# Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates

Vermunt, J.K.; Langeheine, R.; Bockenholt, U.

*Publication date:*
1995

KATHOLIEKE UNIVERSITEIT BRABANT
TILBURG UNIVERSITY

WORC
Work and Organization Research Centre



PAPER

[redacted]

R10
- psychology
- statistical methods
- Panel data

PAPER

**Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates**

Jeroen K. Vermunt, Rolf Langeheine and Ulf Böckenholt

WORC PAPER 95.06.013/7



Paper presented at the 9th European Meeting of the Psychometric Society, July 4-7, 1995, Leiden.

June 1995

ACKNOWLEDGEMENT

Paper presented at the 9th European Meeting of the Psychometric Society, July 4-7, 1995, Leiden. The contribution of Vermunt to this paper is in the context of the WORC Research Programme 'Analysis of Social Change (P7-01).

# Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates

Jeroen K. Vermunt
WORC, Tilburg University, The Netherlands

Rolf Langeheine
Institute for Science Education, University of Kiel

Ulf Böckenholt
Department of Psychology, University of Illinois, Urbana-Champaign

Key words: *panel analysis, categorical data, measurement error, time-varying covariates, log-linear models, logit models, modified path analysis approach, latent class analysis, latent Markov models, modified Lisrel approach, EM algorithm*

**Abstract**

Discrete-time discrete-state Markov chain models can be used to describe individual change in categorical variables. But when the observed states are subject to measurement error, the observed transitions between two points in time will be partially spurious. Latent Markov models make it possible to separate true change from measurement error. The standard latent Markov model is, however, rather limited when the aim is to explain individual differences in the probability of occupying a particular state at a particular point in time. This paper presents a flexible logit regression approach which allows to regress the latent states occupied at the various points in time on both time-constant and time-varying covariates. The regression approach combines features of causal log-linear models and latent class models with explanatory variables. An application is presented in which pupils' interest in physics at different points in time is explained by the time-constant covariate sex and the time-varying covariate physics grade.

## 1 Introduction

Discrete-time discrete-state Markov chain models are well suited for analyzing categorical panel data. They can be used to describe individual change in categorical variables.

However, when the observed states are subject to measurement error, the observed transitions between two points in time will be a mixture of true change and spurious change caused by measurement error in the observed states (Van de Pol and De Leeuw, 1986; Hagenaars, 1992). Therefore, Wiggins (1973) proposed the latent Markov model which makes it possible to separate true change from measurement error (see also Van de Pol and Langeheine, 1990). The latent Markov is strongly related to the latent class model proposed by Lazarsfeld (Lazarsfeld and Henry, 1968).

The standard latent Markov model is, however, rather limited when the aim is to explain individual differences in the probability of occupying a particular state at a particular point in time. The only way that observed heterogeneity can be taken into account is by performing a multiple-group analysis as proposed by Van de Pol and Langeheine (1990). A disadvantage of multiple-group models is, however, that they contain many parameters when several explanatory variables are included in the analysis. Moreover, they can only be used with time-constant covariates, while the availability of information on time-varying covariates is one of the strong points of longitudinal data. Thus, what we actually need is a regression model for the latent states that allows to include both time-constant and time-varying covariates.

Goodman's causal log-linear model (Goodman, 1973) can be used to specify a regression model for the observed states. This model, which uses a priori information on the causal order among a set of categorical variables, consist of a recursive system of logit models in which a variable that appears as a dependent variable in one equation can be used as an independent variable in one of the subsequent equations. Goodman's causal log-linear model assumes, however, that all variables are observed. Also the latent class model has been extended to allow for explanatory variables influencing the latent variable (Haberman, 1979; Dayton and Macready, 1988). These extended latent class models are, however, not very well suited for estimating covariate effects when we have data on more than one occasion.

This paper presents a latent Markov model in which the latent states are regressed on time-constant and time-varying covariates by means of a system logit models. The model is an extension of Goodman's causal log-linear model in that the states occupied at the different points in time are latent variables instead of observed variables. Moreover, it extends Haberman's and Dayton and Macready's latent class models with explanatory variables in that it makes it possible to specify an a priori causal order among the variables included in the model. Hagenaars (1990, 1993) showed how to combine a causal log-linear model with a latent class model, which led to what he called a modified Lisrel approach (see also Vermunt 1993, 1994, 1995). Here, it is demonstrated that this modified Lisrel approach makes it possible to specify latent Markov models with covariates.

Section 2 discusses the manifest Markov model, the latent class model, the latent Markov model and the multiple-group Markov model. Section 3 presents logit regression models for latent states using Hagenaars' extension of Goodman's causal log-linear models. Section 4 discusses maximum likelihood estimation of the extended latent Markov models by means of the EM algorithm and presents the $\ell EM$ program (Vermunt, 1993) which can be used for this purpose. An application using data from a German panel study is presented in Section 5. In this application, pupils' interest in physics at different points in time is explained by the time-constant covariate sex and the time-varying covariate physics grade.

# 2 Markov models

## 2.1 Manifest Markov model

Suppose we have repeated observations on a particular categorical or discrete variable, such as, for instance, marital status, occupational status, the choice among brands, or the grades in English of pupils. This kind of data, which is generally collected to describe individual change in the variable concerned, can very well be analyzed by means of Markov models. When the variable of interest is discrete and when measurements took place at particular points in time, the models are called discrete-time discrete-space Markov models (Bishop, Fienberg and Holland, 1975: Chapter 7).

Let $T$ denote the time variable, $t$ a particular point in time, and $T^*$ the number of discrete time points for which we have observations, or in other words, the number of occasions or panel waves. The variable indicating the state that a person occupies at time point $T = t$ is denoted by $Y_t$, a particular value of $Y_t$ by $y_t$, and the number of states by $Y^*$.

For sake of simplicity, it will be assumed that only information on three occasions is available, or in other words, that $T^* = 3$. The data can be organized in a three-way frequency table with observed frequencies $n_{y_1 y_2 y_3}$. The probability of having $Y_1 = y_1$, $Y_2 = y_2$, and $Y_3 = y_3$ is indicated by $\pi_{y_1 y_2 y_3}$. So, $\pi_{y_1 y_2 y_3}$ denotes the probability of belonging to cell $(y_1, y_2, y_3)$ of the joint distribution of $Y_1$, $Y_2$, and $Y_3$.

When specifying a model for $\pi_{y_1 y_2 y_3}$ it is natural to use the information on the time order, or causal order, among the variables $Y_1$, $Y_2$, and $Y_3$. The most general model for $\pi_{y_1 y_2 y_3}$ is

$$\pi_{y_1 y_2 y_3} = \pi_{y_1} \pi_{y_2|y_1} \pi_{y_3|y_1 y_2} . \tag{1}$$

Here, $\pi_{y_1}$ denotes the probability that $Y_1 = y_1$, $\pi_{y_2|y_1}$ the probability that $Y_2 = y_2$, given that $Y_1 = y_1$, and $\pi_{y_3|y_1 y_2}$ the probability that $Y_3 = y_3$, given that $Y_1 = y_1$ and $Y_2 = y_2$. The model represented in Equation 1 is a saturated model since it contains as many observed cell counts as parameters.

A Markov model is obtained by assuming that the process under study is without memory, that is, the state occupied at $T = t$ depends only the state occupied at $T = t - 1$. Such a model is sometimes also called a first-order Markov model. The general model given in Equation 1 is not a first-order Markov model since $Y_3$ does not only depend on $Y_2$, but also on $Y_1$. Actually, this model is a second-order Markov model because $Y_t$ depends on $Y_{t-2}$. A (first-order) Markov model for $\pi_{y_1 y_2 y_3}$ can be written as

$$\pi_{y_1 y_2 y_3} = \pi_{y_1} \pi_{y_2|y_1} \pi_{y_3|y_2} . \tag{2}$$

As can be seen, in this model it is assumed that $\pi_{y_3|y_1 y_2} = \pi_{y_3|y_2}$.

A more parsimonious Markov model can be obtained by assuming the transition probabilities $\pi_{y_t|y_{t-1}}$ to be independent of $T$. This gives a so-called time-homogeneous or stationary Markov model. The model given in Equation 2 becomes a stationary Markov model by restricting

$$\pi_{y_2|y_1} = \pi_{y_3|y_2} .$$

## 2.2 Latent class model

Above, it was implicitly assumed that the variable of interest is measured without error. But, since in most situations such an assumption is unrealistic, it is important to be able to take measurement error into account when specifying statistical models. The problem of measurement error has given rise to a family of models called latent structure models, which are all based on the assumption of local independence. This means that the observed variables or indicators which are used to measure the unobserved variable of interest are assumed to be mutually independent for a particular value of the unobserved or latent variable.

Latent structure models can be classified according to the measurement level of the latent variable(s) and the measurement level of the manifest variables (Bartholomew, 1987; Heinen, 1993). In factor analysis, continuous manifest variables are used as indicators for one or more continuous latent variables. In latent trait models, normally one continuous latent variable is assumed to underlie a set of categorical indicators. And finally, when both the manifest and the latent variables are categorical, we have a latent class model (Lazarsfeld and Henry, 1968; Goodman, 1974; Haberman, 1979).

Suppose we have a latent class model with one latent variable $W$ with index $w$ and three indicators $A$, $B$, and $C$ with indices $a$, $b$, and $c$. Moreover, let $W^*$ denote the number of latent classes, and $A^*$, $B^*$, and $C^*$ the number of categories of $A$, $B$, and $C$, respectively. The basic equations of the latent class model are

$$\pi_{abc} = \sum_{w=1}^{W^*} \pi_{wabc} , \qquad (3)$$

where

$$\pi_{wabc} = \pi_w \, \pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \qquad (4)$$

Here, $\pi_{wabc}$ denotes a probability of belonging to cell $(w, a, b, c)$ in the joint distribution including the latent dimension $W$. Furthermore, $\pi_w$ is the proportion of the population belonging to latent class $w$. The other $\pi$-parameters are conditional response probabilities. For instance, $\pi_{a|w}$ is the probability of having a value of $a$ on $A$ given that one belongs to latent class $w$.

From Equation 3, it can be seen that the population is divided into $W^*$ exhaustive and mutually exclusive classes. Therefore, the joint probability of the observed variables can be obtained by summation over the latent dimension. The classical parameterization of the latent class model, as proposed by Lazarsfeld and Henry (1968) and as it is used by Goodman (1974), is given in Equation 4. It can be seen that the observed variables $A$, $B$, and $C$ are postulated to be mutually independent given a particular score on the latent variable $W$.

## 2.3 Latent Markov model

By combining the Markov model given in Equation 2 and the latent class model given in Equation 4, one obtains a model which can be used for analyzing change, but in which the states occupied at different points in time may be measured with error. This model, which was originally proposed by Wiggins (1973), is called a latent Markov model. Poulsen (1982), Van de Pol and De Leeuw (1986), and Van de Pol and Langeheine (1990) contributed to its practical applicability.

It is well known that measurement error attenuates the relationships between variables. This means that the relationship between two observed variables which are subject to measurement error will generally be weaker than their true relationship. For the analysis of change, this phenomenon implies that when the observed states are subject to measurement error, the strength of the relationships among the true states occupied at two subsequent points in time will be underestimated, or in other words, the amount of change will be overestimated. When the data are subject to measurement error, the observed transitions are, in fact, a mixture of true change and spurious change resulting from measurement error (Van de Pol and De Leeuw, 1986; Hagenaars, 1992). The latent Markov model makes it possible to separate true change and spurious change caused by measurement error.

To be able to formulate the latent Markov model, the notation has to be extended. Let $W_t$ be the latent or true state at $T = t$ having three indicators which are denoted by $A_t$, $B_t$, and $C_t$. Like above, lower case letters will be used as indices. Assume again that one has observations for three occasions, that is, $T^* = 3$. Note that now the observed data is organized into a nine-way frequency table with cell counts $n_{a_1 b_1 c_1 a_2 b_2 c_2 a_3 b_3 c_3}$. The probability of belonging to a particular cell in the joint distribution of the three latent variables and the nine indicators is denoted by $\pi_{w_1 a_1 b_1 c_1 w_2 a_2 b_2 c_2 w_3 a_3 b_3 c_3}$. The latent Markov model for three points in time and three indicators per occasion can be defined as

$$\pi_{w_1 a_1 b_1 c_1 w_2 a_2 b_2 c_2 w_3 a_3 b_3 c_3} = \pi_{w_1} \pi_{a_1|w_1} \pi_{b_1|w_1} \pi_{c_1|w_1} \pi_{w_2|w_1} \pi_{a_2|w_2} \pi_{b_2|w_2} \pi_{c_2|w_2} \pi_{w_3|w_2} \pi_{a_3|w_3} \pi_{b_3|w_3} \pi_{c_3|w_3} . \quad (5)$$

In contrast to the latent class model, it is also possible to estimate a latent Markov model with only one indicator per occasion. For instance, when we have only $A_t$ as indicator for the latent state $W_t$, the latent Markov model simplifies to

$$\pi_{w_1 a_1 w_2 a_2 w_3 a_3} = \pi_{w_1} \pi_{a_1|w_1} \pi_{w_2|w_1} \pi_{a_2|w_2} \pi_{w_3|w_2} \pi_{a_3|w_3} . \quad (6)$$

To identify the parameters of the multiple indicator latent Markov model represented in Equation 5, it is not necessary to impose further restrictions on the model parameters. The single indicator latent Markov model can, however, not be identified without further restrictions (Van de Pol and Langeheine, 1990). The model for three points in time given in Equation 6 can be identified by assuming the response probabilities to be time-homogenous, in other words, by imposing the following restrictions

$$\pi_{a_1|w_1} = \pi_{a_2|w_2} = \pi_{a_3|w_3} .$$

When there are at least four points in time, a latent Markov model with a single indicator per occasion can also be identified by assuming stationarity.

## 2.4 Heterogeneity

In most cases, it is unrealistic to assume that the process under study is equal for all members of the population under study. For instance, males will not have the same probability of being or becoming employed as females, persons with different educational levels will have different divorce and married rates, the choice of brand in purchasing a particular product will depend on someone's income, and school grades will depend on pupils' social backgrounds. Therefore, it is important to be able to specify latent Markov models which take observed heterogeneity into account.

Analogous to the extension of latent class analysis for dealing with data on several subpopulations (Haberman, 1979; Clogg and Goodman, 1984; Hagenaars, 1990), Van de Pol and Langeheine (1990) proposed multiple-group latent Markov models. These Markov models involve the inclusion of one additional variable indicating a person's subgroup membership. This variable will be denoted by $G$, with index $g$ and $G^*$ categories. In its most general form, the multiple-group version of the latent Markov model with one indicator per occasion given in Equation 6 is

$$\pi_{gw_1a_1w_2a_2w_3a_3} = \pi_g \pi_{w_1|g} \pi_{a_1|w_1g} \pi_{w_2|w_1g} \pi_{a_2|w_2g} \pi_{w_3|w_2g} \pi_{a_3|w_3g} . \quad (7)$$

In this model every parameter is assumed to be subgroup specific. Of course, it is possible to restrict this model by assuming particular parameters to be equal among subgroups. For instance, in most applications, it will be assumed that measurement error is equal among subgroups. But, it is also possible to assume the initial distribution or the transition probabilities to be the same for all subgroups.

Although the multiple-group extension of the latent class model is very valuable, its applicability is limited in several respects. When applying statistical methods, researchers are interested in detecting the effects of a number of independent variables, or covariates, on the phenomenon under study. In the case of latent Markov models, one may be interested in determining the effect of particular covariates on the initial position and on the transition probabilities. When using the multiple-group analysis, the only thing that can be done is crossing all covariates and using this joint covariate as a grouping variable. It will be clear that this approach is only feasible when the number of cells of joint distribution of the independent variables is not too large, because otherwise a huge number of parameters has to be estimated.

Another limitation of the multiple-group approach is that it does not allow to make full use of the dynamic character of the data. A strong point of longitudinal data is that it does not contain only information on the changes in the dependent variable of interest, but also in the independent variables. In other words, variables which may influence the states occupied at the different points in time may be time-varying. It is very difficult to use such time-varying covariates in multiple-group latent Markov models.

What we actually need to be able to explain a person's latent state at $T = t$ is a regression-like model which can deal with both time-constant and time-varying covariates. The next section presents such a model.

# 3 Logit regression models

## 3.1 Causal log-linear models

Several strongly related approaches have been proposed for specifying regression models in the context of Markov modeling (Spilerman, 1972; Muenz and Rubinstein, 1985; Clogg, Eliason, and Grego, 1990; Kelton and Smith, 1991). One of these approaches, which can be used when all variables are categorical, is Goodman's modified path analysis approach (Goodman, 1973). Goodman demonstrated how to specify a causal log-linear model for a set of categorical variables using a priori information on their causal ordering. Because of the analogy with path analysis with continuous data, he called the model a modified path analysis approach.

Goodman's approach will be illustrated by introducing a time-constant covariate $X$ and a time-varying covariate $Z_t$ into the general manifest model described in Equation 1. In its most general form, the modified path model for the relationships among the variables $X$, $Z_1$, $Y_1$, $Z_2$, $Y_2$, $Z_3$, and $Y_3$ can be written as

$$\pi_{xz_1y_1z_2y_2z_3y_3} = \pi_x \, \pi_{z_1|x} \, \pi_{y_1|xz_1} \, \pi_{z_2|xz_1y_1} \, \pi_{y_2|xz_1y_1z_2} \, \pi_{z_3|xz_1y_1z_2y_2} \, \pi_{y_3|xz_1y_1z_2y_2z_3} \, . \quad (8)$$

Thus, the joint distribution of the variables, $\pi_{xz_1y_1z_2y_2z_3y_3}$, is decomposed into a set of conditional probabilities on the basis of the a priori causal order among these variables. Note that in this case, the causal order can almost completely be based on the time order among the variables. Only the order between $Z_t$ and $Y_t$ must determined in another way. Like the general model given in Equation 1, the above model for $\pi_{xz_1y_1z_2y_2z_3y_3}$ is a saturated model which can be restricted in various ways.

As demonstrated by Vermunt (1994, 1995), the general model given in Equation 8 can easily be restricted by assuming particular variables to be (conditionally) independent of some of its preceding variables. Suppose, for instance, that the Markov assumption holds for the dependent variable $Y$, that $Z$ is independent of the previous values of the dependent variable $Y$, and that there are no time-lagged effects of $Z$ on $Y$. These assumptions imply that the general model represented in Equation 8 can be simplified to

$$\pi_{xz_1y_1z_2y_2z_3y_3} = \pi_x \, \pi_{z_1|x} \, \pi_{y_1|xz_1} \, \pi_{z_2|xz_1} \, \pi_{y_2|xy_1z_2} \, \pi_{z_3|xz_1z_2} \, \pi_{y_3|xy_2z_3} \, . \tag{9}$$

When we are not interested in the relationships among the independent variables, it can also be written as

$$\pi_{xz_1y_1z_2y_2z_3y_3} = \pi_{xz_1z_2z_3} \, \pi_{y_1|xz_1} \, \pi_{y_2|xy_1z_2} \, \pi_{y_3|xy_2z_3} \, . \tag{10}$$

Note that the Markov assumption, the assumption of non-existence of time-lagged effects of $Z$ on $Y$, and the assumption of non-existence of direct effects $Y$ and $Z$ can be relaxed and therefore be tested.

The structure of the model given in Equation 10 is similar to a manifest version of the multiple-group latent Markov model given in Equation 7. The main difference is, however, that the grouping variable is composed of two variables, one of which is time-varying. This means that one of the two disadvantages of the multiple-group Markov model, namely, that the grouping variable has to be time-constant, has been overcome. The other weak-point of the multiple-group approach has not been resolved so far since every value of the joint independent variable still has its own set of initial probabilities and transition probabilities.

However, Goodman's modified path analysis approach does not only involve specifying a causal order among the categorical variables which are used in the analysis, but it also involves specifying logit models for the probabilities appearing at the right hand side of the general model represented in Equation 8. Vermunt (1994, 1995) showed that it is also possible to apply the logit parameterization to a restricted model such as the model given in Equation 10. This means that the conditional probability structure cannot only be restricted by assuming particular variables to be conditionally independent of other variables but also by specifying a system of logit models.

Suppose, for instance, that $Y_t$ depends on $Y_{t-1}$, $X$, and $Z_t$, but that there are no interaction effects. This assumption can be implemented by specifying logit models for the probabilities $\pi_{y_1|xz_1}$, $\pi_{y_2|xy_1z_2}$, and $\pi_{y_3|xy_2z_3}$ appearing in Equation 10, i.e.,

$$\pi_{y_1|xz_1} = \frac{\exp\left(u_{y_1}^{Y_1} + u_{y_1x}^{Y_1X} + u_{y_1z_1}^{Y_1Z_1}\right)}{\sum_{y_1}\exp\left(u_{y_1}^{Y_1} + u_{y_1x}^{Y_1X} + u_{y_1z_1}^{Y_1Z_1}\right)}, \quad (11)$$

$$\pi_{y_2|xy_1z_2} = \frac{\exp\left(u_{y_2}^{Y_2} + u_{y_2x}^{Y_2X} + u_{y_2y_1}^{Y_2Y_1} + u_{y_2z_2}^{Y_2Z_2}\right)}{\sum_{y_2}\exp\left(u_{y_2}^{Y_2} + u_{y_2x}^{Y_2X} + u_{y_2y_1}^{Y_2Y_1} + u_{y_2z_2}^{Y_2Z_2}\right)}, \quad (12)$$

$$\pi_{y_3|xy_2z_3} = \frac{\exp\left(u_{y_3}^{Y_3} + u_{y_3x}^{Y_3X} + u_{y_3y_2}^{Y_3Y_2} + u_{y_3z_3}^{Y_3Z_3}\right)}{\sum_{y_3}\exp\left(u_{y_3}^{Y_3} + u_{y_3x}^{Y_3X} + u_{y_3y_2}^{Y_3Y_2} + u_{y_3z_3}^{Y_3Z_3}\right)}, \quad (13)$$

where the $u$ parameters are log-linear parameters which are subject to the well-known ANOVA-like restrictions. Note that the model described in Equations 10-13 gives just one of the possible set of restrictions that can be imposed on the general model presented in Equation 8. It is also possible to specify models containing interaction effects, which relax the Markov assumption, which contain time-lagged effects of $Z$ on $Y$, and which contain direct effects of $Y$ on $Z$.

It is well known that logit models with categorical independent variables are equivalent to log-linear models in which an effect is included to fix the marginal distribution of the independent variables (Goodman, 1972; Agresti, 1990). For instance, the logit model given in Equation 12 is equivalent to the hierarchical log-linear model

$$\log m_{xy_1z_2y_2} = \alpha_{xy_1z_2} + u_{y_2}^{Y_2} + u_{y_2x}^{Y_2X} + u_{y_2y_1}^{Y_2Y_1} + u_{y_2z_2}^{Y_2Z_2}, \quad (14)$$

where $m_{xy_1z_2y_2}$ is an expected cell frequency in the marginal table formed by the variables $X$, $Y_1$, $Z_2$, and $Y_2$, and $\alpha_{xy_1z_2}$ is the parameter that fixes the marginal distribution of the independent variables. The probability $\pi_{y_2|xy_1z_2}$ can simply be obtained from $m_{xy_1z_2y_2}$ by

$$\pi_{y_2|xy_1z_2} = \frac{m_{xy_1z_2y_2}}{\sum_{y_2} m_{xy_1z_2y_2}}$$

Goodman (1973) presented his causal log-linear model by specifying log-linear models for different marginal tables, where every subsequent marginal table had to contain, apart from the dependent variable, all variables of the previous marginal table. More precisely, Goodman's approach involves restricting the general model in Equation 8 by specifying log-linear models for the marginal frequency tables with expected cells counts $m_x$, $m_{xz_1}$, $m_{xz_1y_1}$, $m_{xz_1y_1z_2}$, $m_{xz_1y_1z_2y_2}$, $m_{xz_1y_1z_2y_2z_3}$, and $m_{xz_1y_1z_2y_2z_3y_3}$. These marginal tables can be used to obtain the probabilities appearing at the right hand side of Equation 8. The way we specified the Markov model with covariates is slightly different from Goodman's original formulation of the causal log-linear model because the logit models were specified for the probabilities of the restricted model given in Equation 10 instead of the probabilities of the general model given in Equation 8. The advantage of our approach is that it is computationally more efficient as a result of a reduction of the dimensionality of the marginal tables involved in the analysis (Vermunt, 1994, 1995).

It will be clear that the causal log-linear model provides us with a flexible regression approach which overcomes the limitations of the multiple-group Markov model. However, in Goodman's causal log-linear models it is assumed that all variables are observed, while we are interested in regressing latent states on previous latent states, time-constant covariates, and time-varying covariates.

### 3.2 Causal log-linear models with latent variables

In the context of latent class analysis, models have been proposed which can be used to explain class membership by means of a number of observed covariates. Haberman (1979) parametrized the latent class model as a log-linear model with one or more latent variables. When using this log-linear latent class model it is straightforward to regress the probability of belonging to a particular latent class on a set of categorical covariates by means of a log-linear, or equivalently, a logit model. Dayton and Macready (1988) proposed latent class models with continuous concomitant variables, in which class membership was regressed on the covariates by means of a logistic regression model. Van der Heijden, Mooijaart and De Leeuw (1992) proposed a so-called latent budget model in which a categorical latent variable is explained by a joint independent variable using a logit model.

These strongly related extensions of the standard latent class model, which are all based on specifying a logit model for class membership, are, however, not very well suited to specify logit regression models for repeated observations. What we need here is a regression modeling approach which, like the above-mentioned latent class models, allows to regress a latent variable on a set of covariates, and, like the causal log-linear models discussed above, allows both the dependent variable and the covariates to change with time. Such a model can be obtained by combining Goodman's causal log-linear model with a latent class model. Hagenaars (1990, 1993) showed how to specify simultaneously a system of logit equations for a set of causally ordered latent and manifest variables and a latent class model for the latent variables which are used in the logit models (see also Vermunt, 1993, 1994, 1995). Because of the analogy with the well-known LISREL model for continuous data, he called this causal log-linear model with latent variables a modified Lisrel model. Below, it is shown that this causal log-linear model with latent variables makes it possible to include covariates into a latent Markov model.

Suppose that we have a Markov model for the latent states $W_t$ having the same structure as the manifest Markov model for $Y_t$ given in Equation 10. Moreover, assume that, like in the latent Markov model described in Equation 6, each $W_t$ has only one indicator, $A_t$. In that case, the probability structure of the causal log-linear model with latent variables $W_1$, $W_2$, and $W_3$ is

$$\pi_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3} = \pi_{xz_1z_2z_3}\ \pi_{w_1|xz_1}\ \pi_{a_1|w_1}\ \pi_{w_2|xw_1z_2}\ \pi_{a_2|w_2}\ \pi_{w_3|xw_2z_3}\ \pi_{a_3|w_3}\ . \quad (15)$$

In fact, the only difference with the manifest Markov model given in Equation 10 is that it contains, apart from a structural part, a measurement part in which the relationships between the latent states $W_t$ and the observed states $A_t$ are specified. This measurement part consist of a set of conditional response probabilities $\pi_{a_t|w_t}$. Note that, like in the manifest case, the structural part of the model given in Equation 15 is already a restricted model. In the most general model, the structural part of the model has the same structure as the model given in Equation 8. The measurement part is restricted as well since it is assumed that the relationship between $W_t$ and $A_t$ is independent of $X$, $W_{t-1}$ and $Z_t$. This assumption can easily be relaxed, namely by replacing $\pi_{a_t|w_t}$ by $\pi_{a_t|xw_{t-1}z_tw_t}$. When using such a general specification of the measurement part of the model, $\pi_{a_t|xw_{t-1}z_tw_t}$ has to restricted in some way to avoid identification problems. Note that although the measurement part of the model given in Equation 15 contains only one indicator per occasion, it is straightforward to specify models that, like the latent Markov model given in Equation 5, contain several indicators per occasion.

As mentioned in the discussion of the latent Markov model, when the model contains only one indicator per occasion, the response probabilities have to be assumed to be time-homogeneous, i.e.,

$$\pi_{a|w} = \pi_{a_1|w_1} = \pi_{a_2|w_2} = \pi_{a_3|w_3} \, . \tag{16}$$

Like in the manifest case, the probabilities of the structural part of the model may be parametrized by means of a logit model. For instance, if for the latent states $W_t$ we assume the same kind of model as for the observed states $Y_t$ (see Equations 11-13), $\pi_{w_1|xz_1}$, $\pi_{w_2|xw_1z_2}$, and $\pi_{w_3|xw_2z_3}$ have to be restricted as follows:

$$\pi_{w_1|xz_1} = \frac{\exp\left(u_{w_1}^{W_1} + u_{w_1x}^{W_1X} + u_{w_1z_1}^{W_1Z_1}\right)}{\sum_{w_1}\exp\left(u_{w_1}^{W_1} + u_{w_1x}^{W_1X} + u_{w_1z_1}^{W_1Z_1}\right)}, \tag{17}$$

$$\pi_{w_2|xw_1z_2} = \frac{\exp\left(u_{w_2}^{W_2} + u_{w_2x}^{W_2X} + u_{w_2w_1}^{W_2W_1} + u_{w_2z_2}^{W_2Z_2}\right)}{\sum_{w_2}\exp\left(u_{w_2}^{W_2} + u_{w_2x}^{W_2X} + u_{w_2w_1}^{W_2W_1} + u_{w_2z_2}^{W_2Z_2}\right)}, \tag{18}$$

$$\pi_{w_3|xw_2z_3} = \frac{\exp\left(u_{w_3}^{W_3} + u_{w_3x}^{W_3X} + u_{w_3w_2}^{W_3W_2} + u_{w_3z_3}^{W_3Z_3}\right)}{\sum_{w_3}\exp\left(u_{w_3}^{W_3} + u_{w_3x}^{W_3X} + u_{w_3w_2}^{W_3W_2} + u_{w_3z_3}^{W_3Z_3}\right)}. \tag{19}$$

Although for the sake of simplicity, only hierarchical log-linear models were presented, it is also possible to specify non-hierarchical log-linear models.

## 4 Estimation by means of the EM algorithm

Goodman (1974) showed how to estimate latent class models by means of the EM algorithm (Dempster, Laird and Rubin, 1977). This algorithm was implemented by Clogg (1977) in his MLLSA program. Poulsen (1982) was the first one who showed how to obtain maximum likelihood estimates for the parameters of the latent Markov model by means of the EM algorithm. More recently, Van de Pol, Langeheine and De Jong (1989) implemented this algorithm in their PANMARK program which can be used for estimating latent and mixed Markov models. Hagenaars and Luijkx' (1990) LCAG program, which can be used to estimate both standard latent class models and the causal log-linear model with latent variables discussed above, is based on the EM algorithm as well. More recently, Vermunt (1993) developed a program called $\ell EM$ for estimating causal log-linear models with latent variables. The $\ell EM$ program, which is based on the EM algorithm as well, is more efficient and can therefore handle much bigger problems than LCAG. Moreover, with LCAG only hierarchical log-linear models can be specified for the various marginal subtables, while with $\ell EM$ any type of log-linear model can be specified, including particular types of log-multiplicative models. Specifying the latent Markov models with time-constant and time-varying covariates is straightforward by means of $\ell EM$.

Assuming a multinomial sampling scheme, maximum likelihood estimates for the parameters of the extended latent Markov model described in Equations 15-19 have to be obtained by maximizing the following log-likelihood function:

$$\mathcal{L} = n_{xz_1a_1z_2a_2z_3a_3} \log \sum_{w_1,w_2,w_3} \pi_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3} , \tag{20}$$

where $n_{xz_1a_1z_2a_2z_3a_3}$ denotes an observed cell count in the cross-tabulation of the observed variables. The $n_{xz_1a_1z_2a_2z_3a_3}$ and the above log-likelihood function are sometimes also called the incomplete data and the incomplete data likelihood, respectively.

The EM algorithm (Dempster, Laird and Rubin, 1977) is a general iterative algorithm which can be used for estimating model parameters when there are missing data. In the case of the latent Markov models, the scores on the latent states $W_t$ are missing for all persons. The EM algorithm consists of two separate steps per iteration cycle: an E(xpectation) step and a M(aximization) step. In the E step of the algorithm, auxiliary estimates for the missing data are obtained using the incomplete data and the 'current' parameter estimates, that is, the parameter estimates from the previous EM iteration. For the model concerned, the E step involves

$$\hat{n}_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3} = n_{xz_1a_1z_2a_2z_3a_3} \hat{\pi}_{w_1w_2w_3|xz_1a_1z_2a_2z_3a_3} . \tag{21}$$

Here, $\hat{n}_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3}$ is an estimated cell frequency in the table including the latent dimensions, sometimes also called the completed data. Furthermore, $\hat{\pi}_{w_1w_2w_3|xz_1a_1z_2a_2z_3a_3}$ is the probability of having particular scores on the latent variables, given someone's scores on the observed variables, calculated using the parameter estimates from the last EM iteration.

The M step involves obtaining maximum likelihood estimates for the model parameters using the completed data as if it where observed data, that is, maximizing the complete data log-likelihood function

$$\mathcal{L}^* = n_{xz_1w_1a_1z_2w_2a_2z_3w_2a_3} \log \pi_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3} \, . \tag{22}$$

The simplest situation occurs when no further restrictions are imposed on the (conditional) probabilities appearing in the model for $\pi_{xz_1w_1a_1z_2w_2a_2z_3w_3a_3}$ described in Equation 15. In that case, maximum likelihood estimates of the model parameters can simply be obtained by

$$\begin{aligned}
\hat{\pi}_{xz_1z_2z_3} &= \frac{\hat{n}_{xz_1..z_2..z_3..}}{\hat{n}_{..........}} , \\
\hat{\pi}_{w_1|xz_1} &= \frac{\hat{n}_{xz_1w_1.......}}{\hat{n}_{xz_1........}} , \\
\hat{\pi}_{a_1|w_1} &= \frac{\hat{n}_{..w_1a_1......}}{\hat{n}_{..w_1.......}} , \\
\hat{\pi}_{w_2|xw_1z_2} &= \frac{\hat{n}_{x.w_1.z_2w_2....}}{\hat{n}_{x.w_1.z_2.....}} , \\
\hat{\pi}_{a_2|w_2} &= \frac{\hat{n}_{.....w_2a_2...}}{\hat{n}_{.....w_2....}} , \\
\hat{\pi}_{w_3|xw_2z_3} &= \frac{\hat{n}_{x....w_2.z_3w_3.}}{\hat{n}_{x....w_2.z_3..}} , \\
\hat{\pi}_{a_3|w_3} &= \frac{\hat{n}_{........w_3a_3}}{\hat{n}_{........w_3.}} ,
\end{aligned}$$

where a '.' means that the table with estimated observed frequencies is collapsed over the dimensions concerned.

Particular (conditional) probabilities can be made equal to each other by means of a simple procedure proposed by Goodman (1974). For instance, the restrictions on the response probabilities which are described in Equation 16 can be imposed by

$$\hat{\pi}_{a|w} = \frac{\hat{n}_{..w_1a_1......} + \hat{n}_{.....w_2a_2...} + \hat{n}_{........w_3a_3}}{\hat{n}_{..w_1.......} + \hat{n}_{.....w_2....} + \hat{n}_{........w_3.}} .$$

What is actually done is calculating a weighted average of the unrestricted estimates of the response probabilities. It must be noted that, as demonstrated by Mooijaart and

Van der Heijden (1992), this simple procedure for imposing equality restrictions among conditional probabilities does not always work properly because it does not guarantee that in all situations the probabilities still sum to unity after imposing the equality restrictions (see also Vermunt, 1995). However, in this case, Goodman's procedure, which is also implemented in the above-mentioned MLLSA, PANMARK, and LCAG programs, works properly.

When logit models are specified for particular conditional probabilities, the M step is a bit more complicated. The probabilities $\hat{\pi}_{w_1|xz_1}$, $\hat{\pi}_{w_2|xw_1z_2}$, and $\hat{\pi}_{w_3|xw_2z_3}$, which are restricted as described in Equations 17-19, can be obtained by estimating the log-linear models concerned for the marginal tables with estimated cell counts $\hat{m}_{xz_1w_1}$, $\hat{m}_{xw_1z_2w_2}$, and $\hat{m}_{xw_2z_3w_3}$, respectively. For that purpose, standard algorithms for obtaining maximum likelihood estimates of the parameters of log-linear models can be applied such as the Iterative Proportional Fitting Algorithm (IPF) and the Newton-Raphson algorithm (Goodman, 1973; Hagenaars, 1990; Vermunt, 1993, 1995).

In the $\ell EM$ program (Vermunt, 1993), hierarchical log-linear models are estimated by IPF and non-hierarchical log-linear models by a variant of the one-dimensional Newton algorithm as proposed by Goodman (1979). The latter algorithm differs from the well known Newton-Raphson algorithm in that, like in IPF, parameters are updated subsequently instead of updating them simultaneously (Vermunt, 1995). Therefore, the algorithm implemented in the $\ell EM$ program is actually an ECM algorithm (Meng and Rubin, 1993).

# 5 Application

## 5.1 Data

The data which are used to illustrate the extended latent Markov model presented in the previous sections are taken from a German educational panel study among secondary school pupils. In this panel study by the Institute for Science Education in Kiel, a cohort of pupils was followed during their school career and interviewed once a year with respect to several themes, such as their school grades and their interest in physics and technology.

In the application, the variable interest in physics measured at three points in time is used as the dependent variable. The observed variable interest in physics at $T = t$

Table 1: Test results for the estimated models

| | Model | $X^2$ | $L^2$ | $df$ | $p(X^2)$ | $p(L^2)$ |
|---|---|---|---|---|---|---|
| 1. | Basic (Equations 15-19) | 139.45 | 142.94 | 99 | .005 | .003 |
| 2. | Basic + $u^{W_1W_3}_{w_1w_3}$ | 118.35 | 127.27 | 98 | .079 | .025 |
| 3. | Basic + $u^{W_1Z_2W_2}_{w_1z_2w_2} + u^{W_2Z_3W_3}_{w_2z_3w_3}$ | 141.35 | 140.69 | 97 | .002 | .003 |
| 4. | Basic + $u^{XW_1W_2}_{xw_1w_2} + u^{XW_2W_3}_{xw_2w_3}$ | 137.99 | 142.69 | 97 | .004 | .002 |
| 5. | Basic + $u^{Z_1W_2}_{z_1w_2} + u^{Z_2W_3}_{z_2w_3}$ | 140.08 | 142.52 | 97 | .003 | .002 |
| 6. | Basic + $u^{W_1Z_2}_{w_1z_2} + u^{W_2Z_3}_{w_2z_3}$ | 119.28 | 123.08 | 97 | .062 | .038 |
| 7. | Basic + $u^{W_1W_3}_{w_1w_3} + u^{W_1Z_2}_{w_1z_2} + u^{W_2Z_3}_{w_2z_3}$ | 95.23 | 107.88 | 96 | .503 | .192 |
| 8. | 7 + time-homogeneous effects | 105.38 | 117.19 | 102 | .390 | .144 |

is denoted by $A_t$, while the latent variable interest in physics is denoted by $W_t$. Two covariates are used in the latent Markov models to be specified: the time-constant covariate sex, denoted by $X$, and the time-varying covariate grade in physics, denoted by $Z_t$. Since the time-varying covariate $Z_t$ represents a pupil's grade in physics at the end of the school year preceding the interview date, it can be assumed that $Z_t$ influences $W_t$. What we want to investigate is whether interest in physics at $T = t$ depends on interest in physics at $T = t - 1$, on sex, and on grade in physics at $T = t$.

The total sample size is 541. Because we wanted to avoid sparseness problems to be able to use the Pearson's chi-squared statistic and the likelihood-ratio chi-squared statistic to test the fit of the models to be estimated, the observed variables $A_t$ and $Z_t$ were dichotomized, with the categories 'low' and 'high'. The variable sex has categories 'girls' and 'boys'. The total number of cells in the observed table is $2^7$, 128.

The fact that the variables were dichotomized does not mean that these kinds of models cannot be used with polytomous variables. The problem is that model testing can become very difficult because of sparseness of the observed frequency table. Although in that case nested models can still be compared against each other by means of likelihood ratio tests, models cannot be tested anymore against the data (Haberman, 1977, 1978; Agresti, 1990).

### 5.2 Results

The test result for the models that were estimated by means of the $\ell_{EM}$ program are presented in Table 1. The model selection strategy we followed was starting from a plausible restricted model and subsequently adding parameters to see whether the fit could be improved. Model 1, which we called the basic model, is the model described in Equations 15-19. As already mentioned when presenting the causal log-linear model with latent variables, Model 1 is obtained by imposing some restrictions on the most general model that is possible. That is, it is assumed that someone's interest in physics at a particular point in time ($W_t$) depends only on the interest in physics at the previous occasion ($W_{t-1}$), on sex ($X$), and on the physics grade at the same point in time ($Z_t$), where there are only two-variable effects. In other words, it contains the Markov assumption, it assumes that there are no time-lagged effects, and it assumes that the effects of sex and grade are independent of the previous interest. Another assumption, which is necessary to make a single indicator latent Markov model identifiable, is that the measurement error is the same among time points. And finally, $Z_t$ is postulated not to be influenced by the preceding values of $W$. Below it is demonstrated how to relax some of these assumption.

As can be seen from the test results, Model 1 does not fit. This indicates that at least one of its underlying assumptions has to be rejected. In each of the Models 2-6, one of the above-mentioned assumptions is relaxed. Since Model 1 can be obtained by fixing one or two log-linear parameters of Models 2-6 to zero, conditional likelihood-ratio test between Model 1 and Models 2-6 can be used to test the significance of the additional parameters.

Model 2, which contains a direct effect of $W_1$ on $W_3$, fits significantly better than Model 1 ($\Delta L^2 = 15.67$, $df = 1$, $p < .001$). This means that the Markov assumption does not hold. Models 3 and 4 contain three-variable interactions among $Z_t$, $W_{t-1}$, and $W_t$ and among $X$, $W_{t-1}$, and $W_t$, respectively. The conditional tests of Models 3 and 4 against Model 1 show that neither of these interaction effects are significant: $\Delta L^2 = 2.25$, $df = 2$, $p = .345$, and $\Delta L^2 = .25$, $df = 2$, $p = .882$. This means that the effects of grade and sex on interest at $T = t$ do not depend on the interest at the previous occasion. Model 5, which contains time-lagged effects of $Z$ on $W$, does not fit better than Model 1 neither ($\Delta L^2 = .42$, $df = 2$, $p = .811$). And finally, Model 6 contains an effect of interest at $T = t - 1$ on grade at $T = t$. This model that relaxes the assumption that grade is not influenced directly by interest fits much better than Model 1: $\Delta L^2 = 19.86$, $df = 2$, $p < .001$.

Summarizing, both the Markov assumption and the assumption that $Z_t$ is not influenced by $W_{t-1}$ had to be rejected, while the no three-variable interaction assumptions and the no time-lagged effects assumption were confirmed. Model 7 contains the additional effects that were found to be significant, that is, the effects of $W_1$ on $W_3$, of $W_1$ on $Z_2$, and of $W_2$ on $Z_3$. As can be see from the test results reported in Table 1, this model fits the data very well: $L^2 = 107.88$, $df = 96$, $p < .192$.

Model 7 may still contain more parameters than necessary because so far we did not impose restrictions on the effects among time points. In Model 8, the effects of $W_{t-1}$ on $W_t$, the effects of $X$ on $W_t$, the effects of $Z_t$ on $W_t$, and the effect of $W_t$ on $Z_{t+1}$ are assumed to be time independent. These time-homogeneity restrictions do not deteriorate the fit significantly compared to Model 7: $\Delta L^2 = 9.31$, $df = 6$, $p < .157$.

Table 2 gives the parameter estimates for Models 7 and 8. The $\pi_{a|w}$ are the estimated parameters of the measurement part of the model. It can be seen that in both models, the estimated amount of measurement error is negligible since for $W_t = 1$, the probability that $A_t = 1$ equals 1.000, while for $W_t = 2$, the probability that $A_t = 2$ equals .969. To see whether the measurement error is significant, a model was estimated which is equivalent to Model 8 except for the fact that the response probabilities for a correct response were fix to be equal to zero. This model has an $L^2$ of 177.66 with 104 degrees of freedom. Note that since the parameters are fixed to be equal to their boundary values, it is not allowed to test this model against Model 8 by means of a likelihood-ratio test. Nevertheless, the rather similar $L^2$ values, 117.19 and 177.66, indicate that interest in physics is measured without error. However, it is implausible that the variable interest in physics is really measured without error. Although the results are not reported here, also a number of latent Markov models without covariates and with only sex as covariate were estimated using the same data set. In all these models, the probability of having the same value on an observed state as on a latent state was around .9 for both latent classes. Thus, what happens is that the inclusion of the time-varying covariate grade in physics decreases the amount of 'measurement error'. The reason for this is probably that in the models without $Z$, the 'measurement error' also captured unobserved heterogeneity in the states occupied at the different occasions which disappeared by including $Z$ as a covariate in the model. This indicates that in latent Markov models with a single indicator per occasion it is difficult to distinguish measurement error from unobserved heterogeneity. To detect measurement error it is preferable to use several indicators per occasion since in that case the relationships among the indicators provide information on the reliability of each of the indicators.

Table 2: Estimates of the most important parameters of Models 7 and 8

| Parameter | Model 7 | Model 8 |
|---|---|---|
| $\pi_{a\|w}$ | | |
| $\pi_{1\|1}$ | 1.000 | 1.000 |
| $\pi_{2\|2}$ | .969 | .969 |
| $\pi_{w_1\|xz_1}$ | | |
| $u_{11}^{XW_1}$ | .261 | .251 |
| $u_{11}^{Z_1W_1}$ | .274 | .294 |
| $\pi_{w_2\|xw_1z_2}$ | | |
| $u_{11}^{W_1W_2}$ | .423 | .471 |
| $u_{11}^{XW_2}$ | .320 | .251 |
| $u_{11}^{Z_2W_2}$ | .388 | .294 |
| $\pi_{w_3\|xw_1w_2z_3}$ | | |
| $u_{11}^{W_1W_3}$ | .293 | .305 |
| $u_{11}^{W_2W_3}$ | .551 | .471 |
| $u_{11}^{XW_3}$ | .142 | .251 |
| $u_{11}^{Z_3W_3}$ | .208 | .294 |
| $\pi_{z_2\|xz_1w_1}$ | | |
| $u_{11}^{W_1Z_2}$ | .162 | .180 |
| $\pi_{z_3\|xz_1z_2w_2}$ | | |
| $u_{11}^{W_2Z_3}$ | .202 | .180 |

For the structural part of the model, Table 2 reports the two-variable log-linear effects. Since both the independent variables and the dependent variable appearing in the various logit equations are dichotomous, these parameters are not difficult to interpret. The parameters indicate the effects of belonging to category 1 of the independent variable on the probability of belonging to category 1 of the dependent variable (see Equations 17-19). By taking twice the reported parameters, one obtains the effect for category 1 of the independent variable on the log odds of belonging to category 1 rather than category 2 of the dependent variable. And finally, 4 times the reported log-linear parameter gives the log odds ratio between categories 1 and 2 of the covariate concerned, within the levels of the other covariates.

The parameter estimates show that there is a strong dependence among the interest at subsequent points in time: persons who have a low interest have a high probability of remaining in the category low interest, while persons who have a high interest have a high probability of remaining in the category high interest. Also, the second-order Markov effect from $W_1$ on $W_3$ is quiet strong, and it works in the same direction. The fact that, controlling for $W_2$, $W_1$ has a positive effect on $W_3$ means that persons who moved to another state between $T = 1$ and $T = 2$ tend to move back to their position at $T = 1$ between $T = 2$ and $T = 3$. As can be expected, the effect of the time-varying covariate grade is positive as well, which means that pupils with higher grades are more interested in physics than pupils with lower grades. And finally, the effect of sex on the interest at the different points in time shows that girls are less interested in physics than boys.

Table 2 also reports the effects of $W_1$ on $Z_2$ and $W_2$ on $Z_3$. Note that although the parameters are not reported here, Models 7 and 8 also contain all interaction terms among $X$, $Z_1$, $Z_2$ and $Z_3$. The effects of $W_{t-1}$ on $Z_t$ indicate that interest has a positive effect on the grade at the next point in time. This means that interest at $T = t$ is not only influenced directly by interest at $T = t - 1$, but also indirectly via grade at $T = t$: Pupils who are more interested in physics get higher grades in physics and have therefore a high probability of remaining interested.

In summary, our analysis showed that the first-order Markov assumption does not hold for pupils' interest in physics, that there are time-homogeneous effects of the time-constant covariate sex and the time-varying covariate grade on interest in physics, and that there is an indirect relationship between interest in physics at subsequent points in time via grade in physics. Moreover, the estimated amount of measurement error in the observed states is negligible. Since it is implausible that interest in physics is really measured without error, this may be the result of the fact that only one indicator was used per occasion.

## 6 Discussion

In this paper, an extension of the latent Markov model was presented. It was shown how to specify parsimonious logit regression models for the latent states occupied at the different points in time, in which both time-constant and time-varying categorical covariates can be used as regressors. In fact, the extension, which is based on the use of the causal log-linear modeling approach presented by Goodman (1973) and Hagenaars (1990), leads to a model which is analogous to LISREL models for discrete-time continuous-state panel data.

The causal log-linear modeling framework, which was used to formulate the latent Markov model with covariates and which is implemented in the $\ell EM$ program, can be used to extend the model in several ways. One possible extension is to use more than one indicator per occasion together with a logit parameterization of the conditional response probabilities (Formann, 1992, Vermunt, 1995). This makes it possible to specify measurement models which are discrete approximations of latent trait models (Heinen, 1993, Vermunt and Georg, 1995). In the latent Markov model that was presented, it was assumed that only the dependent variable is subject to measurement error. However, the model can easily be extended to deal with measurement error in the covariates as well. Furthermore, like in the mixed Markov model, an unobserved time-constant covariate can be included in the model to correct for unobserved heterogeneity (Van de Pol and Langeheine, 1990; Vermunt, 1995). Another extension is to use also continuous time-constant covariates (Vermunt, 1995), but it must be noted that in that case the fit of a model cannot be tested anymore by means of chi-squared statistics. Although in latent Markov models it is generally assumed that the measurement error is not correlated among occasions, or, in other words, that the observed states are mutually independent given the joint latent variable, it is possible to specify models with direct effects between indicators (Bassi, Croon, Hagenaars, and Vermunt, 1995). And finally, the approach implemented in $\ell EM$ makes it possible to use partially observed data in the analysis and to specify a model for the mechanism causing the missing data (Vermunt, 1994, 1995).

There are two main limitations with respect to the use of latent Markov models. First, a general problem associated with the analysis of categorical data is that when sparse tables are analyzed, the theoretical $\chi^2$ approximation of the Pearson chi-squared statistic and the likelihood-ratio chi-squared statistic is poor. Although in such situations the significance of parameters can still be tested by means of conditional likelihood-ratio tests, the fit of a model cannot be assessed anymore (Haberman, 1977, 1978; Agresti, 1990). A possible solution for this problem is to use bootstrap procedures for model testing (Langeheine, Pannekoek, and Van de Pol, 1995).

A second limitation is that although much bigger problems can be dealt with than the application that was presented, the size of problems that can be handled with the current computer capacities is limited. In latent Markov models, the size of a problem depends mainly on the number of cells of the joint latent dimension since in the E step of the EM algorithm the contribution to the complete table has to be computed for each non-zero observed cell count. When the latent variables are dichotomous, depending on the internal memory of the computer that is used, the current working version of the $\ell EM$ can deal with eight to ten panel waves. But when each latent variable has five categories, three or four waves is the maximum. A possibility to deal with bigger problems may be the use pseudo-likelihood methods which do not use information on the joint distribution of all variables included in the analysis but only on some marginal distributions (Westers, 1993).

## References


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

Bartholomew, D.J. (1987). *Latent variables models and factor analysis.* London: Griffin.

Bassi, F., M. Croon, J. Hagenaars and J. Vermunt (1995). Estimating latent turnover tables when the data are affected by classification errors. *Statistica Neerlandica.*

Bishop, R.J., Fienberg, S.E., and Holland, P.W. (1975). *Discrete multivariate analysis: theory and practice.* Cambridge, Mass.: MIT Press.

Clogg, C.C. (1977). *Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users.* University Park: PA: Working Paper 1977, Populations Issues Research Center.

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

Clogg, C.C., Eliason, S.R., and Grego, J.M. (1990). Models for the analysis of change in discrete variables. A. Von Eye (ed.), *Statistical methods in longitudinal reseach*, Vol. II, 409-441. Boston: Academic Press.

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

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, Ser. B.*, 39, 1-38.

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

Goodman, L.A. (1972). A modified multiple regression approach for the analysis of dichotomous variables. *American Sociological Review*, 37, 28-46.

Goodman, L.A. (1973). The analysis of multidimensional contingenty tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, 60, 179-192.

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

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.

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

Haberman, S.J. (1978). *Analysis of quantitative data, Vol. 1, Introduction topics.* New York: Academic Press.

Haberman, S.J. (1979). *Analysis of qualitative data, Vol. 2, New developments.* New York: Academic Press.

Hagenaars, J.A. (1990). *Categorical longitudinal data - loglinear analysis of panel, trend and cohort data.*. Newbury Park: Sage.

Hagenaars, J.A. (1992). Exemplifying longitudinal loglinear analysis with latent variables. P.G.M. Van der Heijden, W. Jansen, B. Francis and G.U.H. Seeber (eds.), *Statistical modelling*, 105-120. Amsterdam: Elsevier Science Publishers B.V..

Hagenaars, J.A. (1993). *Loglinear models with latent variables.* Newbury Park: CA: Sage.

Hagenaars, J., and Luijkx, R. (1990). *LCAG: A program to estimate latent class models and other loglinear models with latent variables with and without missing data.* Tilburg: Working Paper #17, Tilburg University, Dep. of Sociology.

Heinen, A. (1993). *Discrete latent variables models.* Tilburg: Tilburg University Press.

Kelton, C.M.L., and Smith, M.A. (1991). Statistical inference in nonstationary Markov models with embedded explanatory variables. *Journal of Statistical Computing and Simulation*, 38, 25-44.

Langeheine, R., Pannekoek, J., and Van de Pol, F. (1995). Bootstrapping goodness-of-fit measures in categorical data analysis (in preparation). Voorburg/Kiel: NCBS/IPN.

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent structure analysis.* Boston: Houghton Mill.

Meng, X.L., and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.

Mooijaart, A., and Van der Heijden, P.G.M. (1992). The EM algorithm for latent class models with constraints. *Psychometrica*, 57, 261-271.

Muenz, L.R., and Rubinstein, L.V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics*, 41, 91-101.

Poulsen, C.A. (1982). *Latent structure analysis with choice modelling.* Aarhus: Aarhus School of Business Administratioon and Economics.

Spilerman, D. (1972). The analysis of mobility processes by the introduction of independent variables into a Markov chain. *American Sociological Review*, 37, 277-294.

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

Van de Pol, F., and Langeheine, R. (1990). Mixed Markov latent class models. C.C. Clogg (ed.). *Sociological Methodology 1990.* Oxford: Basil Blackwell.

Van de Pol, F., Langeheine, R., and De Jong, W. (1989). *PANMARK user manual: PAnel analysis using MARKov chains.* Voorburg: Netherlands Central Bureau of Statistics.

Van der Heijden, P.G.M, Mooijaart, A., and De Leeuw, J. (1992). Constraint latent budget analysis. *Sociological Methodology 1992.*

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.

Vermunt, J.K. (1994). *Causal log-linear modeling with missing data.* WORC PAPER 94.05.021/7, Tilburg University.

Vermunt, J.K. (1995). *Log-linear event history analysis; a general approach with missing data, latent variables, and unobserved heterogeneity.* Tilburg: Tilburg University Press.

Vermunt, J.K, and W. Georg (1995). Analyzing categorical panel data by means of causal log-linear models with latent variables: An application to the change in youth-centrism. *ZA-information.*

Westers, P. (1993). *The solution-error response-error model: A method for the examination of test item bias.* Doctoral dissertation, University of Twente, The Netherlands.

Wiggins, L.M. (1973). *Panel analysis.* Amsterdam: Elsevier.

Work and Organization Research Centre
Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands