ROBERTA SICILIANO (*) - PETER G.M. VAN DER HEIJDEN (**)

Simultaneous latent budget analysis of a set of two-way tables with constant-row-sum data (***)

CONTENTS:

1. INTRODUCTION

In the analysis of contingency tables different types of models can be used depending on the type of relation among the variables. Recently, some attention has been given to models that analyze the dependence of response variables on explanatory variables. Some examples are the logit-bilinear and logit-trilinear models (Lauro and Siciliano, 1989; Siciliano, Lauro and Mooijaart, 1990; Siciliano, 1992), the models of nonsym-

---

(*) Dipartimento di Matematica e Statistica, Università di Napoli Federico II, Italy.
(**) Department of Methodology and Statistics, Utrecht University, The Netherlands.
(***) Some of the contents of this paper have been presented at the 49th Session of the International Statistical Institute (august 25-september 2, 1993) in a contribution-abstract titled "Simultaneous Latent Budget analysis" of Siciliano, R & van der Heijden, P.G.M. (vol. 2, pp. 409-410).

metric correspondence analysis (Siciliano, Mooijaart and van der Heijden, 1992) and the latent budget analysis (van der Heijden, Mooijaart and de Leeuw, 1989).

In this paper we consider some extensions of latent budget analysis. The basic model of latent budget analysis was originally introduced by Clogg (1981) for the analysis of squared social mobility tables. Unaware of this earlier work de Leeuw and van der Heijden (1988) proposed the same model for the analysis of time-budgets. Time-budget data are specific type of constant-row-sum data (also called compositional data) in which the table comprises (group of) individuals in the rows, activities in the columns, and proportions of time spent by individuals on the activities in the cells. De Leeuw, van der Heijden and Verboon (1990) discussed the identifiability of the latent budget model and showed the relation with log-contrast principal components analysis (Aitchinson, 1986). Van der Heijden, Mooijaart ant de Leeuw (1989, 1992) and de Leeuw and van der Heijden (1991) proposed latent budget analysis in the general context of contingency tables when the data are collected under the product-multinomial sampling scheme with fixed row margins. In this respect, compared with latent class analysis (see for example Goodman, 1974), latent budget analysis is most appropriate for the analysis of the dependence when the row variable can be considered as an explanatory variable and the column variable as a response variable. When such a distinction cannot be made then latent class analysis is preferable.

In this paper we introduce a class of models for the simultaneous analysis of the latent budget structure of a set of T tables with constant-row-sum data. Such a row vector is called observed budget: it can include either compositional data (i.e. time-budget data) or conditional proportions adding up to one. The basic model of simultaneous latent budget analysis decomposes the observed budgets of each table by a mixture of unknown, or latent, budgets. The latent budgets are built up of T sets of latent budget parameters, and the mixture is defined by T sets of mixing parameters. To perform a simultaneous analysis of the latent budget structure of the T tables we consider across-group homogeneity constraints that yield three basic types of constrained models.

We present the simultaneous latent budget models for the analysis of a set of time-budgets data as well as for the analysis of a set of multidimensional contingency tables. Special attention is given to the maximum likelihood estimation through a type of EM-algorithm. We prove that the unconstrained and constrained versions of the simultaneous latent budget model can be estimated by using the standard software for the latent budget model. We discuss the identifiability of the model parameters and the degrees of freedom. We present some examples analyzing three-way contingency tables and time-budgets data. In the end we show the relations of the simultaneous latent budget models with the simultaneous latent class models (Clogg and Goodman, 1984, 1985), with the loglinear models with latent variables (Haberman, 1979) and with latent budget analysis for higher-way contingency tables (van der Heijden, Mooijaart and de Leeuw, 1989).

2. Simultaneous latent budget models

2.1. *The general model*

We present a general model for the simultaneous analysis of the latent budget structure of a set of T tables with constant-row-sum data. The T sets are the levels of a categorical variable defining groups, for example, social positions, levels of income, region, time etc.

A type of constant-row-sum data that will be considered in this paper is given by time-budgets data (or compositional data). Time-budgets data summarize how the time of individuals is distributed over a number of activities. For each table t ($t = 1, \ldots, T$) time-budget data are collected in a two-way contingency table of order $I \times J$ with objects i ($i = 1, \ldots, I$) in the rows and activities j ($j = 1, \ldots, J$) in the columns. In each table the number of times that individual i was doing activity j is denoted by $n_{ij(t)}$. When the observations for each individual are made under a product-multinomial sampling scheme, then the total number of observations for individual i is fixed to $n_{i+(t)}$ (an index is replaced by + when we add up over the corresponding way of the matrix). For individual i the J conditional proportions $p_{j|i(t)} = n_{ij(t)}/n_{i+(t)}$ add up to one, and the vector of these conditional proportions is coined an *observed time-budget*. In the following the stratifying variable is denoted by T, the row variable of each two-way table is denoted by A, the column variable is denoted by B whereas the latent variable will be denoted by X. The I observed budgets of each table t are approximated by a model for *theoretical time-budgets*, having elements the conditional probabilities $\pi_{j|i(t)}^{B|A(T)} = \pi_{ij(t)}/\pi_{i+(t)}$ ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $t = 1, \ldots, T$). This model is the simultaneous latent time-budgets model. The idea is to approximate the observed budgets of each table by a mixture of K latent budgets:

$$\pi_{j|i(t)}^{B|A(T)} = \sum_{k=1}^{K} \pi_{k|i(t)}^{X|A(T)} \pi_{j|k(t)}^{B|X(T)}, \tag{1}$$

where the parameters $\pi_{j|k(t)}^{B|X(T)}$ constitute the elements of latent budget k, $k = 1, \ldots, K = \min(I, J)$ in table t. In this model the theoretical budgets $\pi_{j|i(t)}^{B|A(T)}$ of each table t are specified by a mixture of the *latent budgets* $\pi_{j|k(t)}^{B|X(T)}$. The mixture is defined by the *mixing parameters* $\pi_{k|i(t)}^{X|A(T)}$. The total number of latent budgets K is fixed such that the model fits to the data.

The idea behind this model is that in each table t there are K typical or latent time-budgets specified by $\pi_{j|k(t)}^{B|X(T)}$ that determines the behaviour of all individuals. In each table t, these parameters for the activities add up to one for each latent budget and can be interpreted as conditional probabilities of j given k, i.e. $0 \leq \pi_{j|k(t)}^{B|X(T)} \leq 1$ and $\Sigma_j \pi_{j|k(t)}^{B|X(T)} = 1$. Each latent-time budget k "explains" the time spending behaviour of each individual i to some extent, specified by $\pi_{k|i(t)}^{X|A(T)}$. In each table t, these parameters add up over the latent budgets to one for each individual, and can be interpreted as conditional probabilities of k given i, i.e. $0 \leq \pi_{k|i(t)}^{X|A(T)} \leq 1$ and $\Sigma_k \pi_{k|i(t)}^{X|A(T)} = 1$. Notice that for $K = 1$ this model reduces to the independence model. The T sets of parameters $\pi_{j|k(t)}^{B|X(T)}$ and $\pi_{k|i(t)}^{X|A(T)}$ can vary from table to table. In this unrestricted form the simultaneous latent time-budget model comes to the same as the latent time-budgets model with K latent budgets fitted to each of the T tables separately. In order to analyze simultaneously the latent budget structure among the T tables some constraints need to be imposed (see section 2.2).

Another type of constant-row-sum data that will be considered in this paper is given by conditional distributions. Simultaneous latent budget analysis can be used for the analysis of a set of multidimensional contingency tables where we can distinguish between response variables and explanatory variables. Let in each table t ($t = 1, \ldots, T$) the row variable A be a joint explanatory variable and the column variable B be a joint response variable, where a joint variable cross-classifies the categories of more variables. In such a situation we are interested in the dependence of the response variables on the explanatory variables. For this purpose we consider the conditional distributions of the joint response variable on each level of the joint explanatory variable. Such a conditional distribution consisting of conditional proportions adding up to one defines an *observed budget*.

Under a product-multinomial sampling scheme we collect the observed budgets of each table t in a two-way table of order $I \times J$. The observed budgets of each table are approximated by a model for *theoretical budgets*, having elements the conditional probabilities $\pi_{j|i(t)}^{B|A(T)} = \pi_{ij(t)}/\pi_{i+(t)}$ ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $t = 1, \ldots, T$). Then we can define the general model for simultaneous latent budget analysis to be the model (1) but we give a more general interpretation to the model parameters. The K latent budgets are the categories of the latent variable X. The parameters $\pi_{k|i(t)}^{X|A(T)}$ for $k = 1, \ldots, K$ constitute the conditional distribution of the latent variable X given the level i of the joint explanatory variable A. The parameters $\pi_{j|k(t)}^{B|X(T)}$ for $j = 1, \ldots, J$ constitute the conditional distribution of the joint response variable B given the level k of the latent variable X. With the purpose of comparing the latent budget structure of the T tables we assume that the number K of latent budget is fixed in each table although in a more general formulation it could also vary from table to table.

The interpretation is quite simple, as in this unrestricted form simultaneous latent budget analysis comes to the same as a latent budget analysis with K latent budgets fitted to each of the T tables separately. Therefore many properties of simultaneous latent budget analysis follow immediately from the properties of latent budget analysis. The number of degrees of freedom of the model is $T(I-K)(J-K)$ (for the calculation of the degrees of freedom see section 3.3).

2.2. *Constrained models*

The idea of simultaneous latent budget analysis is to analyse simultaneously T different latent budget structures. In the interpretation phase the parameter estimates that correspond to the T different tables will usually be compared. Sometimes it can be interesting to verify whether the T tables have "similar" latent budgets or "similar" mixing parameters. As an example, if there are time budget data collected for T periods in some year, a typical question in simultaneous latent budget analysis could be: "are the typologies of time spending behaviour the same in different periods of the year?"; this is a question about the homogeneity of latent budget parameters over different tables. Another question could be "do the groups make use of the latent budgets in the same way, irrespective of the period of the year?"; this is a question about the homogeneity of mixing parameters over different tables. Consider T tables for T different countries. Then the question could be "Is the latent budget structure in the T tables identical?"; this is a question concerning the homogeneity of both the latent budget parameters and the mixing parameters over different tables.

In order to be able to answer these questions, we consider *across-group homogeneity constraints* which allow us to specify three types of constrained models.

The first constrained model aims to verify whether the latent budget structure across the T tables can be characterized by some common latent budgets, namely $\pi_{j|k(t)}^{B|X(T)} = \pi_{j|k}^{B|X}$ for all t, with mixing parameters that are different for each of the T tables:

$$\pi_{j|i(t)}^{B|A(T)} = \sum_{k=1}^{K} \pi_{k|i(t)}^{X|A(T)} \pi_{j|k}^{B|X}. \tag{2}$$

This constrained model has $(TI - K)(J - K)$ degrees of freedom. For $K = 1$ this model yields the independence model $\pi_{j|i(t)}^{B|A(T)} = \pi_{+j+}$ for the two-way multiple table $TI \times J$.

In the second constrained model, the latent budgets vary from table to table but the mixing parameters are homogeneous over the T tables, namely $\pi_{k|i(t)}^{X|A(T)} = \pi_{k|i}^{X|A}$ for all t,:

$$\pi_{j|i(t)}^{B|A(T)} = \sum_{k=1}^{K} \pi_{k|i}^{X|A} \pi_{j|k(t)}^{B|X(T)}. \tag{3}$$

This model has $(I - K)[T(J - 1) - (K - 1)]$ degrees of freedom. For $K = 1$ this model yields the model of conditional independence $\pi_{j|i(t)}^{B|A(T)} = \pi_{+j(t)}/\pi_{++(t)}$ for the two-way multiple table $TI \times J$, i.e. the variable J does not depend on the variable I within each table t of the set.

Sometimes the T sets of contingency tables have a similar latent budget structure with homogeneous latent budgets and homogeneous mixing-parameters. To validate this assumption we can test the following constrained model:

$$\pi_{j|i(t)}^{B|A(T)} = \sum_{k=1}^{K} \pi_{k|i}^{X|A} \pi_{j|k}^{B|X}. \tag{4}$$

This constrained model has $(TI - K)(J - 1) - (I - K)(K - 1)$ degrees of freedom. For $K = 1$ this model yields again the model of conditional independence $\pi_{j|i(t)}^{B|A(T)} = \pi_{+j(t)}/\pi_{++(t)}$.

Apart from time budgets data the constrained simultaneous latent budget models can be applied in several fields such as for example for marketing purposes. Suppose that we analyze a set of budgets data that summarize how the preferences of individuals are distributed over a number of products in different periods of the year or in different countries. With model (2) we fix some typologies of preferences across the T tables and then we analyze how the compositions of such preferences change from table to table. With model (3) we fix some typologies of consumers across the T tables and then we analyze how the typologies of preferences vary from table to table. With model (4) we can verify whether both the typologies of consumers and the typologies of preferences do not change across the T tables.

We can also impose *partial across-group homogeneity constraints* upon the latent budgets and/or the mixing parameters for "some" of the T tables. For examples, we can assume that the latent budgets are homogeneous across "some" of the T tables, let say the first and the second table: $\pi_{j|k(1)}^{B|X(T)} = \pi_{j|k(2)}^{B|X(T)}$. Furthermore, we can also consider the constraints that can be handled with latent budget analysis (van der Heijden, Mooijaart and de Leeuw, 1992). We can impose *within-group constraints* such as fixed-value constraints, equality constraints for row and column parameters of each table. For example, we can assume that the K latent budgets of the j-th response are equivalent to the k latent budgets of the j-th response across the T tables, namely $\pi_{j|k(t)}^{B|X(T)} = \pi_{j'|k(t)}^{B|X(T)}$ for all k and t. Withing-group constraints can be considered not only in the general model (1) but also in the restricted versions (2), (3) and (4). Another type of constraints uses *additional information* about the row and column categories of each table. In simultaneous latent budget models the parameters are conditional probabilities. The idea is to verify whether such probabilities follow some specific model such as a multinomial logit model (see for example van der Heijden, Mooijaart and the Leeuw, 1992).

3. Maximum likelihood estimation and identifiability

3.1. *The EM-algorithm*

We estimate the simultaneous latent budget model (1) with the EM-algorithm (Dempster, Laird and Rubin, 1977). Similarly, Goodman (1974) uses this algorithm for estimation of latent class analysis. We consider the same type of EM-algorithm used for latent budget models (de Leeuw, van der Heijden and Verboon, 1990; van der Heijden, Mooijaart, de Leeuw, 1992).

The EM-algorithm is used to estimate missing values. In this paper, we consider a four-way table of elements $n_{ijk(t)}$ where the observations on the latent variable are missing and only the three-way margins $n_{ij+(t)}$ are known. For this table we have the latent model $\pi_{ijk(t)}/\pi_{i++(t)} = \pi^{B|X(T)}_{k|i(t)}\pi^{B|X(T)}_{j|k(t)}$. Under the product-multinomial sampling scheme the loglikelihood for the unobserved matrix is

$$L = \sum_i \sum_j \sum_k \sum_t n_{ijk(t)} \log \frac{\pi_{ijk(t)}}{\pi_{i++(t)}} = \sum_i \sum_j \sum_k \sum_t n_{ijk(t)} \log \pi^{X|A(T)}_{k|i(t)} \, \pi^{B|X(T)}_{j|k(t)} \, .$$

We do not known the unobserved proportions $n_{ijk(t)}$, and neither do we known the parameters $\pi^{X|A(T)}_{k|i(t)}$ and $\pi^{B|X(T)}_{j|k(t)}$. The EM-algorithm iterates the expectation step and the maximization step until convergence. In the *E-step* the expectation of the loglikelihood for the unobserved data $n_{ijk(t)}$ is found, conditional on the observed proportions $n_{ij+(t)}$ and the current parameter estimates. The expectations of the sufficient statistics of the unknown data matrix have to be expressed in terms of the model parameters. For this step the current best estimates of $\pi^{X|A(T)}_{k|i(t)}$ and $\pi^{B|X(T)}_{j|k(t)}$ are taken. Thus, updated estimates of the unobserved proportions $\underline{n}_{ijk(t)}$ (we denote the ML estimates by underlining the parameters) are given by

$$\underline{n}_{ijk(t)} = n_{ij+(t)} \frac{\underline{\pi}_{ijk(t)}}{\underline{\pi}_{ij+(t)}} = n_{ij+(t)} \frac{\underline{\pi}_{ijk(t)}}{\underline{\pi}_{i++(t)}} \Big/ \frac{\underline{\pi}_{ij+(t)}}{\underline{\pi}_{i++(t)}} = n_{ij+(t)} \frac{\underline{\pi}^{X|A(T)}_{k|i(t)} \, \underline{\pi}^{B|X(T)}_{j|k(t)}}{\sum_k \underline{\pi}^{X|A(T)}_{k|i(t)} \, \underline{\pi}^{B|X(T)}_{j|k(t)}} \, .$$

In the *M-step* the loglikelihood for the unobserved data is maximized as a function of the model parameters. Considering the conditions upon the model parameters, we can maximize the following functions over $\underline{\pi}^{X|A(T)}_{k|i(t)}$ and $\underline{\pi}^{B|X(T)}_{j|k(t)}$ separately:

$$f\left(\pi^{X|A(T)}_{k|i(t)}, \alpha_{it}\right) = \sum_i \sum_k \sum_t n_{i+k(t)} \log \pi^{X|A(T)}_{k|i(t)} - \sum_i \sum_t \alpha_{it} \left(\sum_k \sum_t \pi^{X|A(T)}_{k|i(t)} - 1 \right),$$

$$f\left(\pi^{B|X(T)}_{j|k(t)}, y_{kt}\right) = \sum_j \sum_k \sum_t n_{+jk(t)} \log \pi^{B|X(T)}_{j|k(t)} - \sum_k \sum_t y_{kt} \left(\sum_j \sum_t \pi^{B|X(T)}_{j|k(t)} - 1 \right),$$

where α_{kt} and γ_{it} are the so-called Lagrange multipliers. This gives as updated estimates:

$$\underline{\pi}^{X|A(T)}_{k|i(t)} = \frac{n_{i+k(t)}}{n_{i++(t)}}, \quad \underline{\pi}^{B|X(T)}_{j|k(t)} = \frac{n_{+jk(t)}}{n_{++k(t)}} \, .$$

These updated estimates are used in the expectation of the loglikelihood as current best estimates in the next E-step of the algorithm.

In each cycle estimates of expected frequencies $m_{ij+(t)}$ are given by $m_{ij+(t)} = n_{i++(t)} \underline{\pi}_{ij+(t)}/\underline{\pi}_{i++(t)}$ so that the usual likelihood ratio statistic G^2 can be evaluated. When the difference between two subsequent G^2 values is smaller than a prespecified value, iterating can stop, and we can test the fit of the model using the current value of G^2.

We discuss now how the EM-algorithm needs to be modified for the estimation of the constrained models (2) and (3) where the homogeneous constraints across the T tables are imposed either on the mixing parameters or on the latent budgets parameters. Whereas the E-step remains unchanged, the M-step will take into account different functions.

Under the model (2) we maximize separately the following functions:

$$f\left(\pi^{X|A}_{k|i}, \alpha_i\right) = \sum_i \sum_k n_{i+k+} \log \pi^{X|A}_{k|i} - \sum_i \alpha_i \left(\sum_k \pi^{X|A}_{k|i} - 1 \right),$$

$$f\left(\pi^{B|X(T)}_{j|k(t)}, y_{kt}\right) = \sum_j \sum_k \sum_t n_{+jk(t)} \log \pi^{B|X(T)}_{j|k(t)} - \sum_k \sum_t y_{kt} \left(\sum_j \sum_t \pi^{B|X(T)}_{j|k(t)} - 1 \right),$$

where $n_{i+k+} = \Sigma_t \, n_{i+k(t)}$. Thus, the updated estimates under the constrained model (2) are given by:

$$\underline{\pi}_{k|i}^{X|A} = \frac{n_{i+k+}}{n_{i+++}}, \quad \underline{\pi}_{j|k(t)}^{B|X(T)} = \frac{n_{+jk(t)}}{n_{++k(t)}}.$$

Similarly, we find the updated estimates under the constrained model (3):

$$\underline{\pi}_{k|i(t)}^{X|A(T)} = \frac{n_{i+k(t)}}{n_{i++(t)}}, \quad \underline{\pi}_{j|k}^{B|X} = \frac{n_{+jk+}}{n_{++k+}}.$$

In conclusion, the across-homogeneity constraints in model (2) imply that the estimate of $\pi_{k|i}^{X|A}$ is obtained as the average of the parameters $\pi_{k|i(t)}^{X|A(T)}$ if these would not have been constrained to be equal, weighted by the proportions $n_{i++(t)}/n_{i+++}$. The across-homogeneity constraints in model (3) imply that the estimate of $\pi_{j|k}^{B|X}$ is obtained as the average of the parameters $\pi_{j|k(t)}^{B|X(T)}$ if these would not have been constrained to be equal, weighted by the proportions $n_{++k(t)}/n_{++k+}$.

In similar way, we can estimate the parameters of the model (4) with across-homogeneity constraints of both mixing parameters and latent budgets.

3.2. *The estimation using constrained latent budget models*

We show how the unconstrained simultaneous latent budget model for T different $I \times J$ tables can be formulated as a certain kind of latent budget model with within-group constraints for a block-diagonal two-way table of dimensions $TI \times TJ$. The constraints specify that certain mixing parameters and certain latent budgets are set to zero. For simplicity we leave out the indication of the variables in the notation of the parameters.

Let π_{rc} $(r = 1, \ldots, TI;\ c = 1, \ldots, TJ)$ denote the probability for cell (r, c). We can write the latent budget model for the table $TI \times TJ$ as

$$\pi_{c|r} = \sum_{u=1}^{U} \pi_{u|r}\, \pi_{c|u}, \tag{5}$$

where the categories u $(u = 1, \ldots, TK)$ of the latent variable can be defined as $u = (t-1)K + k$. This model can be directly related to the simultaneous latent budget model (1) when the model parameters $\pi_{u|r}$ and $\pi_{c|u}$ satisfy the following restrictions

$$\pi_{u|r} = 0,\ \pi_{c|u} = 0,\quad r = (t-1)I + i;\ c = (t-1)J + j;\ u = (t'-1)K + k;\quad t \neq t'. \tag{6}$$

From these restrictions it follows that the theoretical budgets $\pi_{c|r}$ are set to be equal to zero for $r = (t-1)I + i$ and $c = (t'-1)J + j$ if $t' \neq t$. Such restrictions can be understood more clearly if we consider the following matrix notation of the simultaneous latent budget model:

$$\boldsymbol{D}_R^{-1}\boldsymbol{P}_T = \boldsymbol{P}_R\boldsymbol{P}_C^T, \quad \boldsymbol{P}_R\boldsymbol{u} = \boldsymbol{u} \quad \boldsymbol{P}_C\boldsymbol{u} = \boldsymbol{u}, \tag{7}$$

where $\boldsymbol{P}_T$ is a block-diagonal matrix $TI \times TJ$ having as full blocks the T matrices $\boldsymbol{P}_{(t)}$ of the theoretical budgets $\pi_{j|i(t)}$; $\boldsymbol{P}_R$ is a block-diagonal matrix $TI \times TK$ having as full blocks the T matrices $\boldsymbol{P}_{r(t)}$ of the mixing parameters $\pi_{k|i(t)}$; $\boldsymbol{P}_C$ is a block-diagonal matrix $TJ \times TK$ having as full blocks the T matrices $\boldsymbol{P}_{c(t)}$ of the latent budgets $\pi_{j|k(t)}$; $\boldsymbol{D}_R$ is a diagonal matrix $TI \times TI$ with elements $\pi_{i++(t)}$; $\boldsymbol{u}$ is a vector of ones.

This model is fitted to an observed block-diagonal table of dimensions $TI \times TJ$. The full blocks of this table are the T sets of $I \times J$ observed tables. Let n_{rc} $(r = 1, \ldots, TI;\ c = 1, \ldots, TJ)$ denote the general term of the $TI \times TJ$ table. By definition, n_{rc} is equal to zero for $r = (t-1)I + i$ and $c = (t'-1)J + j$ when $t' \neq t$.

From the definition of the observed proportions n_{rc} it follows that the estimates of the unobserved proportions n_{rcu} are equal to zero for $r = (t-1)I + 1, \ldots, tI$ and $c = (t'-1)J + 1, \ldots, tJ$ when $t' \neq t$. Consequently, the estimates $\underline{\pi}_{u|r}$ and $\underline{\pi}_{c|u}$ satisfy restrictions (6). This guarantees that the estimated matrices $\boldsymbol{P}_R$ and $\boldsymbol{P}_C$ are block-diagonal matrices and the estimated matrix of theoretical budgets $\boldsymbol{D}_R^{-1}\boldsymbol{P}_T$ is also block-diagonal.

In conclusion, for the unconstrained simultaneous latent budget model (1) the matrices $\boldsymbol{P}_{r(t)}$ and $\boldsymbol{P}_{c(t)}$ can be estimated separately by applying the ordinary EM-algorithm for the latent budget model $\boldsymbol{P}_{(t)} = \boldsymbol{P}_{r(t)}\boldsymbol{P}_{c(t)}^T$. This is equivalent to fit the latent budget model (5) with the fixed value constraints (6) to the observed block-diagonal table of dimensions $TI \times TJ$.

We consider now the estimation of the constrained simultaneous latent budget models.

As we have previously shown, the estimation of the constrained model (4) can be done using the latent budget model for the table $I \times J$ of elements $n_{ij+} = \Sigma_t n_{ij(t)}$.

The homogeneous constraints in model (2) imply the equality among the T matrices $\boldsymbol{P}_{r(t)}$, whereas the constraints in model (3) imply the equality among the T matrices $\boldsymbol{P}_{c(t)}$. Such constraints could be understood as equality constraints to impose to the parameters of latent budget model (5) apart from the fixed value constraints (6) (see for more details about the type of EM-algorithm to use for the estimation of such constrained models Mooijaart and van der Heijden, 1992).

3.3. *Identifiability and degrees of freedom*

We consider now the calculation of the degrees of freedom for the unconstrained and constrained models. In general, the number of degrees of freedom is equal to the number of independent cells minus the number of independent parameters to estimate. Using the matrix notation (7), the number of independent cells is $TI(J-1)$ because rows of matrix $\boldsymbol{P}_T$ add up to one. The number of independent parameters depends on the conditions, the eventual homogeneity constraints, and the restrictions required to identify the model.

The simultaneous latent budget analysis models are not identified: consider a block-diagonal matrix $\boldsymbol{S}_T$ of dimensions $TK \times TK$ such that rows of each block matrix $\boldsymbol{S}_{(t)}$ of dimension $K \times K$ sum up to one. It results: $\boldsymbol{P}_R \boldsymbol{P}_C^T = \boldsymbol{P}_R \boldsymbol{S}_T \boldsymbol{S}_T^{-1} \boldsymbol{P}_C^T$. Some restrictions can be imposed upon parameters to make them identifiable. The number of restrictions depends on the number of free parameters in $\boldsymbol{S}_T$.

Consider the unconstrained simultaneous latent budget model (1). The number of row parameters and the number of column parameters to estimate are respectively $TI(K-1)$ and $TK(J-1)$ since rows of the matrix $\boldsymbol{P}_R$ and columns of the matrix P_C add up to one. The parameters in $\boldsymbol{P}_R$ and $\boldsymbol{P}_C$ are not independent as they are not identified. For their identification we should use the free parameters of the matrix $\boldsymbol{S}_T$ which are given by $TK(K-1)$, that is there are $K(K-1)$ free elements of each block matrix $\boldsymbol{S}_{(t)}$ and there are T block matrices $\boldsymbol{S}_{(t)}$. The number of degrees of freedom can be obtained as $TI(J-1) - \{TI(K-1) + TK(J-1) - TK(K-1)\} = T(I-K)(J-K)$.

Consider the constrained model (2). The number of row parameters to estimate is $TI(K-1)$ since rows of the matrix $\boldsymbol{P}_R$ add up to one. The number of column parameters to estimate is $K(J-1)$ since the T blocks of the matrix $\boldsymbol{P}_C$ are equal because of the across-homogeneity constraints. The number of free parameters of the matrix $\boldsymbol{S}_T$ is given by $K(K-1)$ since the T matrices $S_{(t)}$ have to be equal such that the across-homogeneity constraints are satisfied and each matrix $\boldsymbol{S}_{(t)}$ contains $K(K-1)$ free parameters (as rows of each matrix $\boldsymbol{S}_{(t)}$ of dimension $K \times K$ sum up to one). So the number of degrees of freedom can be obtained as $TI(J-1) - \{TI(K-1) + K(J-1) - K(K-1)\} = (TI-K)(J-K)$.

Consider the constrained model (3). The number of row parameters to estimate is $I(K-1)$ because of the homogeneity constraints and the number of free parameters of the matrix $\boldsymbol{S}_T$ is given by $K(K-1)$. Analogously, the number of restrictions upon row parameters is equal to $(T-1)I(K-1)$ and the number of unfree parameters is equal to $(T-1)K(K-1)$. The final number of degrees of freedom can be obtained as $T(I-K)(J-K) + (T-1)I(K-1) - (T-1)K(K-1) = (I-K)[T(J-1) - (K-1)]$.

Consider the most constrained model (4). The number of row parameters to estimate is $I(K-1)$ and the number of column parameters to estimate is $K(J-1)$. The number of free parameters of the matrix $\boldsymbol{S}_T$ is given by $K(K-1)$. Therefore, the number of degrees of freedom is obtained as $TI(J-1) - \{I(K-1) + K(J-1) - K(K-1)\} = (TI-K)(J-1) - (I-K)(K-1)$.

4. Examples

4.1. *Secondary school and test for intellectual capacity in males and females*

In the Netherlands children go at the age of 11-12 from primary school to secondary school. There are two main types of secondary education: education for professions and general education. A choice will depend on aspects such as capacities of children, interest, advice of parents, advice of primary school teacher.

In 1977 and 1981 data were collected from more that 37,000 children about their social milieu and aspects regarding their secondary education. Distinct variables were collected (Meester & Leeuw, 1983). Van der Heijden, Mooijaart & de Leeuw (1990) have analysed these data by latent budget analysis. We reanalyse part of these data considering the variables: the scores on an intelligence test, sex and level of education attained in 1981, i.e. after four years of secondary school. The intelligence test used was the Dutch Test for Intellectual Capacity (TIC), a figure exclusion test that consists of 33 items. The TIC scores were recorded as 1 for 1 to 14 items correct, 2 for 15 to 17 correct, 3 for 18 to 20 correct, 4 for 21 to 23 correct, 5 for 24 to 26 correct, 6 for 27 to 29 correct, 7 for 30 to 33 items correct. The response variable is the Level of Education attained after 4 years, and these levels are 1. Dropped out (DO), 2. Junior level of education for professions (LBO), 3. Medium level of general education (MAVO), 4. Senior level of education for professions (MBO), 5. High level of general education (HAVO) and 6. General education preparing for university (VWO). Meester & de Leeuw (1983) have eliminated all children having no TIC scores (16,433 children). According to them, this elimination is not crucial because having no TIC score seemed to have been a random process. Further, children with a missing value on level of education attained (38) or on an education type called extraordinary lower education (646) were eliminated from the sample. Children having a father who is unemployed, or medically unfit for work were also eliminated (6,190). After these selections there was a sample of 16,236 children.

Table 1 gives the three dimensional table with frequencies. This table is formed by two tables that cross-classify the Level of Education and the TIC scores, one for males and the other for females. Mooijaart & van der Heijden (1992) have already analysed this table with log-trilinear models. We apply simultaneous latent budget analysis on table 1.

TABLE 1

CROSS-CLASSIFICATION OF DUTCH SAMPLE ACCORDING TO TEST FOR INTELLECTUAL CAPACITY (TIC) AND LEVEL OF EDUCATION, BY SEX

Sex	Males							Females						
TIC score	1	2	3	4	5	6	7	1	2	3	4	5	6	7
DO	75	77	105	125	89	38	17	51	60	115	123	78	56	9
LBO	216	305	495	522	389	168	34	144	223	382	370	290	107	26
MAVO	67	144	267	368	339	194	54	60	134	288	424	442	266	72
MBO	51	84	239	345	301	208	65	75	167	320	458	428	258	72
HAVO	26	65	200	332	383	258	98	23	68	211	373	450	402	169
VWO	12	27	104	216	325	321	178	5	9	77	183	307	326	209

In Table 2 there are the test statistics for some models. The unconstrained model that fits to the data has $K = 3$ latent budgets. Also the constrained model with homogeneous mixing parameters fits well to the data. The identified parameter estimates of these two models are respectively given in table 3 and in table 4. We notice that the model with homogeneous latent budgets and mixing parameters does not fit to the data.

TABLE 2

MODELS AND TEST STATISTICS FOR DATA IN TABLE 1

Model	K	G^2	d.f.
unconstrained	1	2370.39	60
unconstrained	2	199.86	40
unconstrained	3	25.15	24
homogeneous mixing parameters	2	202.59	45
homogeneous mixing parameters	3	28.97	32
homogeneous latent budgets	3	404.38	48
homogeneous latent structure	3	440.83	55

From the mixing parameter values in table 3 we deduce that for both males and females the mixing parameter values increase in the first latent budget, decrease in the second latent budget and first increase and then decrease in the third latend budget with increasing the TIC scores. Therefore, the first latent budget characterizes males and females with high TIC scores (scores 6 and 7), the second latent budget characterizes males and females with low TIC scores (scores 1 and 2) and the third latent budget characterizes males and females with medium TIC scores (scores 3, 4 and 5 for males and scores 4 and 5 for females). Obviously, the same interpretation follows for the homogeneous mixing parameters in table 4.

The latent budget parameters $\pi_{j|k(t)}^{B|X(T)}$ are compared with their corresponding marginal probabilities $\pi_{+j(t)}$ under independence. Both males and females in the first latent budget typically choose the levels of education VWO and HAVO. Males in the second latent budget typically choose the levels of education DO and LBO. Females in the second latent budget tipically choose the levels of education DO and MBO. Both males and females in the third latent budget typically choose the levels of

TABLE 3

ESTIMATES OF THE PARAMETERS OF THE UNCONSTRAINED MODEL FOR DATA IN TABLE 1

TIC score	Mixing parameters							
	Males				Females			
	$k=1$	$k=2$	$k=3$		$k=1$	$k=2$	$k=3$	
1	.000	1.000	.000	1.000	.000	1.000	.000	1.000
2	.000	.754	.246	1.000	.000	.695	.305	1.000
3	.069	.444	.487	1.000	.122	.512	.366	1.000
4	.175	.273	.552	1.000	.232	.267	.501	1.000
5	.360	.136	.504	1.000	.392	.125	.483	1.000
6	.608	.000	.392	1.000	.632	.000	.368	1.000
7	1.000	.000	.000	1.000	1.000	.000	.000	1.000
Level of education	Latent budgets							
	Males				Females			
	$k=1$	$k=2$	$k=3$	Indep.	$k=1$	$k=2$	$k=3$	Indep.
DO	.038	.149	.026	.066	.021	.122	.048	.059
LBO	.076	.495	.238	.269	.044	.418	.144	.186
MAVO	.121	.166	.232	.181	.129	.163	.298	.203
MBO	.143	.107	.217	.163	.125	.220	.290	.214
HAVO	.225	.055	.219	.172	.313	.063	.208	.204
VWO	.397	.028	.068	.149	.368	.014	.012	.134
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

education MAVO, MBO and HAVO. Moreover, for both males and females the values of the first latent budget increase with increasing the level of education. So, we can conclude that males and females with TIC scores choose high level of education. For the other latent budgets there is not a similar ordering in their values. For instance, in the second latent budget there is a low percentage of females with TIC score 5 but this latent budget is typical for females with a specialized education.

In conclusion, males and females with high TIC scores choose a general education preparing for university, males and females with medium TIC scores choose a general education of medium or high level, or a senior level of education for professions. Males with low TIC scores retire from school or choose a junior level of education for professions, while females with low TIC scores retire from school or choose either a junior or a senior level of education for professions.

TABLE 4

ESTIMATES OF THE PARAMETERS OF THE MODEL WITH HOMOGENEOUS MIXING PARAMETERS FOR DATA IN TABLE 1

TIC score	Homogeneous mixing parameters							
	Males and Females							
	$k=1$	$k=2$	$k=3$					
1	.000	1.000	.000	1.000				
2	.014	.763	.223	1.000				
3	.128	.531	.341	1.000				
4	.241	.337	.422	1.000				
5	.409	.193	.398	1.000				
6	.645	.048	.307	1.000				
7	1.000	.000	.000	1.000				
Level of education	Latent budgets							
	Males				Females			
	$k=1$	$k=2$	$k=3$	Indep.	$k=1$	$k=2$	$k=3$	Indep.
DO	.038	.147	.006	.066	.021	.121	.039	.059
LBO	.077	.495	.217	.269	.044	.419	.104	.186
MAVO	.118	.165	.263	.181	.130	.163	.323	.203
MBO	.144	.109	.243	.163	.126	.220	.305	.214
HAVO	.224	.054	.248	.172	.312	.063	.228	.204
VWO	.399	.030	.023	.149	.367	.014	.001	.134
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

4.2. *Time-budgets of Amazone Indians*

We consider an example of a matrix with time budget data. The data are derived from Gross, Rubin & Flowers (1985) and were analysed earlier by de Leeuw & van der Heijden (1988), de Leeuw, van der Heijden & Verboon (1990). The data concerns the day-time spending behaviour of males, females and children of four Indian tribes from the Amazone (the Mekranoti, the Kanela, the Bororo and the Xavente). Their time can be spent in doing six types of activities: "being idle" (casual conversations, play and sports, cerimonial activities, meals etc.) "sleeping", "caring" (child care: carrying, touching, guiding etc.), "nonsubsistence behavior" (food preparation, wage labor, household maintenance, manufacture, repair and upkeep of tools and dwellings, hygiene such as bathing, etc.), "domestic activities" (domesticated crops and animals), and "wild" (hunting, fishing and gathering). The day-time is divided into seven intervals of two hours starting from 8.00 a.m. until 20.00 p.m. Unfortunately, we are not able to obtain the original data, that were obtained with the random spot check method. In this method, we know at random points in time what an objects is doing. The number of times that we observe a specific activity for some object gives an indication for the total amount of time that this object spends on this activity. For example, a person is working 70 times out of the 100 times that we observe him, we assume that he is working 70% of this time.

De Leeuw, van der Heijden, & Verboon (1990) have analysed with the latent budget model the data set which cross-classifies the objects and the activities without considering the intervals of time.

We consider the simultaneous latent budget analysis of the seven tables which cross-classify the objects and the activities, one table for each interval of time. We adopt the simultaneous latent budget model (2) with homogeneous constraints on mixing parameters considering $K = 3$ latent budgets.

TABLE 5

ESTIMATES OF THE MIXING PARAMETERS OF THE MODEL WITH HOMOGENEOUS MIXING PARAMETERS IN TIME-BUDGETS OF AMAZONE INDIANS

		Mixing parameters for each interval of time			
		$k = 1$	$k = 2$	$k = 3$	
Mekranoti	Males	.935	.065	.000	1.000
	Females	.116	.630	.254	1.000
	Children	.115	.023	.862	1.000
Kanela	Males	.548	.183	.269	1.000
	Females	.000	.750	.250	1.000
	Children	.000	.041	.959	1.000
Bororo	Males	.778	.000	.222	1.000
	Females	.226	.606	.168	1.000
	Children	.236	.000	.764	1.000
Xavente	Males	.549	.002	.449	1.000
	Females	.171	.829	.000	1.000
	Children	.093	.088	.819	1.000

For this model the value of the likelihood ratio statistic is high being $G^2 = 7580.97$ (df = 297). But this is justified by the high sample size of 70,560 objects. The identified parameter estimates are given in tables 5 and 6.

TABLE 6

ESTIMATES OF THE LATENT BUDGETS OF THE MODEL ($K = 3$) WITH HOMOGENEOUS MIXING PARAMETERS IN TIME-BUDGETS OF AMAZONE INDIANS

First latent budget							
Activities \ Time	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
1	.519	.222	.309	.414	.514	.664	.806
2	.123	.000	.025	.052	.056	.015	.021
3	.000	.002	.000	.000	.000	.001	.004
4	.196	.370	.297	.234	.256	.169	.083
5	.089	.167	.156	.075	.092	.068	.028
6	.073	.239	.213	.225	.082	.083	.058
	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Second latent budget							
Activities \ Time	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
1	.328	.310	.294	.315	.380	.429	.588
2	.116	.005	.041	.032	.006	.005	.006
3	.138	.125	.078	.099	.092	.116	.161
4	.350	.375	.400	.373	.398	.372	.245
5	.064	.185	.183	.148	.120	.078	.000
6	.004	.000	.004	.033	.004	.000	.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Third latent budget							
Activities \ Time	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
1	.816	.782	.762	.772	.782	.897	.840
2	.147	.041	.088	.092	.106	.037	.148
3	.000	.006	.010	.000	.002	.002	.000
4	.011	.038	.000	.012	.023	.015	.001
5	.023	.133	.140	.108	.087	.049	.011
6	.003	.000	.000	.016	.000	.000	.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Activities: 1. being idle; 2. sleeping; 3. caring; 4. nonsubsistence; 5. domestic; 6. wild.

From the mixing parameter values we deduce that in each tribe the first latent budget is characterized by males, the second one by females and the third one by children. We interpret the latent budget parameters for males (first latent budget), females (second latent budget), children (third latent budget). Males spend their time from 6.00 a.m. until 8.00 a.m. in nonsubsistence activities, from 8.00 a.m. until 14.00 p.m. in being idle, nonsubsistence activities, wild activities, from 14.00 p.m. until 16 p.m. in being idle and nonsubsistence activities, from 16.00 p.m. until 20.00 p.m. in being idle. So males hardly spend time in caring children and domestic activities. Females spend their time mostly in being idle and nonsubsistence activities, and in particular from 8.00 a.m. until 16.00 p.m. in domestic activities, from 6.00 a.m. until 10.00 a.m. and from 16.00 p.m. until 20.00 p.m. in caring children. Children spend most of the time in being idle or sleeping.

5. Relations with other models

5.1. *The relation with simultaneous latent class analysis*

Simultaneous latent budget analysis is comparable with simultaneous latent class analysis used in simultaneous analysis of several groups (Hagenaars, 1990). Consider first the analysis of a set of T two-way contingency tables. In the following the variables A and B can be understood as two simple variables. The simultaneous latent class analysis model is

$$\pi_{ij|(t)}^{AB|(T)} = \sum_{k=1}^{K} \pi_{k|(t)}^{X|(T)} \pi_{i|k(t)}^{A|X(T)} \pi_{j|k(t)}^{B|X(T)} ,$$

where $\pi_{ij|(t)}^{AB|T}$ is the probability to fall in cell (i, j) of table t; the parameters $\pi_{k|(t)}^{X|(T)}$ are the latent class probabilities for table t and, $\pi_{i|k(t)}^{A|X(T)}$ and $\pi_{j|k(t)}^{B|X(T)}$ are the usual conditional probabilities for table t. Clogg and Goodman (1984, 1985) paid special attention to homogeneous restrictions over the parameters for the T contingency tables, and showed how the model could be estimated by using standard software for latent class analysis.

The unconstrained simultaneous latent budget analysis model can be considered as a reparametrization of the unconstrained simultaneous latent class analysis model if we consider the Bayes rule to obtain the mixing parameters:

$$\pi_{k|i(t)}^{X|A(T)} = \frac{\pi_{k|(t)}^{X|(T)} \pi_{i|k(t)}^{A|X(T)}}{\sum_k \pi_{k|(t)}^{X|(T)} \pi_{i|k(t)}^{A|X(T)}} ,$$

where $\Sigma_k \pi_{k|(t)}^{X|(T)} \pi_{i|k(t)}^{A|X(T)} = \pi_{i+(t)}/\pi_{++(t)}$.

The reparametrization from one model to the other model also holds in the analysis of a set of T higher-way contingency tables. The proof follows in straightforward way taking into account the fact that in latent class models the observed variables are conditionally independent given the latent variable.

In similar way, we can show that the constrained simultaneous latent budget model (2) is also equivalent to a similarly constrained simultaneous latent class model. But the constrained simultaneous latent budget model (3) has not a comparable constrained simultaneous latent class model. Indeed, homogeneity constraints on the mixing parameters imply homogeneity constraints on the products $(\pi_{++(t)} \pi_{k|(t)}^{X|(T)} \pi_{i|k(t)}^{A|X(T)})/\pi_{i+(t)}$ and such restriction cannot be satisfied in simultaneous latent class model.

5.2. *The relation with log-linear analysis*

We show now how the simultaneous latent budget model can be formulated as a constrained version of a log-linear model for the latent table $I \times J \times T \times K$. The simultaneous latent budget model (1) restricts the two observed variables I and J in the four-way latent table to be conditionally independent given the level of the latent variable and the stratifying factor:

$$\pi_{ijk(t)}^{ABX(T)} = \pi_{i++(t)} \pi_{k|i(t)}^{X|A(T)} \pi_{j|k(t)}^{B|X(T)} = \frac{\pi_{i+k(t)} \pi_{+jk(t)}}{\pi_{++k(t)}} .$$

Then, using hierarchically log-linear modeling we can define the following log-linear latent model (compare with Haberman, 1979):

$$\log \pi_{ijkt} = \mu + \mu_i + \mu_j + \mu_k + \mu_t + \mu_{it} + \mu_{jt} + \mu_{kt} + \mu_{ik} + \mu_{jk} + \mu_{itk} + \mu_{jtk} ,$$

where the μ's parameters add up to zero over their indexes. Compared with the saturated log-linear model there are restrictions as $\mu_{ij} = 0$, $\mu_{ijt} = 0$, $\mu_{ijk} = 0$ and $\mu_{ijtk} = 0$.

In straightforward way, we can fix equality constraints upon the interaction parameters in order to impose the corresponding across-homogeneity constraints upon the mixing parameters and the latent budget parameters.

5.3. *The relation with latent budget analysis for higher-way tables*

In this section, we discuss how the simultaneous latent budget models differ from the generalization of the latent budget model to higher-way tables proposed by van der Heijden, Mooijaart and de Leeuw (1989). The higher-way table analysed with the latent budget model is structured into a two-way form that cross-classifies the categories of a joint response variable by the categories of a joint explanatory variable. Instead, the simultaneous latent budget model analyses a set of so-structured higher-way contingency tables and thus a specific role is played by the stratifying variable of the set.

As a result, the simultaneous latent budget model (1) has a more general definition than any type of latent budget model because both the mixing parameters and the latent budget parameters can vary from table to table. Consequently, all the latent budget models for higher-way tables can be also specified as constrained simultaneous latent budget models. But not all the simultaneous latent budget models can be specified by the latent budget models for higher-way tables. As an example, using the latent budget model we cannot define the simultaneous latent budget model (3) with across-group homogeneity constraints on only the mixing parameters since the latent budgets are conditional to both the latent variable and the stratifying variable of the set.

We can only deal with block-diagonal tables as shown in section 3.2 in order to use the standard software of latent budget analysis to estimate the simultaneous latent budget models.

6. Concluding remarks

In this paper we have proposed a model for the analysis of the latent budget structure of a set of T tables with constant-row sum data. Such data can be given either by conditional proportions of multidimensional contingency table or by compositional data such as time-budgets. We have discussed versions of this model with across-group homogeneity constraints. We have shown that these models can be formulated as standard latent budget models with within-group restrictions upon parameters. Clearly, across-group homogeneity constraints can be fitted together with within-group homogeneity constraints. We described how the unconstrained and constrained simultaneous latent budget analysis models can be estimated by a type of EM-algorithm.

We have shown how the unconstrained simultaneous latent budget model can be understood as a reparametrization of the simultaneous latent class model for the analysis of a set of contingency tables: the former seems most useful when the row variable can be considered as an explanatory variable and the column variable as a response variable for each contingency table, the latter when both variables can be considered as response variables for each contingency table. This equivalence does not always hold for constrained simultaneous latent budget models and constrained simultaneous latent class models.

Acknowledgements

This research was supported for the first author by a grant of the Italian MURST-60%. The authors greatefully acknowledge the comments of the referees.

References

Agresti, A. (1990) *Categorical data analysis*, Wiley.

Aitchison, J. (1986) *The statistical analysis of compositional data*, London: Chapman and Hall.

Clogg, C.C. (1981) Latent structure models of mobility, *American Journal of Sociology*, 86: 836-868.

Clogg, C.C. (1982) Some models for the analysis of association in multiway cross-classifications having ordered categories, *Journal of the American Statistical Association*, 77: 803-815.

Clogg, C.C. and Goodman, L.A. (1984) Latent structure analysis of a set of multidimensional contingency tables, *Journal of the American Statistical Association*, 79: 762-771.

Clogg, C.C. and Goodman, L.A. (1985) Simultaneous latent structure analysis in several groups, *Sociological Methodology*, San Francisco: 81-110.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, ser. B*, 39: 1-38.

de Leeuw, J. and van der Heijden, P.G.M. (1988) The analysis of time-budget with a latent time-budget model in E. Diday et al. (eds.): *Data Analysis and Informatics V*, Amsterdam: North Holland, pp. 159-166.

de Leeuw, J. and van der Heijden, P.G.M. (1991) Reduced-rank models for contingency tables, *Biometrika*, 78: 239-242.

de Leeuw, J., van der Heijden, P.G.M. and Verboon, P. (1990) A latent time budget model, *Statistica Nederlandica*, 44: 1-22.

Goodman, L.A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, 61: 215-231.

Groos, D.R., Rubin, J. and Flowers, N.M. (1985) *All in a day's time: random spot check data can describe the texture of a day's activities*. Presented at the Annual Meeting of the American Statistical Association for the Advancement of Science (Los Angeles, may 29, 1985).

Haberman, S.J. (1979) *Analysis of Qualitative Data*, 2 vols. New York: Academic Press.

Hagenaars, J.A. (1989). *Categorical Longitudinal Data*, Sage Publ.

Lauro, N. and Siciliano, R. (1989) Exploratory methods and modelling for contingency tables analysis: an integrated approach, *Statistica applicata*, 1, pp. 5-32.

Meester, A. and de Leeuw, J. (1983) *Intelligence, social milieu and the school career* (in Dutch), Leiden: Department of Data Theory.

Mooijaart, A. and van der Heijden, P.G.M. (1992) *Log-trilinear models for three-way contingency tables*, Presented at SMABS-92, Nijmegen, April 20-24, 1992.

Mooijaart, A. and van der Heijden, P.G.M. (1992) The EM algorithm for latent class analysis with constraints, *Psychometrika*, vol. 57, n. 2, pp. 261-269.

Siciliano, R. (1992) Reduced rank models for dependence analysis of contingency tables, *Italian Journal of Applied Statistics*, 481-502.

Siciliano, R, Lauro, N. and Mooijaart, A. (1990) Exploratory approach and maximum likelihood estimation of models for non symmetrical analysis of two-way multiple contingency tables, *Compstat '90*, Physica Verlag, pp. 157-162.

Siciliano, R, Mooijaart, A. and van der Heijden, P.G.M. (1992) A probabilistic model for non-symmetric correspondence analysis and prediction in contingency tables, *Journal of Italian Statistical Society*, 1, pp. 85-106.

van der Heijden, P.G.M., Mooijaart, A. and de Leeuw, J. (1989) Latent budget analysis, in A. Decarli, B.J. Francis, R. Gilchrist and G.U.H. Seeber (eds.), *Statistical Modelling. Proceedings, Trento 1989*, Berlin: Springer Verlag, pp. 301-313.

van der Heijden, P.G.M., Mooijaart, A. and de Leeuw, J. (1992) Constrained latent budget analysis, in P. Marsden, *Sociological Methodology*, 22, Cambridge: Blackwell, pp. 279-320.

Simultaneous latent budget analysis of a set of two-way tables with constant-row-sum data

Summary

In this paper we introduce a class of models for the simultaneous analysis of the latent budget structure of a set of T tables with constant-row-sum data. Such a row vector is called observed budget: it can include either compositional data (i.e. time-budget data) or conditional proportions adding up to one. In case of multidimensional contingency tables these models allow to analyse the dependence of a joint response variable given the levels of a joint explanatory variable. A further variable is used as stratifying variable of the set.

The basic model of simultaneous latent budget analysis decomposes the observed budgets of each table by a mixture of unknown, or latent, budgets. The latent budgets are built up of T sets of latent budget parameters, and the mixture is defined by T sets of mixing parameters. To perform a simultaneous analysis of the T tables we consider across-group homogeneity constraints upon parameters yielding three basic types of constrained simultaneous latent budget models. We find maximum likelihood estimates of these models with a type of EM-algorithm. At the same time we also show how to estimate all these models by using the standard software for latent budget analysis. We discuss the identifiability of the model parameters and the degrees of freedom. We show the relations of simultaneous latent budget analysis with simultaneous latent class analysis, loglinear analysis and latent budget analysis for higher-way tables. We apply the simultaneous latent budget models in two examples.

Analisi simultanea dei bilanci latenti di un insieme di tabelle a due vie

Riassunto

In questo lavoro introduciamo una classe di modelli per analizzare simultaneamente la struttura latente di un insieme di tabelle con marginale di riga fisso. Si tratta di distribuzioni condizionate, ovvero di bilanci tempo, economici, finanziari che vengono analizzati nel tempo o nello spazio, ovvero secondo le modalità di una variabile esplicativa di stratificazione. In ciascuna tabella i bilanci vengono costruiti rilevando un campione stratificato, dove gli strati sono definiti dalle modalità di un'ulteriore variabile esplicativa.

L'obiettivo è quello di spiegare l'effetto della variabile che stratifica le diverse tabelle sul legame di dipendenza tra le due variabili di ciascuna tabella. Il modello di base decompone i bilanci osservati di ciascuna tabella mediante una mistura di bilanci latenti. Al fine di analizzare in maniera simultanea la struttura latente nelle diverse tabelle, consideriamo tre classi di modelli con vincoli di omogeneità sui parametri. I parametri dei modelli sono stimati con il metodo della massima verosimiglianza utilizzando un algoritmo di tipo EM, dovendo trattare tabelle con proporzioni non osservate o latenti. Dimostriamo inoltre come è possibile ottenere le stesse stime dei parametri definendo dei modelli dei bilanci latenti per una particolare matrice diagonale a blocchi. Descriviamo l'identificazione dei parametri ed il calcolo dei gradi di libertà. Commentiamo i legami e le differenze con i modelli simultanei delle classi latenti, con i modelli log-lineari e con i modelli dei bilanci latenti per tabelle a più vie. Mostriamo infine due esempi di applicazione dei modelli proposti.

KEY WORDS

Latent budget analysis, simultaneous latent class analysis, EM-algorithm, time-budget data.