SERIE RESEARCHMEMORANDA

HYBRID LOG-LINEAR MODELS FOR SPATIAL INTERACTION AND STABILITY ANALYSIS

Floor Brouwer
Peter Nijkamp
Henk Scholten

Researchmemorandum 1983-17 October 1983

VRIJE UNIVERSITEIT
EKONOMISCHE FAKULTEIT
AMSTERDAM

HYBRID LOG-LINEAR MODELS FOR SPATIAL INTERACTION AND STABILITY ANALYSIS

Floor Brouwer [1]
Peter Nijkamp [2]
Henk Scholten [3]

1)
Institute for Environmental Studies
Free University
P.O. Box 7161
1007 MC Amsterdam

2)
Department of Economics
Free University
P.O. Box 7161
1007 MC Amsterdam

3)
National Physical Planning Agency
Willem Witsenplein 6
2596 BK The Hague

This study is partly supported by the Netherlands Organization for the Advancement of Pure Research (ZWO), project number 46-079.

Paper presented at the European Conference of the Regional Science Association, Poitiers, August 1983

Table of Contents

1. Introduction. 1
2. Spatial Interaction Models as Generalized Linear Models 2
3. The Analysis of Incomplete Spatial Interaction Models 5
4. The Analysis of Temporal Stability with Spatial Interaction Models. . . 8
5. An Application to Dutch Regional Migration Data 10
6. Conclusion. 16
References 17

Summary

Hybrid log-linear models are developed for the analysis of spatial interactions with categorical data. Such models may lead to statistically significant results based on a reduced number of parameters.
Special attention is given to the problem of so-called structural zeros and of stability of spatial interaction patterns. Various constraints on the set of interaction parameters emerging from the off-diagonal elements of a spatial interaction matrix are taken into account by means of an entropy function. The analysis is illustrated by means of an empirical application to migration flows.

key-words:

hybrid log-linear models;
generalized linear models;
spatial interaction;
stability analysis;
entropy function.

1. Introduction

In spatial data analysis, much information is frequently non-cardinal (qualitative or categorical) in nature. For instance, surveys often contain categorical data represented in a dichotomous sense, such as: man or woman; married or single persons. These variables can only be distinguished by their names or attributes and are measured on a nominal scale. Categorical data from surveys may be represented in contingency tables. Such a table can be interpreted as a sample from either an independent Poisson distribution or a multinomial distribution. The main focus of this paper will be on interaction analysis of spatial migration flows based on categorical data, followed by an analysis of the stability of such migration patterns.

In section 2 we will treat the well-known spatial interaction models, especially log-linear models, as part of a broad family of generalized linear models. The fit between observed and expected values of the cell-elements in a spatial interaction matrix , conditional to some statistical hypothesis, may be judged by means of an asymptotic chi-squared test-statistic, called the deviance. This interaction analysis will be based on the general gravity model. Special attention will be paid to various possibilities of estimating the spatial distribution function in such a model.

The structure of the above mentioned spatial distribution function will be analysed in section 3. When the diagonal elements are excluded from the analysis, the intra-regional migration flows have zero weights in the estimation procedure, based on prior information. This phenomenon with both theoretical and practical consequences leads to the occurrence of incomplete spatial interaction tables. In this regard, we will discuss four parameter estimation procedures, that have been proposed in the literature on so-called structural zeros. When, on the other hand, the diagonal cell-elements are taken into consideration, the distribution function may be estimated by means of either the well-known general type of the log-linear model or a hybrid type of this model. We will mainly focus the attention on hybrid log-linear models.

In section 4 we will concentrate on the concept of stability of the distribution function , by analyzing both the stability of migration flows themselves and the stability of the underlying spatial structure. The potential of some measures for analysing the stability of the spatial distribution function by means of the log-linear model and a hybrid log- linear model will also be dealt with.

In section 5, the log-linear model is used for analyzing the spatial mobility in the housing market area of the Dutch city of Den Bosch (in the province of Noord-Brabant) for the period 1971-1980 subdivided into five subsequent two-year periods. The intertemporal stability of the spatial pattern is studied by means of a log-linear model that encompasses a dynamic component. Finally, a test-procedure (including empirical results), which is also appropriate for a situation where the number of interaction parameters with a time component becomes large, will be discussed.

2. Spatial Interaction Models as Generalized Linear Models

In this section the analysis of categorical data will be discussed by using so-called generalized linear models as a frame of reference. The generalized linear models consist of a broad family of linear models, made up of the classical regression model, the analysis-of-variance model, the logistic regression model,as well as the log-linear model with contingency table analysis (see also Imrey et al., 1981, 1982; Nelder, 1983; Nelder and Wedderburn, 1972).

A generalized linear model (GLM) can be concisely characterized as follows:

(1) The response (or dependent) variable y is distributed independently, with mean μ and a variance-covariance structure for the response variables determined by the underlying probability distribution function.

(2) A linear relationship, say $E(y)= X\beta$, exists between μ and the set of explanatory (or stimulus) variables; X is a matrix of explanatory variables and β a vector of parameters.

(3) A relationship is defined between y and its mean μ, viz. $\mu = f\ (y)$, where f is called a link-function.

Different types of generalized linear models may be obtained by varying either the distribution function of the response variable and/or the link-function f. The normal regression model for quantitative (cardinal) data will be obtained, for example, if the observations are assumed to be normally distributed with mean μ and variance σ^2, based on a linear link-function. The goodness-of-fit between observed and expected values from such a model, called the deviance D, may be determined by maximizing the log-likelihood function and is defined as:

$$D = \sum_i \frac{(y_i - \mu)^2}{\sigma^2} \qquad (1)$$

The deviance from relationship (1) is the sum of independent standard normal distributed statistics and is an asymptotic chi-squared test statistic with degrees of freedom equal to the number of observations.

The above mentioned GLM for quantitative data can be adjusted in the framework of qualitative data (see also Nelder, 1983). In regard to categorical data for spatial interaction flows, a multiplicative model may be assumed in contingency table analysis, where the logarithm of the expected values of cell-frequencies becomes linear with respect to the classified elements. A GLM for qualitative data is then obtained in the following way. Given categorical data in log-linear models, the normal errors distribution as well as the identity link-function from quantitative models are replaced by the Poisson errors function and the log-link function, respectively. The response variables are assumed to be independently Poisson distributed with mean μ, where $\mu = \exp(y)$.
The likelihood ratio test-statistic $-2 \ln\lambda$, with λ the maximum value of the likelihood function conditional to some given null hypothesis related to the unrestricted maximum value of the likelihood function, or deviance D, is now defined as (see also Mood et al., 1974):

$$D = 2 \sum_i y_i \ln \frac{y_i}{\mu} \qquad (2)$$

In the present section, we will treat the log-linear model as a GLM from a constrained gravity model. The general form of the gravity model is :

$$m_{ij} = k\, O_i\, D_j\, f\,(d_{ij}) \qquad (3)$$

with m_{ij} the volume of flows from region i to region j, O_i and D_j push-factors and attraction-factors respectively, k a constant ensuring the additivity condition , and f (d_{ij}) the spatial distribution (friction or discount) function between region i and region j.
Widely used types of such distance friction functions are the inverse power function, the negative exponential function, or the Tanner function as a combination of the two previous functions (see also Hua and Porrell, 1979).
Model (3) is said to be restricted when the volume of flows leaving a region of origin or arriving at a region of destination are constrained. This restriction is equivalent to the existence of given marginal totals in contingency table analysis. When both the volume of flows leaving a region and those arriving at a region are constrained, we get the double-constrained gravity model, viz.

$$m_{ij} = k\ A_i\ B_j\ O_i\ D_j\ f\ (d_{ij}) \tag{4}$$

where:

O_i = volume of flows leaving region i.

D_j = volume of flows arriving at region j.

A_i, B_j are balancing factors related to the Lagrange multipliers.

k = constant, viz. the geometric average of the flows.

f (d_{ij}) = spatial distribution (friction or discount) function.

When interaction model (4) is linearized by means of a logarithmic transformation we obtain:

$$\log m_{ij} = \log k + \log A_i + \log B_j + \log O_i + \log D_j + \log f(d_{ij}) \tag{5}$$

which can be rewritten equivalently as a conventional log-linear model with

$$\log m_{ij} = u + u^{A}(i) + u^{B}(j) + u^{AB}(i,j) \tag{6}$$

where:

$$u + u^{A}(i) + u^{B}(j) = \log k + \log A_i + \log B_j + \log O_i + \log D_j$$

and

$$u^{AB}(i,j) = \log f\ (d_{ij})$$

If the elements m_{ij} are independently Poisson distributed, model (6) is a log-linear model with a log-link function that belongs to the family of generalized linear models. The parameters are estimated by a maximum likelihood procedure, viz. an iterative weighted least squares approach. The parameter u corresponds to the overall mean-effect, the parameter $u^{A}(i)$ and $u^{B}(j)$ describe the row- and column-effects respectively, and the parameter $u^{AB}(i,j)$ corresponds to the interaction-effect between variable A and variable B.

The multiplicative analogue of the saturated log-linear model in formula (6) becomes:

$$m_{ij} = w\ w_i^{A}\ w_j^{B}\ w_{ij}^{AB} \tag{7}$$

When the structure of the distribution function f (d_{ij}) is analysed, one has to take into account whether or not the diagonal elements are included in the analysis. There are different possibilities for analysing the structure of $f(d_{ij})$ in a log-linear modeling approach:

(1) The intra-regional flows are included; then the classical log-linear model with interaction parameters is employed (see also Brouwer and Nijkamp, 1983; Scholten, 1983).

(2) The intra-regional flows are excluded; then the result is an incomplete spatial interaction model. Then zero-entries called structural zeros occur because of this a priori information. Some theoretical solutions will be discussed in section 3 to obtain parameter estimates for incomplete contingency tables (see also Brouwer and Nijkamp, 1983; Scholten, 1983).

(3) The intra-regional flows are included, while hybrid log-linear models are developed. Hybrid log-linear models arise because of quasi independence restrictions on a set of interaction parameters (see also Willekens and Güvenç, 1983 ; Scholten and Van Wissen, 1983). Hybrid log-linear models include parameters that are not part of the conventional log-linear modeling analysis because of restrictions on a set of parameters. Such models involve a synthesis of (1) and (2), because restrictions arise for the off-diagonal interaction parameters and a simplified interaction structure is obtained . A reduction in the number of interaction parameters can then be obtained by means of the hybrid log-linear model. The goodness- of- fit statistics based on the hybrid log- linear model may also be compared with the classical log-linear model.

We will analyse the hybrid log-linear model by means of an entropy function for the off-diagonal elements. The diagonal elements of the spatial interaction matrix are determined by first-order interaction effects.

$$\begin{aligned} m_{ij} &= w\, w_i^A\, w_j^B\, w_{ij}^{AB} && \text{if } i = j \\ m_{ij} &= w\, w_i^A\, w_j^B \exp[\,-\beta(d_{ij})] && \text{if } i \neq j \end{aligned} \qquad (8)$$

Theoretical results as well as empirical applications of both approaches related to (1) and (3) above will be given in the sequel of this paper.

3. The Analysis of Incomplete Spatial Interaction Models

Multiway contingency tables are called incomplete tables when some cell-entries are ignored or one or more variables are not completely classified. There are two possibilities for the presence of zero elements in contingency tables: fixed and sampling zeros (see among others Bishop et. al., 1975; Fienberg, 1977).

Sampling zeros may occur because of a relative small cell-probability. They may vanish, at least in a theoretical way, when the sample size will be sufficiently enlarged, and need not necessarily be equal to zero. Structural or a priori zeros occur if some cell-elements either do not exist or are excluded from the analysis, and if cell-elements are known to have a priori a zero value.

The a priori knowledge of cell-elements is part of the analysis of incomplete tables. Incomplete tables consist of either some fixed-valued cell-elements or of one or more incompletely classified variables (see also Dempster et al., 1977; Fuchs, 1982).

The analysis of structural zeros in incomplete tables consists of two parts: on the one hand, the computation of the estimated cell-frequencies by means of the marginal totals, and on the other hand the interpretation of the parameter estimates of the log-linear model with incomplete tables.

The first problem can be solved unambiguously when we use an iterative proportional fitting algorithm, the so-called Deming-Stephan procedure with proportional adjustment (see also Bishop et al., 1975). The cell-elements are then partitioned into a set S containing non-zero elements and a set of structural zero elements. The algorithm consists of an initial matrix $N^{(0)}$ with cell-elements n_{ij}:

$$n_{ij}^{(0)} = \begin{cases} 1 & \text{if } (i,j) \in S \\ 0 & \text{in other cases} \end{cases} \quad \text{for all } i,j \tag{9}$$

Restricted maximum likelihood estimates of the cell-frequencies are obtained by bi-proportional adjustment because of the initial condition. Quasi-independency can also be tested (viz. independence of the non-zero cell-elements) (see also Brouwer and Nijkamp, 1983). If the table consists of IJ elements and s structural zeros, only (I-1)(J-1)-s independent elements, called degrees of freedom, remain in the quasi-independence model. The cell-frequencies in the generalized linear model are estimated by means of an iterative weighted least squares approach (based on a maximum likelihood procedure). The main diagonal elements have zero weights in case of structural zeros.

Although the estimation of cell frequencies with the log-linear model of incomplete tables is rather straightforward, the computation and interpretation of the parameter estimates becomes more problematic (see also Willekens, 1983). The interaction-parameters are not computed in a unique way and at least four approaches can be found in the literature that aim at coping with this problem:

(1) Computation of parameters with incomplete tables in an analogous way as with complete tables (see also Willekens, 1983). The restrictions on the parameters are different from those in complete tables and become:

$$\sum_i u^A(i) = \sum_j u^B(j) = 0$$

$$\sum_i S_{ij}\ u^{AB}(i,j) = u^B(j) \qquad (10)$$

$$\sum_j S_{ij}\ u^{AB}(i,j) = u^A(i)$$

The dummy-variable S_{ij} becomes zero if cell-element (i,j) contains a structural zero , and one in all other cases. In order to arrive at the same parameter restrictions for complete and incomplete tables, it has been suggested by Bishop et al.(1975) to give one parameter some arbitrary value.

(2) The cell-frequencies of the table determine the parameter estimates of the log-linear model based on the incomplete table without making use of the cell-elements which contain structural zeros (see also Haberman, 1979).

This analysis is restricted to the unsaturated model. The interaction parameters do not lead to problematic results, because they are zero by definition.

(3) The restrictions from the complete model determine the parameter estimates. The parameter values are obtained by solving a set of equations including the unknown u-parameters. Then a main advantage is that the conventional restrictions hold. A serious disadvantage is the awkward interpretation of the parameter estimates, because the u-terms cannot be simply written as deviations from mean values.

(4) The parameter values with incomplete tables are determined by cross-product ratios (with multiplicative models) or linear contrasts (with linear models) in an analogous way as with complete tables(see also Fienberg and Mason, 1978).

The different parameter estimation procedures with incomplete tables as well as hybrid log-linear models occur because of an identification problem: such models are generally underidentified and an adjusted identification procedure is required to obtain parameter estimates. The number of independent parameters is by definition equal to the number of cell- elements with unrestricted log- linear models so that in that case exactly identified models are obtained. We can conclude that parameter estimates are not obtained in a unique way if in the modeling phase we are confronted with underidentification problems.

4. The Analysis of Temporal Stability with Spatial Interaction Models

The analysis of the structure of $f(d_{ij})$ can also be related to the temporal stability of the spatial pattern of $f(d_{ij})$ by means of the (hybrid) log-linear model.

In this section we will pay attention to the concept of stability and to problems inherent in operationalizing this concept. A literature survey on temporal stability with a wide variety of applications in the field of migration data can be found in Baydar, 1983, where the time dependence of the region of origin and destination is quantified by a log-linear model of migration analysis. A main advantage of the log-linear analysis is its capability to use a dynamic approach, because it allows a simultaneous treatment of migration patterns over a number of periods. The stability of the spatial pattern is analysed by means of the log-linear approach, by focussing the attention on the stability of the temporal interactions with the regions of origin and destination.

There is a difference between temporal stability of migration flows and temporal stability of the spatial pattern. Migration patterns are called stable when all time-dependent interaction parameters are absent.

The spatial pattern may be stable, even if the migration flows themselves are not stable. The stability analysis of flows will now be discussed. The migration matrix in period t and period t + 1 is represented by M_t and M_{t+1} successively. The matrix M_t with elements m_{ijt} consists of the two-dimensional origin-destination elements in period t.

Many distance measures have been developed to analyse stability of migration patterns, for example the mean absolute percentage deviation, defined by:

$$\frac{\Sigma_{i,j} \left| m_{ijt} - m_{ij\ t+1} \right|}{\Sigma_{i,j}\ m_{ijt}} * 100 \tag{11}$$

The distance function in (10) is a real-valued, monotonous distance function which is homogenous of order zero, but a disadvantage is that it has no upper bound (see also Wolpert and Yapa, 1971). The matrices M_t and M_{t+1} are called stable if the distance, defined in formula (10),becomes close to zero. The choice whether two points in time are called stable becomes problematic, because it is based on a subjective choice, viz. the choice of a distance measure.

A coefficient of agreement between M_t and M_{t+1} is defined as (see also Somermeijer, 1961):

$$I = \frac{2\,\Sigma_{i,j}\ m_{ijt}\ m_{ijt+1}}{\Sigma_{i,j}\ (m_{ijt})^2 + \Sigma_{i,j}\ (m_{ijt+1})^2} \tag{12}$$

The coefficient I has values that fall in the range between 0 and 1, because the migration flows are non-negative. A coefficient of agreement has an interpretation different from a coefficient of correlation. When two sets of values are proportional to each other, they are perfectly correlated but they do not agree. A coefficient of agreement indicates whether two matrices take on the same values, i.e. stability of migration flows instead of stability of spatial structure.

The stability of the matrices M_t and M_{t+1} may be tested by means of a chi-squared statistic developed by Pearson and defined as:

$$\chi^2 = \sum_{i,j} \frac{(m_{ijt} - m_{ijt+1})^2}{m_{ijt+1}} \qquad (13)$$

A real disadvantage of the χ^2-value is its proportionality to the sample size $\sum_{i,j} m_{ijt+1}$ and the number of degrees of freedom, which corresponds to the number of cell-elements of the migration-matrix.

The Pearson χ^2-statistic may be partitioned according to regions i of outmigration. The share of the χ^2-value due to this region is then defined as:

$$\chi^2\ (i) = \sum_j \frac{(m_{ijt} - m_{ijt+1})^2}{m_{ijt+1}} \qquad (14)$$

The unstable regions can be identified by relating $\chi^2(i)$ to the value χ^2. Large values of $\chi^2(i)$, when related to the χ^2-value, indicate temporal instability of the migration flows in the i-th region of origin.

In the first part of this section an exploratory phase of stability analysis with migration flows has been developed . A log-linear model of migration flow matrices with a time component will now be proposed in order to analyse stability of a spatial pattern. Suppose a contingency table with cell-elements m_{ijt}. Let variable A and variable B represent the regions of origin and destination respectively, and variable C the time component.

Then a saturated log-linear model with cell-element (i,j,t) becomes (see also Payne, 1977):

$$m_{ijt} = u + u^{A}(i) + u^{B}(j) + u^{C}(t) + u^{AB}(i,j) + u^{AC}(i,t) + u^{BC}(j,t) + u^{ABC}(i,j,t) \qquad (15)$$

The time-dependence of this model consists of the main-effect $u^{C}(t)$ and the interaction effects of time with region of origin and destination, i.e. $u^{AC}(i,t)$, $u^{BC}(j,t)$ and $u^{ABC}(i,j,t)$.

Migration patterns are now called stable when the interaction parameters as well as the main effects have small parameter values and are not related to the time parameter. The separate temporal components are part of the multiplicative analogue of the saturated log-linear model in formula (15), which becomes:

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB}\, w_{it}^{AC}\, w_{jt}^{BC}\, w_{ijt}^{ABC} \tag{16}$$

The temporal dependence of the migration flows is represented by the parameter $w^C(t)$, while the parameters $w^{AC}(i,t)$ and $w^{BC}(j,t)$ measure temporal stability of the out-migration flows and in-migration flows successively. The temporal stability of the spatial interaction pattern is described by the second-order interaction effects $w^{ABC}(i,j,t)$.
The hybrid log-linear model becomes:

$$\begin{aligned} m_{ijt} &= w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB}\, w_{it}^{AC}\, w_{jt}^{BC}\, w_{ijt}^{ABC} && \text{if } i = j \\ m_{ijt} &= w\, w_i^A\, w_j^B\, w_t^C\, w_{it}^{AC}\, w_{jt}^{BC} \exp[-\beta(d_{ij})]\, \exp[-\beta t(d_{ij})] && \text{if } i \neq j \end{aligned} \tag{17}$$

The temporal stability of the different parameters can be tested in two ways. On the one hand the significance of a parameter is tested by means of its standard error. On the other hand the significance of an added parameter is tested by means of the likelihood ratio-test statistic because of its additivity character. The likelihood ratio test statistics will be applied in the following because they give a goodness- of- fit of the log- linear model.
In the following section , we will deal especially with the temporal stability of the spatial pattern because of its relevance with log-linear modeling analysis.

5. An Application to Dutch Regional Migration Data

The foregoing analyses will be applied to migration data to the Dutch city of Den Bosch. The study area consists of 14 municipalities, located as given in maps 1 and 2. A regional master plan for the area was initiated by the provincial government in 1978. The 14 municipalities had to be reduced in number as a consequence of the regional master plan. Some concentrated urban centers and regional centers appear to provide four clusters of municipalities, viz. a) Helvoirt, Esch and Liempde, b) Berlicum and Den Dungen, c) Heeswijk - Dinther and d) St. Michielsgestel. The regional master plan finally gives 10 municipalities. The period 1971-1980 was subdivided into five two-year periods. The inter-municipal migration flows as well as the number of inhabitants on January 1, 1980, are given in table 1.

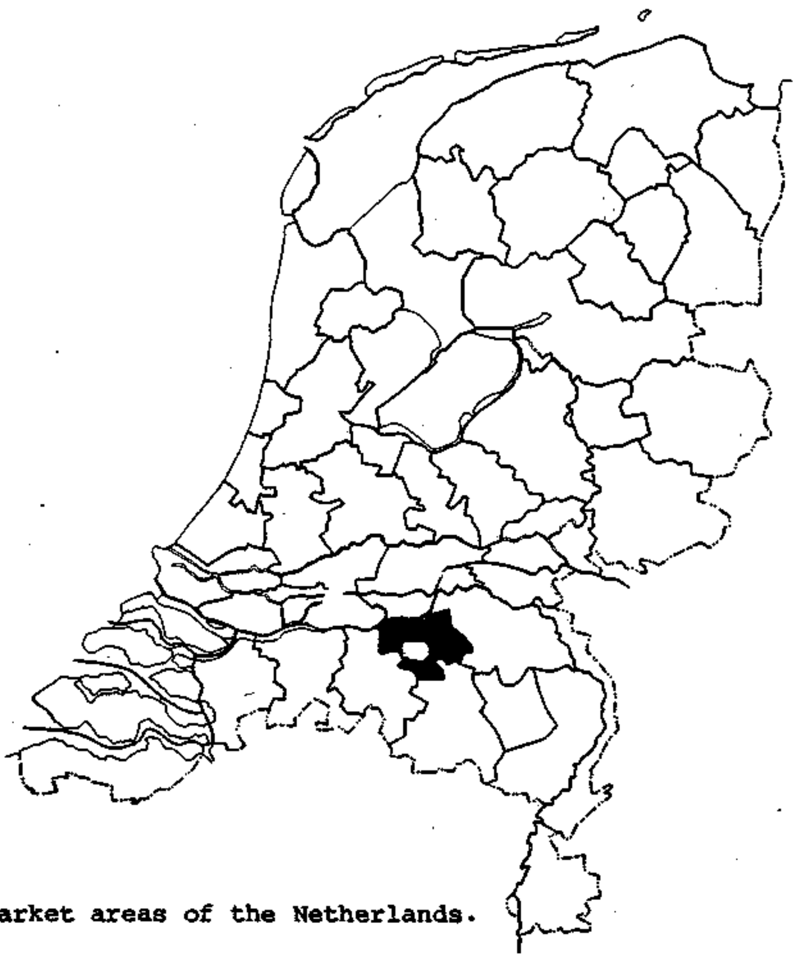

Fig. 1. The housing market areas of the Netherlands.

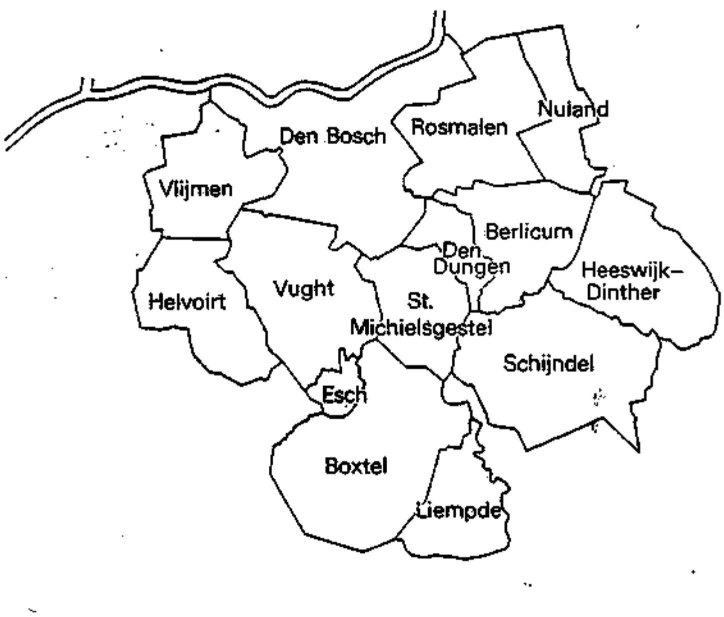

Fig. 2. The municipalities in the housing market area of Den Bosch.

Inhabitants per 1-1-1980		1971-1972 in	1971-1972 out	1973-1974 in	1973-1974 out	1975-1976 in	1975-1976 out	1977-1978 in	1977-1978 out	1979-1980 in	1979-1980 out
1.	12300	771	460	698	583	460	514	482	377	416	405
2.	24030	531	612	843	607	631	535	665	444	475	445
3.	11525	360	296	391	243	388	236	279	206	280	178
4.	10150	377	274	469	346	267	314	323	262	321	287
5.	87900	2138	3507	2042	3646	2251	3066	1886	2522	1756	2157
6.	22550	1595	900	1517	1174	1471	1139	1388	880	1144	796
7.	19365	562	352	662	382	556	398	571	377	372	337
8.	12060	592	529	658	650	516	541	345	572	339	440
9.	13875	454	389	581	311	370	380	349	290	387	269
10.	23580	1199	1269	1393	1312	1332	1119	887	1145	736	912

1 = Berlicum and Den Dungen.
2 = Boxtel.
3 = Heeswijk and Nuland.
4 = Helvoirt, Esch and Liempde.
5 = Den Bosch.
6 = Rosmalen.
7 = Schijndel.
8 = Sint-Michielsgestel.
9 = Vlijmen.
10 = Vught.

Table 1. The number of inhabitants and inter-municipal migration flows within the study-area.

There are different computer packages to analyse conventional log-linear models, e.g., ECTA (Everyman's Contingency Table Analysis) and GLIM (Generalized Linear Interactive Modelling). A disadvantage of these packages is the impossibility of an overall test of the time parameters, due to dimensional problems. For that reason, smaller time dimensions have been chosen to analyse temporal stability, viz. pairs of two-year time periods. The goodness - of-fit which follows from the addition of parameters in a hierarchical log-linear model is tested by means of the likelihood-ratio-test with the additivity characteristic (see also Payne, 1977 for a proof of the addititivity of the likelihood-ratio-test). The empirical results of five log-linear models and three hybrid log- linear models are given in table 2. The models used are:

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C \qquad i,j,t \qquad (i)$$

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB} \qquad i,j,t \qquad (ii)$$

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{it}^{AC} \qquad i,j,t \qquad (iii)$$

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{jt}^{BC} \qquad i,j,t \qquad (iv)$$

$$m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB}\, w_{it}^{AC}\, w_{jt}^{BC}\, w_{ijt}^{ABC} \qquad i,j,t \qquad (v)$$

$$\left.\begin{array}{ll} m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB} & i=j \\ m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C \exp\left[-\beta\ (d_{ij})\right] & i\neq j \end{array}\right. \qquad (vi)$$

$$\left.\begin{array}{ll} m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB}\, w_{it}^{AC}\, w_{jt}^{BC}\, w_{ijt}^{ABC} & i=j \\ m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{it}^{AC}\, w_{jt}^{BC} \exp\left[-\beta\ (d_{ij})\right] & i\neq j \end{array}\right. \qquad (vii)$$

$$\left.\begin{array}{ll} m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{ij}^{AB}\, w_{it}^{AC}\, w_{jt}^{BC}\, w_{ijt}^{ABC} & i=j \\ m_{ijt} = w\, w_i^A\, w_j^B\, w_t^C\, w_{it}^{AC}\, w_{jt}^{BC} \exp\left[-\beta(d_{ij})\right] \exp\left[-\beta t(d_{ij})\right] & i\neq j \end{array}\right. \qquad (viii)$$

Models (i) to (v) are classical log- linear models with parameter estimates computed with GLIM. Models (vi), (vii) and (viii) are hybrid log-linear models. Model (i) shows the deviance values for log-linear models with main effects and their corresponding degrees of freedom.

Model (iii) and model (iv) show the effect of the added temporal interaction effect parameters respectively. The interaction parameters of the out-migration and in-migration components are given in model (ii).

Model (v) is the saturated hierarchical log- linear model with complete fit (by definition), because all possible interaction parameters are added.
Models (vi), (vii) and (viii) are hybrid log- linear models with parameter restrictions on the off- diagonal cell- elements. Model (vi) is a hybrid model with the region of origin and destination as first order interaction parameter for the diagonal elements. Models (vii) and (viii) are both hybrid models with first- order and second- order interaction effects on the diagonal elements and a time- dependent entropy function on the off- diagonal elements.
The results presented in Table 2 can be interpreted in the following way. Model (i) shows the goodness- of- fit statistics for the log- linear models with main effects and their corresponding degrees of freedom. The level of the deviance indicates the decrease of the willingness to migrate during the period 1971- 1980. The empirical results show the poor fit of the mutual independence model which is insufficient to describe the characteristics of the observed migration flows.
Model (ii) in table 2 shows the importance of the interaction effect between the region of origin and the region of destination. When model (i) with only main effects is compared with model (ii), the likelihood- ratio value decreases well over 98 percent for all time periods. The application of GLM as a spatial interaction model in this application has shown that the choice of the interaction parameters for the regions of origin and destination is important. The in-migration component needs special emphasis in migration analysis because of its time-dependence. The sharp reduction from the deviance value in model (ii) is obtained when 81 interaction parameters are added to the model (which is equivalent to a loss of 81 degrees of freedom).
The decrease of the deviance shows the effect of the added interaction effect parameters between the time- components and the out- migration and in- migration in model (iii) and model (iv) respectively. The interaction effects are significant in a statistical sense for all periods at a level of significance of 95 percent (with a critical value of 16.9 for the 9 degrees of freedom). The pattern of the out- migration seems to be more stable than the pattern of in- migration because of smaller deviance values. Only a small reduction in the deviance value (less than 0.5 percent) is obtained when the interaction parameters between time and either the region of origin or the region of destination are added.

Model	period 1 and 2	period 2 and 3	period 3 and 4	period 4 and 5	period 1, 3 and 5
(i)	37412 df = 180	35750 df = 180	31020 df = 180	28276 df = 180	47033 df = 279
(ii)	36897 0.986 df = 81	35292 0.987 df = 81	30657 0.988 df = 81	27990 0.990 df = 81	46123 0.981 df = 81
(iii)	69 0.002 df = 9	24 0.001 df = 9	49 0.002 df = 9	41 0.001 df = 9	141 0.003 df = 18
(iv)	129 0.003 df = 9	165 0.005 df = 9	93 0.003 df = 9	37 0.001 df = 9	195 0.004 df = 18
(v)	317 0.008 df = 81	269 0.008 df = 81	221 0.007 df = 81	208 0.007 df = 81	574 0.012 df = 162
(vi)	35664 0.953 df = 10	34023 0.952 df = 10	29556 0.953 df = 10	26884 0.951 df = 10	44581 0.948 df = 20
(vii)	48 df = 10	60 df = 10	47 df = 10	43 df = 10	86 df = 20
(viii)	5 df= 1	1 df= 1	3 df= 1	7 df= 1	5 df= 2

Table 2. Likelihood ratio-test statistics for 8 log-linear models.

The same number of interaction parameters is necessary in model (v) which only gives a small percentual reduction of the deviance value.
Finally, the empirical results from the hybrid models (vi), (vii) and (viii) will be discussed.

Model (vi) shows the sharp reduction of the deviance value because of the interaction parameter between the region of origin and the region of destination. In period 1 and 2 the reduction equals 35664 with only 10 parameters. Although the deviance value in model (iv) is larger (e.g., 36897-35664 = 1233 in period 1 and 2), it is accompanied by a sharp reduction in the number of parameters (81 parameters in model (ii) and 10 parameters in model (vi)). Model (vii) and model (viii) are both hybrid log-linear models with deviance values for the second order interaction effects. The deviance values in these models are smaller than the one from the second order interaction effects. But this disadvantage is once again accompanied by a reduction in the number of model parameters (81 parameters in model (v) and 10 parameters in model (vii)).

6. Conclusion

Temporal stability and spatial interaction patterns have been analysed in this paper by means of different types of log-linear models. In doing so, the generalized linear modeling approach has been used as a frame of reference and we have discussed the log-linear model as a member of the broad family of GLM.

Restrictions on a set of parameters in spatial interaction tables lead to hybrid log- linear models. Hybrid log- linear models have been linked in this paper with an entropy function for the off-diagonal elements in a migration flow matrix. The goodness- of- fit based on hybrid models is compared with the fit based on conventional log- linear models.

We can conclude from the results with the hybrid model that the interaction parameters of the off-diagonal elements, represented by the entropy function, describe a stable pattern, a conclusion which does not hold for the main diagonal interaction parameters. A reduction in the number of parameters is obtained for that reason.

A main advantage of the hybrid log- linear models is the possibility of a model simplification because statistically significant modeling results can be obtained with a reduced number of parameters.

The log-linear model is easy to apply in stability and spatial interaction research for different types of areas or time-periods as well as further classification procedures of migration flows.

The hybrid log- linear model interpreted in this paper as a generalized linear model gives a useful contribution to the empirical research of spatial interaction patterns.

References

Baydar, N., Analysis of the Temporal Stability of Migration Patterns in the Context of Multiregional Forecasting, Working Paper 38, Netherlands Interuniversity Demographic Institute, Voorburg, 1983.

Bishop, Y.M., S.E. Fienberg, and P.W. Holland, Discrete Multivariate Analysis: Theory and Practice, M.I.T. Press, Cambridge, Massachusetts, 1975.

Brouwer, F., and P. Nijkamp, Log-linear Models and Goodness-of-Fit Statistics for Categorical Data in Contingency Table Analysis: an Application to Dutch Migration Data, Series Research Memoranda 1983-3, Free University, Amsterdam, 1983.

Dempster, A.P., N.M. Laired and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, vol. 37, no. 1., 1977, pp. 1-38.

Fienberg, S.E., The Analysis of Cross-classified Categorical Data, M.I.T. Press, Cambridge, Massachusetts, 1977.

Fienberg, S.M., and W.M. Mason, Identification and Estimation of Age-period-cohort Models in the Analysis of Discrete Archival Data, in: K.F. Schuessler (ed.), Sociological Methodology, San Francisco, Jossey-Bass, 1978, pp. 1-67.

Fuchs, C., Maximum Likelihood Estimation and Model Selection in Contingency Tables with Missing Data, Journal of the American Statistical Association, vol. 77, 1982, pp. 270-278.

Haberman, S.J., Analysis of Qualitative Data vol. 2: New Developments, Academic Press, New York, 1979.

Hua, C., and F. Porell, A Critical Review of the Development of the Gravity Model, International Regional Science Review vol. 4, 1979, pp. 97- 126.

Imrey, P.B., G.G. Koch, and M.E. Stokes, Categorical Data Analysis: Some Reflections on the Log-linear Model and Logistic Regression. Part 1: Historical and Methodological Overview, International Statistical Review vol. 49, 1981, pp. 265-283.

Imrey, P.B., G.G. Koch, and M.E. Stokes, Categorical Data Analysis: Some Reflections on the Log-linear Model and Logistic Regression. Part 2: Data Analysis, International Statistical Review, vol. 50, 1982, pp. 35-63.

Mood, A.M., F.A. Graybill, and D.C. Boes, Introduction to the Theory of Statistics, Mc Graw Hill, London, 1974.

Nelder, J.A., Statistical Models for Qualitative Data, in: P. Nijkamp, H. Leitner, and N. Wrigley (eds.), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1983 (forthcoming).

Nelder, J.A., and R.W.M. Wedderburn, Generalized Linear Models, Journal of the Royal Statistical Society series A, no.135, 1972, pp. 370-384.

Payne, C., The Log-linear Model for Contingency Tables, in: C.Payne and C. O' Muircheartaigh (eds.), The Analysis of Survey Data, vol. 2: Model Fitting, Wiley, New York, 1977, pp. 105-144.

Scholten, H.J., Residential Mobility and Log-linear Modeling, in: G. Bahrenberg, M.M. Fischer, and P. Nijkamp, Recent Developments in Spatial Analysis: Methodology, Measurement, Models, Gower Publishing Company, Aldershot, 1983 (forthcoming).

Scholten, H.J., and L. van Wissen, A Comparison of the Loglinear Interaction Model with other Spatial Interaction Models, in: P. Nijkamp, H. Leitner, and N. Wrigley (eds.), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1983 (forthcoming).

Somermeijer, W.H., Een Analyse van de Binnenlandse Migratie in Nederland tot 1947 en van 1948-1957. Statistische en Economische Onderzoekingen, 1961, pp. 115-174.

Willekens, F., Log-linear Modelling of Spatial Interaction, Papers of the Regional Science Association, vol. 52, 1983 (forthcoming).

Willekens, F., and N. Güvenç, Hybrid Log-linear Models, in: P.Nijkamp, H. Leitner, and N. Wrigley (eds.), Measuring the Unmeasurable, Martinus Nijhoff, The Hague, 1983 (forthcoming).

Wolpert, J., and W. Yapa, The Time Path of Migration Flows in Belgium 1954-1962, Geographical Analysis vol. 2, 1971, pp. 157- 164.