Journal of Classification 14:225-247 (1997)

Latent Class DEDICOM

Yoshio Takane

McGill University

Henk A. L. Kiers

University of Groningen

Abstract: A probabilistic DEDICOM model was proposed for mobility tables. The model attempts to explain observed transition probabilities by a latent mobility table and a set of transition probabilities from latent classes to observed classes. The model captures asymmetry in observed mobility tables by asymmetric latent mobility tables. It may be viewed as a special case of both the latent class model and DEDICOM with special constraints. A maximum penalized likelihood (MPL) method was developed for parameter estimation. The EM algorithm was adapted for the MPL estimation. Two examples were given to illustrate the proposed method.

Keywords: Mobility tables; Latent class models; Maximum penalized likelihood (MPL) method; EM algorithm; RIC.

1. Introduction

We develop a probabilistic model for mobility tables. Denote such a table by $\mathbf{F} = \{f_{ij}\}$, $i,j = 1, \ldots, n$, where f_{ij}, the element in the i-th row and the j-th column of $\mathbf{F}$, represents the observed frequency of moves from origin i to destination j. Although originally designed for mobility tables, the model

---

The work reported in this paper has been supported by grant A6394 to the first author from the Natural Sciences and Engineering Research Council of Canada and by a fellowship of the Royal Netherlands Academy of Arts and Sciences to the second author. We would like to thank anonymous reviewers for their insightful comments.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfiled Avenue, Montréal, Québec, H3A 1B1, Canada.

applies equally well to any square contingency table where there is one-to-one correspondence between rows and columns. Examples of such tables are, among others, brand loyalty data, journal citation data (e.g., Coombs, Dawes, and Tversky 1970, p. 73), discrete panel data on two occasions, stimulus identification data, agreement or association data, e.g., between two pathologists diagnosing a same group of patients using the same set of disease categories, between actual and ideal numbers of children, and between husbands' and wives' religions (e.g., Johnson 1980) or occupations in two-earner families, and so on.

These tables are typically asymmetric, that is, $f_{ij} \neq f_{ji}$. An important element in modelling such a table is how to represent the asymmetry to reveal interesting structures. In this paper we use the DEDICOM model (Harshman, Green, Wind, and Lundy 1982) to capture asymmetry in moblity tables. DEDICOM represents asymmetric relationships among manifest categories by a smaller number of latent categories that have asymmetric relationships between them. We adapt it to give probabilistic accounts of mobility tables.

In the next section we describe the proposed model in some detail and follow with an exposition of related models which either directly or indirectly motivated the proposed model (Section 3). The proposed model is usually not identifiable, and we discuss the range of unidentifiability in Section 4. We develop a maximum penalized likelihood (MPL) method for parameter estimation to deal with the problem of possible unidentifiability in the model and adapt the EM algorithm for the MPL estimation. These are described in Section 5, and further details are given in the Appendix. The MPL method introduces a penalty parameter into the estimation scheme. The problem of choosing an optimal value of the penalty parameter is investigated in Section 6. The paper concludes with two illustrative examples and discussion.

2. The Model

We model $\mathbf{P} = \{p_{ij}\}$, where p_{ij} is the transition probability from origin i to destination j. Let $a_{i|s}$ denote the probability of observed (manifest) class i given latent class s and r_{st} the probability of transition from origin latent class s to destination latent class t (a latent mobility table). We posit that

$$p_{ij} = \sum_{s,t} r_{st} a_{i|s} a_{j|t} + \delta_{ij} q_i, \tag{1}$$

where $s,t = 1, \ldots, S$, δ_{ij} is a Kronecker delta, and q_i the probability of inherent stayers in observed class i (the portion of observation units in a certain observed class who by nature will never move to another class). Note that observation units in this class are not the only ones actually staying in class i. There is another class of observation units in class i who return to class i via transitions through latent classes. The probability of such a class is given by $\Sigma_{s,t}\, r_{st}\, a_{i\,|\,s}\, a_{i\,|\,t}$. The total probability (p_{ii}) of staying in class i is the sum of this probability and q_i, the probability of inherent stayers. We require

$$\sum_{i=1}^{n} a_{i\,|\,s} = 1 \,, \tag{2}$$

for $s = 1, \ldots, S$, and

$$\sum_{s,t} r_{st} + \sum_{i=1}^{n} q_i = 1 \,, \tag{3}$$

and that all $a_{i\,|\,s}$, r_{st} and q_i are between 0 and 1 inclusive. It follows that $\Sigma_{i,j}\, p_{ij} = 1$. The model attempts to explain observed transition probabilities (p_{ij}) by a latent mobility table (r_{st}), and a set of conditional transition probabilities ($a_{i\,|\,s}$) from latent classes to observed classes. The model captures asymmetry in mobility tables by asymmetric latent mobility tables ($r_{st} \neq r_{ts}$).

The model also captures excess probabilities often observed in the diagonal entries of observed mobility tables by postulating probabilities (q_i) of inherent stayers. Quite often, diagonal elements of **F** have some special status not shared by their off-diagonal counterparts, and a special treatment is necessary for them (Clogg 1981). For example, in the trade data between nations, diagonal entries represent domestic trade, which may not be directly comparable with international trade. In the social mobility data discussed below, the q_i may be construed as probabilities of inheritance of social status. This provision is similar to the notion of uniqueness in factor analysis, is useful in dealing with missing diagonal entries, and amounts to providing n additional latent classes, one for each observed class, that do not allow transitions among themselves. Observed transitions among observed classes are accounted for by the latent classes that allow transitions among themselves. While the q_i may optionally be set equal to zero, doing so often leads to a significant deterioration in goodness of fit.

3. Related Models

The proposed model may be viewed as a special case of both the latent class model (LCM; e.g., Hagenaars 1990) and DEDICOM (e.g., Harshman et al. 1982; see, in particular, footnote 2 on page 236 of their paper, where they suggest a similar idea leading to model (1)) with special constraints. Clogg (1981) proposed the following latent class model for mobility tables:

$$p_{ij} = \sum_{u} r_u\, a_{i\,|\,u}\, b_{j\,|\,u} + \delta_{ij}\, q_i \,. \tag{4}$$

This is a special kind of conditional quasi-independence model (Goodman

1968), that is, an independence model except for the diagonal entries, conditional on latent classes. This model is similar to (1) except that two sets of conditional transition probabilities from latent classes to observed classes are differentiated in (4), one on the origin side ($a_{i|u}$) and the other on the destination side ($b_{j|u}$). While $a_{i|u}$ and $b_{i|u}$ are almost always similar, equating them in (4), as in (1), will destroy the model's ability to account for asymmetry in **F**. (See, however, Grover and Srinivasan 1987.) In Clogg's model, asymmetry is captured by differences between $a_{i|u}$ and $b_{i|u}$ for $i = 1, \ldots, n$; $u = 1, \ldots, U$. However, interpreting the differences between the two sets of quantities is not always an easy task. This directly motivated the proposed model. Model (1) singles out asymmetric components and captures them in one set of parameters, r_{st}, which facilitates the interpretation.

Hagenaars (1990) extended LCM so that latent classes have a factorial structure. Suppose there are two factors, origin and destination, that distinguish latent classes. Then, subscript u in (4) is replaced by two indices, s and t, where s ($s = 1, \ldots, S$) indicates the level of the first factor and t ($t = 1, \ldots, T$) the level of the second factor. Model (4) then becomes

$$p_{ij} = \sum_{s,t} r_{st}\, a_{i|st}\, b_{j|st} + \delta_{ij}\, q_i \,. \tag{5}$$

Note that although there are ST latent classes in model (5), we sometimes call levels of the first factor "origin latent classes," and levels of the second factor "destination latent classes." If we further assume that $S = T$, that there is one-to-one correspondence between levels of the two factors (as in the observed mobility table), and that $a_{i|st} = a_{i|s}$ for all t, and $b_{j|st} = b_{j|t}$ for all s, we obtain

$$p_{ij} = \sum_{s,t} r_{st}\, a_{i|s}\, b_{j|t} + \delta_{ij}\, q_i \,. \tag{6}$$

The constraints that $a_{i|st} = a_{i|s}$ and $b_{j|st} = b_{j|t}$ are not so restrictive, help reduce the number of parameters in the model considerably, and make clear that s is specifically related to the origin side and t to the destination side. In model (6) we may assume $r_{st} = r_{ts}$ while capturing asymmetry in **P** by differences between $a_{i|s}$ and $b_{i|s}$, but then we end up with a similar situation as in model (4). A more attractive idea is to capture the asymmetry in **P** by asymmetry in r_{st} while equating $a_{i|s}$ and $b_{i|s}$, which no longer have to be differentiated for the purpose of capturing asymmetry. We then have model (1), which can thus be regarded as a special kind of LCM with special constraints on its parameters (Mooijaart and van der Heijden 1992).

Let $\mathbf{A} = \{a_{i|s}\}$, $\mathbf{R} = \{r_{st}\}$ and $\mathbf{D} = \text{diag}\,\{q_i\}$ where diag $\{q_i\}$ is a diagonal matrix with diagonal elements equal to q_i. Define

$$\mathbf{A}^* = [\mathbf{A}, \mathbf{I}_n],$$

and

$$\mathbf{R}^* = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0}' & \mathbf{D} \end{bmatrix},$$

where $\mathbf{0}$ is an S by n matrix of zeroes. Then,

$$\mathbf{P} = \mathbf{A}^* \mathbf{R}^* \mathbf{A}^{*'} = \mathbf{ARA}' + \mathbf{D}, \tag{7}$$

which is a special case of DEDICOM (Harshman, et al. 1982) with special 0-1 constraints on some portions of $\mathbf{A}^*$, $\mathbf{R}^*$, and $\mathbf{D}$ (as well as constraints (2) and (3) and $0 \le a_{i|s}$, r_{st} and $q_i \le 1$). Model (1) can thus also be regarded as a special case of DEDICOM.

An extensive literature exists on models of mobility tables. Hout (1983) and Breiger (1990) provide concise descriptions of other representative models of mobility tables not discussed above. Also, see Duncan (1979) who proposed association models for social mobility tables, Yamaguchi (1983) who proposed a quasi- independence model for samples of subjects stratified by educational attainment, and Sato and Sato (1995) who presented a model similar to the one proposed in this paper under the name "fuzzy clustering."

4. Identifiability of the Model

The $\mathbf{ARA}'$ part of (7) is usually not unique. Similar unidentifiability properties have been noted elsewhere in similar contexts (Clogg 1981; de Leeuw, van der Heijden, and Verboon 1990). Nonuniqueness derives from the fact that $\mathbf{ARA}' = \mathbf{ATT}^{-1}\mathbf{R}(\mathbf{T}')^{-1}\mathbf{T}'\mathbf{A}' = \mathbf{BCB}'$ for any square nonsingular matrix $\mathbf{T}$, where $\mathbf{B} = \mathbf{AT}$ and $\mathbf{C} = \mathbf{T}^{-1}\mathbf{R}(\mathbf{T}')^{-1}$. However, $\mathbf{B}$ and $\mathbf{C}$ have to satisfy the same constraints as $\mathbf{A}$ and $\mathbf{R}$, which limits the range of admissible $\mathbf{T}$. We have

$$\mathbf{1}'_S = \mathbf{1}'_n\mathbf{B} = \mathbf{1}'_n\mathbf{AT} = \mathbf{1}'_S\mathbf{T}, \tag{8}$$

and

$$1 - \sum_i q_i = \mathbf{1}'_S\mathbf{C}\,\mathbf{1}'_S = \mathbf{1}'_S\mathbf{T}^{-1}\mathbf{R}(\mathbf{T}^{-1})'\,\mathbf{1}'_S. \tag{9}$$

For every $\mathbf{T}$ satisfying (8), we have $\mathbf{1}'_S\mathbf{T}^{-1} = \mathbf{1}'_S$, hence

$$\mathbf{1}'_S\mathbf{T}^{-1}\mathbf{R}(\mathbf{T}^{-1})'\,\mathbf{1}_S = \mathbf{1}'_S\mathbf{R}\,\mathbf{1}_S = 1 - \sum_i q_i. \tag{10}$$

It follows from (10) that, for every $\mathbf{T}$ satisfying (8), constraint (9) is satisfied as well. In addition to (8), however, the matrix $\mathbf{T}$ should be constrained so that the elements of $\mathbf{B}$ and $\mathbf{C}$ also satisfy $0 \le b_{ij} \le 1$ and $0 \le c_{ij} \le 1$, which further restricts the range of admissible **T**. However, these restrictions are usually not sufficient to determine **T** uniquely. That is, **T** other than $\mathbf{I}_S$ is still possible. In general, it is difficult to decide whether **T** is unique in a given situation without actually fitting the model.

In some cases there are strong theoretical reasons to set some of the model parameters to zero. We have already seen that some of the elements in $\mathbf{A}^*$ and $\mathbf{R}^*$ are fixed at zero, but we may also set some of the elements in **A** and **R** to zero. For example, we may postulate that Latent Class 1 is such that $a_{1|1} = 0$, implying that observed category 1 can never arise from this latent class. Or we may posit that latent classes are ordered in some way, for example, according to their economic status, but transitions between them are only in one direction, say, from upper to lower classes. We may then set $r_{st} = 0$ for all $s < t$. These additional constraints help further reduce the degree of nonuniqueness in the model, and in some cases they are sufficient to identify the model. However, it is still not at all easy to decide, on an *a priori* basis, whether a given set of constraints are sufficient to determine the model uniquely.

So whether or not there are additional constraints, a specification of conditions under which the model is unique is extremely difficult, if not impossible, particularly when the number of latent classes is large. A method is, therefore, necessary that works regardless of whether the model is unique. It should be able to produce unique parameter estimates when the model is nonunique and to handle fixed parameters easily.

5. Parameter Estimation

The maximum likelihood (ML) estimation method often used in the analysis of contingency tables is not adequate for our model. When the model is not identified, it leads to nonunique parameter estimates, causing great difficulty, because we do not know which set of possible parameter estimates we should interpret. Presumably, they are all equally good. The ML method also often leads to boundary estimates (i.e., estimates of $a_{i|s}$, r_{st} and q_i which are on the boundary of the parameter space, that is, 0 or 1 in the present case) for which the usual asymptotic properties of maximum likelihood estimates do not hold. The EM algorithm (Dempster, Laird, and Rubin 1977) used in the optimization also becomes infinitely slow in such cases. (Note that the situation just described is fundamentally different from the one in which these parameters are *a priori* fixed.)

We use the maximum penalized likelihood (MPL) method to avoid these difficulties, because it can uniquely determine the parameter estimates and also avoid boundary estimates. By avoiding boundary estimates, the EM algorithm is often faster than when it is used for ML estimation. Like the ML method, the MPL method can handle the fixed constraints quite easily. (We thank an anonymous reviewer for pointing out the importance of this feature.) The MPL estimators are asymptotically equivalent to the ML estimators, because the effect of the penalty term (see below) diminishes relative to the log likelihood term as the sample size increases.

In the MPL method we consider

$$\ln L_\rho = \Sigma_{i,j} f_{ij} \ln p_{ij} - \Sigma_s \lambda_s(\Sigma_i a_{i|s} - 1) - \lambda(\Sigma_{s,t} r_{st} + \Sigma_i q_i - 1) + \rho(\Sigma_{i,s} \delta_{is} \ln a_{i|s} + \Sigma_{s,t} \delta_{st} \ln r_{st} + \Sigma_i \delta_i \ln q_i) , \tag{11}$$

where p_{ij} is given by (1), the λ's are Lagrangean multipliers, ρ is a small positive number representing the penalty parameter, and the δ's indicate whether the corresponding parameters are *a priori* fixed at zero. More precisely, $\delta_{is} = 0$ if $a_{i|s}$ is fixed at zero, and $\delta_{is} = 1$ otherwise, $\delta_{st} = 0$ if r_{st} is fixed at zero, and $\delta_{st} = 1$ otherwise, and $\delta_i = 0$ if q_i is fixed at zero, and $\delta_i = 1$ otherwise. We maximize $\ln L_\rho$ with respect to $a_{i|s}$, r_{st}, and q_i, and obtain its stationary point with respect to λ_s ($s = 1, \ldots, S$) and λ. The first term of (11) is the log likelihood term, and the second and third terms (related to λ and λ_s) are for incorporating side conditions, (2) and (3). The penalty term (related to ρ) introduces nonzero gradients in the direction along which the likelihood function is flat, and thereby uniquely determines the parameter estimates. The uniqueness is assured by nonsingularity of the Hessian matrix derived from $\ln L_\rho$. (The general form of the Hessian matrix is given in the Discussion section). In general, the penalty has the effect of forcing nonfixed parameter estimates away from the boundary of the parameter space. The strength of force is modulated by the penalty parameter. The value of ρ will be chosen to increase the predictive power of the estimates, as will be explained in the next section. When an optimal value of ρ happens to be zero, the MPL method reduces to the ML method.

The EM algorithm for maximum likelihood estimation in LCM (Goodman 1979) can readily be extended to the MPL method (Green 1990). The EM algorithm consists of the following two steps:

E-Step. Evaluate

$$f^*_{ijst} = f_{ij}\, p_{st|ij} , \tag{12}$$

where $p_{st|ij} = r_{st}\, a_{i|s}\, a_{j|t} / p_{ij}$, and

$$f^*_{iii} = f_{ii}\, p_{i|ii} , \tag{13}$$

where $p_{i|ii} = q_i / p_{ii}$. Note that $f^*_{ijst} = 0$ if $\delta_{is} = 0$, $\delta_{jt} = 0$ or $\delta_{st} = 0$, and $f^*_{iii} = 0$ if $\delta_i = 0$.

M-Step. Update each nonfixed parameter by

$$a_{i|s} = f^*_{is} / \sum_k f^*_{ks} , \tag{14}$$

where $f^*_{is} = \Sigma_{j,t} (f^*_{ijst} + f^*_{jits}) + \delta_{is}\, \rho$,

$$r_{st} = f^*_{st} / N^* , \tag{15}$$

where $f^*_{st} = \Sigma_{i,j} f^*_{ijst} + \delta_{st}\, \rho$, and $N^* = \Sigma_{s,t} f^*_{st} = N + n^* \rho$, with N being the sample size (i.e., $N = \Sigma_{i,j} f_{ij} = \Sigma_{i,j} \Sigma_{s,t} f^*_{ijst} + \Sigma_i f^*_{iii}$) and n^* the number of free parameters in $\{r_{st}, q_i\}$ (i.e., $n^* = \Sigma_{s,t}\, \delta_{st} + \Sigma_i\, \delta_i$), and

$$q_i = (f^*_{iii} + \delta_i\, \rho) / N^* . \tag{16}$$

The two steps are alternated until convergence is reached. A detailed derivation of the algorithm is given in the Appendix. We use uniform random numbers for initial estimates of parameters, rescaled to satisfy constraints (2) and (3).

The above algorithm is similar to the iterative proportional fitting algorithm (Bishop, Fienberg, and Holland 1975, pp. 83-102) for log-linear contingency table analyses, and has several advantages over the Newton-Raphson method often used in similar contexts. Constraints (2) and (3) as well as $0 \le a_{i|s}$, r_{st} and $q_i \le 1$ are automatically satisfied. The algorithm is also monotonically convergent. The convergence may, however, be very slow. If necessary, we may use one of various acceleration techniques developed for the EM algorithm (e.g., Jamshidian and Jennrich 1993; Meng and Rubin 1993). In the last few iterations, it may also be helpful to use the score method, which requires the information matrix, but techniques to obtain the observed information matrix have been discussed by Lang (1992) and Louis (1982). The score method also provides asymptotic variance and covariance estimates of estimated parameters. Asymptotic properties of the MPL estimators have been discussed by Cox and O'Sullivan (1990) and Gu and Qiu (1993). Most of the asymptotic properties of ML estimators also hold for the MPL estimators.

6. Choosing the Value of ρ

Various techniques have been proposed for choosing an optimal value of the penalty parameter, ρ and include the generalized cross validation (Craven and Wahba 1979), methods based on marginalization (BAIC; Ishiguro and Sakamoto 1983; Sakamoto 1991), and those based on RIC (Regularlized Information Criterion; Shibata 1989). RIC is an extension of AIC (Akaike 1973) to the MPL method and allows choosing the best value of the penalty parameter. However, the closed-form evaluation of RIC requires the penalty term defined by a summation indexed in the same way as the likelihood term. Unfortunately, this is not the case in the present situation. We, therefore, use a method based on a bootstrap estimate of RIC. Shibata (1995) discusses five asymptotically equivalent ways of obtaining bootstrap estimates of RIC. We use the one least computationally involved, originally proposed by Cavanaugh and Shumway (1994).

Let $\mathbf{F}_k^*$ be the k-th bootstrap sample, and let the maximum penalized likelihood estimate of model parameters based on $\mathbf{F}_k^*$ be denoted by $\hat{\theta}(\mathbf{F}_k^*)$. Let $\ln L_\rho^*(\mathbf{F}, \hat{\theta}(\mathbf{F}_k^*))$ represent the maximum penalized likelihood of $\mathbf{F}$ based on $\hat{\theta}(\mathbf{F}_k^*)$. (Under this notation $\ln L_\rho^*(\mathbf{F}) = \ln L_\rho^*(\mathbf{F}, \hat{\theta}(\mathbf{F}))$.) Then, the bootstrap estimate of RIC by Cavanaugh & Shumway's formula is given by

$$\mathrm{RIC} = 2\ln L_\rho^*(\mathbf{F}) - 4\Big(\sum_k \ln L_\rho^*(\mathbf{F}, \hat{\theta}(\mathbf{F}_k^*))/K\Big). \tag{17}$$

As with AIC, the model associated with the smallest value of RIC is considered the best fitting model. When $\rho = 0$, RIC formally reduces to TIC (Takeuchi Information Criterion; Shibata 1989), which is a generalization of AIC. AIC approximates TIC when the fitted model is reasonably close to the true model and allows comparisons of models across different values of ρ as well as across different models for a specific value of ρ.

RIC assures that the MPL estimators are most predictive of future observations among other alternative estimators such as ML. This statement is true regardless of whether the model is unique, or whether additional constraints are incorporated to reduce the degree of nonuniqueness in the model. The situation is analogous to that of ridge regression where ridge estimators may be biased, but may still have smaller expected mean square errors than the usual least squares estimator. The MPL method with an optimal value of ρ thus has the dual role of obtaining unique and most predictive parameter estimates, regardless of the uniqueness of the model.

7. Illustrative Examples

In this section we report two examples of application. The first draws on the car switching data (Harshman et al. 1982), in which we compare estimates obtained from the proposed method with those obtained by Kiers and Takane (1993). The reported analyses exemplify the constrained estimation feature of the proposed method. The second example analyzes social mobility data from Miller (1960) and demonstrates usefulness of the proposed method in practical data analysis situations.

7.1 Car Switching Data

Harshman et al. (1982, p. 221) presented a contingency table with car switching frequencies among 16 types of cars. The labels for the 16 car types are listed in the first column of Table 1, where the first three characters indicate size (SUB = subcompact, SMA = small specialty, COM = compact, MID = midsize, STD = standard, and LUX = luxury), while the fourth character indicates origin or price (D = domestic, C = captive imports, I = imports, L = low price, M = medium price, and S = specialty). Kiers and Takane (1993) reanalyzed the data by their unconstrained and constrained least squares DEDICOM procedure. In their first analysis, they obtained an unconstrained three-dimensional solution, which was then rotated into a simple structure solution by the normalized varimax rotation. This solution accounted for 86.4% of the total sum of squares (SS). The three dimensions were interpreted as representing clusters of (a) plain large and midsize cars, (b) fancy large cars, and (c) small/specialty cars. It was then hypothesized that each car category was represented by one and only one of these dimensions. For each car category a dimension with the highest loading was identified, and loadings on other dimensions were constrained to be zero. In their second analysis, Kiers and Takane fitted this strictly simple-structured pattern hypothesis by the constrained DEDICOM procedure. The resulting solution (called "constrained nonoverlapping" by Kiers and Takane) accounted for 83.7% of the total SS. In the third analysis, the strictly simple-structured pattern hypothesis was relaxed by allowing multiple nonzero loadings in each car category. Specifically, all loadings in the unconstrained solution larger than .20 in absolute values were left unconstrained in addition to the largest loading within each car type. This quasi-simple-structured pattern hypothesis (called "constrained overlapping" by Kiers and Takane) accounted for 85.3% of the total SS.

For direct comparisons, we replicated the above three analyses using the proposed method. All the results from latent class DEDICOM in this section were obtained with $\rho = 1.0$. This value may not be optimal, since no serious attempt was made to find its optimal value. However, the effect of this value seems relatively minor in this particular instance, because derived estimates are in all cases very close (identical up to three decimal places) to the ML estimates, which happened to be unique. (That the ML estimates were unique in all three cases was confirmed by repeating the ML estimation several times starting from different initial estimates.)

We first fitted the unconstrained three-dimensional latent class DEDICOM. Derived parameter estimates are presented in columns 2-4 of Table 1. The loading pattern (relative sizes of $a_{i|s}$) in this solution is strikingly similar to that of the unconstrained-varimax solution of Kiers and Takane, despite the

Table 1: MPL Estimates of $a_{i|s}$ and r_{st} for the Car Switching Data.

	Unconstrained			Constrained Nonoverlapping			Constrained Overlapping		
Latent Class	1	2	3	1	2	3	1	2	3
$a_{i\|s}$									
SUBD	.05	*.00	.17			.21			.22
SUBC	.01	.00	.02			.02			.03
SUBI	.01	.01	.16			.16			.17
SMAD	.01	.03	.20			.22			.23
SMAC	.00	.00	.00			.00			.00
SMAI	.00	.01	.04			.04			.05
COML	.12	*.00	.06	.19			.17		
COMM	.06	*.00	.02	.09			.08		
COMI	.01	.00	.02			.03			.03
MIDD	.29	.04	.06	.38			.34		
MIDI	.01	.01	.02	.03			.03		
MIDS	.04	.12	.22			.32		.11	.29
STDL	.28	*.03	*.01	.31			.27		
STDM	.13	.37	*.00		.63		.12	.44	*.00
LUXD	*.00	.35	.00		.35		*.00	.42	
LUXI	.00	.02	.00		.03			.03	
r_{st} (multiplied by 1000)									
Dim. 1	284	28	215	201	55	215	262	35	238
Dim. 2	000	174	22	37	109	60	*14	121	26
Dim. 3	000	000	277	67	31	225	72	13	219

The "*" indicates that the corresponding estimate was negative in Kiers and Takane (1993) fact that no rotation was applied in the former. The apparent difference in the overall size of the loadings results from the difference in the scaling convention between the two. Whereas in latent class DEDICOM $a_{i|s}$'s are all nonnegative and satisfy (2), no such constraints are in effect in Kiers and Takane. This means that scaling is essentially arbitrary in the latter, but since some negative loadings occurred, they were scaled so that the sum of squares was unity within each dimension. Although nonnegativity constraints on model parameters are a more restrictive aspect of the proposed model, in most cases their effects on overall goodness of fit are relatively minor, while allowing the precise interpretation of the loadings as conditional probabilities. The unconstrained three-dimensional latent class DEDICOM model accounted for 84.4% of the total SS, which is only slightly smaller than the percent SS accounted for by the corresponding solution in Kiers and Takane. The estimates of conditional probabilities corresponding to the negative loadings in Kiers and Takane's solution (marked by asterisks in Table 1) are all close to zero.

We then fitted the same strictly simple-structured pattern hypothesis as in the second analysis of Kiers and Takane. The results are reported in columns 5-7 of Table 1, where entries left blank indicate the loadings (conditional probabilities) constrained to be zero. Derived estimates are even more similar to those in the corresponding solution of Kiers and Takane. This solution accounted for 82.5% of the total SS. No loadings happened to be negative in Kiers and Takane's solution, so the estimates of loadings were scaled to satisfy the same constraints as in latent class DEDICOM. Kiers and Takane argued that under this normalization convention their estimates of the loadings could be interpreted as approximate conditional probabilities. In the latent class DEDICOM model this interpretation is always possible and exact.

Finally, we fitted the same quasi-simple-structured pattern hypothesis as in the third analysis of Kiers and Takane. Derived estimates are presented in columns 8-10 of Table 1. Again, the overall loading pattern in this solution is very similar to that of Kiers and Takane's constrained DEDICOM solution, in which the same scaling convention was used as in the unconstrained-varimax solution, since there were again negative estimates of loadings. The loadings in Kiers and Takane's solution, therefore, could not be interpreted as conditional probabilities. Conditional probabilities corresponding to the negative estimates in Kiers and Takane's solution (again marked by asterisks) turned out to be close to zero in the latent class DEDICOM solution. This solution accounted for 83.5% of the total SS. The percent SS's accounted for by the latent class DEDICOM model are in all cases somewhat smaller than those in the corresponding solutions by Kiers and Takane. This result stems partly from the fact that in the latter this quantity is explicitly maximized, whereas a different criterion is optimized in latent class DEDICOM. It may also be that the nonnegativity constraints in the latent class DEDICOM model are restrictive (*albeit* only slightly).

Estimates of **R**'s are given at the bottom of the table and are again on different scales from those in Kiers and Takane's solution. However, the overall tendency in the latent mobility tables, that is, the predominant moves from plain large and midsize cars to fancy large or small/specialty cars, remain essentially the same as in Kiers and Takane (1993) and Harshman et al. (1982).

7.2 Social Mobility Data

As a second example, we use intergenerational social mobility data which are a 8×8 joint frequency table of father's social status and son's status in Britain in 1959. The eight social status categories are: 1. professional and high administrative; 2. managerial and executive; 3. inspectional, supervisory and other nonmanual (of high grade); 4. the same as in 3, but of low grade; 5. routine grades of nonmanual; 6. skilled manual; 7. semi-skilled manual; 8. unskilled manual. The data, displayed in Table 2, were taken from Clogg (1981) who presented them with a remark on how he made a correction in the data originally presented by Miller (1960). Several previous authors analyzed the same data set (e.g., Duncan 1979). To benchmark the goodness of fit of the proposed model, the AIC of the saturated model is 538.9 with 63 parameters for this data set, that of the independence model is 1395.0 with 14 parameters, and that of the quasi-independence model fitted to off-diagonal elements is 903.7 with 22 parameters. (Note that in these models no penalty should be introduced to determine the model parameters uniquely, so that the straightforward ML method was used to fit these models. We use AIC as an approximation to RIC(0) where 0 is the value of ρ.)

Table 3 gives bootstrap estimates of RIC as a function of ρ and the number of origin (and destination) latent classes (S). The sample size of the bootstrap study was 100. Numbers in parentheses indicate the bias terms added to minus twice the maximum log penalized likelihood to obtain the value of RIC. These are analogous to twice the number of parameters in AIC but take into account the amount of penalty incorporated in the optimization criterion. The value of ρ was varied from .001 to 1 with an increment factor of 10. The minimum RIC solution is obtained when $\rho = .01$ and $S = 4$. The RIC value of 519.2 compares favorably with both the AIC values of all of the benchmark models mentioned above as well as with the AIC value of 543.1 with 35 parameters associated with the best fitting model obtained by Clogg (1981).

Table 2: The Social Mobility Data from Miller (1960).

Father's Status	Son's Status 1	2	3	4	5	6	7	8	Raw Totals
1	50	19	26	8	7	11	6	2	129
2	16	40	34	18	11	20	8	3	150
3	12	35	65	66	35	88	23	21	345
4	11	20	58	110	40	183	64	32	518
5	2	8	12	23	25	46	28	12	156
6	12	28	102	162	90	553	230	177	1354
7	0	6	19	40	21	158	143	71	458
8	0	3	14	32	15	126	91	106	387
Column Totals	103	159	330	459	244	1185	593	424	3497

Table 3: Bootstrap Estimates of RIC and the Bias for the Social Mobility Data.

ρ	.001	.01	.1	1
Dimensionality				
3	536.1	539.2	572.7	823.0
	(59.6)	(59.1)	(62.2)	(59.9)
4	520.2	*519.2	558.8	900.8
	(81.1)	(75.8)	(76.7)	(65.0)
5	526.7	525.1	580.5	1033.5
	(97.3)	(89.6)	(92.5)	(73.5)

*Minimum RIC solution
The correction factor (bias) to the MPL is given in parentheses.

Table 4: MPL Estimates of $a_{i|s}$ and r_{st} in the 4×4 Latent Classes Solution for the Social Mobility Data.

Conditional Probabilities of Occupational Categories Given Latent Classes

Occupational Category	Latent Class I	II	III	IV	Probability of Stayer (q_i)
1	.000	.005	.182	.000	.011
	(.000)	(.010)	(.067)	(.000)	(.003)
2	.000	.003	.290	.080	.003
	(.003)	(.003)	(.063)	(.045)	(.002)
3	.014	.153	.293	.194	.002
	(.012)	(.034)	(.044)	(.086)	(.002)
4	.065	.294	.104	.195	.007
	(.024)	(.049)	(.048)	(.087)	(.005)
5	.024	.064	.026	.266	.003
	(.014)	(.028)	(.021)	(.072)	(.002)
6	.436	.477	.150	.047	.020
	(.038)	(.072)	(.056)	(.086)	(.014)
7	.253	.001	.000	.186	.011
	(.037)	(.008)	(.004)	(.093)	(.005)
8	.207	.003	.000	.033	.012
	(.023)	(.019)	(.003)	(.039)	(.004)

Joint Probabilities of Origin and Destination Latent Classes

Origin Latent Class	Destination Latent Class I	II	III	IV	Total ($r_{s.}$)
I	.411	.001	.000	.051	.463
	(.055)	(.015)	(.000)	(.025)	
II	.083	.169	.011	.051	.313
	(.033)	(.041)	(.009)	(.025)	
III	.006	.005	.088	.027	.126
	(.005)	(.004)	(.021)	(.012)	
IV	.003	.002	.000	.023	.028
	(.011)	(.016)	(.001)	(.037)	
Total ($r_{.t}$)	.502	.177	.099	.153	

Unique parameter estimates corresponding to the minimum RIC solution are given in Table 4. Origin latent classes are arranged in descending order of their marginal probabilities (i.e., $r_{s.} = \Sigma_t r_{st}$). Numbers in parentheses are estimates of standard errors obtained by the bootstrap method. (They are byproducts of bootstrap estimates of RIC.) Note that zeroes (.000) in the table are not exactly equal, but simply very close to zero. Latent Class I (both origin and destination) represents the low end of social status, while Class III is the opposite. Classes II and IV represent middle strata in the spectrum. Class IV is a somewhat more diffuse class attracting people from many observed categories. Transition probabilities between different latent classes are relatively small with relatively large probabilities concentrated on diagonals. Nonetheless, we see some asymmetry in the table of r_{st}. Latent Classes III and II (both representing relatively high social status) tend to shrink, while I and IV tend to grow, as indicated by the comparison between row and column marginals of r_{st} (i.e., $r_{s.}$ and $r_{.s}$). The probability of Class II $\rightarrow$ I is much larger than the reverse. Also, the probabilities of Classes I, II and III $\rightarrow$ IV are larger than the reverse. It looks as if Class IV (and to a lesser extent, Class I) is an attractor. The stayer probability tends to be large for observed categories 1, 6, 7 and 8 (all representing extreme social status, none in the middle). However, one should note that q_i as well as $a_{i|s}$ are confounded with the size of observed class i.

The above observations can be made with the estimates of original parameters of the model. To characterize the nature of the latent classes, however, it may be better to interpret the conditional probabilities of latent classes given observed classes. These quantities can readily be derived from original parameters of the model by $a^*_{s|i} = a_{i|s} r_{s.}/p_{i.}$, $a^{**}_{t|j} = a_{j|t} r_{.t}/p_{.j}$, $q^*_i = q_i/p_{i.}$ and $q^{**}_i = q_i/p_{.j}$, where $p_{i.} = \Sigma_s a_{i|s} r_{s.}$ and $p_{.j} = \Sigma_t a_{j|t} r_{.t}$ are marginal probabilities of origin observed class i and destination observed class j, respectively. These conditional probabilities, as well as the conditional probabilities of destination latent classes given origin latent classes ($r_{t|s}$) and the conditional probabilities of origin latent classes given destination latent classes ($r_{s|t}$) are given in Tables 5 and 6. Note that because of the differences between $r_{s.}$ and $r_{.s}$, and between $p_{i.}$ and $p_{.i}$, $a^*_{s|i}$ and $a^{**}_{s|i}$ differ, as do q^*_i and q^{**}_i.

Interpretations of the latent classes are clearer, but remain essentially intact. Also, we can now clearly see that the conditional probability of inherent stayers in observed Class 1 (professional & high managerial) is extremely high compared to that in other observed classes. The $r_{t|s}$ is called outflow probability describing the distribution of destination latent classes for given origin latent classes, while $r_{s|t}$ is called inflow probability describing the distribution of origin latent classes for given destination latent classes. Both probabilities confirm the general patterns of asymmetry observed in the table of r_{st}. These quantities may also be useful in modeling outflow and inflow probabilities between observed origin and destination classes, $p_{j|i} = \Sigma_{s,t} a^*_{s|i} r_{t|s} a_{j|t} + \delta_{ij} q^*_i$, and $p_{i|j} = \Sigma_{s,t} a_{i|s} r_{s|t} a^{**}_{t|j} + \delta_{ij} q^{**}_i$, respectively.

Table 5: Conditional Probabilities ($a_{s|i}$ and $r_{t|s}$) Derived from the Estimates in Table 4.

Conditional Probabilities of Origin Latent Classes Given Father's Occupational Categories

Father's Occupational Category	Origin Latent Class I	II	III	IV	Conditional Probability of Stayer
1	.000	.047	.637	.000	.315
2	.006	.025	.843	.051	.075
3	.066	.484	.375	.054	.021
4	.205	.622	.089	.036	.047
5	.253	.444	.073	.164	.066
6	.524	.387	.034	.003	.051
7	.875	.002	.000	.038	.085
8	.871	.009	.000	.008	.112

Conditional Probabilities of Destination Latent Classes Given Origin Latent Classes

Origin Latent Class	Destination Latent Class I	II	III	IV
I	.886	.003	.000	.111
II	.264	.539	.034	.163
III	.050	.037	.696	.217
IV	.094	.060	.003	.843

Table 6: Conditional Probabilities ($a_{t|j}$ and $r_{s|t}$) Derived from the Estimates in Table 4.

Conditional Probabilities of Destination Latent Classes Given Son's Occupational Categories

Father's Occupational Category	Destination Latent Class I	II	III	IV	Conditional Probability of Stayer
1	.000	.032	.592	.001	.375
2	.007	.014	.635	.273	.072
3	.074	.285	.305	.314	.022
4	.247	.394	.078	.226	.053
5	.176	.161	.036	.584	.042
6	.644	.248	.030	.021	.058
7	.760	.001	.000	.170	.068
8	.853	.005	.000	.041	.101

Conditional Probabilities of Origin Latent Classes Given Destination Latent Classes

Origin Latent Class	Destination Latent Class I	II	III	IV
I	.818	.008	.000	.336
II	.165	.956	.108	.333
III	.012	.027	.891	.179
IV	.005	.009	.001	.152

8. Discussion

In this paper we proposed the probabilistic DEDICOM model for square asymmetric contingency tables to capture asymmetry in an easily interpretable manner. Unidentifiability of model parameters was overcome by the use of the MPL (maximum penalized likelihood) method. As in all LCM, the proposed model is fully probabilistic in the sense that all parameters in the model represent some kinds of probabilities, which optionally can be turned into other kinds of probabilities, as desired.

We currently use the bootstrap estimates of RIC for model selection. This approach is somewhat unwieldy, particularly given that each model has to be fitted several times to each of several hundred bootstrap samples to ensure a globally optimal solution. Computing the effective number of parameters taking into account the penalty term is one promising alternative. This is given by $n_p{*} = \mathrm{tr}((\mathbf{H} + \rho\Sigma)^{-1}\,\mathbf{H})$, where $\mathbf{H}$ is the Hessian of the log-likelihood part, and Σ that of the penalty part, of the penalized log likelihood function. The idea is that introducing the penalty term is equivalent to increasing the number of observations, which in turn is equivalent to discounting the number of parameters. Exactly how much is expressed in the formula above, which reduces to $n_p = \mathrm{rank}(\mathbf{H})$ when there is no penalty term (i.e., $\rho = 0$), but in general $n_p{*} \le n_p$. This number is used as the effective number of parameters in AIC.

The proposed model can be generalized in various directions. Extensions to multiway tables are one (e.g., Harshman et al. 1982). Multiway tables take two different forms. One is transition tables at two occasions in several populations, and the other is transition tables at multiple occasions. In either case we may posit multiway latent mobility tables, **R**, possibly with different structures imposed on them. For the second type of multiway tables, this has been done by Böckenholt and Langeheine (1996), albeit in a somewhat different context. Structuring parameters in the model is another interesting possibility even for the two-way model. For example, we may impose a quasi-independence model on **R**. Other kinds of constraints in the DEDICOM model have been discussed by Kiers and Takane (1993).

Appendix

We derive the algorithm described in the parameter estimation section. We also show how the MPL method can be derived from a MAP (maximum *a posteriori*) estimation method in a Bayesian framework. The penalized log likelihood given in (11) is for incomplete (observed) data. The corresponding penalized log likelihood for complete data is given by

$$\ln L_\rho^{(c)} = \sum_{i,j} \sum_{s,t} f_{ijst} (\delta_{is} \ln a_{i|s} + \delta_{jt} \ln a_{j|t} + \delta_{st} \ln r_{st})$$

$$+ \Sigma_i f_{iii}\, \delta_i \ln q_i - (\text{terms related to } \lambda\text{'s in (11)})$$

$$+ \rho\, (\Sigma_{i,s}\, \delta_{is} \ln a_{i|s} + \Sigma_{s,t}\, \delta_{st} \ln r_{st} + \Sigma_i\, \delta_i \ln q_i)\,, \tag{18}$$

where f_{ijst} is the unobserved joint frequency of moves from observed class i to j in latent class st, and similarly, f_{iii} is the unobserved frequency of those staying in observed class i in stayer latent class i.

The E-step requires taking the expected value of $\ln L_\rho^{(c)}$, given the observed data (f_{ij}) and the current values of parameters, $a_{i|s}^{(0)}$, $r_{st}^{(0)}$, and $q_i^{(0)}$. That is,

$$\ln L_\rho^{(c)*} = E\,(\ln L_\rho^{(c)} \mid f_{ij}, a_{i|s}^{(0)}, r_{st}^{(0)}, q_i^{(0)})\,. \tag{19}$$

It amounts to replacing f_{ijst}, f_{iii}, $a_{i|s}$, r_{st}, and q_i in (18) by f_{ijst}^*, f_{iii}^*, $a_{i|s}^{(0)}$, $r_{st}^{(0)}$, and $q_i^{(0)}$, where f_{ijst}^* and f_{iii}^* are given in (12) and (13), respectively.

In the M-step, $a_{i|s}^{(0)}$, $r_{st}^{(0)}$, and $q_i^{(0)}$ are assumed variable, and (19) is maximized with respect to them. Differentiating $\ln L_\rho^{(c)*}$ with respect to a nonfixed $a_{i|s}^{(0)}$ and setting the result equal to zero leads to

$$\lambda_s\, a_{i|s}^{(0)} = \sum_{j,t} (f_{ijst}^* + f_{jits}^*) + \rho\,, \tag{20}$$

so that an update of $a_{i|s}^{(0)}$ will be given by

$$a_{i|s}^{(0)} = [\sum_{j,t} (f_{ijst}^* + f_{jits}^*) + \rho] / \lambda_s\,. \tag{21}$$

The λ_s is obtained by summing both sides of (20) over i. That is,

$$\lambda_s = \sum_i [\sum_{j,t} (f_{ijst}^* + f_{jits}^*) + \delta_{is}\, \rho]\,, \tag{22}$$

since $\Sigma_i\, a_{i|s}^{(0)} = 1$ because of (2). The above (21) with (22) is essentially the update (14). Similarly, differentiating $\ln L_\rho^{(c)*}$ with respect to a nonfixed $r_{st}^{(0)}$ and $q_i^{(0)}$ and setting the results equal to zero lead to

$$\lambda\, r_{st}^{(0)} = \sum_{i,j} f_{ijst}^* + \rho\,, \tag{23}$$

and

$$\lambda\, q_i^{(0)} = f_{iii}^* + \rho\,. \tag{24}$$

Summing both sides of (23) and (24) over st and i, respectively, and adding them together, we obtain

$$\lambda = \sum_{s,t} \sum_{i,j} f^*_{ijst} + \sum_i f^*_{iii} + n^* \rho , \tag{25}$$

because of constraint (3). We obtain (15) and (16) by substituting this expression for λ in (23) and (24), respectively.

We may rearrange (18) into

$$\begin{aligned} \ln L^{(c)}_\rho &= \Sigma_{i,s}\, \delta_{is} [\Sigma_{j,t} (f_{ijst} + f_{jits}) + \rho] \ln a_{i\,|\,s} \\ &+ \Sigma_{s,t}\, \delta_{st} [\Sigma_{i,j} f_{ijst} + \rho] \ln r_{st} \\ &+ \delta_i\, \Sigma_i\, [f_{iii} + \rho] \ln q_i \\ &- (\text{ terms related to } \lambda\text{'s}) . \end{aligned} \tag{26}$$

Doing so indicates that the MPL method can also be construed as the MAP (maximum *a posteriori*) estimation method in a Bayesian framework. We take Dirichlet distributions as prior distributions of $a_{i\,|\,s}$, $s = 1, \ldots, S$ and of r_{st} and q_i. For example, for $a_{i\,|\,s}$ with a specific s, let $\mathbf{a}'_s = (a_{1\,|\,s}, \ldots, a_{n\,|\,s})$, and $\alpha' = (\alpha_1, \ldots, \alpha_n)$, where α_i's, $i = 1, \ldots, n$, are parameters in the Dirichlet distribution,

$$\Pr(\mathbf{a}_s \mid \alpha) = \frac{\Gamma(\Sigma_i(\delta_{is}(\alpha_i - 1) + 1))}{\prod_i \Gamma(\delta_{is}(\alpha_i - 1) + 1)} \prod_i (a_{i\,|\,s})^{\delta_{is}(\alpha_i - 1)} , \tag{27}$$

where we understand $0^0 = 1$. This result, in combination with the multinomial likelihood for frequency data, $\mathbf{f}' = (f_1, \ldots, f_n)$, will lead to a posterior distribution which is proportional to another Dirichlet distribution,

$$L(\mathbf{f} \mid \mathbf{a}_s, \alpha) \propto \frac{\Gamma(\Sigma_i(\delta_{is}(\alpha_i + f_i - 1) + 1))}{\prod_i \Gamma(\delta_{is}(\alpha_i + f_i - 1) + 1)} \prod_i (a_{i\,|\,s})^{\delta_{is}(\alpha_i + f_i - 1)} . \tag{28}$$

This formulation suggests that the $\alpha_i - 1$ above can be regarded as penalty parameters, one for each $a_{i\,|\,s}$. However, in practice it is usually too much to specify different penalty parameters for different $a_{i\,|\,s}$'s, so that we may assume $\alpha_i - 1 = \rho$ for all i. Then, the prior, (27), reduces to

$$\Pr(\mathbf{a}_s \mid \rho) \propto \prod_i (a_{i\,|\,s})^{\delta_{is}\rho} , \tag{29}$$

$s = 1, \ldots, S$, and the posterior, (28), to

$$L(\mathbf{f} \mid \mathbf{a}_s, \delta_{is}\rho) \propto \frac{\Gamma(\Sigma_i(\delta_{is}(f_i + \rho) + 1))}{\prod_i \Gamma(\delta_{is}(f_i + \rho) + 1)} \prod (a_{i\,|\,s})^{\delta_{is}(f_i + \rho)} . \tag{30}$$

The log of (30) is equal to the portion of the penalized likelihood stated in (18) pertaining to the $a_{i\,|\,s}$. Essentially the same holds for the r_{st} and q_i part.

References

AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *The Second International Symposium on Information Theory*, Eds., B.N. Petrov and F. Csaki, Budapest: Academiai Kiado, 267-281.

BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W. (1969), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge: MIT-Press.

BÖCKENHOLT, U. and LANGEHEINE, R. (1996), "Latent Change in Recurrent Choice Data," *Psychometrika, 61*, 285-301.

BREIGER, R. L. (1990), *Social Mobility and Social Structure*, New York: Cambridge University Press.

CAVANAUGH, J., and SHUMWAY, R. (1994), "A Bootstrap Variant of AIC for State-Space Selection," Submitted to *Statistica Sinica*.

CLOGG, C. C. (1981), "Latent Structure Models of Mobility," *American Journal of Sociology, 86*, 836-868.

COOMBS, C. H., DAWES, R. D., and TVERSKY, A. (1970), *Mathematical Psychology: An Elementary Introduction*, Englewood Cliffs, NJ: Prentice-Hall.

COX, D. D., and O'SULLIVAN, F. (1990), "Asymptotic Analysis of Penalized Likelihood and Related Estimators," *The Annals of Statistics, 18*, 1676-1695.

CRAVEN, P., and WAHBA, G. (1979), "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross Validation," *Numerische Mathematik, 31*, 377-403.

DE LEEUW, J., VAN DER HEIJDEN, P. G. M., and VERBOON, P. (1990), "A Latent Time-Budget Model," *Statistica Neerlandica, 44*, 1-22.

DEMPSTER, A. P., LAIRD, N. M., and RUBIN, D. B. (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society, Series B, 37*, 1-38.

DUNCAN, O. D. (1979), "How Destination Depends on Origin in Occupational Mobility Table," *American Journal of Sociology, 84*, 793-803.

GOODMAN, L. A. (1968), "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries," *Journal of the American Statistical Association, 63*, 1091-1131.

GOODMAN, L. A. (1979), "On the Estimation of Parameters in Latent Structure Analysis," *Psychometrika, 44*, 123-128.

GREEN, P. J. (1990), "On the Use of the EM Algorithm for Penalized Likelihood Estimation," *Journal of the Royal Statistical Society, Series B, 52*, 443-452.

GROVER, R., and SRINIVASAN, V. (1987), "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research, 24*, 139-153.

GU, C., and QIU, C. (1993), "Smoothing Spline Density Estimation: Theory," *Annals of Statistics, 21*, 217-234.

HAGENAARS, J. A. (1990), *Categorical Longitudinal Data*, Newbury Park, CA: Sage.

HARSHMAN, R. A., GREEN, P. E., WIND, Y., and LUNDY, M. E. (1982), "A Model for the Analysis of Asymmetric Data in Marketing Research," *Marketing Science, 1*, 205-242.

HOUT, M. (1983), *Mobility Tables*, Beverly Hills, CA: Sage.

ISHIGURO, M., and SAKAMOTO, Y. (1983), "A Baysian Approach to Binary Response Curve Estimation," *Annals of the Institute of Statistical Mathematics, 35*, 115-137.

JAMSHIDIAN, M., and JENNRICH, R. I. (1993), "Conjugate Gradient Acceleration of the EM Algorithm," *Journal of the American Statistical Association, 88*, 221-228.

JOHNSON, R. A. (1980), *Religious Assortative Marriage in the United States*, New York: Academic Press.

KIERS, H. A. L., and TAKANE, Y. (1993), ''Constrained DEDICOM,'' *Psychometrika, 58*, 339-355.

LANG, J. B. (1992), ''Obtaining the Observed Information Matrix from the Poisson Log Linear Model with Incomplete Data,'' *Biometrika, 79*, 405-407.

LOUIS, T. A. (1982), ''Finding the Observed Information Matrix When Using the EM Algorithm,'' *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

MENG, X. L., and RUBIN, D. B. (1993), ''Maximum Likelihood Estimation via the ECM Algorithm: A General Framework,'' *Biometrika, 80*, 267-278.

MILLER, S. M. (1960), ''Comparative Social Mobility,'' *Current Sociology, 9*, 1-89.

MOOIJAART, A., and VAN DER HEIJDEN, P. G. M. (1992), ''The EM Algorithm for Latent Class Analysis with Equality Constraints,'' *Psychometrika, 57*, 261-269.

SAKAMOTO, Y. (1991), *Categorical Data Analysis by AIC*, Dordrecht: Kluwer.

SATO, M., and SATO, Y. (1995), ''On a General Fuzzy Additive Clustering Model,'' *Intelligent Automation and Soft Computing, 1*, 439-448.

SHIBATA, R. (1989), ''Statistical Aspects of Model Selection,'' in *From Data to Model*, Ed., J.C. Willems, Berlin: Springer, 501-522.

SHIBATA, R. (1995), ''Bootstrap Estimate of Kullback-Leibler Information for Model Selection,'' Submitted to *Statistica Sinica*.

YAMAGUCHI, K. (1983), ''The Structure of Intergenerational Occupational Mobility: Generality and Specificity in Resources, Channels, and Barriers,'' *American Journal of Sociology, 88*, 718-745.