STATE-SPACE MODELS FOR CATEGORICAL VARIABLES

Ab Mooijaart
Department of Psychology
University of Leiden

Kees van Montfort
Department of Econometrics
Free University Amsterdam

Abstract

In state-space models there are so-called state variables which are latent variables through which time is channelled. Furthermore, there are input and output variables. In this paper we discuss state-space models in which all variables are categorical. We assume a first order Markov model for the state variables, although extension to higher order models is simple. Furthermore, we assume time homogeneity of the transition probabilities. These models can be conceived as extensions of latent class models with the grouping variable depending on time. This approach can be important in experimental designs, in which, after some period, different groups of persons get different treatments. The multivariate statistical distribution we assume is the (product-) multinomial distribution. Estimation of the model parameters will be carried out by the so-called EM algorithm. A real data example will be discussed.

Key words: state-space models, categorical variables, input variables, sub-populations, multinomial distribution, EM algorithm.

Correspondence should be send to Kees van Montfort, Free University, Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Phone 020-444.6025. Email kvmontfort@econ.vu.nl.

State-space models for categorical variables

1. Introduction

In the case the state and output variables are categorical and there are no input variables, we are dealing with models as discussed by Poulsen (1982), Langeheine (1988), van de Pol and Langeheine (1989) and Langeheine and van de Pol (1990). These models are called Latent Markov (LM) models. An estimation procedure of the parameters in these models, the EM algorithm, was given by van de Pol and de Leeuw (1986).

A restriction we make in this paper, in comparison to the work of Langeheine and van de Pol, is that we do not discuss mixed Markov models. In these models several latent Markov chains are defined. Although this extension is interesting and may improve the fit of the model considerable, it also may introduce identification problems of the model parameters. Because the main issue of this paper is the role of the input variables, we leave the extension to mixed Markov models to the future.

The LM models are closely related to latent class models. These latter models have their origin in the fifties (see Lazarsfeld (1950)) and a general framework of these models was given by Lazarsfeld and Henry (1968). The main breakthroughs of the latent class model was in the seventies, see the work of Goodman (1974a, 1974b) and Haberman (1979). A very interesting and informative paper on the recent developments and prospects for the future is Clogg (1993). The main issue for our state-space model in comparison to the latent class models are the latent class models for multiple groups. In these models the whole population is divided in several subpopulations, where each subpopulation may have different model parameters. In such a setup it is possible to test wether some parameters are invariant or not over the several groups. This approach is analogously to the simultaneous factor analysis in several groups of Jöreskog (1971) and, for categorical variables, to the simultaneous analysis of several subpopulations (see Clogg and Goodman, 1984, 1985, 1986, and Hagenaars, 1990). The groups might correspond to, e.g., gender or different time periods. However, in latent class models and also in the latent Markov models of van de Pol and Langeheine, the samples do not change over time. In the model we propose here, the input may be dependent of time, i.e. after some period some sub-samples may have different treatments (input). This approach can be important in experimental designs, in which, after some period, different groups of persons get different treatments.

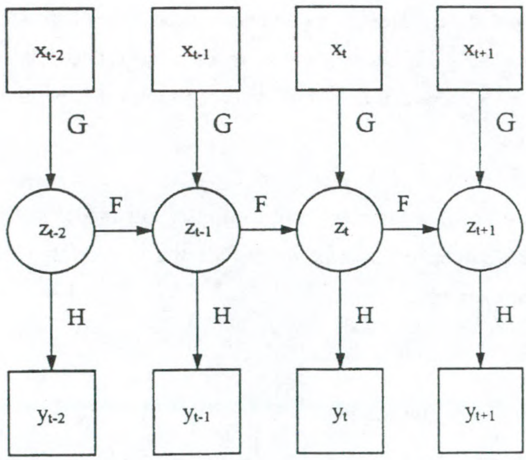
Obviously, our state-space model is a generalization of latent class models for just one time point. For instance, Dayton and Macready (1988a,b) formulate latent class models with concomittant variables. These concommittant variables, or covariates, may be categorical (grouping variables) or continuous. Other papers related to these latent class models with covariates are DeSarbo and Wedel (1994), Formann (1992), and van der Heijden and Dessens (1994). Although, in this paper we discuss categorical variables, only, their models are a special case of the model we propose here, because we have more than one time point. In a future paper we will include ordered categorical and continous variables.

Almost at the end of Clogg (1993) we find the following quote: "Another very fruitful avenue of research only touched on in the above review is formulating LCM's (latent class models) for dynamic processes. The work by Langeheine and van de Pol (1990) on latent Markov models represents this new development very well, and much more can be done". It is our claim that we have done a step forward on this avenue.

In section two of this paper we will formulate our model in detail. An EM estimator of the model parameters will be derived in the next section. Finally we will discuss an empirical example.

2. Formulation of the model

Define y_t as an output variable at time point t, x_t as an input variables at time point t, and z_t as a latent state variable at time point t. The input variable is a fixed variable, whereas the other variables are random variables. We start with one input and one output variable. In figure 1 an example of the state-space model for 4 time points, from t-2 to $t+1$, is given. We see in this figure that the time is channelled only through the state variables z. Furthermore, there are no direct effects from the input variables, x, on the output variables y. In some formulations of the state space model there is a state variable, called z_0, which influences z_1. So z_1 is influenced by the first input variable x_1, and a previous state variable which is completely unknown because it has no input and output variable. Although this state variable z_0 may be important, we assume, just as Langeheine and van de Pol do, that is an empty set. The letters $\boldsymbol{F}$, $\boldsymbol{G}$ and $\boldsymbol{H}$ are matrices which will contain model parameters. For instance, matrix $\boldsymbol{F}$ will contain the transition parameters of the state variables from time point $t-1$ to t. Remark that the matrices $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H}$ do not depend on time. This means that we assume that the model parameters are invariant over time. This assumption is not crucial for our discussion of the model and can be easily relaxed.



In figure 1 there is only one input and output variable at each time point. This can be generalized easily to more input and output variables. Because all variables are categorical and because the input variables are exogeneous fixed variables, it is sufficient to create one new input variable as the Cartesian product of all input variables. For the output variables, such a trick is not possible, because the output variables are non-fixed endogeneous variables. In figure 2 we see a method for dealing with 2 output variables at each time point. At each time point we have two state variables (in general, as much as there are output variables (indicators) at each time point). These state variables are equal to each other for each time point. This can be formulated by restricting the transition matrix between the state variables at each time point to the identity matrix. Furthermore, each output variable is an indicator of one state variable. So in general, the case of multiple indicators is a restricted case of the case with just one output variable for each time point and so it is not necessary to discuss the case with multiple indicators in detail. Note that by this formulation the number of output variables may be different from the actual number of state variables.

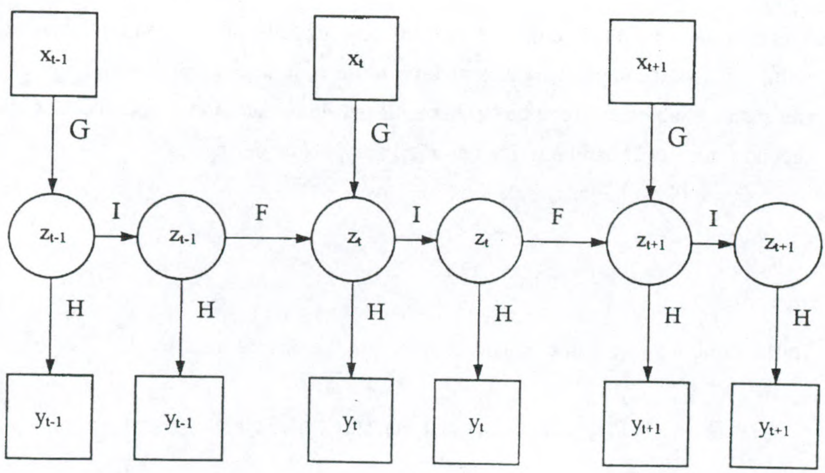


Figure 2

We first derive for three time points the joint probability of the state variables and the output variables conditional the input variables. This joint probability is a crucial part of the complete likelihood function which is needed for the EM algorithm in this paper.

Let J be the set consisting of all variables, i.e. the input, output, and state variables, then the joint probability of these variables can be written as

$$P[J] \equiv P[y_1,y_2,y_3,z_1,z_2,z_3,x_1,x_2,x_3]$$

$$= P[y_3|y_1,y_2,z_1,z_2,z_3,x_1,x_2,x_3]\ P[y_1,y_2,z_1,z_2,z_3,x_1,x_2,x_3]. \qquad (1)$$

The output variable depends on the latent variable only, so the first term in (1) can be written as $P[y_3|z_3]$. Now (1) can be written as

$$P[J] = P[y_3|z_3]\ P[z_3|y_1,y_2,z_1,z_2,x_1,x_2,x_3]\ P[y_1,y_2,z_1,z_2,x_1,x_2,x_3]. \qquad (2)$$

The state variable at time point t depends only directly on the state variables at time point t-1 and the input variables at time point t. It depends indirectly, i.e. through the state variables at time point t-1, on all previous state and input variables. So the second term in (2) can be written as $P[z_3|z_2,x_3]$. Now define

$Q_t \equiv P[y_t|z_t]\ \mathrm{P}[z_t|z_{t-1},x_t]$, then we have

$$P[J] = Q_3\ P[y_1,y_2,z_1,z_2,x_1,x_2,x_3]. \qquad (3)$$

In the same way we can continue to elaborate the second term in (3). This gives

$$P[J] = Q_3\ P[y_2|y_1,z_1,z_2,x_1,x_2,x_3]\ P[y_1,z_1,z_2,x_1,x_2,x_3]$$

$$= Q_3\ P[y_2|z_2]\ P[y_1,z_1,z_2,x_1,x_2,x_3]$$

$$= Q_3\ P[y_2|z_2]\ P[z_2|y_1,z_1,x_1,x_2,x_3]\ P[y_1,z_1,x_1,x_2,x_3]$$

$$= Q_3\ P[y_2|z_2]\ P[z_2|z_1,x_2]\ P[y_1,z_1,x_1,x_2,x_3]$$

$$= Q_3Q_2\ P[y_1,z_1,x_1,x_2,x_3].$$

Repeating this process for y_1 and z_1 gives the result

$$P[J] = Q_3Q_2Q_1P[x_1,x_2,x_3]. \tag{4}$$

From (4) it follows for the joint distribution of the output and state variables, conditional the input variables

$$P[y_1,y_2,y_3,z_1,z_2,z_3|x_1,x_2,x_3] = Q_1Q_2Q_3\ .$$

Obviously, the generalization to T time points becomes

$$P[\boldsymbol{y},\boldsymbol{z}\,|\,\boldsymbol{x}] = \prod_{t=1}^{T} P[y_t\,|z_t]\ P[z_t\,|\,z_{t-1},x_t], \tag{5}$$

where $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{x}$ are vectors with elements y_t, z_t and x_t, for $t= 1,\ldots, T$.

Special case I: no input variables

Because our model, in which there are input variables, is closely related to the model without input variables, we briefly discuss this latter model first and at the same time some notation will be introduced.

With no input variables all x vectors can be eliminated in (5). This gives

$$P[\mathbf{y},\mathbf{z}] = \prod_{t=1}^{T} P[y_t|z_t] P[z_t|z_{t-1}], \tag{6}$$

This defines the dynamic factor analysis, but now for categorical variables, only (see Molenaar (1985)). We start with the simple case for T=3, again. According to (6) we can write

$$P[y_1=i,\ y_2=j,\ y_3=k,\ z_1=\alpha,\ z_2=\beta,\ z_3=\gamma] =$$

$$P[z_1=\alpha]\ \ P[y_1=i|z_1=\alpha]\ \ P[z_2=\beta|z_1=\alpha]$$

$$\text{x}\ P[y_2=j|z_2=\beta]\ \ P[z_3=\gamma|z_2=\beta]\ \ P[y_3=k|z_3=\gamma].$$

Assume the state variables have r categories, and define $\boldsymbol{F}_t$ as the (r x r) transition matrix of the state variables of time t-1 to time point t. Furthermore assume stationarity, then it holds $\boldsymbol{F}_t = \boldsymbol{F}$. So, for instance, we can write

$$P[z_2 = \beta|z_1 = \alpha] = f_{\beta\alpha}.$$

Obviously, it holds

$$\sum_{\beta=1}^{r} f_{\beta\alpha} = 1$$

The conditional probabilities of the output variables given a latent state at time point t, are collected in a (k x r) matrix $\boldsymbol{H}_t$, where k is the number of categories of variable y. Again, we assume a stationary proces, which means that it holds $\boldsymbol{H}_t = \boldsymbol{H}$. So we can write, for instance,

$$P[y_1= i|z_1 = \alpha] = h_{i\alpha}.$$

Furthermore

$$\sum_{i=1}^{k} h_{i\alpha} = 1$$

It follows

$$P[y_1=i,\ y_2=j,\ y_3=k,\ z_1=\alpha,\ z_2=\beta,\ z_3=\gamma] = \mu_\alpha h_{i\alpha} f_{\beta\alpha} h_{j\beta} f_{\gamma\beta} h_{k\gamma},$$

where μ_α is the proportion of subjects in category α of the first state variable. Obviously, it holds

$$\sum_{\alpha=1}^{r} \mu_\alpha = 1$$

Define now the sample size n and the probability $\theta_{\alpha\beta\gamma ijk}$ of a respondent being in cell (α, β, γ, i, j, k). Then under the multinomial distribution $n\theta_{\alpha\beta\gamma ijk}$ are the expected frequencies. So, the logarithm of the complete likelihood function, i.e. for all output and state variables, can be written as

$$\ln L(y,z;U) = \sum_{\alpha\beta\gamma} \sum_{ijk} n\theta_{\alpha\beta\gamma ijk} \ln\left(\mu_\alpha h_{i\alpha} f_{\beta\alpha} h_{j\beta} f_{\gamma\beta} h_{k\gamma}\right),$$

where $\boldsymbol{U}$ is a vector consisting of all unknown parameters, i.e. r parameters μ_α, kr parameters $h_{i\alpha}$, and r^2 parameters $f_{\beta\alpha}$. Note that the θ's play the role of the observations, although they are unknown because the z variables are not observed. This principle of unobserved observations play a crucial role in the EM algorithm, to be discussed later on.

For several subpopulations the logarithm of the complete likelihood function can be written as

$$\ln L(y,z;U)=\sum_{s}\ln L(y_s,z_s;U_s)= \tag{7}$$
$$=\sum_{s}\sum_{\alpha\beta\gamma}\sum_{ijk} n_s\theta_{\alpha\beta\gamma ijk;s}\ln(\mu_{\alpha;s}\, h_{i\alpha;s}\, f_{\beta\alpha;s}\, h_{j\beta;s}\, f_{\gamma\beta;s}\, h_{k\gamma;s}),$$

where $s=1,\ldots,S$ is the index for subpopulation s. Note that all parameters are now defined for each subpopulation.

Special case II: input variables

We start with the simple case for $T=3$, again. According to (6) we can write

$$P[y_1=i,\ y_2=j,\ y_3=k,\ z_1=\alpha,\ z_2=\beta,\ z_3=\gamma \mid x_1=a,\ x_2=b,\ x_3=c]$$

$$= P[z_1=\alpha \mid x_1=a]\ P[y_1=i \mid z_1=\alpha]\ P[z_2=\beta \mid z_1=\alpha,\ x_2=b]$$

$$\text{x}\ P[y_2=j \mid z_2=\beta]\ P[z_3=\gamma \mid z_2=\beta,\ x_3=c]\ P[y_3=k \mid z_3=\gamma].$$

Obviously, it follows

$$P[y_1=i,y_2=j,y_3=k,z_1=\alpha,z_2=\beta,z_3=\gamma]= \tag{8}$$
$$\sum_{abc} p_{abc}\ P[y_1=i,y_2=j,y_3=k,z_1=\alpha,z_2=\beta,z_3=\gamma \mid x_1=a,x_2=b,x_3=c],$$

where p_{abc} is the proportion of sample elements with input scores a, b, c on time points 1, 2 and 3, respectively. Formula (8) can be written as

$$P[y_1=i,y_2=j,y_3=k,z_1=\alpha,z_2=\beta,z_3=\gamma]=$$
$$\sum_{abc} p_{abc}\,\mu_{\alpha;a}\, h_{i\alpha}\, f_{\beta\alpha;b}\, h_{j\beta}\, f_{\gamma\beta;c}\, h_{k\gamma}.$$

So the logarithm of the complete likelihood function can be written as

$$\ln L(x,y,z;U) = \sum_{abc} p_{abc} \sum_{\alpha\beta\gamma} \sum_{ijk} n\theta_{\alpha\beta\gamma ijk;abc} \ln(\mu_{\alpha;a}\, h_{i\alpha}\, f_{\beta\alpha;b}\, h_{j\beta}\, f_{\gamma\beta;c}\, h_{k\gamma}), \tag{9}$$

where $\theta_{\alpha\beta\gamma ijk;abc}$ is the probability of a respondent with input scores a, b and c on time 1, 2 and 3, respectively, being in cell $(\alpha, \beta, \gamma, i, j, k)$.

Let us compare this likelihood function with the likelihood function in the case of no input variables and several subpopulations, see equation (7). Suppose all the input scores are equal for all time points, and suppose this score is "s", then it holds $a=b=c=s$, and $p_{abc}n = p_s n = n_s$. Furthermore, in the case of no input variables, suppose that the factor loadings in matrix $\boldsymbol{H}$ are invariant the subpopulations, i.e. $h_{i\alpha;s} = h_{i\alpha}$, then the likelihood functions in Case I and II are equal. This showes that the case with no input variables and with subpopulations is a special case of the general case with input variables if the factor loadings are invariant over the subpopulations.

3. EM algorithm in the state space model

In this section we discuss how to estimate the model parameters in the general model. We will assune that for all modes stationarity holds, i.e. we assume $\boldsymbol{F}_t = \boldsymbol{F}$, $\boldsymbol{G}_t = \boldsymbol{G}$, and $\boldsymbol{H}_t = \boldsymbol{H}$. In the EM algorithm two steps are defined: the E (Expectation) - step and the M (Maximization) - step.

E-step:

In the E-step the expectation of the sufficient statistics of the complete multinomial distribution, conditional the observed frequencies and the model parameters is formulated. Define

$$\varsigma_{\alpha\beta\gamma ijk;abc} \equiv \mu_{\alpha;a} h_{i\alpha} f_{\beta\alpha;b}\, h_{j\beta} f_{\gamma\beta;c}\, h_{k\gamma},$$

then the conditional expectation of the sufficient statistics can be written as

$$\theta_{\alpha\beta\gamma ijk;abc} = (p_{ijk;abc} / \varsigma_{+++ijk;abc})\, \varsigma_{\alpha\beta\gamma ijk;abc},$$

where $p_{ijk;abc}$ are observed proportions denoting the proportions of subjects in category i, j, and k of the three output variables with, at the same time input scores a, b, and c on the three input variables. Furthermore, the "+"'s in $\varsigma_{+++ijk;abc}$ denotes the summation over the indices α, β, and γ. See for analogous formulations in the latent Markov model van de Pol and de Leeuw (1986).

M-step:

In the M-step the logarithm of the complete likekelihood function (9) is maximized as a function of the unknown parameters. For instance, for estimating $\mu_{\alpha;a}$, we define

$$L^* = \ln L(x,y,z;U) - m(\sum_{\alpha} \mu_{\alpha;a} - 1),$$

where m is a Lagrange multiplier and the side condition is

$$\sum_{\alpha} \mu_{\alpha;a} = 1$$

The derivative of L^* with respect to $\mu_{\alpha;a}$ is

$$\frac{dL^*}{d\mu_{\alpha;a}} = \sum_{bc} p_{abc} \sum_{\beta\gamma} \sum_{ijk} \frac{n\theta_{\alpha\beta\gamma ijk;abc}}{\mu_{\alpha;a}} - m. \tag{10}$$

Equating this derivative equal to zero and summing over α, gives

$$\sum_{bc} p_{abc} \sum_{\alpha\beta\gamma} \sum_{ijk} n\theta_{\alpha\beta\gamma ijk;abc} = m \sum_{\alpha} \mu_{\alpha;a} = m.$$

Now substituting m into (10) gives

$$\mu_{\alpha;a} = \frac{\sum_{bc} p_{abc} \sum_{\beta\gamma} \sum_{ijk} n\theta_{\alpha\beta\gamma ijk;abc}}{\sum_{bc} p_{abc} \sum_{\alpha\beta\gamma} \sum_{ijk} n\theta_{\alpha\beta\gamma ijk;abc}}.$$

In shorter notation this can be written as

$$\mu_{\alpha;a} = \frac{\sum_{bc} p_{abc}\theta_{\alpha+++++;abc}}{\sum_{bc} p_{abc}\theta_{++++++;abc}}.$$

In an analogous way we can estimate $h_{i\alpha}$ and $f_{\beta\alpha;b}$, assuming stationarity. This gives

$$h_{i\alpha} = \frac{\sum_{abc} p_{abc}\left[\theta_{\alpha++i++;abc} + \theta_{+\alpha++i+;abc} + \theta_{++\alpha++i;abc}\right]}{\sum_{abc} p_{abc}\left[\theta_{\alpha+++++;abc} + \theta_{+\alpha++++;abc} + \theta_{++\alpha+++;abc}\right]},$$

$$f_{\beta\alpha;b} = \frac{\sum_{ac} p_{abc}\theta_{\alpha\beta++++;abc} + \sum_{ab} p_{abc}\theta_{+\alpha\beta+++;abc}}{\sum_{ac} p_{abc}\theta_{\alpha+++++;abc} + \sum_{ab} p_{abc}\theta_{+\alpha++++;abc}}.$$

So the EM algorithm runs as follows: first define some start values of the unknown parameters; then compute $\theta_{\alpha\beta\gamma ijk;abc}$ by the E-step; find new estimates for $\mu_{\alpha;a}$, $h_{i\alpha}$, and $f_{\beta\alpha;b}$ as given in the M-step. Repeat the whole procedure as many times as necessary for reaching convergence.

4. An empirical example

This section will apply the introduced method on panel-data intended to market research. These data are put available by the research institute AGB (located in Dongen, The Netherlands) and contain about 300 panel-members (=households). During a year the purchasing-behaviour with respect to several margarine brands is every four months registered for each panel-member. Of each panel-member is known for each timeperiod, among others, the district (two categories) and furthermore whether several margarine brands have been bought (two categories per brand: no purchase or purchase). We have restricted our analysis to the two brands with the largest market-shares (together about 45%).

The authors research questions are:
- To which extent do the two individual brands have the same buyers?
- What is the size of the brand loyalty for each brand?
- Is the factor District of importance in the first two questions?

In our analysis we consider a state space model with three time points of measuring (T=3) and at each time point: two output variables (=indicators; the purchase of brand A and the purchase of brand B), one input variable (= District) and one state variable. All the variables are categorical with two categories. The transition matrix $\boldsymbol{F}$ of the state variables and the matrix $\boldsymbol{H}$, with the conditional probabilities of the output variables given a latent state, have dimension (2 x 2).

To answer the first research question we will use the relation between the output variables and the state variables, i.e. the matrix $\boldsymbol{H}$. Question two can be answered by means of the dynamic part of the model (= the transition matrix $\boldsymbol{F}$). For the last research question the explanatory categorical input variable District can be used.

We create an observed frequency table of size $2^3 * 2^3 * 2^3$, which will be used to fit our state space model. A PC-program using Pascal 7.0 is written to obtain EM-estimates. To be sure that the model is identified we have to investigate whether the corresponding information matrix is of full rank (i.e. none of the eigenvalues of the information matrix is zero). It turned out that the identification of the considered model was guaranteed for our data.

In order to test the model against the data one may look at the loglikelihood ratio statistic (twice the difference between the loglikelihood of the data and the model). This loglikelihood ratio statistic (G^2) is asymptotically chi-quare distributed with degrees of freedom equal to the number of different response patterns of the observed input and output variables, minus the number of independent model parameters. However, the number of cells in the observed frequency table is 512 with most of the cells equal to zero, while the number of independent parameters is 10. Using such a large number of degrees of freedom every model will fit certainly. Furthermore the sparseness of our frequency table makes this test very difficult (Haberman, 1977).

Nevertheless, it is also possible to compare the fit of two nested models by calculating two times the difference in the loglikelihood ratio statistics of the two nested models and using the difference in the number of independent parameters as the number of degrees of freedom. Therefore in this empirical example we also fit, apart from the already mentioned state space model with input variable District (model M_4), some more restrictive models.

- a latent class or latent Markov model without latent change across time and without an input variable (M_1);
- a latent class or latent Markov model without latent change across time and with the input variable District (M_2);
- a state space model without input (M_3; a latent Markov model);

Table 1 gives for model M_2, M_3 and M_4 the G^2-difference with a more restrictive model (respectively model M_1, M_1 and M_2).

Of course, the most restrictive model M_1 has the largest value for G^2. When we introduce the input variable District (model M_2) the fit will increase significantly. Furthermore model M_3 and M_4 are the dynamic versions of model M_1 and M_2. The former models fit significantly better than the latter. So, the dynamic aspect of our model may not be dropped.

The EM-algorithm we use can be programmed easily, iterations are computationally attractive, and convergence is ensured. Otherwise, it may converge to a local maximum of the likelihood function. We should use different starting values for the unknown parameters to increase confidence that the maximum found is indeed the global maximum. In our empirical analysis we indeed used several different starting values, which all converged to the same end values.

Another general handicap of the EM-algorithm is, that it often requires many iterations. In our empirical example we were not confronted with this problem. We needed, apart from the different startings values for the parameters, at most about a hundred iterations.

Table 2 presents the estimates of the unknown parameters $\boldsymbol{H}$, $\boldsymbol{F}$ and μ of our model M_4. The factor loadings (matrix $\boldsymbol{H}$ for brand A and B) indicate that latent class 0 corresponds with no purchase of brand B, while brand A will be bought with probability 0.72. For latent class 1 there will be no purchase of brand A and a possible purchase of brand B (probability 0.53). We see that latent class 0 is the purchase of brand A and latent class 1 is the purchase of brand B. Additionally the two brands have totally different buyers.

The transition matrix $\boldsymbol{F}$ indicates the change (or stability) between successive time points. For district 2, the transition matrix between two sequential time points is equal to the unity matrix. This indicates that the degree of brand loyalty for each of the two brands is high. In district 1 there is more change and the two brands have a smaller degree of loyalty. The values of the transition coefficients (in matrix $\boldsymbol{F}$) show that between two successive time periods the state can go from 0 to 1 with probability 0.09. A state-change in the opposite direction has a probability of 0.03. Therefore the number of buyers of product A will decrease, while the number of buyers of product B will increase slightly with time.

Furthermore, the matrix μ indicates that district 1 contains more buyers of product B and less buyers of product A than district 2.

Once the EM estimates of the model parameters have been computed their variances may be found from the information matrix. This is the inverse of the matrix of second order derivatives of the loglikelihood function toward all independent parameters. Fortunately, the second order derivatives of the multivariate loglikelihood function (see formula (9)) result in rather simple expressions. In our example the variances of the parameter estimates are very small. All the empirical variances are smaller than 0.01.

From our investigations it follows that the state space model for categorical data can answer the research questions of our empirical example. Furthermore the EM-algorithm is computationally easy and works rather quickly.

In our model we assume that the transition matrix and the factor loading matrices do not depend on time. Maybe we can get a significantly better model fit by relaxing this assumption (see Fahrmeir and Kaufmann (1987) which deal with regression models for non-stationary categorical time series). For instance, in the field of marketing research temporary advertising campagnes could be modelled better. However, if we allow for both $\boldsymbol{F}$ and $\boldsymbol{H}$ to change it may be very difficult to interprete results.

Acknowledgement

The authors are greatful to Dr. R. Langeheine (University of Kiel, Germany) for his remarks that led to improvement.

References

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

Clogg, C.C., and L.A. Goodman (1985). Simultaneous latent structure analysis in several groups. In: *Sociological Methodology*, N.B. Tuma (Ed), 81-110, San Francisco: Jossey-Bass.

Clogg, C.C., and L.A. Goodman (1986). On scaling models applied to data from several groups. *Psychometrika*, **51**, 123-135.

Clogg, C.C. (1993). Latent class models: recent developments and prospects for the future. In: G. Arminger, C.C. Clogg, and M.E. Sobel (Eds.), *Handbook of Statistical Modeling in the Social Sciences*. New York: Plenum.

Dayton, C.M. and Macready, G.B. (1988a). Concomittant-variable latent class models. *Journal of the American Statistical Association*, **83**, 173-178.

Dayton, C.M. and Macready, G.B. (1988b). A latent class covariate model with applications to criterion referenced testing. In: R. Langeheine and J. Rost (Eds.). *Latent trait and latent class models*. New York: Plenum.

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

DeSarbo, W.S. and Wedel, M. (1994). A review of recent developments in latent class regression models. In: R. Bagozzi (Ed.) *Handbook of Marketing Reserach*.

Fahrmeir, L. and Kaufmann, H. (1987). Regression models for non-stationary categorical time series. *Journal of Time Series Analysis*, **8**, 2, 147-160.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, **87**, 476-486.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - A modified latent structure approach. *American Journal of Sociology*, **79**, 1179-1259.

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

Haberman, S.J. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics*, **5**, 1148-1169.

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

Hagenaars, J.A. (1990). *Categorical Longitudinal Data. Log-linear, Panel, Trend and Cohort Analysis*. London: Sage.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several groups. *Psychometrika*, **36**, 409-426.

Langeheine, R. (1988). Manifest and latent Markov chain models for categorical panel data. *Journal of Educational Statistics*, **13**, 299-312.

Langeheine, R. and F. van de Pol (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research*, **18**, 416-441.

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. In E.A. Suchman, P.F. Lazarsfeld, S.A. Starr, and J.A. Clausen (Eds.), *Studies in Social Psychology in World War II, Vol 4, Measurement and Prediction,* 362 - 412. Princeton, N.J.: Princeton University Press.

Lazarsfeld, P.F. and Henry, N.W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.

Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50**, 181-202.

Poulsen, C.S. (1982). *Latent Structure Analysis with Choice Modeling Applications*, Ph.D. diss., Wharton School, University of Pennsylvania.

Van de Pol, F., and J. de Leeuw, (1986). A Latent Markov Model to correct for measurement error. *Sociological Methods and Research*, **15**, 118-141.

Van de Pol, F. and R. Langeheine, (1989). Mixed Markov Models, Mover-Stayer models, and the EM Algorithm, with an application to labour market data from the Netherlands Sociometric panel. *Multiway Data Analysis*, R. Coppi and S. Bolasco (Eds), Amsterdam: North-Holland.

Van der Heijden, P.G.M. and Dessens, J. (1994). *Incorporating Continuous Explanatory Variables in Latent Class Models*. Utrecht University, Faculty of Social Sciences / ISOR, Methods Series MS-94-5.

Table 1, The fit of the models

model	G^2	compared to	G^2-difference	df
M_1, no dyn., no input	126.8	-	-	-
M_2, no dyn., D as input	123.1	M_1	3.7	1
M_3, dyn., no input	119.8	M_1	7.0	2
M_4, dyn., D as input	113.0	M_2	10.1	4

Table 2, The parameter estimates $\boldsymbol{H}$, $\boldsymbol{F}$ and μ

Separate factor loading matrices $\boldsymbol{H}$ (parameters $h_{i\alpha}$) for brand A and B:

	no purchase brand A	purchase brand A
state category 0	0.28	0.72
state category 1	0.97	0.03

	no purchase brand B	purchase brand B
state category 0	1.00	0
state category 1	0.47	0.53

Separate transition matrices $\boldsymbol{F}$ (parameters $f_{\beta\alpha}$) for district 1 and district 2:

	district 1	
	state category 0	state category 1
state category 0	0.91	0.09
state category 1	0.03	0.97

	district 2	
	state category 0	state category 1
state category 0	1.00	0
state category 1	0	1.00

Conditional probability $\mu_{\alpha,a}$ in timepoint $t=1$
for state category α given district a:

	state category 0	state category 1
district 1	0.39	0.61
district 2	0.57	0.43

---

ontvangen: 25-9-95
geaccepteerd: 16-9-96