THE SPECIFICATION OF CONSTRAINED INTERACTION MODELS USING THE SPSS LOGLINEAR PROCEDURE

MICHAEL TIEFELSDORF and BARRY BOOTS

Department of Geography—Wilfrid Laurier University, Waterloo, Ontario, Canada, N2L 3C5

Building on the equivalence between the log-linear model and the general spatial interaction model, a straightforward setup of SPSS's LOGLINEAR procedure is presented to calibrate the interaction model by means of maximum likelihood estimation. Several specifications are discussed to estimate the unconstrained, singly constrained (origin or destination) and the doubly constrained interaction model with metric co-variables and structural zeros. The emphasis is on technical issues helpful for utilizing the entire capacity of the LOGLINEAR procedure in calibrating and simulating the general interaction model.

KEY WORDS: interaction models, log-linear models, quasi-log-linear models, SPSS's LOGLINEAR procedure.

1. INTRODUCTION

Among the basic models of spatial interaction familiar to geographers are the simple gravity model and its more elaborate relatives such as augmented models or models with destination-interference effects. The scope of applications for interaction models is extremely broad, including the analysis of aggregated migration systems, the prediction of traffic flows and the modeling of individual shopping trip behavior. For a general overview of the theory and applications see Fotheringham and O'Kelly (1989). More specialized reviews are provided by Stillwell and Congdon (1991) in the context of migration modeling and Ben-Akiva and Lerman (1985) in the context of discrete choice analysis. However, so far analysis of interaction data is not widespread, due, in part, to a lack of generally available and user-friendly software. Often analysis is limited to the application of ordinary least square (OLS) techniques in log-normal gravity models. The statistical drawbacks of this linearized gravity model are well-known and range from the unapplicable assumption that flows are normally distributed to biased estimates and heteroscedastic residuals (see for instance Flowerdew and Aitkin, 1982). Furthermore, due to the limitations of the OLS technique, the concept of constrained interaction models cannot be translated into first hand experience within standard statistical software packages.

Insufficient availability of specialized software applies not only to spatial interaction models but also to the whole field of spatial modeling and spatial statistics. Two strategies have been applied to overcome this problem: (a) writing specialized software, or (b) adapting available software to the special needs. In this paper we will adopt to the second approach: structural linkages between both constrained and unconstrained interaction models and the broad class of log-linear models and Poisson regression allows us to do so.
Log-linear models and Poisson regression can be estimated by maximum likelihood (ML) procedures within specialized statistical packages. GLIM by Baker and Nelder (1978) has been the most used by geographers (see for instance Flowerdew and Aitkin, 1982, Baxter, 1984, Aufhauser and Fischer, 1985, Wrigley, 1985, Upton and Fingleton, 1989, Flowerdew, 1991, and Congdon, 1992, 1993). However, these specialized packages require considerable effort from the analyst in order to become familiar with them and need to be specially setup to handle interaction type models.

This paper demonstrates a straightforward adaptation of constrained and augmented interaction models to the LOGLINEAR procedure embedded in the well-known and widely used statistical package SPSS. The crucial points in applying log-linear models to interaction models are (a) the presence of structural zeros in the interaction matrix due to the exclusion of intraregional flows, leading to quasi-log-linear models, (b) the implementation of constraints via the design-matrix and (c) the flexible inclusion of metric co-variables. Furthermore, an inspection of the residuals, the assessment of the overall goodness of fit, as well as evaluation of the relevance of single factors and co-variables is needed to derive an adequate interaction model. Besides these statistical requirements, the data-handling facilities of the underlying software platform, the stability of the ML-algorithm, efficiency in large problems and the ease of use are of importance. All these properties are included in LOGLINEAR-procedure (see SPSS’s “Advanced Statistics User Guide”, 1990) which will help to foster the application of log-linear models and Poisson regression in general and in interaction analysis in particular.

The layout of this paper is as follows. The next section reviews the equivalence between mathematical, behavioral and statistical modeling approaches towards the general spatial interaction model. Then basic relevant issues of the log-linear model are summarized and several specifications of interaction models within the framework of the log-linear model are discussed. This is followed by an illustration of the implementation of interaction models within SPSS’s LOGLINEAR procedure. Also a SPSS input program to simulate Poisson distributed interaction flows is given. Finally, by comparing the cornered and the centered effect coding, some guidance is given on how to interprete the estimated model parameters.

2. APPROACHES TO INTERACTION MODELING

Suppose that we have a spatial system consisting of $I$ origins and $J$ destinations and that $m_{ij}$ is the number of individuals moving from $i$ to $j$ ($i=1, \ldots, I; j=1, \ldots, J$). Usually, this information is displayed in the form of an interaction matrix $M_{IJ}$ defined by

$$M_{IJ} = \begin{pmatrix}
  m_{11} & \cdots & m_{1j} & \cdots & m_{1J} \\
  \vdots & \ddots & \vdots & \ddots & \vdots \\
  m_{i1} & m_{ij} & m_{ij} & \ddots & \vdots \\
  \vdots & \ddots & \ddots & \ddots & \vdots \\
  m_{I1} & \cdots & m_{Ij} & \cdots & m_{IJ}
\end{pmatrix}. \tag{1}$$

Note the close resemblance between $M_{IJ}$ and a contingency table. In some applications, for example migration modeling, the sets of origins and destinations are the same and thus $M_{IJ}$ is a square matrix. However, in such cases $M_{IJ}$ is not usually symmetric since it is unlikely that $m_{ij}$ will equal $m_{ji}$. The interpretation of the main-diagonal of $M_{IJ}$ will
depend on the specific application. For instance, it might represent internal interaction within region \( i \) or it could simply be a count of individuals in region \( i \) not involved in the interaction process. Often such values are not recorded. In other applications, for example shopping trips from residential areas to individual shopping malls, the number of origins and destinations will differ and \( M_{p	imes j} \) will not be square. For all applications, the \( i \)-th row of this matrix describes the outflow from region \( i \) to each of the \( J \) destinations, whereas inflows from each of the \( I \) origins into destination \( j \) are described by the \( j \)-th column.

From \( M_{p	imes j} \) we can calculate the number of individuals originating from region \( i \) or terminating in region \( j \), i.e. \( m_i^0 = \sum_{j=1}^{J} m_{ij} \) \((i = 1, \ldots, I)\) and \( m_j^0 = \sum_{i=1}^{I} m_{ij}\) \((j = 1, \ldots, J)\), respectively. In turn these marginal sums can be used to calculate the overall level of interaction which is defined by \( m = \sum_{i=1}^{I} \sum_{j=1}^{J} m_{ij} \). Often several attributes of this interaction system are also available to the analyst. In general, these attributes relate to the characteristics of origin \( i \) and destination \( j \) as well as measures of differences between both of them. Note that in the classical gravity model these attributes are the population sizes of origin \( i \) and destination \( j \), respectively, and the geographic distance between the two regions.

The distribution of interactions within a regional system can be described by the generic interaction model written in its general form as

\[
\mu_{ij} = g_{ij} \cdot u_i \cdot v_j \cdot f_{ij} \tag{2}
\]

where
- \( \mu_{ij} \) is the estimated number of interactions between region \( i \) and region \( j \)
- \( g_{ij} \) is a balancing factor with varying subscripts depending on which constraints \( \{ \mu_{ij} \} \) has to obey concerning \( m_i^0, m_j^0 \), or \( m \)
- \( u_i \) is a function characterizing the origin \( i \)
- \( v_j \) is a function characterizing the destination \( j \)
- \( f_{ij} \) is a function reflecting interregional differentials.

Alternative forms of the general interaction model can be specified by imposing different constraints on \( \mu_{ij} \). For the relationship between the constraints and the balancing factor \( g_{ij}, \) see, for instance, Ledent (1985). In the unconstrained case the only condition specified is that the total estimated interaction equals the observed interaction, i.e.

\[
m = \sum_{i=1}^{I} \sum_{j=1}^{J} \mu_{ij}. \tag{3}
\]

In the singly constrained case either the estimated outflows from each region (origin constraint) or the estimated inflows to each region (destination constraint) have to match the observed outflow totals \( m_i^0 \) or inflow totals \( m_j^0 \), respectively, i.e.

\[
m_i^0 = \sum_{j=1}^{J} \mu_{ij} \quad \text{or} \quad m_j^0 = \sum_{i=1}^{I} \mu_{ij}. \tag{4}
\]

Since the origin and destination constraints are isomorphic, we deal here only with the origin constrained model. Note that in general the origin characteristic function \( u_i \) is linearly dependent with the origin specific balancing factor \( g_{in} \), and therefore cannot be estimated in the origin constrained model (see Section 3.2). Finally, in the doubly constrained case both the estimated inflows and outflows must equal their observed counter-
parts, i.e. both equations in (4) must be satisfied simultaneously. In this case both \( u_i \) and \( v_j \) are linearly dependent with the balancing factor \( g_{ij} \) and so cannot be estimated.

There are three different approaches to estimating the general interaction model of equation (2): the maximum entropy approach, the maximum utility approach leading to the destination choice model, and the log-linear approach, which is a special case of Poisson regression. All three approaches give identical estimates of the interaction flows in the case where the interacting units are measured on the same level of aggregation and identical sets of independent variables are used to calibrate the model. Nevertheless, the underlying assumptions and functional forms of the three models are quite different, as can be seen from Table 1. These differences are summarized below. Anas (1983) provides a detailed demonstration of the equivalence of the maximum entropy and the maximum utility approaches. The latter approach leads to the statistical multinomial logit model which shares with the log-linear model and Poisson regression the assumption that the dependent variable is sampled from a Poisson distribution. Bishop et al. (1980, pp 446–448) give a proof on the equivalence between estimators based either on the Poisson distribution or the multinomial logit distribution. Thus all three approaches to the general interaction model can be tackled using the log-linear model. In the remainder of this paper, maximum entropy parameter estimates are distinguished by a tilde while the other two are indicated by a hat.

The maximum entropy approach developed by Wilson (1967) involves considering the number of ways \( W(\{\mu_{ij}\}) \) in which a specific set of interactions \( \{\mu_{i1}, \ldots, \mu_{ij}\} \) can occur within an interaction system. This is given by \( W(\{\mu_{ij}\}) = m!/\Pi_{i=1}^{m} \Pi_{j=1}^{m} \mu_{ij}! \). By maximizing this expression with respect to all possible sets \( \{\mu_{i1}, \ldots, \mu_{ij}\} \) which satisfy the appropriate constraint (either one equation from (4) for the singly constrained case or both for the doubly constrained case) we can determine the most likely state \( \{\tilde{\mu}_{i1}, \ldots, \tilde{\mu}_{ij}\} \) of the system. This state is least biased with respect to the missing information and is thus chosen as the maximum entropy estimate, yielding the equations given in Table 1 (for more details see Senior, 1979).

In contrast the maximum utility approach starts from a microeconomic perspective and is based on the behavioral assumption that individuals attempt to maximize their utilities with respect to prospective moves. In the destination choice model each individual is assumed to assign utilities to moving to each of the \( j \) destinations \( (j \neq i) \) and to staying at origin \( i \). Because of individual taste variations and missing relevant explanatory vari-

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Unconstrained Case</th>
<th>Singly Constrained Case (Origin Constraint)</th>
<th>Doubly Constrained Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy Maximization</td>
<td>( \tilde{\mu}_{ij} = \tilde{g} \cdot \tilde{u}_i \cdot \tilde{v}<em>j \cdot \tilde{f}</em>{ij} )</td>
<td>( \tilde{\nu}_{jii} = \tilde{v}<em>j + \tilde{f}</em>{ij} )</td>
<td>( \tilde{\nu}<em>{jii} = \tilde{\lambda}</em>{ij} + \tilde{f}_{ij} )</td>
</tr>
<tr>
<td>Maximum Utility (Destination Choice)</td>
<td>( \text{not applicable} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-linear Model (Poisson Regression)</td>
<td>( \ln(\tilde{\mu}_{ij}) = \tilde{\lambda} + \tilde{u}_i \cdot \tilde{v}<em>j + \tilde{f}</em>{ij} )</td>
<td>( \ln(\tilde{\mu}<em>{ij}) = \tilde{\lambda} + \tilde{\lambda}</em>{ij} + \tilde{f}_{ij} )</td>
<td>( \ln(\tilde{\mu}<em>{ij}) = \tilde{\lambda} + \tilde{\lambda}</em>{ij} + \tilde{f}_{ij} )</td>
</tr>
</tbody>
</table>
ables, the utilities of all destinations conditional upon a given origin \( i \) can be parameterized only up to a random component \( \varepsilon_{ji} \), i.e. \( U_{ji} = V_{ji} + \varepsilon_{ji} \) with the random utility \( U_{ji} \) and the systematic utility \( V_{ji} = \tilde{v}_j + \tilde{f}_j \). Note that in this specification, the function of origin characteristic \( \tilde{u}_i \) is irrelevant because it is constant over all destinations given a specific origin. If the random error terms \( \varepsilon_{ji} \) are independently and identically extreme value Gumbel distributed, then the expected probability \( \tilde{\pi}_{ji} \) that \( U_{ji} \) maximizes the utilities of the individuals in region \( i \), i.e. they choose destination \( j \), can be written as

\[
\tilde{\pi}_{ji} = \Pr(U_{ji} > U_{j'i}, \forall j' \neq j) = \frac{\exp[V_{ji}]}{\sum_{j' = 1}^{J} \exp[V_{j'i}]} \forall i \in I.
\]

(5)

This is the so-called multinomial logit specification (for more details see Ben-Akiva and Lerman, 1985). The expected interactions are calculated by means of \( \hat{\mu}_{iq} = \hat{\pi}_{iji} \cdot m^0 \). For each origin, the destination choice probabilities in equation (5) sum to 1 so that the origin constraint of the singly constrained case is automatically satisfied, and therefore the unconstrained case is not applicable (see Table 1). The destination characteristic function \( \tilde{v}_j \) is replaced by a set of destination specific constants \( \hat{\lambda}_j^0 \) in the doubly constrained case.

The third approach involves making a purely statistical assumption about the distribution of the \( m_{ij} \)'s in \( M_{i \times J} \). If we assume that the cell entries in \( M_{i \times J} \) are independently Poisson distributed with \( \mu_{ij} \) being the expected value, i.e.

\[
\Pr(m_{ij}) = \frac{\mu_{ij}^{m_{ij}} \cdot \exp(\mu_{ij})}{m_{ij}!},
\]

(6)

the model parameters can be estimated by means of the maximum likelihood method (Kirby, 1974). Since the same assumptions and estimation procedures are used in log-linear models, the generic types of general interaction model can also be specified in the log-linear form as can be seen in Table 1. The terms \( \hat{\lambda}, \hat{\lambda} + \hat{\lambda}_j^0, \) and \( \hat{\lambda} + \hat{\lambda}_j^0 + \hat{\lambda}_j^0 \) specify the balancing factors of the three different generic types.

3. SPECIFICATION OF THE GENERAL INTERACTION MODEL AS A LOG-LINEAR MODEL

Since both the general interaction model and the broad statistical class of log-linear models rest on the same assumption that the cell entries are Poisson distributed, the interaction matrix \( M_{i \times J} \) can be conceived of as a contingency table. Thus all statistical concepts, methodology and terminology of log-linear models are also applicable to interaction models.

Log-linear models can be applied to cross-tabulated data from two different perspectives: (a) testing whether a particular categorical classification scheme of the data is in any respect meaningful, or (b) modeling the cell-entries as dependent variables by means of factors and explanatory co-variables which parallels regression analysis. Our attention in this paper is directed towards the second viewpoint. We treat the set of \( \{m_{ij}\} \) as dependent variables, the vector \( \hat{\lambda}_j^0 \) of origin-specific constants and the vector \( \hat{\lambda}_j \) of destination-specific constants as balancing factors, and the explanatory variables as co-variables.

This section starts off with the specification of the doubly constrained interaction model. The only co-variable associated with the cross-regional function \( \tilde{f}_j \) is the distance \( d_{ij} \). 

between the regions. The design-matrix of this model allows examination of the linear dependency problem. Also the log-linear model's fulfillment of the constraints demanded by the doubly constrained interaction model can be verified in the design-matrix. Structural zeros leading to quasi-log-linear models can be modeled by assigning a zero weight to specific cells of the interaction matrix and therefore excluding these cells from the analysis. With these prerequisites we are ready to present the code on how to estimate the general interaction model in SPSS using the LOGLINEAR procedure.

3.1 The Design Matrix for the Doubly Constrained Interaction Model

To stay in line with the proposed regression perspective, the interaction matrix $M_{I\times J}$, the matrix of the expected interaction volume, and the distance matrix need to be vectorized. This will be done in a row by row fashion, i.e. $m = (m_{i1}, \ldots, m_{iJ}, \ldots, m_{l1}, \ldots, m_{lJ})^T$, $\hat{\mu} = (\hat{\mu}_{i1}, \ldots, \hat{\mu}_{iJ})^T$, and $x_i^T = (d_{i1}, \ldots, d_{iJ})$. Thus each vector consists of $K = I \cdot J$ elements with the alternative index $k = (i - 1) \cdot J + j$ defining these elements. Furthermore, we explicitly specify the co-variable functions as linear combinations with a common parameter structure shared by all origins and destinations, i.e. $\hat{u}_i = x_i^T \cdot \hat{\lambda}$; $\hat{v}_j = x_j^T \cdot \hat{\lambda}$; and $\hat{f}_{ij} = x_i^T \cdot \hat{\lambda}$ where $x_i$ and $x_j$ are the characteristics of the origin $i$ and destination $j$, respectively, and $x_{ij}$ measures the separation of both regions.

To set up the doubly constrained interaction model, each of the $K$ origin-destination combinations is assigned to a set of indicator variables for the relevant factor levels $i$ and $j$. In addition, the distance $d_{ij}$ needs to be included as a metric co-variable to calibrate the cross-regional function $\hat{f}_{ij}$, i.e. $\hat{f}(d_{ij}) = \hat{\lambda}_d \cdot d_{ij}$. The interaction literature uses different forms of this function to estimate the impact of the interaction costs (see Kirby, 1974, Scholten and van Wissen, 1985, and Fotheringham and O’Kelly, 1989, for overviews). The co-variable, together with a unity vector for the overall reference parameter $\lambda$ and the indicator variables, define the $K \times L$ dimensional design matrix $Z$ where $L$ is the number of columns of the design matrix. For a 3 by 3 interaction system the regression model for $\ln(\hat{\mu}) = Z \cdot \hat{\lambda}$, with design matrix $Z = (1, \beta_1^T, \beta_2^T, d_{ij})$ and the parameter vector $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5, \hat{\lambda}_6, \hat{\lambda}_7, \hat{\lambda}_8)$, gives

$$
\begin{pmatrix}
\ln(\hat{\mu}_{i1}) \\
\ln(\hat{\mu}_{i2}) \\
\ln(\hat{\mu}_{i3}) \\
\ln(\hat{\mu}_{j1}) \\
\ln(\hat{\mu}_{j2}) \\
\ln(\hat{\mu}_{j3}) \\
\end{pmatrix} = 
\begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & d_{12} \\
1 & 1 & 0 & 0 & 0 & d_{13} \\
1 & 0 & 1 & 1 & 0 & d_{21} \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & d_{23} \\
1 & 0 & 0 & 1 & 0 & d_{31} \\
1 & 0 & 0 & 0 & 1 & d_{32} \\
1 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix} \cdot
\begin{pmatrix}
\hat{\lambda}_1 \\
\hat{\lambda}_2 \\
\hat{\lambda}_3 \\
\hat{\lambda}_4 \\
\hat{\lambda}_5 \\
\hat{\lambda}_6 \\
\hat{\lambda}_7 \\
\hat{\lambda}_8 \\
\end{pmatrix}.
$$

(7)

A single row of this design matrix is denoted by the vector $z_k = (1, \beta^T, d_{ij})$ with $\beta$ being an indicator vector having 1 at the $i$-th position and zeros otherwise. The natural logarithm of the expected cell entry of the interaction matrix $M_{I\times J}$ is expressed linearly by $\ln(\hat{\mu}_{ki}) = z_k^T \cdot \hat{\lambda}$.

Without loss of generality we assume that the last indicator variable for either the origin or destination factors has been suppressed. Since the third origin and destination indicator variables $\beta_3$ and $\beta_3$ are missing in the design matrix $Z$, the overall reference parameter
\( \hat{\lambda} \) is identical to \( \ln(\hat{\mu}_{ij}) \). This matrix reflects the so-called cornered effect coding. This specification of the design matrix will prove to be especially helpful in the discussion of the linear dependency problem in section 3.2 and the explicit satisfaction of the interaction constraints in section 3.3. For the purpose of parameter interpretation, we prefer the contrast specification introduced in section 5 where both specifications of the design matrix are also compared.

### 3.2 Potential Linear Dependencies Within the Design Matrix \( \mathbf{Z} \)

Two potential situations of linear dependencies lead to an unestimable model: (a) factor level dependencies and (b) more important, factor/co-variable dependencies. Factor level dependency deals with the well-known fact that the reference factor associated with the unity vector and a completely specified set of factor levels are linearly dependent. This can be readily shown by introducing the third factor level for either the origin or destination specific constants. For instance, let \( \mathbf{j}_o^o = (0, 0, 0, 0, 0, 1, 1, 1)^T \) be the origin specific vector for the third origin. Then the third origin specific vector is perfectly linearly dependent with the unity vector and the other two origin specific vectors by means of \( \mathbf{j}_o^o = \mathbf{1} - 1 \cdot \mathbf{j}_o^o - 1 \cdot \mathbf{j}_o^o \). The same arguments hold for a fully specified set of destination specific constants. Thus one level of each origin or destination balancing factors is redundant and must be dropped.

Factor/co-variable dependencies arise in situations where the origin characteristic function \( \hat{u}_i \) is used together with origin specific constants. The same applies for the combination of destination characteristic function \( \hat{v}_j \) and destination specific constants. To illustrate this linear dependency let us define the origin characteristic function \( u_i \) in terms of the population size \( \hat{u}_i = \hat{\lambda}_i \cdot p_i \). To model \( \hat{u}_i \) we have to expand the design matrix by the vector \( \mathbf{x}_o^o = (p_1, p_1, p_1, p_2, p_2, p_2, p_3, p_3)^T \), which reflects the interaction potential of each origin. This vector is perfectly linearly dependent with the unity vector and a proper specified set of origin specific constants because \( \mathbf{x}_o^o = p_3 \cdot \mathbf{1} + (p_1 - p_3) \cdot \mathbf{j}_o^o + (p_2 - p_3) \cdot \mathbf{j}_o^o \).

Often only origin characteristics and destination attributes are available to the analyst. As outlined above, the doubly constrained model does not allow use of singly indexed variables per se. Thus the analyst might be tempted to define a doubly indexed variable by using cross-regional differences. For instance, the differential population size may be defined by \( \mathbf{x}_p^o - \mathbf{x}_p^o \), with \( \mathbf{x}_p = (p_1, p_2, p_3, p_1, p_2, p_3, p_1, p_2, p_3)^T \). This specification does not work because when \( \mathbf{x}_p^o \) is linearly dependent on the set of origin specific constants and \( \mathbf{x}_p^o \) is linearly dependent on the set of destination specific constants, then their difference must be linearly dependent with the combination of both factor levels. Nevertheless, a cross-regional combination of the origin characteristics vector \( \mathbf{x}_p^o \) and the destination characteristics vector \( \mathbf{x}_p^o \) is estimable if these vectors are linked by elementwise division \( \mathbf{x}_p^o = \mathbf{x}_p^o \div \mathbf{x}_p^o \).

The specifications of the generic types of general interaction model are limited due to the factor-co-variable linear dependencies. For proper specification in terms of the balancing factors, and the co-variables functions \( \hat{u}_i \), \( \hat{v}_j \), and \( \hat{f}_{ij} \) see Table 1. The unconstrained model, which only includes the unity vector, is equivalent to a Poisson regression model. Since here neither origin nor destination specific constants are part of the model, the problem of factor/co-variable linear dependency does not apply.
3.3 The Implementation of the Constraints Within the Log-Linear Model

Let the expected interactions $\hat{\mu}$ be specified at the maximum of the log-likelihood function for a Poisson regression model, i.e.

$$\ln L(\lambda^1, \ldots, \lambda^K | m_1, \ldots, m_K) = c - \sum_{k=1}^{K} \exp(z_k^T \lambda) + \sum_{k=1}^{K} m_k \cdot z_k^T \lambda$$ (8)

with a constant term $c$. Inspecting the first derivatives of this log-likelihood function at the maximum with respect to the single parameters $\hat{\lambda}_k$, i.e. $\Sigma_{k=1}^{K} \hat{\mu}_k \cdot z_{k\ell} = \Sigma_{k=1}^{K} m_k \cdot z_{k\ell}$, it can be verified that the doubly constrained interaction model, set up by equation (7), satisfies both constraints (4) defined in section 2. Actually, the outflow and inflow totals of each region can be written as vector products, i.e. $m^0 = m^T \cdot \Phi^0$ and $m^\ell = m^T \cdot \Phi^\ell$, respectively. Thus the first derivatives of the log-likelihood function ensure that the following specific constraints

$$m^0 = \hat{\mu}^T \cdot \Phi^0$$ for the origin constraint, and

$$m^\ell = \hat{\mu}^T \cdot \Phi^\ell$$ for the destination constraint

hold and in general the equivalence

$$m^T \cdot z_\ell = \hat{\mu}^T \cdot z_\ell$$ (9)

for any co-variable $z_\ell$ is satisfied. The outflow and inflow totals constraints for the dropped balancing factor levels are implicitly satisfied due to the factor level dependency. Furthermore, by rearranging the equation of the first derivatives it can be shown that the residuals $(m_{ij} - \hat{\mu}_{ij})$ over each origin and destination are unbiased, i.e. $(m - \hat{\mu})^T \cdot \Phi^0 = 0$ and $(m - \hat{\mu})^T \cdot \Phi^\ell = 0$, respectively.

3.4 Quasi-Log-Linear Models and Structural Zeros

Quasi-log-linear models are defined by the property that for theoretical reasons specific cells of an interaction or contingency matrix are of no interest to the analyst or not observable on conceptional grounds. This leads to an analysis of a subtable generated from the complete table by a priori suppression of these cells, which are called structural zeros. Structural zeros should be clearly differentiated from random zeros, which arise due to sampling variations of the underlying Poisson process. Thus observed zero flows, i.e. $m_{ij} = 0$, should remain in the analysis. Cells in an interaction matrix may be excluded from analysis in several situations. Intra-regional flows are often not available to the analyst and even if they were, he/she might prefer to exclude intra-regional flows because otherwise their magnitude would dominate the whole interaction system. Also off-diagonal flows can be irrelevant. For instance, flows measured along a given network are only available between adjacent nodes of this network.

The effect of structural zeros can be easily demonstrated by use of equation (7). The related rows are excluded from the analysis. Thus the estimated parameters and all statistics are no longer affected by the information in these rows. The degrees of freedom have to be reduced accordingly by one for each structural zero since these observations are no longer part of the analysis. Internally, the LOGLINEAR procedure uses a weight function $\varphi_{ij}$ specific to each observation. By default this function is set to $\varphi_{ij} = 1$ for all $K$ observations. To exclude an observation $\varphi_{ij}$ has to be set to zero. For instance, to exclude all observations on the main-diagonal of $M_{ij}$ from the analysis a weight function $\varphi_{ij}$ with
\[ \varphi_{ij} = \begin{cases} 1 & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases} \] (10)

has to be used.

Nevertheless, more flexible weighting functions \( \varphi_{ij} \) are also conceivable and can be implemented in SPSS's LOGLINEAR procedure. They can deal with situations where the variances of the \( m_{ij} \)'s are larger than their expectations. This may stem from violations of the independence assumption inherent in the underlying Poisson process generating the single observations \( m_{ij} \). This kind of overdispersion can be modeled by the compound Poisson distribution. For discussions of this overdispersion from the interaction modeling point of view see Congdon (1992, 1993).

4. ADAPTON OF SPSS'S LOGLINEAR PROCEDURE TO CALIBRATE INTERACTION MODELS

This section illustrates the estimation and specification of the different generic types of interaction models within SPSS's LOGLINEAR procedure. For a general interpretation of results from log-linear models see section 5 or Chapter 6 of SPSS's "Advanced Statistics User Guide" (1990). Here it is our intention to enable the reader to realize the full potential of the LOGLINEAR procedure in interaction modeling and to gain some experience of the model behaviour. Consequently, we offer no substantial interpretation of the real data set which is examined below and instead provide some simple simulation experiments.

4.1 Annotated Program Setup

Here we make use of a data set given by Upton and Fingleton (1989, pp. 148–152; see also Ledent, 1985) which describes the interprovincial migration flows between the 10 Canadian provinces for the accounting period 1976 to 1981. The intra-provincial flows are not given and therefore the main-diagonal of this migration matrix has to be excluded from the analysis. The interprovincial distance matrix is based on the distance in miles between the main provincial cities. Since the provincial populations for 1976 are not given in Upton and Fingleton we provide them here (NF: 557,725; PEI: 118,229; NS: 828,571; NB: 677,250; QU: 6,234,445; ON: 8,264,465; MA: 1,021,506; SA: 921,323; AL: 1,838,031; BC: 2,466,608). For an initial interpretation of the doubly constrained interaction model see Upton and Fingleton (1989). They used GLIM to calculate the parameters of this model with Newfoundland as the reference category. Their results and the estimates of SPSS’s LOGLINEAR procedure are identical up to rounding differences.

For the following discussion we assume some knowledge of the SPSS system. Figure 1 gives the program syntax to analyze several specifications of interaction models. Line 2 defines the input variables: i is an origin and j is a destination index, both ranging from 1 to 10 for the 10 provinces; the variable \( m_{ij} \) is the actual interaction flow between provinces \( i \) and \( j \); \( d_{ij} \) reflects the interprovincial distance, and the number of people living at the origin or destination is denoted by \( p_i \) and \( p_j \), respectively. Enclosed by a BEGIN DATA statement and an END DATA statement, the vectorized migration and distance matrices, the origin and destination indices, and population sizes are read in between lines 3–104 in a free format. Each of the 100 input lines represents aggregated data on one specific origin and destination combination. Lines 106–109 document the variables
and label the nominal scaled province indicators i and j. Several co-variables are generated in lines 112–114. Line 112 specifies regional population differentials pij as a ratio of the origin to the destination population. This will later be used to doubly constrained interaction model. Line 113 defines the distance variable in a negative exponential form whereas in line 114 the distance is specified in its Pareto form. At the moment this transformation is commented out.

Lines 117–118 contain the statements needed to generate the indicator function \( \varphi_{ij} \) defined by equation (10) which will be used later to exclude from the analysis the cells on the main-diagonal of the migration matrix. The WEIGHT BY statement in line 121 is of paramount importance. Because we are dealing with aggregated interaction data, we have to inform the LOGLINEAR procedure how many individual cases there are for each origin-destination
combination. This is done by means of the WEIGHT BY mij statement which tells the SPSS system that each data record in lines 4–103 exists \( m_j \) times. For data input on a disaggregated basis, the co-variables are automatically averaged within each cell by the LOGLINEAR procedure (see Chapter 6.22 of SPSS'S "Advanced Statistics User Guide", 1990).

The LOGLINEAR procedure for five different interaction model specifications is given in lines 123–132. The LOGLINEAR command in line 123 is followed by the basic definition of the model. The origin index \( i \) and the destination index \( j \) with 10 categories each set up the \( 10 \times 10 \) interaction matrix. In statistical terms these factors define the cross-classification table. Separated by the keyword WITH, a list of potential co-variables follows. Note that the LOGLINEAR procedure can actually be tricked into doing simple Poisson regression analysis by assigning to each data record an individual class index; i.e. by use of the system variable $CASENUM an index for each case can be computed with the statements COMPUTE index=CASENUM. Let us assume that we know a priori that there are \( K \) records then the LOGLINEAR command to do Poisson regression is specified as LOGLINEAR index

\[ (1, K) \text{ WITH} \ldots \text{ .} \]

The information about cell-weights is passed to the Loglinear procedure in the subcommand CWEIGHT. In line 124 the cells on the main-diagonal are excluded from the analysis by means of the pre-defined variable cwt. If the subcommand CWEIGHT is omitted all cells are included automatically. In the PRINT subcommand (line 125) additional output information can be requested by the user. For this analysis a print-out of the estimated interactions \( \hat{\mu}_{ij} \) and their associated residuals, as well as the estimated parameters \( \hat{\lambda} \) and the design matrix \( Z \) are given. With the CONTRAST subcommands (lines 126 and 127) a specific form of the design matrix is requested for each categorical factor \( i \) and \( j \) from the basic model definition in line 123. To be able to compare our results for the doubly constrained interaction model with those given by Upton and Fingleton, the cornered effect coding indicated by the keyword SIMPLE has been chosen with Newfoundland (category 1) set as the anchor cell.

The three generic types of interaction models are specified by the DESIGN subcommands. The first model in line 128 is the plain doubly constrained interaction model which was also analyzed by Upton and Fingleton. Besides having origin and destination specific factors the only variable making up the cross-regional function \( f_{ij} \) is the distance between the origin and destination. The equivalent log-linear model is

\[ \ln(\hat{\mu}_{ij}) = \hat{\lambda} + \hat{\lambda}_1 \cdot d_{ij} \]

The second model in line 129 is a so-called augmented doubly constrained interaction model because in addition to the origin and destination distances it includes the interprovincial population differentials \( p_{ij} \) as an additional co-variable. Recall from section 3.2 that population differences cannot be used because they would be mutually collinear with the origin and destination specific factors. These two variables setup the cross-regional function \( f_{ij} \) and the equivalent log-linear model is

\[ \ln(\hat{\mu}_{ij}) = \hat{\lambda} + \hat{\lambda}_1 \cdot d_{ij} + \hat{\lambda}_2 \cdot p_{ij} \]

The third model in line 130 defines a simple origin constrained gravity model. In this model the destination specific balancing factor has been suppressed and therefore singly indexed variables relating to the destinations, like the destination population \( p_i \), can model the pulling power the destinations. In log-linear terms this model is written as

\[ \ln(\hat{\mu}_{ij}) = \hat{\lambda} + \hat{\lambda}_1 \cdot d_{ij} + \hat{\lambda}_2 \cdot p_i \]

The fourth model in line 131 is a refinement of the third model with origin specific distance decay parameters. The keyword BY generates \( I - 1 \) origin specific distance variables \( d_{iii} \) by elementwise multiplication of the distance variable \( x_{ij} \) with the \( I - 1 \) origin indicator variables \( \delta_i^j \), i.e. the single elements \( d_{iii} \) of \( x_{ij}^j \) are zero except for origin \( i \). Thus, in combination with a common distance parameter, the effect of the distance for each origin is modelled separately by its associated origin specific distance parameter. Its specification in log-linear terms is
\[ \ln(\hat{\mu}_{ij}) = \hat{\lambda} + \hat{\lambda}^o + \hat{\lambda}^d \cdot d_q + \hat{\lambda}^p \cdot p_l + \hat{\lambda}^r \cdot p_j. \] 

An interpretation of this row-effect model is given in SPSS’s “Advanced Statistics User’s Guide” (1990, Chapter 6.14, pp. 189–191). In addition, for an introductory discussion of this “spatial structure effect in spatial interaction modelling” reflected by spatial variation of origin specific distance coefficients, see Fotheringham (1991). The last model in line 132 is the simple unconstrained interaction model. This model does not have origin or destination specific balancing factors and therefore both the singly indexed origin and destination attributes are allowed as explanatory variables. The equivalent log-linear model is \[ \ln(\hat{\mu}_{ij}) = \hat{\lambda} + \hat{\lambda}^d \cdot d_q + \hat{\lambda}^p \cdot p_l + \hat{\lambda}^r \cdot p_j. \]

This is the general Poisson regression model which is free of internally defined categorical indicator variables. Note that this dichotomy between Poisson regression and log-linear models parallels that of ordinary regression analysis and the analysis of co-variance.

To clarify the relationship between the expected interactions \( \{\hat{\mu}_{ij}\} \), the underlying interactions \( \{\mu_{ij}\} \) of the statistical population and the Poisson distributed observed interactions \( \{m_{ij}\} \) we give a small SPSS input program\(^1\) (see Fig. 2). It allows the simulation of a \( 5 \times 5 \) random interaction matrix \( M_{\times J} \) which has a pre-defined parameter structure \( \lambda \).

```
1: SET MXLOOPS = 1000.      /* MXLOOPS must be greater than mu_ij */
2: INPUT PROGRAM.
3: LOOP #i = 1 TO 5.
4:   + LOOP #j = 1 TO 5.
5:     + COMPUTE i=#i.
6:     + COMPUTE j=#j.
7:     + COMPUTE cwt = 1.
8:     + IF (i = j) cwt = 0.
9:     + COMPUTE dij = ABS(i-j). /* Distances along a line */
10:    
11:     + COMPUTE #lambda = 1. /* Calculate the expectation mu_ij */
12:     + COMPUTE mu_ij = exp(#lambda + 0.5*i + 0.5*j - 0.25*dij).
13:     
14:     + COMPUTE m_ij = 0. /* Generate random m_ij = Poisson(mu_ij) */
15:     + COMPUTE #sum = 0.
16:     + LOOP.
17:     + COMPUTE #sum = #sum - LN(UNIFORM(1)).
18:     + IF (#sum LT mu_ij) m_ij = m_ij + 1.
19:     + END LOOP IF (#sum GE mu_ij).
20:     
21:     + END CASE.
22:     + END LOOP.
23:     + END FILE.
24:     + END INPUT PROGRAM.
25:     
26:     + COMPUTE m_ij = m_ij + cwt. /* Exclude main-diagonal */
27:     + WEIGHT BY m_ij. /* Generate cases for cells */
28:     
29:     + CROSSTABS /* Print interaction table */
30:     /TABLES= i BY j
31:     /CELLS= COUNT ROW COLUMN.
```

Figure 2  SPSS input program to generate Poisson distributed interaction flows.

\(^1\) Note that this code will not work in the SPSS/PC+ version because it does not permit input programs.
Note that for simplicity, the five interacting regions are arranged at equally spaced locations along a line. This permits the interregional distances to be calculated as the absolute differences of the regional indices in line 9. Line 12 calculates the interaction $\mu_{ij}$ of the statistical population on the basis of the regional indices and a distance variable. In lines 14–19 random Poisson distributed variables $m_{ij}$ are generated which vary around $\mu_{ij}$ of the underlying statistical population. The generation of the random interaction $m_{ij}$ uses the inverse relationship between the expectation of the Poisson distribution and the expectation of the exponential distribution. Here the exponential distribution function $Pr(Y \leq y)$ is randomly generated by a uniformly distributed variable $U(0, 1)$, i.e. $Pr(Y \leq y) = U(0, 1)$. Then this is solved for the exponentially random distributed variable $y$ with expectation 1 by means of the inverse distribution function. A detailed description of this random number algorithm can be found in the SPSS's "Keywords" (1992). In Line 27 the main-diagonal is set to zero and in line 28 the actual number of cases is generated by means of the WEIGHT BY statement. Finally, in lines 29–31 the generated random interaction matrix $M_{1x1}$ is depicted by means of the CROSSTABS procedure.

To get a feeling on how to interpret the results from a log-linear interaction model we suggest starting off by looking at the statistics based on the systematic component $\mu_{ij}$, i.e. use WEIGHT BY $m_{1x1}$. One can vary the parameters in line 12 to see how the estimates of the origin and destination specific balancing factors respond. Also the effects of different choices of the coding scheme, the anchor cell and the inclusion or exclusion of the main-diagonal can be investigated. In addition, deliberately misspecified models with suppressed factors and co-variables can be calibrated to see the impact on the estimated parameters and the goodness of fit. After one has developed some experience in this deterministic setting the impact of the stochastic variation around $\mu_{ij}$, i.e. WEIGHT BY $m_{1x1}$, on the estimated parameters and interactions can be investigated under different model specifications.

### 4.2 Special Features of SPSS's Loglinear Procedure

The Loglinear procedure of SPSS uses maximum likelihood theory to estimate the parameters and statistics of the log-linear model and to generate the expected cell frequencies. The underlying assumption is that the cell frequencies are independently Poisson distributed. Internally the Loglinear procedure is implemented as a Newton-Raphson algorithm. Thus the estimation procedure is embedded in a well developed theory and its performance is very robust. Therefore random zeros, which are not tolerated by estimation procedures such as weighted least squares or the log-normal gravity model (see Flowerdew and Aitkin, 1982), do not cause problems in the Loglinear procedure. Only on the first iteration do they have to be replaced by some positive value in order to get an initial estimate for the variance. Any subsequent iteration uses intermediate estimates of $\mu_{ij}$ for the variance (see "SPSS Statistical Algorithms", 1986). Nevertheless, the analyst should be extra careful in evaluating a model which has many expected cell frequencies $\hat{\mu}_{ij}$ below 5, since the asymptotic properties of the statistics rely on the sample size.

Under modern operating systems with virtual memory management even large problems can be handled by the Loglinear procedure of SPSS. This overcomes the physical constraints faced less than a decade ago, for instance, by Aufhäuser and Fischer (1985). The statistics in Table 2 refer to a 486 personal computer running with 33 MHz clock speed and having 16 MB RAM on board. SPSS's release 4.1 for OS/2 and the operating system OS/2 2.0 were used. All calculations were performed in the core memory with
Table 2  Performance of SPSS’s LOGLINEAR Procedure

<table>
<thead>
<tr>
<th>Model</th>
<th>Observations (Model Structure)</th>
<th>Parameters</th>
<th>Memory Requirements</th>
<th>Computing Time (Iterations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100 (10 × 10)</td>
<td>20</td>
<td>27 k</td>
<td>10.7 sec (6)</td>
</tr>
<tr>
<td>B</td>
<td>2,500 (50 × 50)</td>
<td>100</td>
<td>2.15 MB</td>
<td>37 min (4)</td>
</tr>
<tr>
<td>C</td>
<td>6,400 (80 × 80)</td>
<td>160</td>
<td>8.43 MB</td>
<td>4 h (4)</td>
</tr>
<tr>
<td>D</td>
<td>2,200 (10 × 10 × 22)</td>
<td>440</td>
<td>8.28 MB</td>
<td>15 h (4)</td>
</tr>
<tr>
<td>E</td>
<td>16,900 (130 × 130)</td>
<td>260</td>
<td>40 MB</td>
<td>model not calibrated</td>
</tr>
<tr>
<td>F</td>
<td>90,000 (300 × 300)</td>
<td>600</td>
<td>420 MB</td>
<td>model not calibrated</td>
</tr>
</tbody>
</table>

the system working only on the SPSS task. Each model calibrates a doubly constrained interaction model which includes distance as the only metric co-variable. Model D consists of 10 × 10 interaction matrices for 22 consecutive time periods. As can be seen from Table 2 the memory requirements are proportional to the size of the design matrix calculated by the number of observations $K$ times the number of parameters $L$, whereas the computing time per iteration is affected mostly by the number of parameters, as long as everything can be done in the core memory. Medium sized spatial configurations can be analyzed without problem by SPSS/PC+ which allows students to get classroom experience in modeling interaction models. We expect similar performance for the recently released Windows version of SPSS.

The outstanding features of SPSS’s LOGLINEAR procedure are its ability to handle structural zeros and co-variables in a maximum likelihood framework. As far as the authors are aware, the combination of these features in the LOGLINEAR procedure is only reproduced in GLIM. For a comparison of features of log-linear and logit procedures in common statistical packages see Agresti (1990, Appendix A). Another important property of the LOGLINEAR procedure is its ability to calculate adjusted residuals. Adjusted residuals take into account that the standard errors of the residuals themselves are estimates. In general they approximate the standard normal distribution more accurately than Pearson residuals. Since adjusted residuals are larger than the Pearson residuals, they allow detection of outliers more precisely and in turn adjustment of the model structure accordingly. For details about residuals in log-linear models see Wrigley (1985) and Agresti (1990, pp. 432–433).

5. ALTERNATIVE CODING SCHEMES AND PARAMETER INTERPRETATION

Besides the cornered effect coding of the design matrix $Z$ in equation (7), which is also known as dummy variable coding, the parameterization of the design matrix in the centered coding leads to a different interpretation of the estimated parameter vector $\hat{\beta}$. For a detailed discussion of both coding schemes see Wrigley (1985). The design matrix of the centered coding scheme
looks very similar to that of the cornered coding scheme, except for the implementation of contrasts \((-1\)) between all levels of a factor and its excluded third level. In SPSS's LOGLINEAR procedure this coding scheme is the default and is denoted by the keyword DEVIATION in the CONTRAST subcommand. While both coding schemes give different estimates for the parameter vector \(\hat{\lambda}\) they also share some common properties: (a) most important, their estimated interaction vectors \(\hat{\mu}\) are identical; and (b) the parameter differentials within each factor, i.e. either \(\hat{\lambda}_i^p - \hat{\lambda}_i^o\) with \(i \neq i'\) or \(\hat{\lambda}_j^p - \hat{\lambda}_j^o\) with \(j \neq j'\), are the same in both coding schemes.

Implicitly implemented in the cornered coding scheme is the restriction that the excluded factor levels equal zero, i.e.
\[
\hat{\lambda}_i^o = \hat{\lambda}_j^o = 0,
\]
(12)
whereas in the centered coding scheme the suppressed factor levels can be calculated from the included factor levels
\[
\hat{\lambda}_i^p = -\sum_{j=1}^{i-1} \delta_j \cdot \hat{\lambda}_j^p \quad \text{and} \quad \hat{\lambda}_j^o = -\sum_{i=1}^{j-1} \delta_i \cdot \hat{\lambda}_i^o,
\]
(13)
with
\[
\delta_j = \begin{cases} 1 & \text{if at least one cell in column } j \text{ of } M_{i \times j} \text{ is not a structural zero} \\ 0 & \text{otherwise}, \end{cases}
\]
and \(\delta_i\) being analogously defined to control for the situation that a complete row of \(M_{i \times j}\) has been excluded due to structural zeros. Thus \(\delta_i\) and \(\delta_j\) ensure the existence of either \(\hat{\lambda}_i^p\) or \(\hat{\lambda}_j^o\).

Closely associated with these alternative parameter restrictions (12) and (13) is the interpretation of the reference factor \(\hat{\lambda}\) related to the unity vector \(\mathbf{1}\) as well as the single factor levels \(\hat{\lambda}_i^p\) and \(\hat{\lambda}_j^o\).

In the cornered coding scheme \(\hat{\lambda}\) is anchored to cell \((IJ)\) and equals \(\hat{\lambda} = \ln(\hat{\mu}_{ij})\) in the case that this cell has not been suppressed by means of a structural zero. Otherwise would \(\hat{\lambda}\) measure a hypothetical base-level. For this reason we prefer not to use a cell defined as a structural zero as anchor cell. If \(\hat{\lambda}_i^p\) is positive then the outflow from origin \(i\) is greater than that of region \(I\) which is the region attached to the reference cell \((IJ)\). The reverse statement holds if \(\hat{\lambda}_i^o\) is negative. On the other hand if \(\hat{\lambda}_j^p\) is positive then destination \(j\)
faces more inflow than the reference region \( J \) and if it is negative it experiences less inflow. The magnitudes of \( \hat{\lambda}_r^o \) and \( \hat{\lambda}_r^p \) are absolute measures of the flow differences with respect to the reference origin and destination. This means that the addition of new regions to a spatial system does not change these parameters. Since the choice of the anchor cell is at the discretion of the analyst, it is wise to select a well-known combination of origin and destination as reference. When the internal flows are excluded from the analysis we suggest selecting the interaction with the largest magnitude, otherwise we suggest choosing the largest region. Then estimated parameters of the log-linear model are easier to interpret.

In the centered coding scheme \( \hat{\lambda} \) reflects the average interaction within the regional system. Consequently it is called the grand-mean and is defined by \( \hat{\lambda} = 1/(IJ) \sum_{i=1}^{I} \sum_{j=1}^{J} \ln(m_{ij}). \) Both \( \hat{\lambda}_r^o \) and \( \hat{\lambda}_r^p \) model the deviation of the out- and inflows from the grand-mean. Again a positive sign means that either the outflow from a region or inflow to a region is higher than average and a negative sign means that these flows are below average. The magnitudes of \( \hat{\lambda}_r^o \) and \( \hat{\lambda}_r^p \) are relative measures since no real reference is available. By adding other regions to the system the grand-mean will change and consequently so will all the parameters of the balancing factors.

The property of \( \hat{\lambda}_r^o \) and \( \hat{\lambda}_r^p \) being relative measures in the centered coding scheme has actually a major advantage over the cornered coding scheme. The magnitudes of both parameters might either be related to the relative sizes of the origins and destinations or to their attractiveness. These two components cannot be disentangled by looking at the parameters for either factor alone. But under the assumption that size-effects operate the same on region \( i \) regardless of whether it is an origin or a destination, the magnitude of the difference \( \hat{\lambda}_i^o - \hat{\lambda}_i^p \), which reflects the netflow, can be used as an indicator for the attractiveness. A negative differential indicates a region where emissivity dominates whereas a positive differential is associated with an attractive region. Such cross-factor comparison for a common region is only permitted in the centered coding scheme since the parameters do not change by selecting other than the \( I \) and \( J \) factor levels as suppressed categories. When there is a set of interaction matrices available reflecting a time series, both sets of parameters \( \hat{\lambda}_r^o \) and \( \hat{\lambda}_r^p \) with \( t = 1, \ldots, T \) can be plotted simultaneously against time. Scholten and van Wissen (1985) give an example. This allows the display of changes in the attractiveness and interaction volume of region \( i \) over time.

A note of caution must be issued on testing the significance of single \( \lambda \)-parameters within each factor. Single t-values for each parameter are meaningless and the significance of each factor level must be assessed in the context of the other levels of the same factor because the estimated parameters are interdependent. For instance, due to an arbitrary change of the suppressed factor level in the cornered coding scheme, levels may become statistically not different from zero simply because the reference value of the scale has been shifted. Thus one should not drop insignificant factor levels. On the other hand if all factor levels are insignificant one should drop the factor completely.

6. CONCLUSION

The previous discussion has shown two major things: (a) the statistical equivalence of the different theoretical approaches towards the generic types of the general interaction model (see Table 1), and (b) the easy implementation of constrained interaction models in SPSS's LOGLINEAR procedure (see Fig. 1). Also more advanced specifications of the general interaction model like generation-allocation stage models, symmetry models, or
time-series and spatial contiguity models can be calibrated by means of SPSS's LOGLINEAR procedure. Here the design matrix has to be coded explicitly by the analyst. For a discussion of these more advanced models see for instance Snickers and Weibull (1977), Aufhauser and Fischer (1985), Willekens and Baydar (1985), and SPSS's "Advanced Statistics User Guide" (1990, Chapter 6.18, pp. 193–196).

With this powerful estimation procedure, simulations can be done to study the effects of spatial distance, spatial size, spatial configuration and spatial substitution (see for instance Lo, 1991). Also misspecification effects of the interaction model can be directly evaluated by means of simulations. However, research is still needed for situations where the underlying stochastic mechanisms of a Poisson process are violated, i.e. when interacting units within $m_i$ are either clustered (see for instance Congdon, 1993, for a discussion on the overdispersion effect) or the set of {$m_i$} is spatially autocorrelated. In particular, spatially autocorrelated Poisson distributed variables are difficult to handle because there is no closed form of the likelihood function available (see Cressie, 1991). Despite these serious statistical problems, with the proposed methodology, an efficient and broadly available procedure is provided for calibrating constrained interaction models.

Acknowledgements

We are grateful for the comments of an anonymous referee on an earlier draft of this paper.

References


