Latent class analysis with panel data: developments and applications

Reinecke, Jost

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Reinecke, J. (1999). Latent class analysis with panel data: developments and applications. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 44, 137-157. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-199765

Latent Class Analysis with Panel Data: Developments and Applications[1]

by Jost Reinecke[2]

Zusammenfassung:

In der vorliegenden Arbeit wird das statistische Modell der Analyse latenter Klassen nach der Parametrisierung von Lazardsfeld vorgestellt. Den Schwerpunkt bilden Entwicklungen und Anwendungen der Analyse latenter Klassen auf Paneldaten. Das latente Markov Modell erlaubt sowohl Restriktionen über zeitbezogene Gleichsetzungen von konditionalen Wahrscheinlichkeiten als auch Restriktionen der Übergangswahrscheinlichkeiten zwischen den latenten Variablen. Die allgemeinste Variante ist das latente mixed Markov Modell. Dieses Modell verfügt über zusätzliche Spezifikationsmöglichkeiten der unbeobachteten Heterogenität mit Markov Ketten. Empirische Beispiele, durchgeführt mit PANMARK, verdeutlichen die jeweiligen Modellierungstechniken.

Abstract:

The present paper discusses the statistical model of the latent class analysis according to the parametrization of Lazarsfeld. Developments and applications of latent class analysis with panel data are the main topic of this paper. The latent Markov model allows time-specific restrictions of the conditional probabilities as well as restrictions of the transition probabilities between the latent variables. The most general model, the latent mixed Markov model, has additional opportunities to specify unobserved heterogeneity via different Markov chains. Empirical examples, calculated with PANMARK elucidate the relevant modeling techniques.

---

1 The article is a revised version of a guest lecture given during the 28th Spring Seminar at the Zentralarchiv (February 22 – March 12, 1999).

2 PD Dr. ***Jost Reinecke***, Westfälische Wilhelms-Universität Münster, Institut für Soziologie, Scharnhorststr. 121, 48151 Münster, E-Mail: reineck@uni-muenster.de

1 Definition of Latent Class Analysis and Relation to other Models

Many theoretical propositions or concepts in the social sciences cannot be observed directly. For example, external control as a relevant concept of personality can be measured by items such as *I have little influence over the things that happen to me* and *The world is to complicated for me to understand*. If we expect a reasonable covariation between those two items, we can argue that these observed variables are measurements of an underlying latent variable (in this example external control). We can also argue that the latent variable explains the covariation between both items which results in correspondence hypotheses or a theory of measurement. Regarding the example, a corresponding hypothesis can be formulated as follows: The higher the level of external control, the higher the score of the item indicating little influence over things that happen to me.

Lazarsfeld and ***Henry*** (1968) used the term *latent structure analysis* to describe the application of statistical models for characterizing latent variables in the analysis of observed variables. They included factor analysis as a latent structure method to characterize continuous latent variables based on continuous observed variables. Latent class analysis can be defined as a qualitative data analog to factor analysis which enables researchers to empirically identify discrete latent variables from two or more discrete observed variables (cf. Table 1). The categories of the latent variable are called *latent classes*.

Table 1: Models based on Lazarsfeld *Latent Structure Analysis*

Model	Scaling of Variables	
	Latent Variable	Observed Variable
Factor Analysis	continuous	continuous
Latent Trait Analysis	continuous	discrete
Latent Profile Analysis	discrete	continuous
Latent Class Analysis	discrete	discrete

If the latent variable is continuous and the observed variable discrete, the latent structure analysis will be called latent trait analysis. Latent profile analysis characterizes the use of discrete latent variables and continuous observed variables.

Until the mid seventies, only a few applications of latent class analysis could be found in the social sciences. ***Goodman's*** and ***Haberman's*** publications showed a capable way in obtaining maximum liklihood estimates of the parameters of the latent class model and made applications more feasible (***Goodman*** 1974a, 1974b; ***Haberman*** 1978, 1979).

The next section will give a description of the formal latent class model with one latent variable including estimation of parameters, goodness-of-fit statistics and an example.

2 The latent Class Model with One latent Variable

2.1 Model and Assumptions

The latent class model may be represented in terms of ***Lazarsfeld's*** original parametrization or in terms of a log-linear model (cf. ***Haberman*** 1979). The notation of ***Goodman*** (1974a, 1974b) will be used here.

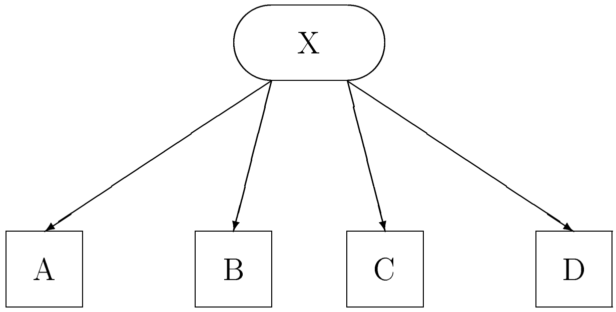
Assuming the case of four manifest variables A, B, C and D and one latent variable X, the latent class model is (cf. Figure 1):

$$m_{ijkl} = N \sum_{g=1}^{G} \pi_g^X \pi_{i,g}^{\overline{A}X} \pi_{j,g}^{\overline{B}X} \pi_{k,g}^{\overline{C}X} \pi_{l,g}^{\overline{D}X} \tag{1}$$

with

- m_{ijkl} as the *expected frequency* of the multivariate contingency table with size $ijkl$,
- π_g^X as the *latent class probability* for latent variable X and class g,
- $\pi_{i,g}^{\overline{A}X}$ as the *probability* for the observed variable A under the *condition* of X and class g,
- $\pi_{j,g}^{\overline{B}X}$ as the *probability* for the observed variable B under the *condition* of X and class g,
- $\pi_{k,g}^{\overline{C}X}$ as the *probability* for the observed variable C under the *condition* of X and class g,
- $\pi_{l,g}^{\overline{D}X}$ as the *probability* for the observed variable D under the *condition* of X and class g.

Figure 1: Path Diagram of a Latent Variable with four Observed Variables



The latent class probabilities (π_g^X) describe the distribution of classes (levels) of the latent variable within which the observed variables are independent of one another. The two important aspects of the latent class probabilities are the *number of classes* and the *relative sizes* of these classes. The number of classes G in the latent variable X represents the number of latent types defined by the latent class model for the observed contingency table. The minimum number of identifiable classes in a latent variable is two, since a latent variable with only a single latent class ($G = 1$) is equivalent with the independence among the observed variables. The size of each of the G classes also provides information for the interpretation of the latent class probabilities. The latent class probabilities over all G latent classes of the latent variable (X) sum to 1:

$$\sum_g \pi_g^X = 1 \tag{2}$$

For each of the G classes of the latent variable there is a set of conditional probabilities for each of the observed variables. In Equation 1 four observed variables have been used to define the latent classes, each of the classes will have four sets of conditional probabilities ($\pi_{i,g}^{\bar{A}X}$, $\pi_{j,g}^{\bar{B}X}$, $\pi_{k,g}^{\bar{C}X}$, $\pi_{l,g}^{\bar{C}X}$). Since each of the observed variables can be either dichotomous or polytomous, the number of distinct conditional probabilities for each of the observed variables is equal to the number of levels measured for that variable. With four observed variables, there are $i + j + k + l$ distinct conditional probabilities for each of the G latent classes of the latent variable X. Within each of the latent classes the conditional probabilities for each of the observed variables sum to 1:

$$\sum_i \pi_{i,g}^{\bar{A}X} = \sum_j \pi_{j,g}^{\bar{B}X} = \sum_k \pi_{k,g}^{\bar{C}X} = \sum_l \pi_{l,g}^{\bar{D}X} = 1.00 \tag{3}$$

Latent class analysis assumes that G classes exist in the population related to the given sample which are disjunctive and exhaustive. If the four observed variables A, B, C and D are items measuring the latent variable external control, the low scorers would represent a latent class called *low external control.* The high scorers would represent a latent class called *high external control*. In each latent class exists homogeneity of the conditional response probabilities. Between latent classes exist heterogeneity of the conditional response probabilities. Associations between the observed variables disappear within each latent class and are explained by the existence of the latent variable (cf. Figure 1). This criterion is called "local independence" (***McCutcheon*** 1987: 14). The relationships between latent and observed variables are probabilistic.

2.2 Estimation of Parameters

The procedure to fit the model to the data was first outlined by ***Goodman*** (1974a, 1974b). He obtained latent class and conditional probabilities with maximum likelihood estimates. The primary goal of the iterative procedure is to minimize the difference between observed and expected frequencies.

Equation (1) can be modified in that way that the parameters of the model are Maximum-Likelihood(ML) estimators:

$$\hat{\pi}_{ijklg}^{ABCDX} = \hat{\pi}_{ig}^{\overline{A}X}\hat{\pi}_{jg}^{\overline{B}X}\hat{\pi}_{kg}^{\overline{C}X}\hat{\pi}_{lg}^{\overline{D}X}\hat{\pi}_{g}^{X} \tag{4}$$

If Equation 4 is summed over all G classes of the latent variable, we obtain the ML conditional probability associated with each of the $ijkl$-levels of the observed variables:

$$\hat{\pi}_{ijkl} = \sum_{g=1}^{G} \hat{\pi}_{ijklg}^{ABCDX} \tag{5}$$

If Equation 4 is divided by Equation 5, we obtain the ML probability that an observation at level $ijkl$ of the observed variables will be at level g of the latent variable:

$$\hat{\pi}_{ijklg}^{ABCD\overline{X}} = \frac{\hat{\pi}_{ijklg}^{ABCDX}}{\hat{\pi}_{ijkl}} \tag{6}$$

The EM-Algorithm produces ML estimates of the unknown parameters (cf. ***Dempster et al.*** 1977). In the so-called E-Step of the algorithm starting values $\overline{\pi}_{g}^{X}$, $\overline{\pi}_{ig}^{\overline{A}X}$, $\overline{\pi}_{jg}^{\overline{B}X}$, $\overline{\pi}_{kg}^{\overline{C}X}$ and $\overline{\pi}_{lg}^{\overline{D}X}$ have to be provided to estimate conditional and latent class probabilities of the model in Equation 4:

$$\overline{\pi}_{ijklg}^{ABCDX} = \overline{\pi}_{g}^{X}\overline{\pi}_{ig}^{\overline{A}X}\overline{\pi}_{jg}^{\overline{B}X}\overline{\pi}_{kg}^{\overline{C}X}\overline{\pi}_{lg}^{\overline{D}X} \tag{7}$$

$\overline{\pi}_{ijklg}^{ABCDX}$ is used to obtain initial values $\overline{\pi}_{ijkl}$ and $\overline{\pi}_{ijklg}^{ABCDX}$ for $\hat{\pi}_{ijkl}$ and $\hat{\pi}_{ijklg}^{ABCDX}$:

$$\overline{\pi}_{ijkl} = \sum_{g=1}^{G} \overline{\pi}_{ijklg}^{ABCDX} \tag{8}$$

$$\overline{\pi}_{ijklg}^{ABCD\overline{X}} = \frac{\overline{\pi}_{ijklg}^{ABCDX}}{\overline{\pi}_{ijkl}} \tag{9}$$

Using the observed frequencies p_{ijkl} a new value $\overline{\pi}_g^X$ for $\hat{\pi}_g^X$ is obtained:

$$\overline{\pi}_g^X = \sum_{ijkl} p_{ijkl} \overline{\pi}_{ijklg}^{ABCD\overline{X}} \tag{10}$$

In the so-called M-step, the new value of $\overline{\pi}_g^X$ is used to calculate new conditional probabilities:

$$\overline{\pi}_{ig}^{\overline{A}X} = \frac{\sum_{jkl} p_{ijkl} \overline{\pi}_{ijklg}^{ABCDX}}{\overline{\pi}_g^X} \tag{11}$$

$$\overline{\pi}_{jg}^{\overline{B}X} = \frac{\sum_{ikl} p_{ijkl} \overline{\pi}_{ijklg}^{ABCDX}}{\overline{\pi}_g^X} \tag{12}$$

$$\overline{\pi}_{kg}^{\overline{C}X} = \frac{\sum_{ijl} p_{ijkl} \overline{\pi}_{ijklg}^{ABCDX}}{\overline{\pi}_g^X} \tag{13}$$

$$\overline{\pi}_{lg}^{\overline{D}X} = \frac{\sum_{ijk} p_{ijkl} \overline{\pi}_{ijklg}^{ABCDX}}{\overline{\pi}_g^X} \tag{14}$$

The new values of conditional probabilities are for Equation 7 to get a new value of $\overline{\pi}_{ijklg}^{ABCDX}$.

The expected frequencies are calculated from the summed products of the conditional and latent class probabilities:

$$\hat{m}_{ijkl}^{ABCD} = N \sum_{g=1}^{G} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} \hat{\pi}_i^X \hat{\pi}_{ig}^{\overline{A}X} \hat{\pi}_{jg}^{\overline{B}X} \hat{\pi}_{kg}^{\overline{C}X} \hat{\pi}_{lg}^{\overline{D}X} \tag{15}$$

2.3 Goodness-of-Fit

The question whether the model fits the data can be answered via the comparison between observed and expected frequencies of the multivariate contingency table with χ^2-statistics:

$$Pearson - \chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{(n_{ijkl}^{ABCD} - \hat{m}_{ijkl}^{ABCD})^2}{\hat{m}_{ijkl}^{ABCD}} \tag{16}$$

$$Likelihood - Ratio - \chi^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} n_{ijkl}^{ABCD} ln \frac{n_{ijkl}^{ABCD}}{\hat{m}_{ijkl}^{ABCD}} \tag{17}$$

with

- n_{ijkl} as the *observed frequency* of the multivariate contingency table with size $ijkl$
- m_{ijkl} as the *expected frequency* of the multivariate contingency table with size $ijkl$

The number of degrees of freedom is calculated via the difference between the number of response patterns ($ijkl$) minus 1 minus the number of parameters (N_{par}) to be estimated:

$$df = (ijkl - 1) - N_{par} \tag{18}$$

Small differences between observed and expected frequencies result in small χ^2-values which can be compared to the theoretical χ^2-distribution given the model's degrees of freedom. Of these two, the Likelihood-Ratio-χ^2 (abbreviated L^2) is generally preferred, since it permits statistical comparison of nested models. Sparse tables, meaning a large number of empty cells in the multivariate contingency table lead to interpretation problems with both χ^2-tests. Two indices, the Akaike Information Criterion (AIC) and ***Schwarz'*** Bayes Information Criterion (BIC) are alternatives for goodness-of-fit (***Bozdogan*** 1987; ***Schwarz*** 1978):

$$AIC = -2ln(L) + 2N_{par} \tag{19}$$

$$BIC = -2ln(L) + ln(N_{pers})N_{par} \tag{20}$$

with L as the Likelihood of the data (i. e. the product of the pattern probability over all persons), N_{par} as the number of parameters to be estimated and N_{pers} as the number of persons in the data. N_{par} and N_{pers} are so-called penalty terms. While the AIC selects more complex models, the BIC "corrects" the AIC in so far as it weights the number of parameters with $ln(N_{pers})$. The smaller the values of AIC and BIC, the better fits the model with the data.

2.4 Example

Langeheine and ***Rost*** (1996) discuss an ordinary latent class example about four dichotomous items from an "Arithmetic Reasoning Test" with data from 776 persons. They test the hypothesis that a model with two latent class (so-called masters and nonmasters) is sufficient to describe the data.

Table 2 gives an overview about differences between observed and expected frequencies for every cell in the multivariate contingency table. 16 different response patterns exist. Large differences between the observed frequencies and the sum of the expected frequencies over the classes would result in large χ^2-values and a model rejection. But with $L^2 = 4.94$ (see Equation 17) and $df = 16 - 1 - 9 = 6$ (see Equation 18) the model fit is acceptable and confirms the latent variable with two classes. The number of parameters (N_{par}) to be estimated contains one latent class probability and 2×4 conditional probabilities.

Table 2: Observed and Expected Frequencies of the 4 dichotomous Items under a Model with 2 Latent Classes

				Frequencies			
Items				observed	expected		
A	B	C	D		Class 1	Class 2	total
0	0	0	0	99	98.46	.03	98.49
0	0	0	1	26	28.11	.08	28.19
0	0	1	0	48	41.36	.12	41.48
0	0	1	1	10	11.81	.38	12.19
0	1	0	0	66	66.11	.17	66.28
0	1	0	1	18	18.87	.54	19.41
0	1	1	0	27	27.77	.76	28.53
0	1	1	1	11	7.93	2.50	10.43
1	0	0	0	74	78.38	1.27	79.65
1	0	0	1	34	22.37	4.17	26.54
1	0	1	0	35	32.92	5.88	38.80
1	0	1	1	28	9.40	19.27	28.67
1	1	0	0	65	52.63	8.36	60.99
1	1	0	1	40	15.02	27.43	42.45
1	1	1	0	61	22.11	38.67	60.78
1	1	1	1	134	6.31	126.81	133.12

According to Equation 1 the expected frequency of a certain response pattern (1111) is calculated with the estimated parameters (see Table 3) as follows:

$$m_{ijkl} = N \sum_{g=1}^{G} \pi_g^X \pi_{i,g}^{\overline{A}X} \pi_{j,g}^{\overline{B}X} \pi_{k,g}^{\overline{C}X} \pi_{l,g}^{\overline{D}X} \tag{21}$$

$$\begin{aligned} 133.12 = \quad & 776 * (.695 * .443 * .402 * .296 * .222 \\ + \quad & .305 * .981 * .868 * .822 * .766) \\ = \quad & 776 * (.00813 + .16353) \end{aligned}$$

Table 3: Estimated Latent Class and Conditional Probabilities of the Latent Class Model with 2 Latent Classes

			Conditional Probabilities			
Class	Class Prob.	Category	A	B	C	D
1	.695	0	.557	.598	.704	.778
		1	.443	.402	.296	.222
2	.305	0	.019	.132	.178	.234
		1	.981	.868	.822	.766

Estimated parameters of the latent class model are given in Table 3. About 70% of the sample are classified as nonmasters (Class 1), the other 30% are classified as masters (Class 2). The conditional probabilities for category 1 in class 2 are higher on the average than the conditional probabilities for category 0 in class 1. This means that the observed variables reflect a somewhat better measurement for the masters than for the nonmasters.

In difference to cross-sectional data a panel design contains repeated responses to one or more items at two or more time points from the same respondent. Latent class models with panel data can be formulated with different types of Markov models. Here, the assumption of local independence, i. e., the joint probability of several responses is the product of the marginal response probabilities given a latent class, will be relaxed. The following section discusses and applies several types of Markov models beginning with manifest Markov models. In the last part of the section we continue to latent Markov models. Three time points are used for the following formal descriptions and the examples. All type of models can be extended to subsequent panel waves.

3 Markov Models

3.1 Manifest Mixed Markov Model

Markov models are used to analyze categorical panel data. They are able to specify a Markov process in each latent class describing a pattern of repeated responses. In order to introduce the typical notation for these models (see ***Langeheine*** and ***van de Pol*** 1990), the case of one item x at three panel waves is considered. An ordinary Markov chain describes the frequency of the response pattern $m_{x_{ijkl}}$:

$$m_{x_{ijk}} = N\delta_i^1 \tau_{j|i}^{2\,1} \tau_{k|j}^{3\,2} \tag{22}$$

with

- δ_i^1 as the *initial probabilities* for t_1,
- $\tau_{j|i}^{2\,1}$ as the *transition probabilities* from t_1 to t_2, given a particular category at time t_1,
- $\tau_{k|j}^{3\,2}$ as the *transition probabilities* from t_2 to t_3, given a particular category at time t_2,

Subcripts of the τ's refer to item categories, superscripts of the τ's refer to panel waves.

The relationsship to latent class models becomes obvious when the model is generalized to a *mixed Markov model*. This model assumes that G markov chains (classes) describe the data:

$$m_{ijk} = N \sum_{g=1}^{G} \pi_g \delta_{i,g}^{1} \tau_{j|ig}^{2\,1} \tau_{k|jg}^{3\,2} \tag{23}$$

with π_g as the *latent class probabilities* for chain g.

If the probability of being in some category at time $t+1$ is independent of the probability of being in some category at the most recent point at time t, the model in Equation 23 reduces to Equation 24:

$$m_{ijk} = N \sum_{g=1}^{G} \pi_g \delta_{i,g}^{1} \tau_{j|g}^{2} \tau_{k|g}^{3} \tag{24}$$

The latent class model in Equation 24 can be considered as a special case of the mixed Markov model in Equation 23 with all transition probabilities assumed to be independent of the responses at the preceding time point. Or, the latent class model in Equation 24 is extended with a Markov model to Equation 23.

Markov models can only be estimated with certain restrictions on the parameters. For example, one needs dichotomous variables in four panel waves to estimate Equation 23 with two chains. Plausible restrictions are explored with the Mover-Stayer model and the Black and White model.

Mover-Stayer Model

The Mover-Stayer model assumes two classes of persons:

1. Class of persons *moving* between panel waves, and
2. Class of persons *staying* between panel waves.

This special case of the mixed Markov model can be modelled for the two classes with two Markov chains. Equation 25 models the class of *movers*:

$$m_{ijk} = N \pi_1 \delta_{i,1}^{1} \tau_{j|i\,1}^{2\,1} \tau_{k|j\,1}^{3\,2} \tag{25}$$

with π_1 as the *latent class probability* to be a "mover". Time-homogeneous transition probabilities can be applied for the movers:

$$\tau_{j|i\,1}^{2\,1} = \tau_{k|j\,1}^{3\,2} \tag{26}$$

Equation 27 models the class of *stayers*:

$$m_{ijk} = N \pi_2 \delta_{i,2}^{1} \tau_{j|i\,2}^{2\,1} \tau_{k|j\,2}^{3\,2} \tag{27}$$

with π_2 as the *latent class probability* to be a “stayer“. The “stayer“ remain with a probability of 1.0 in the same category as in previous waves:

$$\begin{array}{ll} \tau_{j|i,2} = 1 \text{ for } i = j; & \tau_{j|i,2} = 0 \text{ for } i \neq j \\ \tau_{k|j,2} = 1 \text{ for } j = k; & \tau_{k|j,2} = 0 \text{ for } j \neq k \end{array} \quad (28)$$

The Mover-Stayer model is a mixed Markov model with two chains. Transition probabilities are restricted to time-homogeneity for the class of “movers“ and restricted to the identity matrix for the class of “stayers“. For example, a population of voters can be classified into “movers“ and “stayers“ meaning that one part of the population tends to move in their votes from one election to the other and the other part of the population tends to vote the same party in every election.

Black and White Model

The Black and White model assumes two classes of persons:

1. Class of persons *staying* between panel waves, and
2. Class of persons *randomly moving* between panel waves.

Equation 27 applies for the Black and White model as follows:

$$m_{ijk} = N\pi_2\delta^1_{i,2}\tau^{2\,1}_{j|i\,2}\tau^{3\,2}_{k|j\,2} \quad (29)$$

Restrictions according to Equation 28 applies for the class of “stayers“. For the class of “random movers“ the following restrictions can be stated:

$$\begin{array}{rl} \delta^1_{i,2} & = 0.5 \text{ for all } i \\ \tau_{j|i,2} = \tau_{k|j,2} & = 0.5 \text{ for all } i, j, k \end{array} \quad (30)$$

The Black and White model differs only in one part from the Mover-Stayer model. A random process underlies the process of moving between the time points. Regarding the example from the Mover-Stayer model this means a random change of votings from one election to the other. The tendency to vote the same party in every election describes the class of “stayers“.

3.2 Latent Markov Model

The main weakness of the manifest Markov model is that it takes no measurement error into account. Because measurement errors are ubiquitous for social science data in general, it seems advisable to adhere to latent Markov models allowing for this notion.

The latent Markov model is a combination of the classical latent class model (cf. Equation 1) and the manifest Markov model (cf. Equation 22). Again, three panel waves are assumed. The expected frequencies m_{ijk} are calculated as follows:

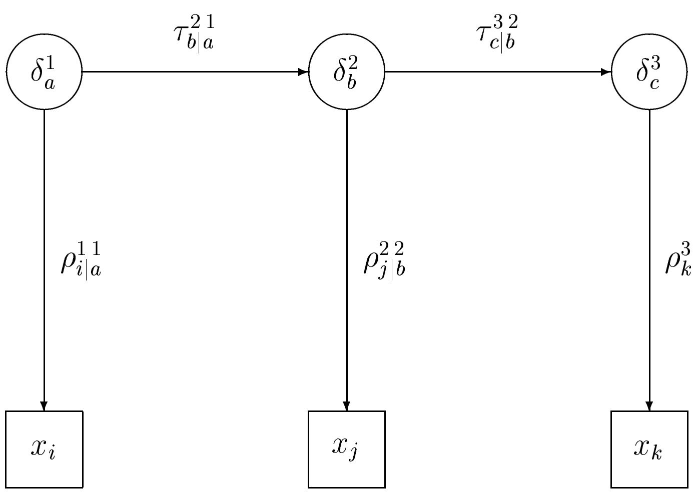
$$m_{ijk} = N \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} \delta_a^1 \rho_{i|a}^{1\,1} \tau_{b|a}^{2\,1} \rho_{j|b}^{2\,2} \tau_{c|b}^{3\,2} \rho_{k|c}^{3\,3} \tag{31}$$

with

- δ_a^1 as the *latent class probabilities* of time t_1,
- $\rho_{i|a}^{1\,1}$ as the *conditional probabilities* of item x_i for class a at time t_1,
- $\rho_{j|b}^{2\,2}$ as the *conditional probabilities* of item x_j for class b at time t_2,
- $\rho_{k|c}^{3\,3}$ as the *conditional probabilities* of item x_k for class c at time t_3,
- $\tau_{b|a}^{2\,1}$ as the *transition probabilities* from t_1 to t_2,
- $\tau_{c|b}^{3\,2}$ as the *transition probabilities* from t_2 to t_3.

Figure 2 shows the graphical representation of the observed variable x measured at three time points (indices i, j and k) mapped onto the latent one (indices a, b and c) by conditional response probabilities ($\rho_{i|a}^{1\,1}$, $\rho_{j|b}^{2\,2}$, $\rho_{k|c}^{3\,3}$). Conditional probabilities show the reliability of the observed variables whereas the transition probabilities show the amount of change between the classes over time considering random measurement error.

Figure 2: Latent Markov Model with One Observed Variable and Three Panel Waves



A latent Markov model with one indicator needs four panel waves to estimate the unknown parameters. But substantive restrictions can reduce the amount of parameter estimates. The conditional probabilities can be set equal over time ($\rho^{1\,1}_{i|a} = \rho^{2\,2}_{j|b} = \rho^{3\,3}_{k|c}$ in Equation 31). This restriction assumes equal reliabilities of the measurements over time. Furthermore, the transition probabilites can be set equal over time if $t > 2$ ($\tau^{2\,1}_{b|a} = \tau^{3\,2}_{c|b}$). This restriction assumes that the amount of stability and change is equal from panel wave to the other. Both restrictions are applied in the following example.

Empirical Example

Data and variables are from a longitudinal study of adolescents' stress and risk behavior with four panel waves (cf. ***Engel*** and ***Reinecke*** 1994). Three panel waves are used in the following example. The number of respondents is $N = 574$. Two items measuring negative or positive feelings are considered: "Sucessful experience" ($x_1/x_2/x_3$) and "Feeling to set up something" ($y_1/y_2/y_3$). Categories are dichotomous: "no, seldom" (1) and "sometimes, often" (2). The latent Markov model with the observed variables $x_1/x_2/x_3$ and $y_1/y_2/y_3$ representing three panel waves is specified with two latent classes in each panel wave:

$$m_{ijklmn} = N \sum_{a=1}^{2} \sum_{b=1}^{2} \sum_{c=1}^{2} \delta^1_a \rho^{1\,1}_{x_1|a} \rho^{1\,1}_{y_1|a} \tau^{2\,1}_{b|a} \rho^{2\,2}_{x_2|b} \rho^{2\,2}_{y_2|b} \tau^{3\,2}_{c|b} \rho^{3\,3}_{x_3|c} \rho^{3\,3}_{y_3|c} \qquad (32)$$

The program PANMARK (***van de Pol et al.*** 1991) was used to estimate the different model variants of the latent Markov model. Goodness-of-fit statistics is summarized in Table 4.

Table 4: Model Variants and Goodness-of-Fit Statistics of the Latent Markov Model

Model variants	L^2	df	AIC	BIC
LM0 (not restricted)	65.43	46	2128.67	2199.38
LM1 (same ρ's for same variables)	79.54	54	2126.78	2164.21
LM2 (same τ's)	**80.91**	**56**	**2124.15**	**2153.26**
LM3 (same ρ's for all variables)	120.42	58	2159.66	2180.45

Goodness-of-Fit statistics of the accepted model are in boldfaced type.

Model variant LM0 does not contain any restrictions on the parameters. Model variant LM1 estimates equal conditional probabilities over time for the same observed variables. In addition, model variant LM2 assumes equal transition probabilities between the panel waves. The last model variant LM3 estimates equal conditional probabilities over time for every observed variable testing for equal reliabilities.

The difference between the L^2 of model variant LM0 and LM1 does not lead to a significant decrease in model fit. This confirms the assumption of equal reliabilities of the same observed variables. The difference between the L^2 of model variant LM1 and LM2 does also not lead to a significant decrease in model fit. In addition, this confirms the assumption of equal transition probabilities. Only the assumption that all observed variables have the same reliability lead to a significant difference between model and data comparing model variants LM2 and LM3. Table 5 shows the estimated parameters of the accepted model variant LM2.

Table 5: Estimated Parameters of the Latent Markov Model with Restrictions of Model Variant LM2

	t_1		t_2		t_3	
δ^1_a	$\rho^{1\,1}_{x_1\mid a}$	$\rho^{1\,1}_{y_1\mid a}$	$\rho^{2\,2}_{x_2\mid b}$	$\rho^{2\,2}_{y_2\mid b}$	$\rho^{3\,3}_{x_3\mid c}$	$\rho^{3\,3}_{y_3\mid c}$
not successful	*.807*	*.567*	*.807*	*.567*	*.807*	*.567*
.155	(.063)	(.054)	(.063)	(.054)	(.063)	(.054)
(.029)	.193	.433	.193	.433	.193	.433
	(.063)	(.054)	(.063)	(.054)	(.063)	(.054)
successful	.074	.040	.074	.040	.074	.040
.845	(.012)	(.009)	(.012)	(.009)	(.012)	(.009)
(.029)	*.926*	*.960*	*.926*	*.960*	*.926*	*.960*
	(.012)	(.009)	(.012)	(.009)	(.012)	(.009)
		$\tau^{2\,1}_{b\mid a}$		$\tau^{3\,2}_{c\mid b}$		
		δ^2_1	δ^2_2	δ^3_1	δ^3_2	
not successful	δ^1_1	*.710*	.290	*.710*	.290	
		(.060)	(.060)	(.060)	(.060)	
successful	δ^1_2	.051	*.949*	.051	*.949*	
		(.012)	(.012)	(.012)	(.012)	

Conditional response probabilities characterizing the latent classes are emphasized as well as transition probabilities indicating stability between panel waves. Standard errors are given in brackets.

The first latent class contains the “not successful“ adolescents (16%), the second latent class contains the “successful“ adolescents (84%). Characteristic conditional probabilities of the first class (Category 1) are .807 for the first item (x) in each wave and .567 for the second item (y). The probability to have no successful experience is higher than the probability to have no feeling to set up something. Characteristic conditional probabilities of the second class (Category 2) are .926 for the first item (x) in each wave and .960 for the second item (y). The probability to have successful experience is almost equal to the probability to have a feeling to set up something.

Transition probabilities (τ's) show a high stability for the second ("successful") class (.949) and a somewhat lower stability for the first ("not successful") class (.710). The transition rate from the first class to the second is .290 and from the second to the first class only .051. This means, that a change over the time period from the class of "not successful" adolescents to the class of "successful" adolescents is higher than vice versa. It implicates that the latent class probability of the first class decreases and the latent class probability of the second class increases over time. The distribution of the latent class probabilities of the second wave is calculated as follows:

$$\delta^{(1)'} * T = \delta^{(2)'} \tag{33}$$

$$\begin{pmatrix} .155 & .845 \end{pmatrix} * \begin{pmatrix} .710 & .290 \\ .051 & .949 \end{pmatrix} = \begin{pmatrix} .153 & .847 \end{pmatrix}$$

The distribution of the latent class probabilities of the third wave is calculated as follows:

$$\delta^{(2)'} * T = \delta^{(3)'} \tag{34}$$

$$\begin{pmatrix} .153 & .847 \end{pmatrix} * \begin{pmatrix} .710 & .290 \\ .051 & .949 \end{pmatrix} = \begin{pmatrix} .152 & .848 \end{pmatrix}$$

Because of the time-homogeneous transition rates ($\tau^{2\,1}_{b|a} = \tau^{3\,2}_{c|b}$) the matrix T has the same values for both multiplications. Only a slight change of latent class probabilities can be observed. The Markov process tends to reach an equillibrium. This tendency is also confirmed by the expected frequencies of the transitions (cf. Table 6).

Table 6: Contingency Table of Expected Frequencies of the Latent Classes between t_1 and t_3

t_1	t_3		
	δ^3_1	δ^3_2	Total
δ^1_1	38.02	35.20	73.22
	(0.519)	(0.481)	(1.000)
δ^1_2	33.67	366.02	399.78
	(0.084)	(0.916)	(1.000)
Total	71.68	401.32	473.00
	(0.152)	(0.848)	(1.000)

Row proportions are given in brackets

38 adolescents are expected to remain in the first class while 399 are expected to remain in the second class over the time period. 35 persons are expected to change from the first to the second class while around 34 persons are expected to change vice versa. To prevent interpration problems, it should be noted that transition probabilities reflect the *relative* extent of stability and change in the latent class model while expected frequencies of the transitions reflect the *absolute* extent of stability and change.

The disadvantage of the latent Markov model is the restriction to one Markov chain. If one wants to model unobserved heterogeneity with different Markov chains (for example movers and stayers) and wants to take random measurement error into account, a latent Markov model has to be extended to a latent mixed Markov model.

3.3 Latent Mixed Markov Model

The latent mixed Markov model is a generalization of the latent Markov model (cf. Equation 31) and the manifest mixed Markov model (cf. Equation 23; ***Langeheine*** and ***van de Pol*** 1990). As with the latent Markov model three panel waves are assumed. The expected frequencies m_{ijk} are calculated as follows:

$$m_{ijk} = N \sum_{g=1}^{G} \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} \pi_g \delta^{1}_{a|g} \rho^{1\,1}_{i|ag} \tau^{2\,1}_{b|ag} \rho^{2\,2}_{j|bg} \tau^{3\,2}_{c|bg} \rho^{3\,3}_{k|cg} \tag{35}$$

with

- π_g as the *latent class probabilities* for chain g
- $\delta^{1}_{a|g}$ as the *latent class probabilities* of time t_1 in chain g
- $\rho^{1\,1}_{i|ag}$ the *conditional probabilities* of item x_i for class a at time t_1 in chain g
- $\rho^{2\,2}_{j|bg}$ as the *conditional probabilities* of item x_j for class b at time t_2 in chain g
- $\rho^{3\,3}_{k|cg}$ as the *conditional probabilities* of item x_k for class c at time t_3 in chain g
- $\tau^{2\,1}_{b|ag}$ as the *transition probabilities* from t_1 to t_2 in chain g
- $\tau^{3\,2}_{c|bg}$ as the *transition probabilities* from t_2 to t_3 in chain g

The difference to Equation 31 is only the specification of the parameter π which allows different markov chains in the latent Markov model. The difference to Equation 23 is the level of the Markov chains. Markov chains in the latent mixed Markov model are on the latent level instead of the manifest level.

Empirical Example

Data and variables are also from the longitudinal study of adolescents' stress and risk behavior. Two stress items from three panel waves with $N = 574$ respondents are considered: "Nervous, restlessness" ($x_1/x_2/x_3$) and "Difficulties to concentrate" ($y_1/y_2/y_3$). Categories are dichotomous: "no" (1) and "yes" (2).

A general latent mixed Markov model with two observed variables in each of the three panel waves are specified. Specification of two chains are according to the restrictions of the Mover-Stayer model (cf. Equations 25 and 26). Model specification of the chain of "movers" is as follows:

$$m_{ijklmn} = N \sum_{a=1}^{2} \sum_{b=1}^{2} \sum_{c=1}^{2} \pi_1 \delta^{1}_{a,1} \rho^{1\,1}_{x_1|a1} \rho^{1\,1}_{y_1|a1} \tau^{2\,1}_{b|a1} \rho^{2\,2}_{x_2|b1} \rho^{2\,2}_{y_2|b1} \tau^{3\,2}_{c|b1} \rho^{3\,3}_{x_3|c1} \rho^{3\,3}_{y_3|c1} \tag{36}$$

with π_1 as the *latent class probabilities* for chain 1 (mover). Equal conditional probabilities for the same variables are specified across waves but transition probabilities are unequal, meaning that $\tau^{2\,1}_{b|a1} \neq \tau^{3\,2}_{c|b1}$. Model specification of the chain of "stayers" is as follows:

$$m_{ijklmn} = N \sum_{a=1}^{2} \sum_{b=1}^{2} \sum_{c=1}^{2} \pi_2 \delta^{1}_{a,2} \rho^{1\,1}_{x_1|a2} \rho^{1\,1}_{y_1|a2} \tau^{2\,1}_{b|a2} \rho^{2\,2}_{x_2|b2} \rho^{2\,2}_{y_2|b2} \tau^{3\,2}_{c|b2} \rho^{3\,3}_{x_3|c2} \rho^{3\,3}_{y_3|c2} \tag{37}$$

with π_2 as the *latent class probabilities* for chain 2 (stayer). Similar to Equation 28 the following restrictions are specified for the transition probabilities:

$$\tau^{2\,1}_{b|a,2} = 1 \ for \ a = b; \ \tau^{2\,1}_{b|a,2} = 0 \ for \ a \neq b \tag{38}$$

$$\tau^{3\,2}_{c|b,2} = 1 \ for \ b = c; \ \tau^{3\,2}_{c|b,2} = 0 \ for \ b \neq c \tag{39}$$

The latent Mover-Stayer model is equal to a latent two-chain mixed Markov model. The model fit is acceptable with $L^2 = 68.51$ and $df = 48$. Specifications and restrictions of Equations 36 to 39 are confirmed.

Table 7 gives the estimated parameters of the latent mixed Markov model. The first column of the table shows the latent class probabilities of the Markov chains. 59% of the respondents are classified as "movers" and 41% as "stayers". The second column contains the latent class probabilities of the latent variable in each chain. In the mover chain 59% perceive no stress and 41% perceive stress in form of restlessness or concentration problems. In the stayer chain 79% of the respondents indicate no stress and 21% stress. The following columns contains the conditional probabilities for two variables in each wave and each chain. The first item has less measurement errors in the stayer chain, the second item shows a better reliability in the mover chain.

The lower half of Table 7 shows the transition probabilities estimated for the mover chain and the fixed values for the stayer chain. Those fixed values indicate perfect stability within the latent classes across the waves and no change. In the mover chain,

Table 7: Estimated Parameters of the Latent Mixed Markov Model with two chains (Mover/Stayer)

		t_1		t_2		t_3	
π_g	δ^1_{ag}	$\rho^{1\,1}_{x_1\|ag}$	$\rho^{1\,1}_{y_1\|ag}$	$\rho^{2\,2}_{x_2\|bg}$	$\rho^{2\,2}_{y_2\|bg}$	$\rho^{3\,3}_{x_3\|cg}$	$\rho^{3\,3}_{y_3\|cg}$
Mover	no	.687	1.000	.687	1.000	.687	1.000
.589	stress	(.044)	(–)	(.044)	(–)	(.044)	(–)
(.048)	.590	.313	.000	.313	.000	.313	.000
	(.088)	(.044)	(–)	(.044)	(–)	(.044)	(–)
	stress	.360	.154	.360	.154	.360	.154
		(.055)	(.139)	(.055)	(.139)	(.055)	(.139)
	.410	.640	.846	.640	.846	.640	.846
	(.088)	(.055)	(.139)	(.055)	(.139)	(.055)	(.139)
Stayer	no	1.000	.909	1.000	.909	1.000	.909
.411	stress	(–)	(.025)	(–)	(.025)	(–)	(.025)
(.048)	.794	.000	.091	.000	.091	.000	.091
	(.088)	(–)	(.025)	(–)	(.025)	(–)	(.025)
	stress	.042	.114	.042	.114	.042	.114
		(.079)	(.071)	(.079)	(.071)	(.079)	(.071)
	.206	.958	.886	.958	.886	.958	.886
	(.088)	(.079)	(.071)	(.079)	(.071)	(.079)	(.071)
			$\tau^{2\,1}_{b\|ag}$		$\tau^{3\,2}_{c\|bg}$		
			δ^2_{11}	δ^2_{21}	δ^3_{11}	δ^3_{21}	
Mover	no	δ^1_{11}	.750	.250	.773	.227	
	stress		(.047)	(.047)	(.050)	(.050)	
	stress	δ^1_{21}	.439	.561	.313	.687	
			(.150)	(.150)	(.140)	(.140)	
			δ^2_{12}	δ^2_{22}	δ^3_{12}	δ^3_{22}	
Stayer	no	δ^1_{12}	1.000	.000	1.000	.000	
	stress		(fixed)	(fixed)	(fixed)	(fixed)	
	stress	δ^1_{22}	.000	1.000	.000	1.000	
			(fixed)	(fixed)	(fixed)	(fixed)	

75% of the respondents classified as persons who perceive no stress remain in the latent class from the first to the second wave and 77% remain from the second to the third wave. 25% of those persons change to the latent class indicating stress whereas about 44% change in the other direction. The difference of change in both directions lowers to approximately 31% and 23% from the second to the third wave.

The change of the latent class probabilities are calculated according to Equation 33. For the chain of stayers no calculation is necessary. The distribution of latent class probabilities is the same in all panel waves because the matrix of transition probabilities T is equal to an identity matrix I:

$$\delta^{(1)'} * T = \delta^{(2)'} * T = \delta^{(3)'} \tag{40}$$

with $T = I$. For the chain of movers the distribution of the latent class probabilities of the second wave is calculated as follows:

$$\delta^{(1)'} * T = \delta^{(2)'} \tag{41}$$

$$\begin{pmatrix} .590 & .410 \end{pmatrix} * \begin{pmatrix} .750 & .250 \\ .439 & .561 \end{pmatrix} = \begin{pmatrix} .622 & .378 \end{pmatrix}$$

The distribution of the latent class probabilities of the third wave is calculated as follows:

$$\delta^{(2)'} * T = \delta^{(3)'} \tag{42}$$

$$\begin{pmatrix} .622 & .378 \end{pmatrix} * \begin{pmatrix} .773 & .227 \\ .313 & .687 \end{pmatrix} = \begin{pmatrix} .599 & .401 \end{pmatrix}$$

Transition rates are not time-homogeneous ($\tau^{2\,1}_{b|a} \neq \tau^{3\,2}_{c|b}$) which means that the matrix T has different values for both multiplications.

Comparing the latent class distributions in each panel wave, only a slight change can be observed. The expected frequencies of the transitions confirm this tendency. In Table 8 those expected frequencies between the first and last panel wave are shown seperately for the mover and the stayer chain.

Table 8: Contingency Table of Expected Frequencies of the Latent Classes between t_1 and t_3 (seperate for mover and stayer)

	"mover"			"stayer"		
t_1	t_3			t_3		
	δ^3_1	δ^3_2	Total	δ^3_1	δ^3_2	Total
δ^1_1	112.27 (0.658)	58.24 (0.342)	170.51 (1.000)	160.43 (1.000)	0.00 (0.000)	160.43 (1.000)
δ^1_2	61.07 (0.516)	57.40 (0.485)	118.47 (1.000)	0.00 (0.000)	41.59 (1.000)	41.59 (1.000)
Total	173.34 (0.600)	115.64 (0.400)	288.98 (1.000)	160.43 (0.794)	41.59 (0.206)	202.02 (1.000)

Row proportions are given in brackets

In the stayer chain 160 adolescents are expected to remain in the first class while 42 are expected to remain in the second class over the time period. According to the restrictions no one is expected to move between the classes.

In the mover chain 112 adolescents are expected to remain in the first class while 57 are expected to remain in the second class over the time period. Around 58 persons are expected to change from the first to the second class while 61 persons are expected to change vice versa.

4 Conclusions

Latent class analysis is a flexible statistical tool either to explore the latent structure of a set of observed variables, or to confirm substantive hypotheses. Latent class analysis makes it possible to identify the categories of a latent variable characterizing the latent types of an item battery, i. e., indicators measuring stress. Depending on the hypotheses, restrictions on the parameters may apply.

For panel analysis, the latent variables reflecting the panel waves could be combined to one discrete latent variable. With latent Markov models, the latent variables are not combined to a joint latent variable, rather they are linked by transition rates. This extension allows simultaneous analysis of latent class probabilities, conditional probabilities and transition probabilities in panel models. Several kinds of restrictions (i. e. equal conditional probabilities across panel waves) reduces the number of parameters and makes models more parsimonious.

The latent mixed Markov model is a generalization of the latent Markov model and the manifest mixed Markov model which allows specification of unobserved heterogeneity via different Markov chains. Instead of one observed variable for every panel wave also multiple indicators as a set of observed variables for one or more latent variables may be analyzed. These latent variables should be related by a Markov chain or a mixture of Markov chains. This kind of modeling can be extended to several populations resulting in simultaneous multiple group comparisons (for an example see ***Engel*** and ***Reinecke*** 1994: 240). If multiple indicators are used in the described Markov models, no substantive differences exist in comparison to the well-known structural equation models (see ***Jöreskog*** and ***Sörbom*** 1993).

A common assumption in latent class analysis is that respondents do not differ within latent classes. All respondents of a given class are identical with respect to their response probabilities. ***Rost*** (1990) developed the mixed Rasch model which allows quantitative differences among subjects within a class by means of the ability parameter of the Rasch model. Further descriptions are also given in ***Rost*** (1996). An extension of mixed models to the discussed Markov models is beyond the scope of the present article.

A last remark should be given to the available software products. The classical program is MLLSA (Maximum Likelihood Latent Structure Analysis) from ***Clogg*** (1977). Extensions to log-linear parameterizations are possible with LCAG (Latent Class Analysis Models and other Log-linear Models with latent Variables) from ***Hagenaars*** and ***Luijkx*** (1990). This program was further developed and extended by ***Vermunt*** (1993) to ℓEM (Log-linear and Event History Analysis with Missing Data using the EM Algorithm). PANMARK developed by ***van de Pol***, ***Langeheine*** and ***de Jong*** (1991) is a program especially developed for Markov Models as discussed and applied in the present article. The new program M*plus* (The Comprehensive Modeling Program for Applied Researchers) from ***Muthen*** and ***Muthen*** (1998) should be considered for further model developments and applications. It allows combinations of categorical and continuous latent variables. Latent class models with covariates can be estimated within a framework of mixture models.

5 References

Bozdogan, H. 1987: Model Selection and Akaikes Information Criterion (AIC): The General Theory and its Analytical Extensions, in: Psychometrika, 52: 345 - 370.

Clogg, C. C. 1977: Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users (Working Paper 1977-09), University Park: Population Issues Research Center.

Dempster, A. P., N. M. Laird and ***D. B. Rubin*** 1977: Maximum Likelihood From Incomplete Data Via the EM Algorithm, in: Journal of the Royal Statistical Society, Ser. B, 39: 1 - 38.

Engel, U. and ***J. Reinecke*** 1994: Panelanalyse. Grundlagen - Techniken - Beispiele, Berlin: DeGruyter.

Goodman, L. A. 1974a: The Analysis of Qualitative Variables when Some of the Variables are Unobservable. Part I - A Modified Latent Structure Approach, in: American Journal of Sociology, 79: 1179 - 1259.

Goodman, L. A. 1974b: Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiably Models, in: Biometrika, 61: 215 - 231.

Haberman, S. J. 1978: Analysis of qualitative data. Volume 1. Introductory topics, New York: Academic Press.

Haberman, S. J. 1979: Analysis of qualitative data. Volume 2. New developments, New York: Academic Press.

Hagenaars, J. A. 1990: Categorical Longitudinal Data, Newbury Park: Sage.

Hagenaars, J. A. and ***R. Luijkx*** 1990: LCAG. A Program to Estimate Latent Class Models and Other Loglinear Models with Latent Variables with and without Missing Data. Working Paper Series 17, University of Tilburg.

Jöreskog, K. G. and ***D. Sörbom*** (1993): LISREL8: Structural Equation Modeling with the SIMPLIS Command Language, Chicago: Scientific Software.

Langeheine, R. and ***J. Rost*** 1996: Latent-Class-Analyse, in: ***E. Erdfelder, R. Mausfeld, T. Meiser*** and ***G. Rudinger*** (Hrsg.), Handbuch Quantitative Methoden, Weinheim: Beltz: 315 - 332.

Langeheine, R. and ***F. van de Pol*** 1990: A Unifying Framework for Markov Modeling in Discrete Space and Discrete Time. Sociological Methods and Research, 18: 416 - 441.

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

McCutcheon, A. L. 1987: Latent Class Analysis, Newbury Park: Sage.

Muthen, L. K. and ***B. O. Muthen*** 1998: M***plus.*** The Comprehensive Modeling Program for Applied Researchers. User's Guide, Los Angeles.

Rost, J. 1990: Rasch models in Latent Classes: An Integration of Two Approaches to Item Analysis, in: Applied Psychological Measurement, 14: 271 - 282.

Rost, J. 1996: Testtheorie - Testkonstruktion, Bern: Huber.

Rost, J. and ***R. Langeheine*** 1997: Applications of Latent Trait and Latent Class Models in the Social Sciences. Münster: Waxmann.

Schwarz, G. 1978: Estimating the Dimension of a Model, in: The Annals of Statistics, 6: 461 - 464.

van de Pol, F. and ***R. Langeheine*** 1990: Mixed Markov Latent Class Models, in: ***C. C. Clogg*** (ed.), Sociological Methodology. Oxford: Blackwell: 213 - 247.

van de Pol, F., R. Langeheine and ***W. de Jong*** 1991: PANMARK User Manual, Voorburg: Netherlands Central Bureau of Statistics.

Vermunt, J. K. 1993: ℓEM: Log-linear and Event History Analysis with Missing Data using the EM Algorithm. WORC paper 93.09.015/7. Tilburg University, Tilburg.