi^2}{\partial a(i)} &= \sum_{ij} \frac{\partial}{\partial a(i)} \left[\frac{N_{ij}}{A(i)^{1/2} B(j)^{1/2} F_{ij}^{1/2}} - A(i)^{1/2} B(j)^{1/2} F_{ij}^{1/2} \right]^2 \\
&= -\sum_{j} \left[\frac{N_{ij}}{A(i)^{1/2} B(j)^{1/2} F_{ij}^{1/2}} - A(i)^{1/2} B(j)^{1/2} F_{ij}^{1/2} \right] \\
&\qquad \cdot \left[\frac{N_{ij}}{A(i)^{3/2} B(j)^{1/2} F_{ij}^{1/2}} + A(i)^{-1/2} B(j)^{1/2} F_{ij}^{1/2} \right] \frac{\partial A(i)}{\partial a(i)} \qquad (6.143) \\
&= -\sum_{j} \left[\frac{N_{ij}}{T_{ij}^{1/2}} - T_{ij}^{1/2} \right] \left[\frac{N_{ij}}{T_{ij}^{1/2}} + T_{ij}^{1/2} \right] A(i)^{-1} A(i) \\
&= -\sum_{j} \left[\frac{N_{ij}^2}{T_{ij}} - T_{ij} \right]
\end{aligned}
$$

since $\partial A(i)/\partial a(i) = A(i)$. Recall that we have been using the notation that $a(i) = \log[A(i)]$ and $b(j) = \log[B(j)]$. It may also be shown that

$$
\frac{\partial \chi^2}{\partial b(j)} = -\sum_{i} \left[\frac{N_{ij}^2}{T_{ij}} - T_{ij} \right] \qquad (6.144)
$$

and

$$
\begin{aligned}
\frac{\partial \chi^2}{\partial \theta_k} &= -\sum_{ij} \left[\frac{N_{ij}^2}{T_{ij}} - T_{ij} \right] F_{ij}^{-1} \frac{\partial F_{ij}}{\partial \theta_k} \\
&= -\sum_{ij} c_{ij}{}^{(k)} \left[\frac{N_{ij}^2}{T_{ij}} - T_{ij} \right].
\end{aligned} \qquad (6.145)
$$

From (6.143), (6.144) and (6.145), one gets, as necessary conditions for a minimum of (6.141):

$$
\begin{aligned}
&\sum_{i=1}^{I} N_{ij}^2 = \sum_{i=1}^{I} T_{ij}^2, \quad \sum_{j=1}^{J} N_{ij}^2 = \sum_{j=1}^{J} T_{ij}^2 \\
\text{and} \quad &\sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij}{}^{(k)} N_{ij}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij}{}^{(k)} T_{ij}^2.
\end{aligned} \qquad (6.146)
$$

One method of obtaining MCS estimates consists of solving these equations. Notice that the equations are very similar to (5.12) and (5.13) — only N_{ij}'s and T_{ij}'s are replaced with N_{ij}^2's and T_{ij}^2's. Although, we have not tried them out, we conjecture that satisfactory solution algorithms can be obtained by adapting ML algorithms [replace N_{ij}'s in the ML algorithms by N_{ij}^2's and obtain estimates for $2\boldsymbol{\theta}$].

The following theorem corresponds to Theorem 5.1 of last chapter.

Theorem 6.2 *Under Condition ML1 of Theorem 5.1, there exists a unique value $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^t$ of $\boldsymbol{\theta}$, for which positive numbers $A(1), \ldots, A(I)$ and $B(1), \ldots, B(J)$ can be found which minimize (6.141).*

PROOF: Consider the set of values of

$$(A(1), \ldots, A(I), B(1), \ldots, B(J), \theta_1, \ldots, \theta_k)^t$$

such that $A(i) \geq 0$ and $B(j) \geq 0$ for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$ and $\chi^2 \leq Q$ for a large enough Q. This set is compact; hence, the minimum of χ^2 occurs at point(s) in this set and, as in the proof of Lemma 5.3, it may be seen that such point(s) are interior to the set. Thus, at these points (6.146) holds.

Now, notice that [recalling our previous notation $a(i) = \log[A(i)]$ and $b(j) = \log[B(j)]$]

$$\begin{aligned}
\frac{\partial^2 \chi^2}{\partial b(j) \partial a(i)} &= -\sum_j \frac{\partial}{\partial b(j)} \left[\frac{N_{ij}^2}{A(i)B(j)F_{ij}} - A(i)B(j)F_{ij} \right] \\
&= - \left[-\frac{N_{ij}^2}{A(i)B(j)^2 F_{ij}} - A(i)F_{ij} \right] \frac{\partial B(j)}{\partial b(j)} \\
&= \left[\frac{N_{ij}^2}{A(i)B(j)F_{ij}} + A(i)B(j)F_{ij} \right] B(j)^{-1} B(j) \\
&= \left[\frac{N_{ij}^2}{A(i)B(j)F_{ij}} + A(i)B(j)F_{ij} \right] = \nu_{ij}, \text{ say.}
\end{aligned} \tag{6.147}$$

Since $T_{ij} = A(i)B(j)F_{ij} > 0$, it follows that $\nu_{ij} > 0$.

As for (6.147), it may be shown that

$$\begin{aligned}
&\frac{\partial^2 \chi^2}{\partial^2 a(i)} = \nu_{i\oplus}, \quad \frac{\partial^2 \chi^2}{\partial^2 b(j)} = \nu_{\oplus j} \\
&\frac{\partial^2 \chi^2}{\partial \theta_k \partial a(i)} = \sum_j \nu_{ij} c_{ij}{}^{(k)}, \quad \frac{\partial^2 \chi^2}{\partial \theta_k \partial b(j)} = \sum_i \nu_{ij} c_{ij}{}^{(k)}, \\
\text{and} \quad &\frac{\partial^2 \chi^2}{\partial \theta_k \partial \theta_\ell} = \sum_{ij} \nu_{ij} c_{ij}{}^{(k)} c_{ij}{}^{(\ell)}.
\end{aligned} \tag{6.148}$$

Recall that the replacement of a subscript by a $\oplus$ denotes that we are adding over that subscript — e.g., $\nu_{i\oplus} = \sum_{j=1}^{J} \nu_{ij}$.

If, as in the case of ML estimation, one $A(i)$ or $B(j)$ is kept fixed, it may be verified (by matrix multiplication) from (6.148) that the Hessian matrix of (6.141) is $M_{(2)}^t \mathcal{D} M_{(2)}$, where

$$\mathcal{D} = \operatorname{diag}(\nu_{11}, \ldots, \nu_{1J}, \ldots\ldots, \nu_{I1}, \ldots, \nu_{IJ})$$

is a diagonal matrix with, from (6.147), positive elements and $M_{(2)}$ is as in Chapter 5 — see (5.215). Since, under Condition ML1, matrix $M_{(2)}$ is of full rank, $M_{(2)}^t \mathcal{D} M_{(2)}$ is seen to be positive definite. Hence, (6.141) is convex. The theorem follows. □

A similar result for MMCS estimates is easily shown.

6.7 Notes and Concluding Remarks

As in the last chapter, we take this final opportunity to relate the concepts and results of the chapter to the literature and to emphasize the principal findings. Since, the immediate ancestry of the results in the chapter have been mentioned in the course of presentation of the results, only broader issues of general setting are addressed. Also some points are made below which did not conveniently fit into earlier discussions.

Section 6.1

In this chapter we focused mainly on estimation methods and did not pay much attention to model diagnostics or the assessment of fit. Since, diagnostic procedures for least squares models are well described in books on regression, and many of them were discussed in Chapter 5, there seemed to be little point in discussing them further. However, we stress again that one of the principal reasons for doing least squares analysis is the opportunity of taking advantage of such techniques. It is also important to note that those diagnostic procedures that include variances or covariances of estimates might need to be appropriately modified.

Since the model being considered is the same as in Chapter 5, the discussion of model fit also does not need to be repeated. In our own previous work we have relied on the chi-square statistic. We also find the R^2 and s^2 (i.e., the sum of squares of residuals divided by degrees of freedom) to be useful. Note that since we are weighting by approximately the reciprocal of the variance of the dependent variable, s^2 for a well fitting model should be close to one, although for reasons given in Section 5.9, 'close' should be interpreted broadly.

Section 6.2

Outside of urban transportation planning, least squares methods are possibly the most common method used for gravity model parameter estimation. Just as least squares preceded the development of most of the theoretical bases of present-day statistics, in the gravity model literature, least squares methods preceded most of the theoretical investigations of the model. Indeed, the formulations of some of the earlier models were probably motivated by convenience of use of LS procedures. Examples include several of the models in the Introduction; e.g., (4), where taking logarithms of both sides yielded a form suitable for LS estimation. Since there are very few parameters involved, no parameter reduction is necessary. However, the difficulties associated with a logarithmic transformation remain. Weighting was seldom performed, with the result that comments like 'it is the large traffic generators for which the gravity model trip interchanges must be adjusted,' implying the need some adjustment after estimation, are seen in the literature [this quotation from (FHWA, 1973, p. V-35) refers to a least squares procedure other than the one involving (4)]. Authors who wrote such models out in full often wrote the error term multiplicatively — i.e., $N_{ij} = T_{ij}\epsilon_{ij}$ — and made the further assumption that $\log[\epsilon_{ij}]$ had the same variance for all i and j. While this was convenient for LS analysis, since it justified the use of ordinary least squares, it does not seem to be realistic. The model is not Poisson, nor is it Poisson-like [i.e., the variance is not proportional to the mean — indeed, if $\log[\epsilon_{ij}]$ has constant variance, the implication is that the variance of N_{ij} be roughly proportional to the square of the mean, a phenomenon seldom encountered in practice].

Procedure 1 was first presented in Sen and Sööt (1981), and was motivated by Tukey's 'two-way analysis' [also called 'elementary analysis;' see Tukey (1970, 1977)]. Computationally, Procedure 1 appears to be the tied with the Gray and Sen (1983) Procedure as the simplest of any gravity model estimation procedure. However, it is not clear whether its properties are any better than the Gray and Sen (1983) Procedure. This latter procedure is based on the logarithm of the odds-ratio. Both the odds-ratio and its logarithm have been widely studied and used in the biostatistics and the contingency table literature (see Bishop, *et al.*, 1975).

Section 6.3

The results in the section are presented for the first time. While asymptotic properties of least squares estimates have been examined before [e.g., in Theil, 1971, Judge *et al.*, 1985], the problem considered here has several elements that do not appear to have been examined. First, asymptotic here meant either $n \to \infty$ or $T_{\#} \to \infty$. Second, we had a correlated error structure with a singular covariance matrix. Finally, we weighted by random variables which were observations, an approach which in itself does not appear to have been examined before. The combination of these factors makes this a unique problem. Indeed, it is the simultaneous presence of the last two factors that made the bias acceptably low!

SECTION 6.4

This discussion is also new. It might be pointed out the simplification of GLIM by the use of some parameter reduction technique like the one used in Procedure 1, is perhaps worthy of further investigation.

The hand computation of Procedure 1 estimates is strongly influenced by Tukey (1970, 1977).

SECTION 6.5

Much of the work presented in this section is taken from an unpublished paper by Sen and Matuszewski (1985).

SECTION 6.6

While non-linear least squares has been suggested for gravity models before, the discussion given in this section, particularly the results, are new.

6.7.1 CONCLUSIONS

As a cheaper (in terms of computer time) alternative to the maximum likelihood procedures, the best of the linear least squares techniques perform very well. However, a number of caveats are in order.

Some linear least squares procedures can have large appetites for computer time and/or space. Some can give severely biased results. Fortunately, one of the easiest of the procedures also turns out to be an excellent performer.

Procedure 1 gives excellent estimates both when samples are large and, as the simulations show, when sample sizes are fairly small. However, when sample sizes were particularly small, modifications, which we have called Procedures 1A and 1B, yielded better results. Indeed the latter performed not too poorly even when our samples were based on about 250 trips (for the '$T_{ij}/100$' data sets)! But we remain sufficiently unclear of the performance for these very small sample sizes that at the present time we do not recommend the procedures presented in this chapter for sample sizes much smaller than that in the Skokie data.

Procedure 1 is simple enough that it can even be used by hand in moderate sized practical application (see end of Section 6.3.5). However, if a standard package is used to compute estimates of θ_k's, the standard errors will not be correct. Instead methods like those given in Section 6.3.5 need to be used. Moreover, Procedure 1 does not have anything close to the minimum variance property of maximum likelihood procedures.

However, linear least squares packages have highly developed diagnostic tools. This recommends the use of Procedure 1 for exploratory use followed by ML methods. Moreover, savings in computer time might occur if linear least squares methods are used to obtain initial values for maximum likelihood procedures.

As far as NLS methods are concerned, we conjecture that, in all respects, their performance would be somewhat similar to those of ML procedures. We have already conjectured that the MCS algorithms could be obtained by adapting ML algorithms and would have similar computer time and space requirements. MCS and ML estimates are asymptotically equivalent. We also conjecture that, for reasons similar to those for ML estimates, small sample MCS estimates are quite good. However, there is a tie-breaker: the already mentioned criterion of second order efficiency (Rao, 1973, p. 352) which favors maximum likelihood.

Thus, the overall recommendation is for the use of maximum likelihood methods and, in particular, for the modified scoring procedure. However, where computational ease is particularly important, as, for example, when computing resources are scarce or diagnostic procedures need to be run or initial rough estimates need to be found, LS procedures, more specifically, Procedure 1, could have a role to play. All in all, the 'tool-box' of methods for statistical analysis using the gravity model, is rather complete. The methods, for the most part, are quick and convenient.

Appendix: Skokie Data

The Skokie data-set, which underlies most of the Monte Carlo runs and many of the numerical examples in the book, is given in this appendix. As mentioned earlier, we are grateful to Prof. Foerster of the School of Urban Planning and Policy, University of Illinois at Chicago, for making this data available to us. After deleting rows and columns consisting entirely of zeros, the original 42×42 origin-destination table became one with 40 origins and 18 destinations. This shortened table is presented as Exhibit A1. Quite obviously, the rows correspond to origins and the columns to destinations. The origin and destination zones are not numbered, so that all numbers shown are trips. Exhibit A2 gives travel times and Exhibit A3 gives distances. The latter table was constructed by Zbigniew Matuszewsky from a map. The distances are shortest road (as opposed to airline) distances. In order to reduce multicollinearity a permuted version of these distances was also used. These are presented in Exhibit A4.

1	1	0	24	8	3	0	0	1	2	0	3	0	4	4	1	2	9
3	13	2	94	34	32	4	1	7	11	1	16	1	24	28	8	17	68
2	23	2	55	21	59	5	1	8	8	1	12	1	20	27	8	22	69
2	25	6	66	25	65	15	3	16	12	2	18	2	33	56	16	50	99
2	14	11	50	18	37	26	4	27	9	2	13	3	26	45	13	63	63
1	8	5	21	13	21	13	8	50	6	2	9	3	16	39	11	119	43
5	9	1	247	61	23	3	0	3	15	1	22	1	40	35	10	7	57
8	15	2	276	87	40	5	1	6	22	1	33	1	51	51	14	21	102
7	29	4	218	79	76	9	3	16	25	2	38	2	55	66	18	41	157
5	59	6	143	53	152	14	3	21	21	2	31	2	53	70	20	58	176
5	62	15	162	61	159	36	7	40	30	4	45	5	82	138	39	122	241
4	35	26	123	44	89	64	11	67	22	5	33	6	65	109	31	155	155
1	8	5	21	13	20	13	8	48	6	2	9	3	16	37	11	114	41
0	1	0	1	1	1	1	1	6	0	0	1	0	1	3	1	21	2
0	3	2	7	3	9	4	6	34	2	1	3	1	6	17	5	129	15
2	7	1	78	25	18	3	1	4	31	1	47	1	63	48	14	14	53
2	10	2	86	27	26	5	1	8	27	1	40	1	56	66	19	28	77
6	90	10	208	73	232	25	7	45	53	5	79	7	175	262	74	135	211
1	13	1	30	11	34	4	1	7	8	1	11	1	26	38	11	20	31
5	30	12	152	53	78	29	11	69	31	12	47	16	116	239	67	208	162
2	13	6	117	21	34	14	7	45	14	8	21	11	55	107	30	158	34
0	1	1	1	1	2	1	1	4	1	1	1	1	3	8	2	15	5
0	5	3	5	5	12	7	4	22	4	4	6	6	16	47	13	87	29
3	10	2	118	38	27	4	1	6	47	1	71	1	95	72	20	22	79
4	19	4	160	51	49	9	3	16	50	2	74	2	104	122	34	51	142
2	35	4	80	28	89	9	3	17	20	2	30	3	67	100	28	52	81
2	26	3	61	21	68	7	2	13	15	2	23	2	51	76	21	39	61
3	21	8	106	37	54	20	8	48	22	8	33	11	81	166	47	145	113
2	18	8	155	28	45	19	10	60	19	11	28	15	73	142	40	210	46
1	8	5	9	9	22	13	6	38	7	7	11	10	28	83	23	153	52
1	3	1	37	10	8	1	0	3	11	0	17	1	107	40	11	6	21
1	5	1	46	14	12	3	1	5	13	1	20	1	78	109	31	18	33
2	8	2	62	20	20	5	2	12	16	2	24	2	75	401	113	34	53
3	16	4	100	33	41	11	4	25	24	4	37	5	120	491	138	75	100
2	12	4	58	26	31	10	5	28	19	4	29	5	95	213	60	92	78
2	17	6	37	24	43	15	7	45	20	8	30	11	85	191	54	158	96
0	2	1	17	6	6	1	1	3	5	0	7	1	21	113	32	10	15
2	12	3	76	25	31	8	3	19	18	3	28	4	90	370	104	56	76
3	14	5	66	29	35	12	5	31	22	4	33	6	108	241	68	103	89
1	11	4	25	16	29	10	5	30	13	5	20	7	56	128	36	105	64

Exhibit A1: Trip table for Skokie

1	9	14	5	8	12	14	22	22	10	19	10	20	12	16	17	20	14
6	6	15	8	8	11	11	21	21	12	16	12	18	14	14	15	17	14
9	4	12	11	11	8	8	18	18	15	13	14	15	17	11	12	14	12
12	6	9	13	12	7	5	15	15	15	10	15	12	17	10	11	11	11
15	12	5	15	16	12	7	12	11	17	10	18	12	19	14	15	12	18
17	11	5	16	16	11	6	10	9	19	9	18	11	20	13	14	9	17
5	11	13	1	7	11	13	22	21	10	18	10	20	12	16	17	19	14
8	11	15	7	1	10	13	22	22	6	16	6	17	8	12	13	17	12
10	13	16	9	5	10	12	21	20	7	15	7	15	9	11	12	16	10
12	8	12	11	10	3	8	18	17	14	12	13	13	15	8	9	13	10
13	8	8	13	12	6	4	14	13	14	9	14	10	16	9	10	10	13
14	8	7	13	13	8	3	13	13	16	8	15	10	17	10	11	9	14
19	14	8	19	17	12	9	11	10	19	7	18	9	21	13	14	7	18
16	16	12	15	16	18	13	3	3	18	12	18	13	20	18	19	9	22
15	16	11	15	16	17	13	3	2	17	11	17	12	19	17	18	9	22
10	15	17	10	6	14	16	24	24	2	18	6	18	8	13	14	20	14
11	13	16	11	6	10	12	21	20	6	15	6	15	8	11	12	16	10
14	8	12	13	9	6	8	16	16	12	10	11	9	12	5	6	11	10
14	8	11	13	9	6	7	16	16	12	10	11	10	12	5	6	11	10
16	10	11	15	12	8	7	15	15	14	9	13	9	15	8	9	11	13
19	13	10	18	16	12	8	12	11	18	2	18	6	19	11	12	7	17
20	17	15	19	20	16	12	7	7	21	9	20	9	21	14	15	9	21
20	17	15	19	20	16	12	7	7	21	9	20	9	21	14	15	9	21
10	14	17	10	6	13	15	24	23	6	18	2	15	5	11	12	19	13
13	15	18	13	9	12	14	22	22	9	16	6	13	7	8	9	18	14
15	10	13	15	11	7	9	18	17	13	12	11	9	12	5	6	13	12
17	12	12	17	13	9	8	17	16	16	11	13	8	14	6	7	12	14
18	12	9	17	14	10	7	14	14	17	8	15	6	16	8	9	9	15
20	15	12	20	17	13	10	13	12	18	6	15	1	16	9	10	10	17
20	15	12	20	18	14	10	10	9	20	6	19	9	21	14	15	9	19
12	17	19	12	8	15	17	25	25	8	19	5	16	2	11	11	21	16
13	15	18	13	9	12	14	22	22	9	16	6	13	7	8	9	18	14
16	11	14	16	12	8	10	18	17	13	11	11	9	11	4	4	14	13
18	13	13	18	14	10	10	16	16	15	10	12	7	13	6	7	13	15
20	14	12	20	16	12	9	15	14	17	9	15	6	15	8	9	12	17
20	15	12	20	17	13	10	13	12	18	6	15	4	16	9	10	10	17
17	12	15	17	13	9	11	19	18	14	12	12	10	11	4	1	15	14
19	14	14	19	15	11	11	17	17	16	11	13	8	13	7	6	14	16
21	15	13	21	17	14	10	16	15	19	9	16	7	15	9	8	13	18
23	17	15	22	18	15	12	15	15	19	9	17	8	15	9	8	12	19

Exhibit A2: Travel Times in Minutes for Skokie Data

05	16	26	14	10	18	30	38	44	16	40	21	45	27	35	40	42	22
10	11	21	17	13	13	25	33	39	17	35	22	40	28	30	35	37	17
16	05	15	23	19	07	19	27	33	23	29	28	34	34	24	29	31	11
21	10	10	28	24	12	14	22	28	28	24	33	29	39	29	34	26	16
26	15	05	33	29	17	09	17	23	33	19	38	24	44	34	39	21	21
31	20	10	38	34	22	14	12	18	38	14	43	19	49	39	44	16	26
14	23	33	05	11	21	31	41	45	17	41	22	46	28	36	41	43	23
10	19	29	11	05	17	25	37	39	11	35	16	40	22	30	35	37	17
15	14	24	16	10	12	20	32	34	14	30	19	35	25	25	30	32	12
18	07	17	21	17	05	17	25	31	21	27	26	32	32	22	27	29	09
25	14	14	26	20	12	10	22	24	24	20	29	25	35	25	30	22	12
30	19	09	31	25	17	05	17	19	29	15	34	20	40	30	35	17	17
35	24	14	36	30	22	10	12	14	34	10	39	15	45	35	40	12	22
38	27	17	41	37	25	17	05	11	41	17	46	22	52	42	47	09	29
44	33	23	45	39	31	19	11	05	43	19	48	24	54	44	49	11	31
16	23	33	17	11	21	29	41	43	05	29	10	34	16	24	29	37	17
20	19	29	21	15	17	25	37	39	09	25	14	30	20	20	25	33	13
25	14	24	26	20	12	20	32	34	14	20	19	25	25	15	20	28	08
30	19	19	31	25	17	15	27	29	19	15	24	20	30	20	25	23	13
35	24	14	36	30	22	10	22	24	24	10	29	15	35	25	30	18	18
40	29	19	41	35	27	15	17	19	29	05	34	10	40	30	35	13	23
45	34	24	46	40	32	20	12	14	34	10	39	15	45	35	40	08	28
49	38	28	50	44	36	24	16	10	38	14	43	19	49	39	44	12	32
21	28	38	22	16	26	34	46	48	10	34	05	29	11	19	24	42	22
25	24	34	26	20	22	30	42	44	14	30	09	25	15	15	20	38	18
30	19	29	31	25	17	25	37	39	19	25	14	20	20	10	15	33	13
35	24	24	36	30	22	20	32	34	24	20	19	15	25	15	20	28	18
40	29	19	41	35	27	15	27	29	29	15	24	10	30	20	25	23	23
45	34	24	46	40	32	20	22	24	34	10	29	05	35	25	30	18	28
50	39	29	51	45	37	25	17	19	39	15	34	10	40	30	35	13	33
27	34	44	28	22	32	40	52	54	16	40	11	35	05	15	18	48	28
30	29	39	31	25	27	35	47	49	19	35	14	30	10	10	15	43	23
35	24	34	36	30	22	30	42	44	24	30	19	25	15	05	10	38	18
40	29	29	41	35	27	25	37	39	29	25	24	20	20	10	15	33	23
45	34	24	46	40	32	20	32	34	34	20	29	15	25	15	20	28	28
50	39	29	51	45	37	25	27	29	39	15	34	10	30	20	25	23	33
40	29	39	41	35	27	35	47	49	29	35	24	30	18	10	05	43	23
45	34	34	46	40	32	30	42	44	34	30	29	25	23	15	10	38	28
50	39	29	51	45	37	25	37	39	39	25	34	20	28	20	15	33	33
55	44	34	56	50	42	30	32	34	44	20	39	15	33	25	20	28	38

Exhibit A3: Distances in tenths of a mile for Skokie Data

05	15	15	29	44	15	34	15	18	35	25	20	24	25	10	24	31	20
15	15	10	19	24	20	20	15	27	29	20	05	20	25	34	24	20	24
21	30	20	15	15	05	25	29	44	15	44	25	18	35	25	30	34	25
10	14	41	30	15	15	20	29	24	20	20	15	37	39	20	15	10	35
44	34	30	24	21	30	20	25	15	25	05	39	54	25	24	25	28	45
35	10	24	15	20	34	21	20	25	15	10	29	34	30	20	25	17	19
10	15	30	15	24	14	10	34	31	40	10	15	29	29	39	05	20	19
30	19	16	11	09	34	20	49	24	28	33	24	39	39	34	15	18	14
44	19	37	43	44	29	34	29	38	48	44	10	13	14	44	29	44	44
54	20	05	34	45	34	31	14	24	49	35	64	39	35	48	39	54	54
49	30	25	29	59	34	52	58	59	44	49	44	53	63	59	25	28	19
59	44	15	15	25	19	34	05	34	15	08	25	15	20	24	35	10	24
31	20	25	25	20	19	14	10	30	15	27	29	30	15	20	25	34	34
30	14	11	20	30	15	34	44	24	30	45	34	05	24	31	36	34	19
15	34	39	53	08	19	44	34	29	20	43	39	39	44	12	18	29	34
49	14	13	23	19	25	28	39	29	24	15	25	25	19	34	15	24	05
12	25	15	20	14	35	20	34	21	10	25	25	20	09	24	20	30	25
23	29	30	15	30	15	24	34	30	14	11	20	30	15	18	18	28	16
31	08	31	12	05	22	12	23	21	38	13	27	28	17	28	28	23	16
17	13	33	18	26	32	33	18	23	22	31	37	33	11	08	17	33	18
35	35	45	11	14	25	36	25	22	05	15	40	26	55	30	26	39	30
45	45	40	21	16	20	50	25	43	49	50	35	40	35	44	54	50	16
19	10	50	35	25	25	35	09	24	15	34	15	12	15	05	30	24	45
20	24	31	20	35	35	30	19	14	10	40	15	33	39	40	25	30	25
34	44	40	14	11	10	40	25	20	30	10	34	49	20	19	20	23	40
30	05	19	20	25	39	16	15	30	20	15	24	29	25	25	30	12	14
15	20	35	10	19	19	15	29	26	35	15	10	24	34	24	20	35	24
15	14	21	26	24	19	05	34	29	43	18	09	34	24	19	10	33	29
29	34	22	28	29	24	39	14	23	33	29	15	18	29	29	14	25	25
15	49	64	35	34	35	38	55	45	20	34	05	30	34	31	30	15	15
20	39	44	40	10	35	27	29	10	25	20	25	34	24	20	44	41	50
10	25	10	10	20	24	39	10	39	20	13	30	20	25	29	30	05	19
36	25	20	20	15	24	19	15	25	10	32	34	25	10	15	30	39	29
25	19	16	25	25	20	24	14	34	28	35	24	53	34	27	26	24	39
43	34	19	05	50	39	24	24	29	38	15	19	29	14	46	48	29	24
19	44	53	43	39	33	30	21	29	34	31	41	21	33	48	31	08	21
28	39	31	16	18	31	36	50	05	16	41	31	26	23	40	36	36	41
09	15	26	31	46	11	10	20	16	28	25	36	26	21	20	30	20	24
39	20	19	10	17	30	20	15	09	30	25	39	16	05	30	20	15	14
29	25	25	30	18	24	25	20	35	10	19	29	25	19	16	25	25	10

Exhibit A4: Permuted distances for Skokie Data

References

Alcaly, R.E. (1967). Aggregation and Gravity Models: Some Empirical Evidence. *Journal of Regional Science* **7** 61–73.

Allen, W.B. (1972). An Economic Derivation of the 'Gravity Law' of Spatial Interaction: A Comment. *Journal of Regional Science* **12** 119–126.

Amemiya, T. (1976). The Maximum Likelihood, the Minimum Chi-Square and the Non-linear Least Squares Estimator in the General Qualitative Response Model. *Journal of the American Statistical Association* **71** 347–351.

Anas, A. (1981). Discrete Choice Theory. Information Theory and the Multinomial Logit and Gravity Models. *Transportation Research* **17B** 13–23.

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

Anderson, T.R. (1956). Potential Models and the Spatial Distribution of Population. *Papers of the Regional Science Association* **2** 175–182.

Anderson, T.W. and S. Dasgupta (1963). Some Inequalities on Characteristic Roots of Matrices. *Biometrika* **50** 522–524.

Andersson, P-A. (1981). On the Convergence of Iterative Methods for the Distribution Balancing Problem. *Transportation Research* **15B** 173–201.

Anscombe, F.J. (1948). The Transformation of Poisson, Binomial and Negative-Binomial Data. *Biometrika* **35** 246–254.

Aufhauser, E. and M.M. Fischer (1985). Log-linear Modelling and Spatial Data. *Environment and Planning A* **17** 931–951.

Bachrach, M. (1970). *Biproportional Matrices and Input-Output Change.* Cambridge University Press, Cambridge, England.

Bapat, R. and T.E.S. Raghavan (1980). On Diagonal Products of Doubly Stochastic Matrices. *Linear Algebra and its Applications.* **31** 71–75.

Bapat, R. and T.E.S. Raghavan (1989). An Extension of a Theorem of Darroch and Ratcliff in Loglinear Models and its Application to Scaling Multidimensional Matrices. *Linear Algebra and its Applications.* **114/115** 705–715.

Batsell, R.R. (1981). A Multiattribute Extension of the Luce Model which Simultaneously Scales Utility and Substitutability. Working paper, J.H. Jones Graduate School of Administration, Rice University.

Batten, D.F. and D.E. Boyce (1987). Spatial Interaction, Transportation, and Interregional Commodity Flow Models. In *Handbook of Regional and Urban Economics.* P. Nijkamp, ed. **1** Chapter 9.

Batty, M. (1976). *Urban Modelling — Algorithms, Calibration, Predictions.* Cambridge University Press, Cambridge, England.

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

Batty, M. and P.K. Sikdar (1982). Spatial Aggregation in Gravity Models: An Information Theoretic Framework. *Environment and Planning A* **14** 377–405.

Baxter, M.J. (1978) A Note on the Claibration of Maximum-Performance Spatial Interaction Models. *Environment and Planning A* **10** 1151–1154.

Baxter, M.J. (1984). A Note on the Estimation of a Nonlinear Migration Model Using GLIM. *Geographical Analysis* **16** 282–286.

Baxter, M.J. (1985). Quasi-Likelihood Estimation and Diagnostic Statistics for Spatial Interaction Models. *Environment and Planning A* **17** 1627–1635.

Baxter, M.J. (1987). Testing for Misspecification in Models of Spatial Flows. *Environment and Planning A* **19** 1152–1160.

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

Beardwood, J.E. and H.R. Kirby (1975). Zone Definition and the Gravity Model: the Separability, Excludability and Compressibility Properties. *Transportation Research* **9** 363–369.

Beckmann, M.J. and T.F. Golob (1972). A Critique of Entropy and Gravity in Travel. In *Traffic Flow and Transportation.* G.F. Newell, ed. American Elsevier, New York.

Beckmann, M.J. and J.P. Wallace (1969). Evaluation of User Benefits arising from Changes in Transportation Systems. *Transportation Science* **3** 344–351.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Applications to Travel Demand.* MIT Press, Cambridge, Massachusetts.

Ben-Akiva, M. and J. Swait (1987). Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research* **21B** 91–102.

Berge, C. (1963). *Topological Spaces.* Macmillan, New York.

Birch, M. W. (1964). A New Proof of the Pearson-Fisher Theorem. *Annals of Mathematical Statistics* **35** 817-824.

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, Massachusetts.

Black, W.R. (1973) An Analysis of Gravity Model Distance Exponents. *Transportation* **2** 299–312.

Black, W.R. and R. Larson (1972) A Comparative Evaluation of Alterante Friction Factors in the Gravity Model. *The Professional Geographer* 334–337.

Blommestein, H. and P. Nijkamp (1986). Testing the Spatial Scale and the Dynamical Structure in Regional Models. *Journal of Regional Science* **26** 1–18.

Boots, B.A. and P. Kanaroglou (1988). Incorporating the Effects of Spatial Structure in Discrete Choice Models of Migration. *Journal of Regional Science* **28** 495–510.

Borgers, A. and H. Timmermans (1987). Choice Model Specification, Substitution and Spatial Structure Effects: A Simulation Experiment, *Regional Science and Urban Economics* **17** 29–48.

Boyce, D.E., K.S. Chon, Y.J. Lee, K.T. Lin and L.J. LeBlanc (1983). Implementation and Computational Issues for Combined Models of Location, Destination, Mode and Route Choice. *Environment and Planning A* **15** 1219–1230.

Boyle, P.J. and R. Flowerdew (1993). Modelling Sparse Interaction Matrices: Interward Migration in Hereford and Worcester, and the Underdispersion Problem. *Environment and Planning A* **25** 1201–1209.

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

Carey, H.C. (1858). *Principles of Social Science.* J. Lippincott. Philadelphia, Pennsylvania.

Carroll, J.D. (1955) Spatial Interaction and the Urban Metropolitan regional Description. *Papers and Proceeddings of the Regional Science Association* **1** D1–D14.

Carroll, J.D. and H.B. Bevis (1957). Predicting Local Travel in Urban Regions. *Papers of the Regional Science Association* **3** 183–197.

Carrothers, G.A.P. (1956a). Discussion: Gravity and Potential Models of Spatial Interaction. *Papers of the Regional Science Association* **2** 196–198.

Carrothers, G.A.P. (1956b). An Historical Review of the Gravity and Potential Concepts of Human Interaction. *Journal of the American Institute of Planners* **22** 94–102.

Carter, D.S. and P.M. Prenter (1972). Exponential Spaces and Counting Processes. *Z. Wahrscheinlichkeittheorie* **21** 1–19.

Catton, W.R., Jr. (1965). The Concept of 'Mass' in the Sociological Version of Gravity. In *Mathematical Explorations in Behavioral Science.* F. Massarik and P. Ratoosh, eds. Irwin and Dorsey Press, Homewood, Ill.

Cesario, F.J. (1973). A Generalized Trip Distribution Model. *Journal of Regional Science* **13** 233–248.

Cesario, F.J. (1974a). The Interpretation and Calculation of Gravity Model Zone-to-Zone Adjustment Factors. *Environment and Planning A* **6** 247–257.

Cesario, F.J. (1974b). More on the Generalized Trip Distribution Model. *Journal of Regional Science* **14** 389–397.

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

Charnes, A., W.M. Raike and C.O. Bettinger (1972). An Extremal and Information Theoretic Characterization of some Interzonal Transfer Models. *Socio-Economic Planning Sciences* **6** 531–537.

Chatterji, S.D. (1963). Some Elementary Characterizations of the Poisson Distribution. *American Mathematical Monthly* **70** 958–964.

Chicago Area Transportation Study (1960). *Final Report. Vol.II.*

Choi, S-C., W.S. DeSarbo and P.T. Harker (1990) Product Positioning under Price Competition. *Management Science* **36** 175–199.

Chojnicki, Z. (1965) *Zastosowanie Modeli Grawitacji I Potencjalu w Badaniach Przestrzenno-Ekonomicznych.* Polski Akademii Nauk, Warsaw.

Choukroun, J.M. (1975). A General Framework for the Development of Gravity-Type Trip Distribution Models. *Regional Science and Urban Economics* **5** 177–202.

Colwell, P.F. (1982). Central Place Theory and the Simple Economic Foundations of the Gravity Model. *Journal of Regional Science* **22** 541–546.

Cinlar, E. (1972). Superposition of Point Processes. In *Stochastic Point Processes: Statistical Analysis. Theory and Applications.* P.A. Lewis, ed. Wiley, New York. 549–606.

Cox, T.F. and V. Isham (1980). *Point Processes.* Chapman and Hall, London.

Daganzo, C.F. and Y. Sheffi (1977). On Stochastic Models of Traffic Assignment. *Transportation Science* **11** 253–274.

Daley, D.J. and D. Vere-Jones (1972). A Summary of the Theory of Point Processes. In *Stochastic Point Processes: Statistical Analysis, Theory, and Applications.* Lewis. P.A., ed. Wiley. New York. 299–383.

Dantzig, G.B. (1963). *Linear Programming and Extensions.* Princeton University Press, Princeton, New Jersey.

Davies, R.B. and C.M. Guy (1987). The Statistical Modeling of Flow Data when the Poisson Assumption is Violated. *Geographical Analysis* **19** 300–314.

Deming, W.E. and F.F. Stephan (1940). On a Least Squares Adjustment of Sampled Frequency Tables when the Expected Marginal Tables are Known. *Annals of Mathematical Statistics* **11** 427–444.

DePalma, A. and C. Lefevre (1983). Individual Decision-Making in Dynamic Collective Systems. *Journal of Mathematical Sociology* **9** 103–124.

Dieudonné, J. (1960) *Foundations of Modern Analysis.* Academic Press, New York.

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press, New York.

Dodd, S.C. (1950). The Interactance Hypothesis: A Gravity Model Fitting Physical Masses and Human Groups. *American Sociological Review* **15** 245–256.

Eichhorn, W. (1978). *Functional Equations in Economics.* Addison-Wesley, Reading, Massachusetts.

Ellis, R.S. (1985). *Entropy, Large Deviations, and Statistical Mechanics.* Springer-Verlag, New York.

Erlander, S. (1980). *Optimal Spatial Interaction and the Gravity Model.* Lecture Notes in Economics and Mathematical Systems, No. 173, Springer Verlag, Berlin.

Erlander, S. (1985). On the Principle of Monotone Likelihood and Log-Linear Models. *Mathematical Programming Study 25.* North-Holland, Amsterdam. 108–123.

Erlander, S. and N.F. Stewart (1989). *The Gravity Model in Transportation Analysis: Theory and Practice.* VSP, Utrecht, The Netherlands.

Erlander, S. and T.E. Smith (1990). General Representation Theorems for Efficient Population Behavior. *Applied Mathematics and Computation* **36** 173–217.

Evans, A. W. (1970). Some Properties of Trip Distribution Models. *Transportation Research* **4** 19–36.

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.

Evans, S.P. and H.R. Kirby (1974). A Three-dimensional Furness Procedure for Calibrating Gravity Models. *Transportation Research* **8** 105–122.

Feller, W. (1957). *An Introduction to Probability Theory and its Applications.* Volume 1 (2nd edition). Wiley, New York.

FHWA (1973). *Urban Transportation Planning System.* Federal Highway Administration, U.S. Department of Transportation.

FHWA (1974). *Urban Trip Distribution Friction Factors.* Federal Highway Administration, U.S. Department of Transportation.

Fienberg, S.E. (1970). An Iterative Procedure for Estimation in Contingency Tables. *Annals of Mathematical Statistics* **41** 907–917 [Corrig. (1971). **42** 778.]

Findlay, A. and P. Slater (1981). Functional regionalization of Spatial Interaction Data. *Environment and Planning A* **13** 1357–1368.

Fingleton, B. (1981). Log-linear Modelling of Geographical Contingency Tables. *Environment and Planning A* **13** 1539–1551.

Fingleton, B. (1983). Log-linear Models with Dependent Spatial Data. *Environment and Planning A* **15** 801–813.

Fishburn, P.C. (1970). *Utility Theory for Decision Making.* New York, Wiley.

Fisk, C.S. (1985). Entropy and Information Theory: Are We Missing Something? *Environment and Planning A* **17** 679–710.

Fisk, C.S. and D.E. Boyce (1984). A Modified Composite Cost Measure of Probabilistic Choice Modeling. *Environment and Planning A* **16** 241–248.

Fisk, C.S. and G.R. Brown (1975). A Note on the Entropy Formulation of Distribution Models. *Operational Research Quarterly* **26** 755–758.

Flowerdew, R. and M. Aitkin (1982). A Method of Fitting the Gravity Model based on the Poisson Distribution. *Journal of Regional Science* **22** 191–202.

Fotheringham, A.S. (1983a). A New Set of Spatial Interaction Models: The Theory of Competing Destinations. *Environment and Planning A* **15** 15–36.

Fotheringham, A.S. (1983b). Some Theoretical Aspects of Destination Choice and Their Relevance to Production-Constrained Gravity Models. *Environment and Planning A* **15** 1121–1132.

Fotheringham, A.S. (1986). Modelling Hierarchical Destination Choice. *Environment and Planning A* **18** 401–418.

Fotheringham, A.S. (1988). Consumer Store Choice and Choice Set Definition. *Marketing Science* **7** 299–310.

Fotheringham, A.S. and M.E. O'Kelly (1989). *Spatial Interaction Models: Formulations and Applications.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Fotheringham, A.S. and P.A. Williams (1983). Further Discussion of the Poisson Interaction Model. *Geographical Analysis* **15** 343–347.

Furness, K. P. (1965). Time Function Iteration. *Traffic Engineering and Control* **7** 458–460.

Gale, D. (1960). *The Theory of Linear Models.* McGraw-Hill, New York.

Gatrell, A.C. (1983). *Distance and Space.* Clarendon Press, Oxford, England.

Genesco, J. and R. Hasl (1981). Reconsidering the Theory of Social Gravity: A Comment. *Journal of Regional Science* **21** 551–554.

Giles, D.E.A. and P. Hampton (1981). Interval Estimation in the Calibration of Certain Trip Distribution Models. *Transportation Research B* **15B** 203–219.

Ginsberg, R.B. (1972). Incorporating Causal Structure and Exogenous Information with Probabilistic Models: With Special Reference to Choice, Gravity, Migration, and Markov Chains. *Journal of Mathematical Sociology* **2** 83–104.

Gokhale, D.V. and S. Kullback (1978). *The Information in Contingency Tables.* Marcel Dekker, New York.

Golob, T.F. and M.J. Beckmann (1971). A Utility Model for Travel Forecasting. *Transportation Science* **5** 79–90.

Good, I.J. (1950). *Probability and the Weighing of Evidence* Griffin, London.

Good, I.J. (1963). Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables. *Annals of Mathematical Statistics* **34** 911–934.

Goodman, L.A. (1968). The Analysis of Cross-Classified Data: Independence, Quasi-Independence and Interaction in Contingency Tables With or Without Missing Cells. *Journal of the American Statistical Association* **63** 1091–1131.

Govindarajulu, Z. and R.T. Leslie (1970). Elementary Characterization of Discrete Distributions. In *Random Counts in Scientific Work: Volume 1.* Patil. G.P., ed. Pennsylvania State University Press, University Park, Pennsylvania.

Gnedenko, B.W. and A.N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Variables.* (Translated from Russian). Addison-Wesley, Reading, Massachusetts.

Gray, R.H. and A. Sen (1983). Estimating Gravity Model Parameters: A Simplified Method Based on the Odds Ratio. *Transportation Research B* **17B** 117–131.

Griesinger, D.W. (1979). Reconsidering the Theory of Social Gravity. *Journal of Regional Science* **19** 291–302.

Griesinger, D.W. (1981). Reconsidering the Theory of Social Gravity: A Reply. *Journal of Regional Science* **21** 555–556.

Haag, G. (1989). *Dynamic Decision Theory: Applications to Urban and Regional Topics.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Haberman, S.J. (1974). *The Analysis of Frequency Data.* University of Chicago Press, Chicago, Illinois.

Haberman, S.J. (1978). *Analysis of Qualitative Data. Volume 1: Introductory Topics.* Academic Press, New York.

Haberman, S.J. (1979). *Analysis of Qualitative Data. Volume 2: New Developments.* Academic Press, New York.

Halmos, P.R. (1950). *Measure Theory.* Von Nostrand, New York.

Hansen, W.G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* **25** 73–76.

Harris, B. (1964). A Note on the Probability of Interaction at a Distance. *Journal of Regional Science* **5** 31–36.

Hastie, T.J. and R.J. Tibshirani (1990). *Generalized Additive Models.* Chapman and Hall, London.

Hathaway, P.J. (1975). Trip Distribution and Disaggregation. *Environment and Planning A* **7** 71–97.

Haynes, K.E. and F.Y. Phillips (1982). Constrained Minimum Discrimination Information: A Unifying Tool for Modeling Spatial and Individual

Choice Behavior. *Environment and Planning A* **14** 1341–1354.

Haynes, K.E. and A.S. Fotheringham (1984). *Gravity and Spatial Interaction Models.* Sage Publications, Beverly Hills, California.

Hensher, D.A. and L.W. Johnson (1981). *Applied Discrete Choice Modeling.* Croom Helm, Beckenham, Kent, England.

Hirst, M.A. (1977) Hierarchical Aggregation Procedures for Interaction Data: A Comment. *Environment and Planning A* **9** 99–103.

Hoeffding, W. (1965). Asymptotically Optimal Tests for Multinomial Distributions. *Annals of Mathematical Statistics* **36** 369–408.

Hua, C-I. and F. Porell (1979). A Critical Review of the Development of the Gravity Model. *International Regional Science Review* **4** 97–126.

Huang, K. (1963). *Statistical Mechanics.* Wiley, New York.

Huber, P.J. (1967). The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability* **1** 221-234. University of California Press, Berkeley/Los Angeles.

Huff, D.L. (1959). Geographical Aspects of Consumer Behavior. *University of Washington Business Review* **18** 27–37.

Hurwitz, L., J.S. Chipman, M.L. Richter and H.F. Sonnenschein (1971). *Preferences, Utility and Demand.* Harcourt Brace Jovanovich, New York.

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

Iklè, F.C. (1954). Sociological Relationship of Traffic to Population and Distance. *Traffic Quaterly* **8** 123–136.

Isard, W. (1975a). *An Introduction to Regional Science.* Prentice-Hall, Englewood Cliffs, New Jersey.

Isard, W. (1975b). A Simple Rationale for Gravity Model Type Behavior. *Papers of the Regional Science Association* **35** 25–30.

Isard, W. and D.F. Bramhall (1960). Gravity, Potential, and Spatial Interaction Models. In *Methods of Regional Analysis.* W. Isard. MIT Press, Cambridge, Massachusetts. 493–568.

Isham, V. (1975). On a Point Process with Independent Locations. *Journal of Applied Probability* **12** 435–436.

Johnson, N.L. and S. Kotz (1969). *Discrete Distributions.* Houghton Mifflin, Boston.

Johnson, N.L. and S. Kotz (1970). *Continuous Univariate Distributions* (two volumes). Houghton Mifflin Company, New York.

Judge, G. G., W.E. Griffiths, R. Carter Hill, H. Lütkepohl and T-C. Lee (1985). *The Theory and Practice of Econometrics* (Second Edition). Wiley, New York.

Kagan, A.M., Y.V. Linnik and C.R. Rao (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

Kallenberg, O. (1976). *Random Measures.* Academic Press, New York.

Kanaroglou, P., K-L. Liaw and Y.Y. Papageorgiou (1986). An Analysis of Migratory Systems: 1. Theory. *Environment and Planning A* **18** 913–928.

Kanaroglou, P., K-L. Liaw and Y.Y. Papageorgiou (1986). An Analysis of Migratory Systems: 2.Operational Framework. *Environment and Planning A* **18** 1039–1060.

Karlin, S. and H.M. Taylor (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.

Karni, E. and A. Schwartz (1977). Search Theory: The Case of Search with Uncertain Recall. *Journal of Economic Theory* **16** 38–52.

Kau, J.B. and C.F. Sirmans (1979). The Functional Form of the Gravity Model. *International Regional Science Review* **4** 127–136.

Kelly, F.P. and B.D. Ripley (1976). A note on Strauss's Model for Clustering. *Biometrika* **63** 357–360.

Kendall, M.G. and A. Stuart (1963). *The Advanced Theory of Statistics, Vol 1: Distribution Theory* (Second Edition). Hafner, New York.

Kendall, M.G. and A. Stuart (1967) *The Advanced Theory of Statistics, Vol 2: Inference and Relationship* (Second Edition). Hafner, New York.

Kim, H. (1994). *Trip-Chaining Behavior and Spatial Interaction Models.* Ph.D. Dissertation. School of Urban Planning and Policy, University of Illinois at Chicago, Chicago, Illinois.

Kim H. and A. Sen (1993) Testing Goodness of Fit of Gravity Models. Unpublished Manuscript. Urban Transportation Center, University of Illinois at Chicago. Chicago, Illinois.

Kim, H., A. Sen and S. Sööt (1993). Factoring Household Travel Surveys. *Transportation Research Record* **1412** 17–22.

Kim, H., S. Sööt, A. Sen and E. Christopher (1994). Shopping Trip Chains: Current Patterns and Changes since 1970. *Transportation Research Record* **1443** 38–44.

Kindermann, R. and J.L. Snell (1980). *Markov Random Fields and their Applications.* Contemporary Mathematics Series. **1** American Mathematical Society, Providence, RI.

Kingman, J.F.C. (1978). The Uses of Exchangeability. *Annals of Probability* **6** 183–197.

Kirby, H.R. (1970). Normalizing Factors of the Gravity Model: An Interpretation. *Transportation Research* **4** 37–48.

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

Kirby, H.R. and M.N. Leese (1978). Trip Distribution Calculations and Sampling Error: Some Theoretical Aspects. *Environment and Planning A* **10** 837–851.

Krantz, D.H., R.D. Luce, P. Suppes and A. Tversky (1971). *Foundations of Measurement: Volume I.* Academic Press, New York.

Kruithof, J. (1937). *Calculation of Telephone Traffic.* Translation No. 2663. Post Office Research Department Library, London.

Kullback, S. and R.A. Leibler (1951). On Information and Sufficiency. *Annals of Mathematical Statistics* **22** 79–86.

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

Kulldorf, G. (1955). *Migration Probabilities.* Lund Studies in Geography, Series B: No.14. Department of Geography, Lund University, Lund, Sweden.

Laumann, E.O. (1966). *Prestige and Association in an Urban Community.* Bobbs-Merrill Company, New York.

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

Lehmann, E.L. (1959) *Testing Statistical Hypotheses.* Wiley, New York.

Lill, E. (1891). *Das Reisegesetz und Seine Anwendung auf den Eisenbahnverkehr.* Vienna, Austria.

Lowe, J.M. (1993) *Gravity Model Analysis of Hospital Patient Flows.* Ph.D. Dissertation. School of Urban Planning and Policy, University of Illinois at Chicago, Chicago, Illinois.

Luce, R.D. (1959). *Individual Choice Behavior.* Wiley, New York.

Luce, R.D. and P. Suppes (1965). Preference, Utility, and Subjective Probability. In *Handbook of Mathematical Psychology, Volume III.* R.D. Luce. *et al.*, eds. Wiley, New York.

Luce, R.D. (1977). The Choice Axiom after Twenty Years. *Journal of Mathematical Psychology* **15** 215–233.

Luoma, M. and M. Palomaki (1983). A New Theoretical Gravity Model and its Application to a Case with Drastically Changing Mass. *Geographical Analysis* **15** 14–27.

Macgill, S.M. (1977). Theoretical Properties of Biproportional Matrix Adjustments. *Environment and Planning A* **9** 687–701.

Macgill, S.M. (1979). Convergence and Properties for a Modified Biproportional Matrix Problem. *Environment and Planning A* **11** 499–506.

Madansky, A. (1988). *Prescriptions for Working Statisticians.* Springer-Verlag, New York.

Manski, C. (1977). The Structure of Random Utility Models. *Theory and Decision* **8** 229–254.

Manski, C. and S.R. Lerman (1977). The Estimation of Choice Probabilities from Choice-Based Samples. *Econometrica* **45** 1977–1988.

Manski, C. and D. McFadden (1981). Alternative Estimators and Sample Designs for Discrete Choice Analysis. In *Structural Analysis of Discrete Data with Econometric Applications.* C. Manski and D. McFadden, eds. MIT Press, Cambridge, Massachusetts.

March, L. (1971). Urban Systems: A Generalized Distribution Function. In *Urban and Regional Planning: London Papers in Regional Science 2.* A.G. Wilson, ed. Pion Press, London.

March, L. and M. Batty (1975). Generalized Measures of Information. Bayes' Likelihood Ratio and Jaynes' Formalism. *Environment and Planning B* **2** 99–105.

Masser, I (1977). A Comparative Analysis of Spatial Representation in Doubly Constrained Interaction Models. *Environemnt and Planning A* **9** 759–769.

Masser, I. (1979) A Note on the Treatment of Flows Across System Boundaries in Spatial Interaction Models *Environemnt and Planning A* **11** 447–453.

Masser, I. (1981) The Analysis of Spatial Interaction Data. Paper presented the North American Meetings of the Regional Science Association. Montreal. Published as TRP 31, Department of Town and Regional Planning, University of Sheffield, U.K.

Masser, I., P. Batey and P. Prown (1973). Design of Zoning Systems for Interaction Models. Paper presented to the British Section of The Regional Science Association, London.

Masser, I. and P.J.B. Brown (1975) Hierarchical Aggregation Procedures for Interaction Data. *Environemnt and Planning A* **7** 509–523.

Masser, I. and P.J.B. Brown (1977) Spatial Representation and Spatial Interaction. *Papers of the Regional Science Association* **38** 71–92.

Masser, I. and P.J.B. Brown (1978) Spatial Representation and Spatial Interaction: an Overview. In *Spatial Representation and Spatial Interaction*. I Masser and P.J.B. eds. Pion, London.

Masser, I. and J. Scheurwater (1980). Functional Regionalization of Spatial INteraction Data: an Evaluation and Some Suggested Strategies. *Environemnt and Planning A* **12** 1357–1382.

Mathur, V.K. (1970). An Economic Derivation of the 'Gravity Law' of Spatial Interaction: A Comment. *Journal of Regional Science* **10** 403–405.

Matthes, K., J. Kerstan, and J. Mecke (1978). *Infinitely Divisible Point Processes*. Wiley, New York.

McCullagh, P. and J.A. Nelder (1989). *Generalized Linear Models* (Second Edition). Chapman and Hall, London.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers of Econometrics*. P. Zarembka, ed. New York, Academic Press. 105–142.

Meyer, R.J. and T.C. Eagle (1981). A Parsimonious Multinomial Choice Model Recognizing Alternative Interdependence and Context-Dependent Utility Functions. Working paper 26–80–81, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA.

Meyer, R.J. and T.C. Eagle (1981). Context Induced Parameter Instability in a Disaggregate-Stochastic Model of Store Choice. *Journal of Marketing Research* **19** 62–71.

Moran, P.A.P. (1952). A Characteristic Property of the Poisson Distribution. *Proceedings of the Cambridge Philosophical Society* **48** 206–207.

Morrill, R.L. and F.R. Pitts (1967). Marriage, Migration, and the Mean Information Field. *Annals of the Association of American Geographers*

57 401–422.

Moyal, J.E. (1962). The General Theory of Stochastic Population Processes. *Acta Mathematica* **108** 1–31.

Niedercorn, J.H. and B.V. Bechdolt (1969). An Economic Derivation of the 'Gravity Law' of Spatial Interaction. *Journal of Regional Science* **9** 273–281.

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

Niedercorn, J.H. and B.V. Bechdolt (1972). An Economic Derivation of the 'Gravity Law' of Spatial Interaction: A Further Reply and Reformulation. *Journal of Regional Science* **12** 127–136.

Niedercorn, J.H. and J.D. Moorehead (1974). The Commodity Flow Gravity Model: A Theoretical Reassessment. *Regional and Urban Economics* **4** 69–75.

Nijkamp, P. and J. Poot (1987). Dynamics of Generalized Spatial Interaction Models. *Regional Science and Urban Economics* **17** 367–390.

Okabe, A. (1976). A Theoretical Comparison between the Opportunity and Gravity Models. *Regional Science and Urban Economics* **6** 381–397.

Olsson, G. (1965). *Distance and Human Interaction.* Regional Science Research Institute, Philadelphia, Pennsylvania.

Openshaw, S. (1976). An Empirical Study of Some Spatial Interaction Models. *Environment and Planning A* **8** 23–41.

Openshaw, S. (1977). Optimal Zoning Systems for Spatial Interaction. *Environment and Planning A.* **9** 169–184.

Openshaw, S. (1978). An Empirical Study of Some Zone Design Criteria. *Environment and Planning A.* **10** 781–794.

Openshaw, S. (1979). Alternative Methods for Estimating Spatial Interaction Models and their Performance in Short-Term Forecasting. In *Exploratory and Explanatory Statistical Analysis of Spatial Data.* C.P.A. Bartels and R.H. Ketellaper, eds. Martinus Nijhoff, Boston.

Openshaw, S. and C.J. Connolly (1977). Empirical Derived Deterrence Functions for Maximum Performance Spatial Interaction Models. *Environment and Planning A* **9** 1067–1079.

Parlett, B.N. and T.L. Landis (1982). Methods for Scaling to Doubly Stochastic Form. *Linear Algebra and its Applications.* **48** 53–79.

Papageorgiou, Y.Y. and T.R. Smith (1983). Agglomeration as Local Instability of Spatially Uniform Steady-States. *Econometrica* **51** 1109–1119.

Patil, G.P. and V. Seshadri (1964). Characterization Theorems for some Univariate Probability Distributions. *Journal of the Royal Statistical Society, Series B* **26** 286–292.

Pfanzagl, J. (1968). *Theory of Measurement.* Wiley, New York.

Philbrick, A.T. (1973). A Short History of the Development of the Gravity Model. *Australian Road Research* **5** 40–54.

Phipps, A. and W. Laverty (1983). Optimal Stopping and Residential Search Behavior. *Geographical Analysis* **15** 187–204.

Porell, F.W. (1980). Intermetropolitan Migration and the Quality of Life. Paper presented at the 27th North American Meetings of the Regional Science Association, Milwaukee, WI.

Preston, C.J. (1976). *Random Fields.* Springer-Verlag, Berlin.

Putman, S.H. and F.W. Ducca (1978a) Calibrating Urban Residential Models 1: Procedures and Strategies. *Environment and Planning A* **10** 633–650.

Putman, S.H. and F.W. Ducca (1978b) Calibrating Urban Residential Models 2: Empirical Results. *Environemnt and Planning A* **10** 1001–1014.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

Ravenstein, E.G. (1885). The Laws of Migration. *Journal of the Royal Statistical Society* **48** 167–235.

Ravenstein, E.G. (1889). The Laws of Migration. *Journal of the Royal Statistical Society* **52** 241–305.

Reilly, W.J. (1929). *Methods for the Study of Retail Relationships.* Bulletin No.2944. University of Texas, Austin, Texas.

Reilly, W.J. (1931). *The Law of Retail Gravitation.* W.J. Reilly, New York.

Ripley, B.D. and F.P. Kelly (1976). Markov Point Processes. *Journal of the London Mathematical Society* **15** 188-192.

Rogers, A. (1974). *Statistical Analysis of Spatial Dispersion.* Pion, London.

Rogerson, P. (1982). Spatial Models of Search. *Geographical Analysis.* **14** 217–228.

Roy, J.R. and P.F. Lesse (1981). On Appropriate Microstate Descriptions in Entropy Modelling. *Transportation Research* **15B** 85–96.

Ruiter, E.R. (1967). Toward a Better Understanding of the Intervening Opportunities Model. *Transportation Research* **1** 47–56.

Rummel, R.J. (1972). *The Dimensions of Nations.* Sage Press, Beverley Hills, California.

SAS (1985a). *SAS User's Guide: Basics, Version 5 Edition.* SAS Institute, Cary, NC.

SAS (1985b). *SAS User's Guide: Statistics, Version 5 Edition.* SAS Institute, Cary, NC.

Schneider, M. (1959). Gravity Models and Trip Distribution Theory. *Papers and Proceedings of the Regional Science Association* **5** 51–58.

Sen, A. (1985). Research Suggestions on Spatial Interaction Models. *Transportation Research A* **19A** 432–435.

Sen, A. (1986). Maximum Likelihood Estimation of Gravity Model Parameters. *Journal of Regional Science* **26** 461–474.

Sen, A. and Z. Matuszewski (1991). Properties of Maximum Likelihood Estimates of Gravity Model Parameters. *Journal of Regional Science* **31** 469–486.

Sen, A. and Z. Matuszewski (1985). Linear Least Squares Estimates of Gravity Model Parameters. Unpublished Manuscript. Urban Transportation Center. University of Illinois at Chicago.

Sen, A. and R.K. Pruthi (1983). Least Squares Calibration of the Gravity Model when the Intrazonal Flows are Unknown. *Environment and Planning A* **15** 1545–1550.

Sen, A. and S. Sööt (1981). Selected Procedures for Calibrating the Generalized Gravity Model. *Papers of the Regional Science Association* **48** 165–176.

Sen, A. and M. Srivastava (1990). *Regression Analysis: Theory, Methods and Applications*. Springer-Verlag, New York.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Analytical Methods*. Prentice-Hall, Engelwood Cliffs, New Jersey.

Shepard, R.N. (1974). Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika* **39** 373–421.

Shocker, A.D. and V. Srinivasan (1979). Multiattribute Approaches to Product Concept Evaluation and Generation: A Critical Review. *Journal of Marketing Research* **16** 159–180.

Simon, H.A. (1957). *Models of Man*. Wiley, New York.

Sinkhorn, R. (1967). Diagonal Equivalence to Matrices with Prescribed Row and Column Sums. *American Mathematical Monthly* **74** 402–405.

Smith, T.E. (1975). A Choice Theory of Spatial Interaction. *Regional Science and Urban Economics* **5** 137–176.

Smith, T.E. (1976a). A Spatial Discounting Theory of Interaction Preferences. *Environment and Planning A* **8** 879–915.

Smith, T.E. (1976b). Spatial Discounting and the Gravity Hypothesis. *Regional Science and Urban Economics* **6** 331–356.

Smith, T.E. (1978a). A Cost-Efficiency Principle of Spatial Interaction Behavior. *Regional Science and Urban Economics* **8** 313–337.

Smith, T.E. (1978b). A General Efficiency Principle of Spatial Interaction. In *Spatial Interaction Theory and Planning Models*. Karlqvist. A. *et al.*, eds. North-Holland, Amsterdam. 97–118.

Smith, T.E. (1983). A Cost-Efficiency Approach to the Analysis of Congested Spatial-Interaction Behavior. *Environment and Planning A* **15** 435–464.

Smith, T.E. (1984). Testable Characterizations of Gravity Models. *Geographical Analysis* **16** 74–94.

Smith, T.E. (1986a). An Axiomatic Foundation for Poisson Frequency Analyses of Weakly Interacting Populations. *Regional Science and Urban Economics* **16** 269–307.

Smith, T.E. (1986b). Two Extension Theorems for Log-Linear Representations of Positive Probabilities. *Applied Mathematics and Computation* **18** 239–255.

Smith, T.E. (1987a). Poisson Gravity Models of Spatial Flows. *Journal of Regional Science* **27** 315–340.

Smith, T.E. (1987b). A Threshold Theory of Discretionary Spatial Interaction Behavior. *Regional Science and Urban Economics* **17** 495–517.

Smith, T.E. (1988). A Cost-Efficiency Theory of Dispersed Network Equilibria. *Environment and Planning A* **20** 231–266.

Smith, T.E. (1990). Most-Probable-State Analysis: A Method for Testing Probabilistic Theories of Population Behavior. In *New Frontiers in Regional Science*. M. Chatterji and R.E. Kuenne, eds. MacMillan, London. 75–94.

Smith, T.E., M. Harwitz, B. Lentnek and P. Rogerson (1992) Optimal Search on Spatial Paths with Recall. Regional Science Working Paper No. 154, Regional Science Department, University of Pennsylvania, Philadelphia, PA.

Smith, T.E. and W. Yu (1982). A Prominence Theory of Context-Sensitive Choice Behavior. *Journal of Mathematical Sociology* **8** 225–249.

Smith, T.R. and P.B. Slater (1981). A Family of Spatial Interaction Models Incorporating Information Flows and Choice Set Constraints Applied to U.S. Interstate Labor Flows. *International Regional Science Review* **6** 15–32.

Snickars, F. and J.W. Weibull (1977). A Minimum Information Principle. *Regional Science and Urban Economics* **7** 137–168.

Sööt, S. and A. Sen (1991). A Spatial Employment and Economic Development Model. *Papers in Regional Science* **70** 149–166.

Stewart, J.Q. (1941). An Inverse Distance Variation for Certain Social Influences. *Science* **93** 89–90.

Stewart, J.Q. (1948). Demographic Gravitation: Evidence and Application. *Sociometry* **11** 31–58.

Stewart, J.Q. (1950). The Development of Social Physics. *American Journal of Physics* **18** 239–253.

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

Strauss, D.J. (1975). A Model for Clustering. *Biometrika* **62** 467–475.

Southworth, F. (1979). Spatial Structure and Parameter Disaggregation in Trip Distribution Models. *Regional Studies* **13** 381–394.

Tanner, J.C. (1961). *Factors Affecting the Amount of Travel.* Road Research Technical Paper No. 51. Department of Scientific and Industrial Research, London, England.

Tellier, L.N. and D. Sankoff (1975). Gravity Models and Interaction Probabilities. *Journal of Regional Science* **15** 317–322.

Thakuriah, P., A. Sen, S. Sööt and E. Christopher (1993). Non-Response Bias and Trip Generation Models. *Transportation Research Record* **1412** 64–70.

Theil, H. (1971). *Principles of Econometrics.* Wiley, New York.

Thomas, R.W. (1977). An Interpretation of the Journey-to-Work in Merseyside Using Entropy Maximizing Methods. *Environment and Planning A* **9** 817–834.

Tobler, W. (1983). An Alternative Formulation for Spatial-Interaction Modeling. *Environment and Planning A* **15** 693–703.

Tukey, J.W. (1970). *Exploratory Data Analysis.* Addison-Wesley, Reading, Massachusetts. Three volumes. [In manuscript form]

Tukey, J.W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, Massachusetts.

Tversky, A. (1977). Features of Similarity. *Psychological Review* **84** 327–352.

Volodin, I.N. (1965). On Distinguishing between Poisson and Polya Distributions on the basis of a Large Number of Samples. *Theory of Probability and its Applications* **10** 335–338.

Weber, J.S. (1987). Elasticities of Constrained Gravity Models. *Journal of Regional Science* **27** 621–640.

Weber, J.S. and A. Sen (1983). On the Sensitivity of Maximum Likelihood Estimates of Gravity Model Parameters. In *Optimization and Discrete Choice in Urban Systems.* B.G. Hutchinson, P. Nijkamp and M. Batty, eds. Springer-Verlag, Berlin.

Weber, J.S. and A. Sen (1985). On the Sensitivity of Gravity Model Forecasts. *Journal of Regional Science* **25**, 317–336.

Webber, M.J. (1979). *Information Theory and Urban Spatial Structure.* Croom Helm, Beckenham, Kent.

Weibull, J.W. (1978). A Search Model for Microeconomic Analysis — With Spatial Applications. In *Spatial Interaction Theory and Planning Models.* A. Karlqvist *et al.*, eds. North-Holland, Amsterdam. 47–73.

West, M. (1994). *Statistical Inference for Gravity Models in Transportation Flow Forecasting.* Unpublished Manuscript. Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina.

White, M.J. (1988). Location Choice and Commuting Behavior in Cities with Decentralized Employment. *Journal of Urban Economics* **24** 129–152.

Williams, H.C.W.L. (1978). On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A* **9** 285–344.

Williams, I.N. (1976). A Comparison of Some Calibration Techniques for Doubly Constrained Models with an Exponential Cost Function. *Transportation Research* **10** 91–104.

Williams, I.N. (1977). Algorithm 1: Three Point Rational Function Interpolation for Calibrating Gravity Models. *Environment and PLanning A* **9** 215–221.

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

Wilson, A.G. (1970). *Entropy in Urban and Regional Planning.* Pion Limited, London, England.

Wilson, A.G. (1971). A Family of Spatial Interaction Models and Associated Developments. *Environment and Planning A* **3** 1–32.

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

Wills, M.J. (1986). A Flexible Gravity-Opportunities Model for Trip Distribution. *Transportation Research B* **20B** 89–112.

Young, E.C. (1924). *The Movement of Farm Population.* Bulletin 426. Cornell Agricultural Experiment Station, Ithaca, New York.

Yun, S. (1992). *Maximum Likelihood Estimation Procedures for Generalized Gravity Model Parameters*, Ph.D. Dissertation. School of Urban Planning and Policy, University of Illinois at Chicago, Chicago, Illinois.

Yun, S. and A. Sen (1994). Computation of Maximum Likelihood Estimates of Gravity Model Parameters. *Journal of Regional Science* **34** 199–216.

Zaryouni, M.R. and J.S. Liebman (1976). Quantifying the Space Separation Function Using Existing Locational Patterns. *Journal of Regional Science* **16** 73–82.

Zipf, G.K. (1946). The P1P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review* **11** 677–686.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Cambridge, Massachusetts.

List of Principal Definitions and Results

Subject Index

Author Index