Analyzing Categorical Panel Data by Means of Causal Log-linear Models with Latent Variables: An Application to the Change in Youth-Centrism

Jeroen K. Vermunt
Tilburg University, The Netherlands

Werner Georg
University of Konstanz, Germany

Jeroen Vermunt
Department of Methodology and Statistics
Tilburg University
PO Box 90153
5000 LE Tilburg
The Netherlands.
E-mail: J.K.Vermunt@KUB.NL
phone: +3113 4662748

This paper was written in 1995/1996. The German version was published by ZA Infomation. The English version was submitted but not revised.

Abstract

This paper presents a general approach to the analysis of categorical panel data which is based on using causal log-linear models with latent variables. Like the well-known LISREL model, these models consist of a structural and a measurement part. In the structural part, a system of logit equations is used to explain changes which occur in the dependent variable of interest. An unrestricted or restricted latent class model is used in the measurement part of the model. It is demonstrated that the measurement model can be used to specify discrete variants of latent trait models, such as the Rasch model and the Lord-Birnbaum model.

The approach is illustrated by means of an application on youth-centrism. Several measurement models are tested for a scale which is assumed to measure youth-centrism. In addition, the influence of covariates on a person's initial position and on the transition probabilities between time points is studied.

Analyzing Categorical Panel Data by Means of Causal Log-linear Models with Latent Variables: An Application to the Change in Youth-Centrism

Panel data is, together with event history data, the best suited kind of data for detecting determinants of social change. In this paper, we will present a general approach to the analysis of categorical panel data. The approach, which was originally proposed by Hagenaars (1990, 1993), is similar to the well-known LISREL model for continuous variables (Jöreskog and Sörbom, 1988). Like LISREL, it combines a path model in which the relationships among the structural variables are specified with a measurement model for the structural variables which are latent or measured indirectly. Because of the analogy with the LISREL model, Hagenaars called the model a modified LISREL approach.

The structural part of the modified LISREL model is a modified path model proposed by Goodman (1973). A modified path model is a log-linear model in which an a priori causal order can be imposed on a set of categorical variables. Such a causal log-linear model consists of a system of logit equations in which a variable that appears as a response variable in one logit equation can be used as an explanatory variable in a logit equation for one of its posterior variables.

The measurement part of the modified LISREL model is a latent class model, a factor-analytic model for categorical data (Lazarsfeld and Henry, 1968; Goodman, 1974; Haberman, 1979). Latent class models can be used to correct for measurement error in categorical observed variables or, in other words, to construct measurement models with categorical indicators. Although unrestricted latent class models treat the latent variable and its indicators as nominal variables, several kinds of restricted latent class models have been developed which allow researchers to impose restrictions on the log-linear parameters describing the relationship between the latent variable and its indicators (Rost and Georg, 1991; Formann, 1992; Heinen, 1996). Recently, Heinen (1996) demonstrated that when the latent distribution is discretized, most latent trait models can be formulated as restricted latent class models.

Hagenaars (1990, 1993) demonstrated how to combine the modified path model and the latent class model. More precisely, he showed how to specify a model for the structural relationships among observed and latent categorical variables by means of a causal log-linear model. When analyzing panel data, such a model can be used to separate true change from spurious change resulting from measurement error (Van de Pol and De Leeuw, 1986; Hagenaars, 1990, 1993). Vermunt (1996b) extended the model proposed by Hagenaars by using a much more general type of logit parameterization for the conditional probabilities appearing in the model.

The causal log-linear model with latent variables is strongly related to latent and mixed Markov models (Wiggins, 1973; Van de Pol and Langeheine, 1990). Actually, these models for the analysis of panel data are special cases of the approach to be presented in this paper (Vermunt, 1996b). However, when a latent or mixed Markov model is specified as a causal log-linear model, it is possible to impose logit restrictions on the conditional probabilities. For the structural part of the model, this means that parsimonious logit regression models can be specified for the (latent) transitions. In the measurement part of the model, the logit models can be used, for instance, to specify discrete approximations of latent trait models, such as the semi-parametric Rasch model (Lindsay, Clogg, and Grego, 1991; Heinen, 1996).

Section I presents the modified path model. Unrestricted and restricted latent class models and the modified LISREL model are discussed in Section II. An application using two-wave panel data on youth-centrism from a subsample of the 1992 Shell Youth Survey is presented in Section III.

I. Modified path models

This section presents a path-analytic extension of the well-known logit model. Goodman (1973) called this log-linear model which takes a priori information on the causal ordering of the variables into account a modified path analysis approach. As is demonstrated below, the modified path model is very well suited for analyzing categorical panel data. Note that the term categorical is not synonymous to nominal: Ordinal variables and discrete interval variables are categorical variables too (Agresti, 1990).

Specifying the probability structure

Suppose that we have data from a three-wave panel study, and that we want to explain individual transitions in a particular categorical variable. Let W, Y, and Z denote the dependent variable at the first, second, and third point in time. Let R, S, and T indicate three categorical independent variables which are used to explain the value of W, the transitions between W and Y, and the transitions between Y and Z. Thus, the variables R, S, and T are exogenous variables, while W, Y, and Z are endogenous, where Y is assumed to be posterior to W, and Z is assumed to be posterior to Y. For the moment, we will assume that all variables can be observed directly.

Let π_{rstwyz} denote the probability that $R = r$, $S = s$, $T = t$, $W = w$, $Y = y$, and $Z = z$. Using the a priori information on the causal order among the variables, π_{rstwyz} can be written as

$$\pi_{rstwyz} = \pi_{rst}\,\pi_{w|rst}\,\pi_{y|rstw}\,\pi_{z|rstwy}\,. \tag{1}$$

Thus, the information on the causal ordering of the variables is used to decompose the joint probability into a product of marginal and conditional probabilities (Goodman, 1973). This is a straightforward way to express that the value on a particular variable can only depend on the preceding variables but not on the posterior ones. For instance, Y is assumed to depend only on the preceding variables R, S, T, and W, but not on the posterior variable Z. Therefore, the probability that $Y = y$ depends only on the values of R, S, T, and W, and not on the value of Z. Note that the model given in Eq. (1) is a recursive model. Although non-recursive models for categorical data, which have recently been proposed by Mare and Winship (1991), can also be handled within our approach, here we will restrict ourselves to recursive models.

Decomposing the joint probabilities into a set of marginal and conditional probabilities is only the first step in describing the causal relationships among the variables under study. Generally, we also want to reduce the number of parameters in some way, while the right-hand side of Eq. (1) contains as many unknown (conditional) probabilities as observed cell frequencies. In other words, the model given in Eq. (1) is a saturated model in which it is assumed that a particular dependent variable is influenced by all its posterior variables, including all their interactions terms.

The simplest way to specify more parsimonious models is to restrict directly the conditional probabilities appearing in Eq. (1). Suppose that W depends on R and S, but not on T, that Y depends on S, T, and W, but not on R, and that Z depends on S, T, and Y, but not on R. In other words, the initial position depends on R and S and the transition probabilities depend on S and T. These restrictions can simply be implemented by replacing the unrestricted model given in Eq. (1) by

$$\pi_{rstwyz} = \pi_{rst} \, \pi_{w|rs} \, \pi_{y|stw} \, \pi_{z|sty} \; . \qquad (2)$$

Application of this type of conditional independence assumptions is the simplest procedure for specifying more parsimonious models. These types of restrictions are also used in, for instance, discrete-time Markov models (Bishop, Fienberg and Holland, 1975; Van de Pol and Langeheine, 1990). Note that the model described in Eq. (2) is a Markov model as well.

The above-mentioned rather simple procedure for obtaining more restricted models has, however, one important disadvantage: the dependent variable must always be related to the joint independent variable. For instance, the variable Y depends on the joint variable STW. That is, for each combination of S, T, and W, there is a parameter describing the probability that $Y = y$. Thus, when a particular variable is thought to influence the dependent variable concerned, all interactions with the other independent variables are automatically included in the model as well. The result will generally be that the model contains more parameters than necessary.

Logit parameterization of the probabilities

By using a log-linear or logit parameterization of the marginal and conditional probabilities appearing in Eq. (1), it is possible to specify and test more parsimonious causal models for categorical data. This leads to what Goodman called a modified path analysis approach (Goodman, 1973). A causal model for the relationships among the variables used in the example would consist of four modified path steps or submodels: a saturated log-linear model for the exogenous variables R, S, and T, and three logit models in which W, Y, and Z appear as dependent variables.

The logit model in which Y appears as a response variable could, for instance, be of the form

$$\pi_{y|rstw} = \frac{\exp\left(u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY}\right)}{\sum_y \exp\left(u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY}\right)},$$

where the u terms are log-linear parameters which are subject to the well-known ANOVA-like restrictions. The logit model for Y contains the two-variable interaction terms of S, T, and W and Y, and the three-variable interaction term of S, W, and Y. The same kinds of logit models can be specified for the other two conditional probabilities appearing in Eq. (1), $\pi_{w|rst}$ and $\pi_{z|rstwy}$. It will be clear that such a system of logit equations makes it possible to specify more parsimonious models than with the simple restrictions on conditional probabilities applied in Eq. (2).

Specifying a logit model for a set of conditional probabilities is equivalent to specifying a log-linear model for a frequency table where the marginal distribution of the independent variables is treated as fixed (Goodman, 1972; Haberman, 1979; Agresti, 1990). For instance, the logit model we specified for $\pi_{y|rstw}$ is equivalent to log-linear model $\{RSTW, TY, SWY\}$ for (marginal) frequency table $RSTWY$, or

$$\log m_{rstwy} = \alpha_{rstw}^{RSTW} + u_y^Y + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY}. \qquad (3)$$

where m_{rstwy} denotes an expected frequency in the marginal table $RSTWY$, and α_{rstw}^{RSTW} denotes the effect which fixes the marginal distribution of the dependent variables.

Thus, specifying a causal log-linear model for a set of categorical variables can simply be accomplished by specifying separate log-linear models for different marginal tables, or subtables. The marginal tables are formed by the variables used in the previous marginal table and the dependent variable concerned (Goodman, 1973). In this case, log-linear models must be specified for tables RST, $RSTW$, $RSTWY$, and $RSTWYZ$.

Note that the probabilities in Eq. (1) can be obtained from the expected frequencies of the subtables. For instance,

$$\pi_{y|rstw} \quad = \quad \frac{m_{rstwy}}{\sum_y m_{rstwy}} \; . \tag{4}$$

Modifications of Goodman's approach

Above, it was shown that the conditional probabilities of a modified path model can be restricted either by means of simple restrictions on the probabilities or by means of a logit parameterization. But actually, it is simpler and computationally more efficient to combine these two ways of restricting the conditional probabilities. More precisely, first we can restrict the model in the way we did in Eq. (2), and then we can restrict the conditional probabilities appearing in this equation with a logit parameterization. This leads to a (small) modification of the procedure proposed by Goodman (1973).

Let us look again at the model for dependent variable Y. Because Y does not depend on R, we can replace $\pi_{y|rstw}$ by $\pi_{y|stw}$. Therefore, the log-linear restrictions imposed on $\pi_{y|rstw}$ can also be imposed directly on $\pi_{y|stw}$. This is equivalent to specifying the log-linear model $\{STW, TY, SWY\}$ for marginal table $STWY$, that is, Eq. (3) can be replaced by

$$\log m_{stwy} \quad = \quad \alpha_{stw}^{STW} + u_y^{Y} + u_{sy}^{SY} + u_{ty}^{TY} + u_{wy}^{WY} + u_{swy}^{SWY} \; ,$$

where m_{stwy} denotes an expected frequencies in the marginal table $STWY$. By imposing restrictions in two steps, the log-linear parameters are estimated in the marginal table which includes only the dependent variables which are really used. This procedure may dramatically reduce the size of a problem, and is therefore computationally more efficient. Moreover, this two-step procedure has another important advantage: it prevents fitted zeroes when the observed table contains zeroes in the margin $RSTW$, but not in the margin STW.

So far, only simple hierarchical log-linear models have been used to parameterize the conditional probabilities. Furthermore, we have assumed that all variables included in the modified path model are categorical. Actually, these two features are in agreement with the way Goodman presented his modified path model. However, by using a more general type of logit model to parameterize the conditional probabilities, it is possible to specify non-hierarchical log-linear models, and to use continuous exogenous variables in the modified path model. Suppose that k is the index of the dependent variable in a particular logit equation, and that i denotes the index of the joint independent variable.

In this case, the logit model may be of the general form,

$$\pi_{k|i} \quad = \quad \frac{\exp\left(\sum_j u_j x_{ikj}\right)}{\sum_k \exp\left(\sum_j u_j x_{ikj}\right)}\,, \tag{5}$$

where u_j is a log-linear parameter, and x_{ikj} is an element of the design matrix. This logit model is equivalent to the multinomial response model proposed by Haberman (1979). When the index i is used to denote a particular individual instead of a level of the joint independent variable, the model given in Eq. (5) becomes a logistic regression model (Agresti, 1990). In that case, it can used with continuous independent variables, where x_{ikj} denotes the value of person i on variable j for level k of the response variable.

Estimation and testing

Maximum likelihood estimates for the log-linear parameters and the expected frequencies in the various subtables can be obtained by means of standard programs for log-linear analysis. Actually, the various submodels can be estimated separately, and the estimated cell probabilities for the overall model can be computed by means of (4) and (1). Model testing can be performed, for instance, by means of the likelihood-ratio chi-squared statistic L^2. The submodel-specific L^2 values and degrees of freedom can be added to obtain a test of the overall model fit. It must be noted that when continuous exogenous variables are used in the model, the L^2 statistic cannot be used anymore. Like in logistic regression models, we can use likelihood-ratio tests only to compare the fit of nested models.

A program called ℓEM has been developed to estimate the log-linear path models discussed in this section without the necessity to set up the different marginal tables (Vermunt, 1993). In ℓEM, specifying a log-linear path model is the standard way of modeling an observed frequency table. The experimental version of the program (ℓEM 0.11) uses the procedure proposed by Goodman (1973). But, in the most recent working version of ℓEM, the more efficient two-step procedure for restricting the probabilities which was presented above is implemented (Vermunt, 1996b).

The standard estimation procedure implemented in ℓEM to estimate the hierarchical log-linear models discussed above is the iterative proportional fitting algorithm (IPF). However, ℓEM can also be used to estimate more complex log-linear or logit models of the form given in Eq. (5). This is accomplished by allowing the user to specify a design matrix for particular log-linear effects. In ℓEM, it is also possible to use log-multiplicative effects, such as the type II association models developed by Goodman (Goodman, 1979; Clogg, 1982; Xie, 1992). Non-hierarchical log-linear models are estimated by means of the variant of the uni-dimensional Newton algorithm proposed by Goodman (1979).

This algorithm differs from the well-known Newton-Raphson algorithm in that only one parameter is updated at a time instead of updating all parameters simultaneously (Vermunt, 1996b).

Other options which are implemented in the current working version of ℓEM are the possibility to impose equality restrictions on log-linear parameters appearing in different logit equations and the possibility to impose fixed-value, equality, and inequality restrictions on the conditional probabilities.

II. Models with latent variables

So far, it has been assumed that all variables used in the causal log-linear model can be directly observed. However, since in social sciences many concepts are difficult or impossible to measure directly, often we use several directly observable variables, or indicators, as indirect measures for the concepts we want to measure. The value on an indicator is assumed to be determined by the unobservable value of the underlying variable we are interested in. In latent structure models, this principle is implemented statistically by the assumption of local independence. This means that the indicators are assumed to be independent of each other within the levels of the unobserved or latent variable. In other words, they are only correlated because of their common cause.

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, 1996). 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).

Unrestricted latent class models

Suppose there is 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. The basic equations of the latent class model are

$$\pi_{abc} \quad = \quad \sum_{w=1}^{W^*} \pi_{wabc}\,, \tag{6}$$

where

$$\pi_{wabc} \quad = \quad \pi_w \, \pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \tag{7}$$

Here, π_{wabc} denotes a probability in the joint distribution including the latent dimension W. Furthermore, π_w is the proportion of the population belonging to latent class w. The other π parameters appearing in Eq. (7) are conditional response probabilities. For instance, $\pi_{a|w}$ is the probability of having value a on A given that one belongs to latent class w.

From Eq. (6), 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 proposed by Lazarsfeld and Henry (1968) and used by Goodman (1974) is described in Eq. (7). 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. Note that Eq. (7) is very similar to the modified path model discussed in the previous section. Actually, it is a modified path model in which one variable is unobserved.

Haberman (1979) demonstrated that the latent class model can also be formulated as a log-linear model with latent variables. For instance, the latent class model given in Eq. (7) is equivalent to the hierarchical log-linear model $\{WA, WB, WC\}$. Thus, it can also be written as

$$\log m_{wabc} \quad = \quad u + u_w^W + u_a^A + u_b^B + u_c^C + u_{wa}^{WA} + u_{wb}^{WB} + u_{wc}^{WC} \,, \tag{8}$$

where $m_{wabc} = \pi_{wabc}N$. Eq. (8) contains, apart from the overall mean and the one-variable terms, only the two-variable interaction terms between the latent variable W and the manifest variables. The fact that no interactions between the manifest variables are included indicates that they are assumed to be conditionally independent.

The relationship among the parameters of the two parameterizations of the latent class model, that is, between the conditional probabilities of Eq. (7) and the log-linear parameters of Eq. (8) is (Haberman, 1979; Heinen, 1996):

$$\pi_{a|w} \quad = \quad \frac{\exp\left(u_a^A + u_{wa}^{WA}\right)}{\sum_a \exp\left(u_a^A + u_{wa}^{WA}\right)}. \tag{9}$$

Note that this is the same kind of logit parameterization of a conditional probability as we used in the modified path model. Formann (1992) used this parameterization in this linear logistic latent class model.

Restricted latent class models

Restricted latent class models may be specified either by imposing restrictions on the probabilities of the classical latent class model or by imposing restrictions on the log-linear parameters of the log-linear latent class model. Typical constraints applied in the classical latent class model are fixed-value and equality restrictions on the latent and conditional response probabilities (Goodman, 1974). More recently, Croon (1990) proposed an ordinal latent class model with inequality constrains on the conditional response probabilities. In the log-linear latent class model it is possible to impose all kinds of linear constraints on the log-linear parameters. Here, we will focus on the latter type of restrictions.

Recently, Heinen (1996) demonstrated that when the latent variable is discretized, most latent trait models can be formulated as latent class models by restricting the two-variable interactions in a similar manner as in the linear-by-linear, row, column, and row-column association models proposed by Goodman (1979) and Clogg (1982). In other words, the log-linear latent class model can be used to specify discretized variants of well-known latent trait models, such as the Rasch model (Rasch, 1960), the Lord-Birnbaum model (Lord and Novick, 1968), the Nomimal Response model (Bock, 1972), and the Partial Credit model (Masters, 1982).

Suppose we want to construct a Rasch scale using three dichotomous items A, B, and C. The basic assumption of the Rasch model is that all items have the same discrimination parameter or, in other words, that the item characteristic curves are parallel (Rasch, 1960; Rost and Georg, 1991). The probability of a 'correct' answer is postulated to depend only on a person's ability and on the difficulty of the item concerned. According to Heinen (1996), a discrete variant of the Rasch model can be obtained by imposing particular kinds of restrictions on the two-variable interactions terms appearing in the conditional response probabilities of the latent class model (see Eq (9)). The only difference between the standard Rasch model and the discrete Rasch model is that in the latter the latent ability distribution is discretized. More precisely, the number of different ability levels equals the number of latent classes. A discrete Rasch model can be obtained by restricting

$$\begin{aligned} u_{wa}^{WA} &= \theta_w \, \alpha \, x_a \, , \\ u_{wb}^{WB} &= \theta_w \, \alpha \, x_b \, , \\ u_{wc}^{WC} &= \theta_w \, \alpha \, x_c \, . \end{aligned} \tag{10}$$

The parameter α is the discrimination parameter, which is assumed to be equal for all items. Furthermore, x_a, x_b, and x_c are the scores for the categories of the items, and θ_w denotes the score for category w of W. The category scores of the items are fixed quantities. The standard scoring is 0 for the 'incorrect' answer and 1 to the 'correct' answer. However, if one wants to preserve the ANOVA-like restrictions on the two-variable interactions terms, one may also score the two categories of A, B, and C as 1 and -1. The scores for the latent variable, which are sometimes also called the latent nodes, can be either fixed or random quantities. The model with random nodes is equivalent to the semi-parametric Rasch model proposed by Lindsay, Clogg, and Grego (1991). It should be noted that when the latent scores are fixed and equidistant, the interaction terms in (10) have the form of linear-by-linear interactions. With random latent nodes, they have the form of row-association terms, where W operates as the row variable.

A discretized Lord-Birnbaum model is obtained by the following set of restrictions:

$$\begin{aligned} u_{wa}^{WA} &= \theta_w \, \alpha^A \, x_a \, , \\ u_{wb}^{WB} &= \theta_w \, \alpha^B \, x_b \, , \\ u_{wc}^{WC} &= \theta_w \, \alpha^C \, x_c \, . \end{aligned} \tag{11}$$

The only difference with the Rasch model is that now the discrimination parameters are item specific. In other words, apart from the item difficulty parameters, which depend mainly on the one-variable effects, the Lord-Birnbaum model contains item-specific parameters indicating the strength of the association between the latent variable and the item concerned. With fixed and equidistant values for the θ_w's, the two-variable effects have the form of a linear-by-linear interaction. On the other hand, with random scores for W, the two-variable interactions have multiplicative structures as in the log-multiplicative row-column association models proposed by Goodman (1979) and Clogg (1982).

The most general item response model for polytomous items is the Nominal Response model. Assuming that A, B, and C are polytomous, a discrete variant of the Nominal Response model can be obtained by

$$\begin{aligned} u_{wa}^{WA} &= \theta_w \, \alpha_a^A \, , \\ u_{wb}^{WB} &= \theta_w \, \alpha_b^B \, , \\ u_{wc}^{WC} &= \theta_w \, \alpha_c^C \, . . \end{aligned} \tag{12}$$

Thus, the Nominal Response model contains one association parameter for each item category. It can be seen that the latent variable is assumed to be an interval level variable, while the items are assumed to be measured on a nominal level. With fixed and equidistant latent scores, the interactions are column associations, where the latent variable acts as the row variable. If the scores for the latent variable are parameters to be estimated, we obtain a model with log-multiplicative row-column association terms.

A more restrictive latent trait model for polytomous items is the Partial Credit model. Like the Rasch model, the Partial Credit model is obtained by imposing the restrictions given in (10) on the two-variable interaction terms, where now the items may have more than two categories. Note that when the items are polytomous, it is also possible to specify a measurement model with the restrictions described in (11). This yields a variant of the Partial Credit model with item-specific discrimination parameters.

Modified LISREL models

Several extensions of the standard latent class model have been proposed, such as models for more than one latent variable (Goodman, 1974; Haberman 1979), models with external variables (Clogg, 1981), models for multiple-group analysis (Clogg and Goodman, 1984; McCutcheon, 1987), and local dependence models (Hagenaars, 1988). A limitation of these extensions is, however, that they are all developed within the framework of either the classical or the log-linear latent class model. As a result, it is not always possible to postulate the wanted a priori causal order among the structural variables incorporated the model. This subsection demonstrates that this problem can be resolved by using the general formulation of the modified path model with latent variables proposed by Hagenaars (1990, 1993) and extended by Vermunt (1996b). This model combines a structural model with a measurement model for the latent variables, that is, a modified path model with a latent class model. Because of the analogy with the LISREL model for continuous variables (Jöreskog and Sörbom, 1988), Hagenaars called it a modified LISREL approach.

Suppose that the endogenous variables in the causal model given in Eq. (1), W, Y, and Z, are latent variables, and that each of them is measured indirectly using three observed variables. The indicators for W are denoted by A, B, and C, for Y by D, E, and F, and for Z by G, H, and I. As we saw above, the latent variables can be related to their indicators by means of latent class models. Specifying the same model as in Eq. (1), but now assuming the endogenous variables to be latent, yields the following modified LISREL model:

$$\begin{aligned}\pi_{abcdefghirstwyz} &= \pi_{rst}\,\pi_{w|rst}\,\pi_{y|rstw}\,\pi_{z|rstwy}\,\pi_{a|w}\,\pi_{b|w}\,\pi_{c|w}\,\pi_{d|y}\,\pi_{e|y}\,\pi_{f|y} \\ &\quad \pi_{g|z}\,\pi_{h|z}\,\pi_{i|z}\,. \end{aligned} \tag{13}$$

Like in a modified path model, all conditional probabilities appearing in this equation can be restricted with a logit parameterization. Although in Eq. (13) it is implicitly assumed that the measurement models for W, Y, and Z do not depend on R, S, and T, it is not a problem to relax this assumption. For example, by replacing $\pi_{d|y}$ by $\pi_{d|ry}$, it can be tested whether R influences the relationship between Y and D.

Thus, including latent variables in a modified path model only involves specifying a number of additional modified path steps in which the relationships among the latent variables and their indicators are specified. These additional modified path steps have the same structure as the latent class model described in (7).

In the model that was used to illustrate the modified LISREL approach, the latent variables were used as indirect measures for some variables which cannot be observed directly. Actually, we specified a kind of latent Markov model with covariates. However, the possibility to include latent variables in the model can also be used to specify all kinds of finite mixture models (Titterington, Smith and Makov, 1985), such as mixed Markov models (Van de Pol and Langeheine, 1990), mixed Rasch models (Rost, 1990) and mixed logit models (Formann, 1992).

Estimation

Obtaining maximum likelihood estimates of the parameters of log-linear models with latent variables is slightly more complicated than for log-linear models in which all variables are observed. Maximum likelihood estimation can be performed, for instance, by means of the EM algorithm (Dempster, Laird and Rubin, 1977), a general iterative algorithm to estimate models with missing data. The EM algorithm consists of two separate steps per iteration cycle: an E(xpectation) step and a M(aximization) step.

The E step of the EM algorithm involves estimating the missing data. In our case, we must obtain estimates for the unobserved frequencies of the complete table $ABCDEFGHIRSTWYZ$, denoted by $\hat{n}_{abcdefghirstwyz}$, conditional on the observed data and the parameter estimates from the last EM iteration. This is accomplished using the observed incomplete data and the parameter estimates from the last iteration by

$$\hat{n}_{abcdefghirstwyz} \quad = \quad n_{abcdefghirst} \, \hat{\pi}_{wyz|abcdefghirst} \, . \tag{14}$$

Here, $n_{abcdefghirst}$ denotes an observed frequency, and $\hat{\pi}_{wyz|abcdefghirst}$ denotes the probability that $W = w$, $Y = y$, and $Z = z$, given the observed variables, evaluated using the 'current' parameter estimates.

In the M step, standard estimation procedures for log-linear models, such as IPF or Newton-Raphson, can be used to obtain improved parameter estimates using the completed data as if it were the observed data. In fact, the likelihood function in which $\hat{n}_{abcdefghirstwyz}$ appears as data, sometimes also referred to as the complete data likelihood, is maximized. The improved parameter estimates are used again in the E step to obtain new estimates for the complete table, and so on. The EM iterations continue until some convergence criterion is reached, for instance, a minimum increase in the likelihood function.

The ℓEM program (Vermunt, 1993) is especially developed for estimating modified path models with latent variables. Actually, in the model specification, latent and observed variables are treated in exactly the same way by the program. By means of ℓEM, it is also possible to use partially observed data in the analysis (Vermunt, 1996a, 1996b).

The algorithm used in ℓEM is a modified version of the original EM algorithm because the M step consists of only one iteration: The complete data likelihood is not maximized but only improved within a particular M step. This is a special case of the GEM algorithm which states that each increase in the complete data likelihood also leads to an increase of the incomplete data likelihood we actually want to maximize (Dempster, Laird and Rubin, 1977; Little and Rubin, 1987). In fact, the algorithm which is used in ℓEM is also a version of the ECM algorithm (Meng and Rubin, 1993). In the ECM algorithm, the M step is replaced by a conditional maximization (CM) step. Conditional maximization means that instead of improving all the parameters simultaneously, subsets of parameters are updated fixing the other ones at their previous values. This is just what is done by IPF and by the uni-dimensional Newton algorithm which are used in ℓEM. Meng and Rubin (1993) state that such simple and stable linear convergence methods are often more suitable for the M (or CM) step of the EM (or ECM) algorithm than superlinear converging, but less stable, algorithms like Newton-Raphson.

III. Application to the change in youth-centrism

The approach to the analysis of panel data that was presented in the previous sections is illustrated by means of an application on the change in youth-centrism. The data are taken from the 1992 Shell Youth Survey carried out in the summer of 1991, and from a second wave which took place between July and September 1993. In this follow up study, which is part of a research project financed by the German Research Foundation (DFG), a subsample of 288 persons was interviewed together with their parents.

Data

Background of the Shell Studies

The first Shell Study was carried out in 1953, with the aim to produce a yearly report on the situation of German youth (Zinnecker, 1985: pp. 409). The initial studies which were financed by the German Shell Foundation draw a picture of the post-war youth in West Germany (see for instance Schelsky, 1957). Since 1980, three Shell Studies have been conducted, each with its own special theme. In the 1981 Survey, attention was paid to the future perspectives and life-course orientations of young people. The 1985 Shell Study, surveyed in 1984, focussed on the comparison of contemporary youth with the generation of their parents. As a result of German unification, the 1992 Shell Study focussed on the comparison between the East and West German youth. In the latter study, data were collected by means of a quota sample with community size, age, educational attainment, and sex as quota criteria.

An important common theoretical construct of the last three Shell Studies is the concept of youth-centrism. The attitude 'Youth-Centrism' describes the orientation of young people to confine their own world and concept of life against the one of adults and to mistrust societal authorities which are determined by adults. In addition, young people persist in the right of having their own experiences, because the experiences of adults are not appropriate to solve their problems. This results in feelings of powerlessness and aggression against the adult world (Jugendwerk der Deutschen Shell, 1981: pp. 38). Five sub-scales were developed for the concept 'Youth-Centrism', referring to the following dimensions: (1) feeling of being discriminated against by social authorities, (2) acceptance of adult experience and privileges, (3) personal trust and gratitude towards adults and parents, (4) alienation and independence from adults, and (5) assessment of adults as powerful but lacking understanding.

In the 1985 Shell Survey, the two 'best' items (according to the discrimination parameters) of each sub-scale were selected to construct a 'Super Scale Youth-Centrism' out of these five dimensions. This ten item scale consisting of items with four answering categories (disagree completely, disagree, agree, agree completely) was replicated in the 1992 Shell Youth Survey.

The properties of the ten item scale have been analyzed in different ways. By means of ordinal latent class analysis Rost and Georg (1991) found, for the 1985 Shell Study, four latent classes with class-specific thresholds for the item categories. Since only three of these four classes could be ordered, a mixed Rasch model was applied. Two Rasch scalable latent classes, consisting of 80 and 20 percent of the sample, were identified.

Georg (1992) used the last four items of the ten item scale (see below) in a cross-cultural comparison between East and West Germany, and in analysis of change in youth-centrism between 1984 and 1991. For that purpose, he estimated a multiple group confirmatory factor model with latent means by means of LISREL. Common congeneric measurement models were established for the four subgroups of East and West German boys and girls in 1991, and for the 1984 and 1991 samples. It was found, among other things, that the mean of the latent variable youth-centrism decreased between 1984 and 1991.

Hypotheses

As already mentioned, the data for the application that is presented below were taken from a subsample of the 1992 Shell Youth Survey (N=288) which was interviewed again in 1993. For the example, the last four items of the youth-centrism scale were selected. The wording of these items is as follows: 1) Very few adults really understand the problems of young people, 2) I do not think much of the experience of adults; I would rather rely on myself, 3) I learn more from friends of my own than from my parents, and 4) Parents are always interfering in things that are none of their business. As covariates we used the dichotomous variables East/West Germany, sex, and age (15-17,18-20).

In East Germany, a substantive amount of social change took place between 1991 and 1993. The breakdown of traditional institutions and norms, in conjunction with a process of de-industrialization, led to a situation of identity-threat and unemployment for young people. It can be expected that this has resulted in a more youth-centered orientation for East German young people between 1991 and 1993. Note that such a change must be interpreted as a period effect and not as an age effect.

For sex, no specific hypotheses can be formulated. The age effect on the change in youth-centrism does not have a substantial interpretation on its own, but only in connection with status passages from youth to adulthood. Age is strongly correlated with these kinds of status passages which can be expected to diminish someone's youth-centristic orientation.

Summarizing, for the three covariates it is hypothesized that boys and girls do not differ with respect to the change in youth-centrism, that in East Germany there is an increase of youth-centrism between 1991 and 1993 as result of substantive social change, and that youth-centrism diminishes with age because of the gradual transition from youth to adulthood.

Results

The background variables East/West, sex, and age are denoted by R, S, and T, respectively. Furthermore, the time-specific latent youth-centrism variables are denoted by W and Y, the indicators for W by A, B, C and D, and the indicators for Y by E, F, G, and H. For the sake of simplicity and also because of sparseness of the table to be analyzed, the youth-centrism items are dichotomized (disagree, agree). Although this leads to some loss of information, the loss of information is not larger than when performing a LISREL analysis, where the observed relationships among variables are assumed to be described completely by means of some bivariate association measure. Here, the data used to estimate the model parameters consists of the multivariate frequency table with observed cells entries $n_{abcdefghrst}$.

In its most general form, the modified path model with latent variables for the above-mentioned variables is given by

$$\pi_{abcdefghrstwy} = \pi_{rst}\, \pi_{w|rst}\, \pi_{y|rstw}\, \pi_{a|w}\, \pi_{b|w}\, \pi_{c|w}\, \pi_{d|w}\, \pi_{e|y}\, \pi_{f|y}\, \pi_{g|y}\, \pi_{h|y}\, .$$

[INSERT FIGURE 1 ABOUT HERE]

This model is quite complex, especially if one wants to impose all kinds of logit restrictions on conditional probabilities describing the structural part and the measurement part of the model. Figure 1 gives an example of a more restricted modified LISREL model that can be specified on the basis of the variables included in the model. It can be seen that in this model W and Y are latent variables, each with four indicators. Furthermore, R and T are assumed to influence the initial position (W), where there is an interaction between these two covariates. The latent transition from $W = w$ to $Y = y$ is assumed to be influenced by R, S, and T. Actually, the model depicted in Figure 1 is similar to the final model that was obtained for the Shell panel data.

Because of the complexity of the model, a step-wise procedure was followed. First, the time-specific measurement models were investigated, that is,

$$\begin{aligned} \pi_{abcdw} &= \pi_w \, \pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \, \pi_{d|w} \, , \\ \pi_{efghy} &= \pi_y \, \pi_{e|y} \, \pi_{f|y} \, \pi_{g|y} \, \pi_{h|y} \, . \end{aligned}$$

Then, the stability of the measurement model over time was investigated by performing an analysis using information on both time points, but without using the background variables. In other words, we estimated models of the form

$$\pi_{abcdefghwy} = \pi_w \, \pi_{y|w} \, \pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \, \pi_{d|w} \, \pi_{e|y} \, \pi_{f|y} \, \pi_{g|y} \, \pi_{h|y} \, .$$

This model also provides us with information on the latent change between the two points in time. And finally, the relationships among the structural variables were investigated. More precisely, we tested whether R (East/West), S (sex), and T (age) influence the amount of youth-centrism at the first point in time and the transition probabilities between the first and the second point in time.

Separate measurement models

Table 1 reports the test results for the estimated measurement models for the two points in time. As can be seen, we estimated unrestricted latent class models, Rasch models, and Lord-Birnbaum models.

[INSERT TABLE 1 ABOUT HERE]

Let us first look at the two-class models. For both time points, the unrestricted two-class models (Models 1a and 2a) perform quite well ($L^2 = 12.05, df = 6, p = .061$, and $L^2 = 9.40, df = 6, p = .152$). The two-class Rasch models (Models 1b and 2b) are obtained by restricting the two-variable interaction terms to be equal among items. Note that when a model has only two latent classes, it does not matter whether one assumes fixed or random latent nodes. The conditional likelihood ratio tests between Models 1a and 1b ($\Delta L^2 = 6.98, df = 3, p = .073$) and between Model 2a and 2b ($\Delta L^2 = 6.89, df = 3, p = .075$) show that for neither of the two time points, the two-class Rasch models performs worse than the unrestricted two-class model.

The unrestricted three-class models (Models 1c and 2c) fit almost perfect for both time points. It should be noted that an unrestricted three-class model with four dichotomous indicators is not identified (Goodman, 1974). An identified model with the same L^2 value as the unrestricted three-class model can be obtained by imposing one arbitrary restriction on the model parameters. For this reason, the reported number of degrees of freedom for Models 1c and 2c is 2 instead of 1. As a result of difficulties associated with parameter space boundaries (Titterington, Smith and Makov, 1985), the two- and three-class models cannot be tested against each other by means of an L^2 test.

Several restricted three-class models were estimated, namely: Rasch models with random latent nodes (Models 1d and 2d), Rasch models with fixed latent nodes (Models 1e and 2e), and Lord-Birnbaum models with fixed latent nodes (Models 1f and 2f). For the first point in time, both three-class Rasch models perform very badly. Models 1d and 1e fit much worse than the unrestricted three-class model, and moreover, they have almost the same L^2 value as the two-class Rasch model. The Lord-Birnbaum model does not fit significantly better than the Rasch model ($\Delta L^2 = 6.68, df = 3, p = .083$), and moreover, it fits significantly worse than the unrestricted three-class model ($\Delta L^2 = 10.19, df = 3, p = .017$). Thus, for the first point in time, the best fitting three-class model is the unrestricted model (Model 1c).

From the restricted three-class models for the second point in time, only the Rasch model with fixed latent models (Model 2d) fits worse than the unrestricted model (Model 2c). This can be seen from the conditional tests between Models 2c and 2d ($\Delta L^2 = 12.87, df = 6, p = .045$), between Models 2c and 2e ($\Delta L^2 = 10.27, df = 5, p = .067$), and between Models 2c and 2e ($\Delta L^2 = 3.31, df = 3, p = .364$). Furthermore, the Rasch model with fixed latent nodes does not fit worse than the Rasch model with random latent nodes ($\Delta L^2 = 2.51, df = 1, p = .113$). And finally, the Lord-Birnbaum fits better than the Rasch model with fixed nodes ($\Delta L^2 = 9.47, df = 3, p = .024$). Thus, for the second point in time, the Lord-Birnbaum is the best fitting three-class model.

The last column of Table 1 reports the BIC values for all models, where BIC is defined as: $-2\,\text{log-likelihood} + \log(N)\,\text{npar}$. When using this definition of BIC, the model with the lower BIC value is the model that should be preferred. It can be seen that, according to the BIC criterion, the two-class Rasch model should be preferred for both time points.

Simultaneous measurement models

The second part of Table 1 reports the test results for the simultaneously estimated measurement models for the two points in time. The main purpose of this analysis was to test whether the measurement model is stable between the two time points. It should be noted that when the measurement model cannot be assumed to be stable, inference on latent change may be problematic.

The stability of the measurement model between the two points in time was tested using unrestricted two-class models (Models 3a-3f). Comparison of the L^2 values of the completely homogeneous two-class model (Model 3a) and the completely heterogeneous two-class model (Model 3b) shows that the differences between the two time points are just significant ($\Delta L^2 = 16.09, df = 8, p = .041$). To see which items are responsible for the differences in the time-specific measurement models, Models 3c-3f relax one by one the homogeneity assumption for the four items. Conditional tests of these models against Model 3a show that only the fourth item (D/H) behaves differently at the two points in time ($\Delta L^2 = 10.26, df = 2, p = .006$). However, as is demonstrated below, the transition probabilities are not strongly influenced by assuming homogeneity for all items. Therefore, we continued assuming homogeneity of the measurement models.

Like in the separate measurement models, the two-class Rasch model (Model 3g) does not fit worse than the unrestricted latent class model ($\Delta L^2 = 4.26, df = 3, p = .235$). Furthermore, some three-class models were estimated. From the three-class models (Models 3h-3k), the Rasch model with fixed latent nodes performs best. It does not fit worse than the unrestricted model ($\Delta L^2 = 10.61, df = 7, p = .157$) nor the Lord-Birnbaum model ($\Delta L^2 = 4.67, df = 3, p = .198$). Since the two- and three-class Rasch model cannot be compared by means of the likelihood-ratio statistic, we have to compare them using, for instance, the BIC criterion. According to this criterion, the two-class Rasch model should be preferred.

[INSERT TABLE 2 ABOUT HERE]

Table 2 gives the parameter estimates for the homogeneous two-class Rasch model (Model 3g). Let us first consider the measurement part of the model. It must be noted that since the measurement model for Y is identical to the measurement model for W, its parameters are not reported in Table 2. The conditional response probabilities indicate that the first class is the non-youth-centristic class, while class two is the youth-centristic class: Persons belonging to class one have a much higher probability of disagreeing with the youth-centrism items than do persons belonging to class two. For instance, respondents belonging to class one have a probability of .7338 to disagree with item D, while for class two this probability is only .2364. It can be seen that items B, C, and D have almost the same difficulty, while item A is much easier than the other three items: Even persons belonging to the non-youth-centristic class have a probability of .5167 to agree with item A.

The initial distribution of the latent variable youth-centrism, π_w, shows that at the first point in time the two latent classes have an almost equal size. The transition probabilities, $\pi_{y|w}$, demonstrate that a great deal of change occurred between the two time points, especially among persons belonging to the youth-centristic class at time point one. Almost all respondents who are non-youth-centristic remain non-youth-centristic. But, respondents who are youth-centristic at the first point in time, have a probability of .5786 to change to the non-youth-centristic position. As a result, the population becomes less youth-centristic: At the second point in time, the probability of being non-youth-centristic is .7626 ($= .4699 \cdot .9596 + .5301 \cdot .5786$).

The homogeneous nonrestricted two-class model (Model 3a) gives nearly the same transition probabilities as the ones presented in Table 2. The model in which the response probabilities for the last item (D/H), the item that did not pass the homogeneity test, are allowed to be different for the two points in time (Model 3g), gives a lower transition probability from $W = 2$ to $Y = 1$, namely, .4981 instead of .5786. Thus, not taking into account that the reliability of the fourth item differs between the two time points leads to an overestimation of the amount of true change.

Structural models

[INSERT TABLE 3 ABOUT HERE]

Table 3 gives the test results for the estimated structural models. In all these models, a two-class homogeneous Rasch model was assumed for the relationships among the latent variables W and Y and their indicators. First, we estimated some reference models. In Model 1, both W and the transitions from W to Y are assumed to be independent of R, S, and T, while Model 2 is the saturated structural model. In other words, Model 1 gives the upper limit L^2 value and Model 2 the lower limit L^2 value which can be obtained by including covariate effects in the model. Models 3 and 4 postulate a saturated model for $\pi_{w|rst}$ and $\pi_{y|rstw}$, respectively, assuming independence of R, S, and T for the other part of the model. These models give the lower limit L^2 values that can be obtained by separately explaining W and the transition from W to Y by R, S, and T. Note that as a result of sparse data, the L^2 statistic reported in Table 3 will not be chi-squared distributed. Therefore, the reported p values are, in fact, meaningless. However, conditional L^2 tests can still be used to compared nested models.

Then, we tested several unsaturated hierarchical logit models for $\pi_{w|rst}$. The only two-variable interaction that improved the model fit significantly was the effect of T on W. Although the two-variable interaction RW was not significant, including the three-variable term u^{RTW}_{rtw} in the model, improved the fit again. Model 5 includes these three interaction terms. Inspection of the parameters of Model 5 showed that the three-variable interaction term was needed to describe the fact that age (T) influences W only for persons living in East Germany ($R = 2$). Model 6, is a non-hierarchical model that includes only this effect. It does not fit significantly worse than Model 3 ($\Delta L^2 =$

$4.31, df = 6, p = .635$) and, moreover, it fits significantly better than Model 1 ($\Delta L^2 = 5.52, df = 1, p = .014$). So, only in East Germany there is some difference in youth-centrism between the two age groups. This was the only significant covariate effect on W.

Next, we tested some hierarchical models for the effects of R, S, and T on the transition from W to Y. Model 7, which contains the three-variable interactions RWY, SWY, and TWY, was the best fitting model. Like for $\pi_{w|rst}$, we tried to specify a more parsimonious model by allowing the logit model for Y to be non-hierarchical. This resulted in Model 8 in which East/West, sex, and age influence the transition from $W = 1$ to $Y = y$, but not the transition from $W = 2$ to $Y = y$. This model does not fit worse than Model 4 ($\Delta L^2 = 6.25, df = 11, p = .856$), and moreover, it does fit better than Model 1 ($\Delta L^2 = 9.02, df = 3, p = .029$). So, the transition out of the state non-youth-centristic depends on covariate values, while the transition out of the state youth-centristic does not.

Model 9 is a combination Models 6 and 8. It describes a large part of the variation in the dependent variables W and Y that could be explained by the independent variables. The difference between the L^2 values of Models 1 and 9 is 15.00 using 4 additional parameters, while the maximum increase in L^2 by including covariate effects was 24.97 using 21 parameters.

The effect of age on W for East Germans is -0.2444. This indicates that in East Germany, the youngest age group ($T = 1$) has a lower probability of being non-youth-centristic ($W = 1$) than the oldest age group: that is, the younger group (15-17) is more youth-centristic than the older group (18-20). The effects of R, S, and T on the transition from $W = 1$ to $Y = y$ are 1.9646, -2.3430, and -2.0496, respectively. Note that these parameters are quite extreme. The parameters indicate that persons with $R = 1$, $S = 2$, and $T = 2$, that is, West-German older females have the highest probability of remaining in the position non-youth-centristic, while East-German younger males have the highest probability of moving from non-youth-centristic ($W = 1$) to youth-centristic $Y = 2$.

Inspection of the conditional probabilities $\pi_{y|rstw}$ obtained from Model 9 showed that for all combination of R, S, and T the probability of staying in $W = 1$ is near to 1, except for East-German males belonging to the youngest age group. This group has a probability of almost 50 percent of changing from non-youth-centristic to youth-centristic. Apparently, the rather extreme logit parameters of Model 9 only describe this 'outlier'. Therefore, logit model was specified for $\pi_{y|rstw}$ in which the probability of leaving $W = 1$ was assumed to be equal for all levels of R, T, and S, except for the group with $R = 2$, $S = 1$, and $T = 1$. As can been seen from the test results for Model 10, this model has exactly the same L^2 value as Model 9. Note that according to the BIC criterion, Model 10 is the best model. Thus, to describe the covariate effects on W and on the transition from W to Y, only two additional parameters are needed compared to Model 1, namely: a parameter describing the fact that among East-Germans the younger age group is more youth-centristic than the older age group and a parameter describing the fact that East-German younger males have a much higher probability of moving from non-youth-centristic to youth-centristic than do other persons.

Discussion

This paper presented a general approach for the analysis of categorical panel data using causal log-linear model with latent variables. In the application on youth-centrism we used a restricted latent class model to construct a semi-parametric Rasch model for the latent variable youth-centrism which was measured at two points in time. The change in youth-centrism between the two waves was described by means of a latent turnover table. Furthermore, both the initial state and the transition probabilities between the two time points were regressed on a set of covariates using quite complex logit models.

It is interesting to compare the information on latent change obtained using the approach presented here with the results that could have been obtained by means of a LISREL analysis. It can be expected that the low stability in youth-centrism can be detected using a LISREL model. This is confirmed by a LISREL analysis we performed on the Shell panel data. Moreover, using a model with latent means it is not a problem to detect the tendency towards less youth-centrism. However, the description of the process of change by the causal log-linear models with latent variables is more informative since the latent turnover table gives exact information on the direction of change for persons with different initial positions. In the current application, it was found that, except for East-German younger males, persons with a non-youth-centered attitude remained in the same position, while many youth-centristic persons moved towards the non-youth-centered position. The possibility to detect such complicated interaction effects is specific for the log-linear approach.

The modified LISREL approach has, of course, also weak points compared to the LISREL model. Although the current working version of the ℓ_{EM} program is much more efficient than the previous version, the number of variables that can be included in a model is still much smaller than in LISREL. So, the size of the problem that can be dealt with is limited. Another problem of log-linear analysis is that when sparse tables are analyzed, the theoretical χ^2 approximation of the likelihood-ratio chi-squared statistic is poor. Although in such situations the model fit can no longer be assessed, the significance of parameters can still be tested by means of likelihood-ratio tests (Agresti, 1990).

References

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

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

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

Clogg, C.C. (1981). "New developments in latent structure analysis," *in* Factor analysis and measurement in sociological research (D.J. Jackson and E.F. Borgotta, Eds.), pp. 215-246. Sage Publications, Beverly Hills.

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

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

Croon, M. (1990). "Latent class analysis with ordered latent classes," British Journal of Mathematical and Statistical Psychology **43**, 171-192.

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.

Georg, W. (1992). "Die skala Jugendzentrismus im Zeitreihen- und Kulturvergleich," *in* Jugend '92, Lebenslagen, Orientierungen und Entwicklungsperspectiven im vereinten Deutschland, Vol. 4: Methodenberichte, Tabellen, Fragebogen (Jugendwerk der deutschen Shell, Ed.), pp. 15-16, Leske und Budrich, Opladen.

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 contingency 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. (1979). Analysis of qualitative data, Vol 2, New developments, Academic Press, New York.

Hagenaars, J.A. (1988). "Latent structure models with direct effects between indicators: local dependence models," Sociological Methods and Research **16**, 379-405.

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

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

Heinen, A. (1996). Latent class and discrete latent trait models: similarities and differences, Sage, Thousand Oaks.

Jöreskog, K.G., and Sörbom, D. (1988). Lisrel 7: a guide to the program and applications.

Jugendwerk der deutschen Shell (1995). Jugend ’81. Lebenswürfe, Alltagkulturen, Zukunftsbilder, Vol. 1, Jugendwerk der deutschen Shell, Hamburg.

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

Lindsay, B., Clogg, C.C., and Grego, J. (1991). “Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis,” Journal of the American Statistical Association **86**, 96-107.

Little, R.J., and Rubin, D.B. (1987). Statistical analysis with missing data, Wiley, New York.

Mare, R.D., and Winship, C. (1991). “Loglinear models for reciprocal and other simultaneous effects,” *in* Sociological Methodology 1991 (C.C. Clogg, Ed.), pp. 199-234, Basil Blackwell Oxford.

Masters, G.N. (1982). “A Rasch model for partial credit scoring,” Psychometrika **47**, 149-174.

McCutcheon, A.L. (1988). “Sexual morality, pro-life values and attitudes toward abortion,” Sociological Methods and Research **16**, 256-275.

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

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,” *in* Sociological Methodology 1990 (C.C. Clogg, Ed.), Basil Blackwell, Oxford.

Schelsky, H. (1957). Die skeptisch Generation. Eine Soziologie der deutschen Jugend, Eugen Diederichs Verlag, Düsseldorf, Köln.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests, Nielson and Lydicke, Copenhagen.

Rost, J. (1990). “Rasch models in latent classes: an integration of two approaches to item analysis,” Journal of Applied Psychological Measurement **14**, 271-282.

Rost, J., and Georg, W. (1991). “Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am beisplie1 zur Skala Jugendzentrismus,” ZA-Information **28**, 52-74.

Titterington, D.M., Smith, A.F., and Makov, U.E. (1985). Statistical analysis of finite mixture distributions, John Wiley & Sons, Chichester.

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. (1996a). “Causal log-linear modeling with latent variables and missing data,” *in* Analysis of change: advanced techniques in panel data analysis (U.Engel and J. Reinecke, Eds.), pp. 35-60. Walter de Gruyter, Berlin/New York.

Vermunt, J.K. (1996b). Log-linear event history analysis; a general approach with missing data, latent variables, and unobserved heterogeneity, Tilburg University Press, Tilburg.

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

Xie, Yu (1992). “The log-multiplicative layer effects model for comparing mobility tables,” American Sociological Review **57**, 380-395.

Zinnecker, J. (1985). “Die Jugendstudien von EMNID/Shell 1953-1955. Zur Archäologie repräsentativer Jugendforschung im Nachkriegsdeutschland und zugleich zu einigen Schwierigkeiten der

Wiederholung solcher Studien," *in* Jugendliche und Erwachsene '85, Generation im Vergleich, Vol. 3: Jugend der fünfziger Jahre un heute (Jugendwerk der deutschen Shell, Ed.), pp. 408-480. Leske und Budrich, Opladen.

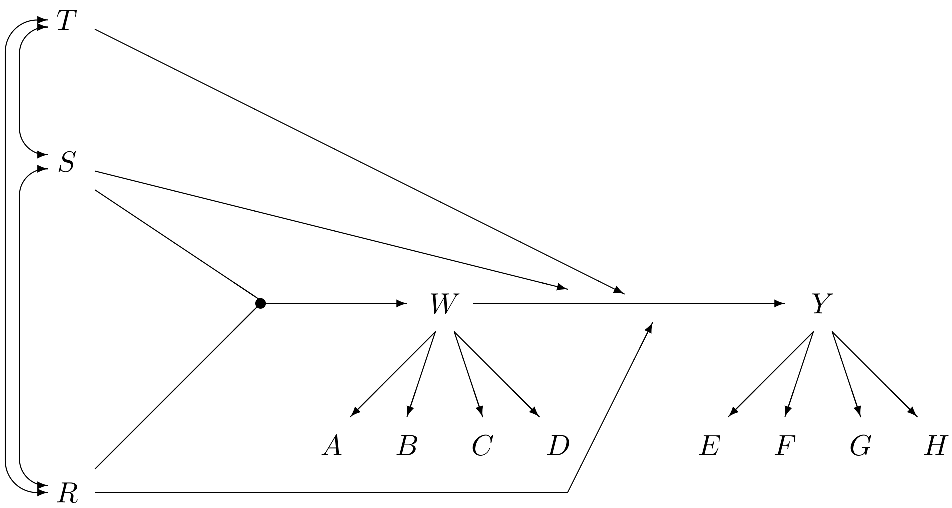

Figure 1: Modified Lisrel model

Table 1: Test results for the estimated measurement models

Model		L^2	*df*	$p(L^2)$	npar	*BIC*
separate models for time point one						
1a.	2 class	12.05	6	.061	10	1493
1b.	2 class Rasch	19.03	9	.025	7	1483
1c.	3 class (1 identifying restriction)	1.13	2	.288	14	1504
1d.	3 class Rasch (random nodes)	19.00	7	.008	9	1494
1e.	3 class Rasch	19.00	8	.015	8	1488
1f.	3 class Lord-Birnbaum	11.32	5	.045	11	1498
separate models for time point two						
2a.	2 class	9.40	6	.152	10	1552
2b.	2 class Rasch	16.29	9	.061	7	1542
2c.	3 class (1 identifying restriction)	2.71	2	.258	14	1568
2d.	3 class Rasch (random nodes)	12.98	7	.073	9	1550
2e.	3 class Rasch	15.49	8	.050	8	1547
2f.	3 class Lord-Birnbaum	6.02	5	.304	11	1555
simultaneous models for time points one and two						
3a.	2 class homogeneous	294.36	244	.015	12	2998
3b.	2 class heterogeneous	278.27	236	.031	20	3027
3c.	2 class $\pi_{a\|w} \neq \pi_{e\|y}$	292.07	242	.015	14	3007
3d.	2 class $\pi_{b\|w} \neq \pi_{f\|y}$	291.21	242	.017	14	3006
3e.	2 class $\pi_{c\|w} \neq \pi_{g\|y}$	294.10	242	.012	14	3009
3f.	2 class $\pi_{d\|w} \neq \pi_{h\|y}$	284.10	242	.033	14	2999
3g.	2 class Rasch homogeneous	298.62	247	.014	9	2985
3h.	3 class homogeneous	271.30	235	.052	21	3026
3i.	3 class Rasch homogeneous (random nodes)	281.39	241	.038	15	3002
3j.	3 class Rasch homogeneous	281.91	242	.040	14	2997
3k.	3 class Lord-Birnbaum homogeneous	277.24	239	.045	17	3009

Table 2: Parameter estimates for the homogeneous two-class Rasch model

π_w	$W=1$	$W=2$
	0.4699	0.5301
$\pi_{y\|w}$	$Y=1$	$Y=2$
$W=1$	0.9596	0.0404
$W=2$	0.5786	0.4214
$\pi_{a\|w}$	$A=1$	$A=2$
$W=1$	0.4833	0.5167
$W=2$	0.0951	0.9049
$\pi_{b\|w}$	$B=1$	$B=2$
$W=1$	0.6911	0.3089
$W=2$	0.2008	0.7992
$\pi_{c\|w}$	$C=1$	$C=2$
$W=1$	0.7229	0.2771
$W=2$	0.2267	0.7733
$\pi_{d\|w}$	$D=1$	$D=2$
$W=1$	0.7338	0.2662
$W=2$	0.2364	0.7636

Table 3: Test results for the estimated structural models assuming a homogeneous two-class Rasch measurement model

Model		L^2	df	$p(L^2)$	npar	BIC
1.	$\{W\}\{WY\}$	963.76	2032	1.000	16	4125
2.	$\{RSTW\}\{RSTWY\}$	938.79	2011	1.000	37	4219
3.	$\{RSTW\}\{WY\}$	953.93	2025	1.000	23	4155
4.	$\{W\}\{RSTWY\}$	946.49	2018	1.000	30	4187
5.	$\{RTW\}\{WY\}$	957.67	2029	1.000	19	4136
6.	$\{RTW^1\}\{WY\}$	958.24	2031	1.000	17	4125
7.	$\{W\}\{RWY, SWY, TWY\}$	949.82	2026	1.000	22	4145
8.	$\{W\}\{RWY^2, SWY^3, TWY^4\}$	952.74	2029	1.000	19	4131
9.	$\{RTW^1\}\{RWY^2, SWY^3, TWY^4\}$	948.76	2028	1.000	20	4133
10.	$\{RTW^1\}\{RSTWY^5\}$	948.76	2030	1.000	18	4121

1: only an effect of T on W for $R=2$

2, 3 and 4: only effects of R, S and T on $\pi_{y|rst1}$

5: only an effect on $\pi_{y|rst1}$ for $R=2$, $S=1$, and $T=1$