Longitudinal data analysis using log-linear path models with latent variables[1]

Jeroen K. Vermunt
Tilburg University, The Netherlands

Werner Georg
University of Konstanz, Germany

Resumen

Este trabajo muestra cómo analizar datos categóricos longitudinales por medio del programa de computador ℓEM . En paralelismo con los conocidos modelos LISREL, los modelos que se proponen consisten de una parte estructural y una parte de medida, donde la parte estructural es un sistema de ecuaciones logit y la parte de medida es un modelo de clase latente que puede ser no restringido o restringido. Los modelos de transición de clase discreta que explican los cambios que ocurren a nivel latente se estiman simultáneamente con modelos de medida sofisticados como, por ejemplo, los modelos de rasgo latente con variables discretas. El enfoque se ilustra utlizando una aplicación empírica. Varios modelos de medida se someten a prueba para una escala que se supone que mide el centrismo juvenil. También se analiza la influencia de covariantes sobre la posición inicial y las probabilidades de transición entre puntos temporales.
PALABRAS CLAVE: *Análisis de clase latente, TRI con variables discretas, análisis de datos categóricos, modelo de Markov latente, modelo de transición latente, análisis de datos de panel, modelo de Rasch, enfoque Lisrel modificado*

Abstract

This paper shows how to analyze categorical longitudinal data by means of the log-linear path modeling approach implemented in the ℓEM computer program. Like the well-known LISREL models, the proposed models consist of a structural and a measurement part, where the structural part is a system of logit equations and the measurement part is an unrestricted or restricted latent class model. Discrete-time transition models explaining changes occurring at the latent level are estimated simultaneously with sophisticated measurement models like, for example, discretized latent trait models.
The approach is illustrated by means of an empirical application. Several measurement models are tested for a scale that is assumed to measure youth-centrism. In addition, the influence of covariates on the initial position and the transition probabilities between time points is studied.
KEY WORDS: *Latent class analysis, discretized IRT, categorical data analysis, latent Markov model, latent transition model, panel data analysis, Rasch model, modified Lisrel approach*

Introduction

Longitudinal data obtained via panel studies is, together with event history data, the best suited kind of data for detecting determinants of individual change. This paper demonstrates how to analyze categorical longitudinal data by means of the log-linear path modeling approach implemented in the ℓEM computer program (Vermunt 1997a, 1997b). The approach was originally proposed by Hagenaars (1990, 1993) and is, in fact, a categorical data variant of the well-known LISREL model for continuous variables (Jöreskog and Sörbom, 1988). A path model specifying the relationships among the structural variables is combined with a measurement model for the latent variables. Because of the analogy with the LISREL model, Hagenaars referred to it as a modified LISREL approach.

The structural part of a modified LISREL model consists of a system of logit equations in which a variable that appears as a response variable in one logit equation

---

[1]Address for correspondence: Jeroen K. Vermunt. Department of Methodology and Statistics. Tilburg University. PO Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: <J.K.Vermunt@kub.nl>

can be used as an explanatory variable in the logit equation for one of its posterior variables (Goodman, 1973). The measurement part has the form of a latent class model (Lazarsfeld and Henry, 1968; Goodman, 1974; Haberman, 1979). It is used to correct for measurement error in certain observed variables. As explained by Van de Pol and De Leeuw (1986) and Hagenaars (1990, 1993), in panel data analysis, it is extremely important to separate true change from spurious change resulting from measurement error.

The log-linear path model with latent variable is strongly related to the discrete-time latent Markov model (Wiggins, 1973; Van de Pol and Langeheine, 1990). Actually, the latter model for the analysis of longitudinal data is a special case of the approach presented in this paper (see Vermunt, 1997b). As was shown by Vermunt, Langeheine, and Böckenholt (1999), specification of the latent Markov model as a log-linear path model may yield parsimonious logit regression models for latent transitions. New in this paper is that it is shown how to combine such latent transition models with sophisticated latent class measurement models like the discretized latent trait models proposed by Lindsay, Clogg, and Grego (1991) and Heinen (1996).

The next section discusses the log-linear path models without latent variables. Then, unrestricted and restricted latent class models, as well as the log-linear LISREL model are presented. The proposed methods are illustrated with an application to a real world two-way panel study.

Log-linear 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 or time ordering of the variables into account a modified path analysis approach. As is demonstrated below, this model is very well suited for analyzing categorical longitudinal data.

Specifying the probability structure and simple constraints

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.

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 or time order among the variables, π_{rstwyz} can be written as

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

As can be seen, the joint probability is decomposed into a product of marginal and conditional probabilities, which is a straightforward manner to express that the value

on a particular variable can only depend on the preceding variables but not on the posterior ones (Goodman, 1973). 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 Equation (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 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 Equation (1) contains as many unknown (conditional) probabilities as observed cell frequencies. In other words, the model given in Equation (1) is a saturated model in which it is assumed that a particular dependent variable is influenced by all its preceding variables, including all higher-order interactions. One of the problems associated with such a saturated model is that some of its probabilities may be inestimable because there are no cases in certain categories of the conditioning variables.

The simplest way to specify more parsimonious models is via conditional independence assumptions. 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 and W. These restrictions can implemented by replacing the unrestricted model given in Equation (1) by

$$\pi_{rstwyz} = \pi_{rst}\,\pi_{w|rs}\,\pi_{y|stw}\,\pi_{z|sty}\,. \tag{2}$$

Conditional independence restrictions are well-known from the field of graphical modeling (Whittaker, 1990). The model described in Equation (2) is a first-order discrete-time Markov model with covariates. Note that these are all assumptions that should be tested.

Specifying logit constraints

The above-mentioned procedure for obtaining more restricted models has one important disadvantage: The resulting model may still contain many higher-order interactions. Higher-order can be excluded by using logit parameterization for the marginal and conditional probabilities appearing in Equation (1), yielding what Goodman (1973) called a "modified path analysis approach". In our example, we would have four submodels: an unrestricted model for the exogenous variables R, S, and T, as well as three logit models with W, Y, and Z 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 subject to, for instance, ANOVA-like restrictions. This model contains the effect of S, T, and W on Y, as well as the

second-order interaction of S and W. Note that, as above, we are assuming that Y does not depend on R; that is, $\pi_{y|rstw} = \pi_{y|stw}$. Similar logit constraints could be specified for $\pi_{w|rst}$ 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 conditional independence restrictions of Equation (2).

By using a more general class of logit models, it becomes possible to specify non-hierarchical models and to use continuous exogenous variables in the log-linear path model. Suppose that k is the index for the dependent variable in a particular logit equation, and that i denotes the index for the joint independent variable. In this case, the logit model may be of the general form,

$$\pi_{k|i} = \frac{\exp\left(\sum_j u_j x_{ikj}\right)}{\sum_k \exp\left(\sum_j u_j x_{ikj}\right)}, \tag{3}$$

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 Equation (3) becomes a multinomial logistic regression model (Agresti, 1990). In that case, it can be used with continuous independent variables, where x_{ikj} denotes the value of person i on variable j for level k of the response variable.

The ℓEM program

A program called ℓEM has been developed to estimate the log-linear path models discussed in this section without the necessity to set up different marginal tables (Vermunt, 1997b). In ℓEM specifying a log-linear path model is the standard way of modeling an observed frequency table.

The standard estimation procedure implemented in ℓEM for hierarchical log-linear models is the iterative proportional fitting algorithm (IPF). However, ℓEMcan also be used to estimate more complex log-linear or logit models of the form given in Equation (3), in which case the user has to specify a design matrix. 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 and log-multiplicative models are estimated by means of the uni-dimensional Newton algorithm (Goodman, 1979; Vermunt, 1997b).

Models with latent variables

So far, it has been assumed that all variables used in the analysis can be observed directly. In social and behavioral sciences, however, we often use several observable or manifest variables as indicators for concepts that are difficult or impossible to measure directly. It is assumed that an individual's score on an indicator is determined by the unobservable value of the underlying latent variable of interest. In latent structure models, this principle is implemented statistically by the assumption of local independence; that is, that indicators are independent of each other within levels of the latent variable. The latent class model (LCM) is a latent structure model

in which manifest and latent variables are categorical (Lazarsfeld and Henry, 1968; Goodman, 1974; Haberman, 1979).

Latent class measurement models

Suppose there is a LCM 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 LCM are

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

where

$$\pi_{wabc} = \pi_w \, \pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \tag{4}$$

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 Equation (4) 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.

The LCM was proposed by Lazarsfeld and Henry (1968) for dichotomous indicators and extended by Goodman (1974) to polytomous variables. 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 Equation (4) is very similar to the log-linear path model discussed in the previous section: In fact, it is a log-linear path model in which one of the variables is unobserved.

Restricted LCMs may be specified by imposing restrictions on the probabilities of the LCM. Typical constraints are fixed-value and equality restrictions on the latent and conditional response probabilities (Goodman, 1974; Vermunt, 1997b). Croon (1990) proposed an ordinal LCM with inequality constrains on the conditional response probabilities (see also Vermunt, 2001).

It is also possible to specify restricted LCMs by writing the conditional response probabilities of Equation (4) as logit models (Formann, 1992; Heinen, 1996):

$$\pi_{a|w} = \frac{\exp\left(u_a^A + u_{wa}^{WA}\right)}{\sum_a \exp\left(u_a^A + u_{wa}^{WA}\right)}. \tag{5}$$

Heinen (1996) showed how to obtain discretized variants of the most important latent trait models by restricting the two-variable terms in a similar manner as in Goodman (1979) and Clogg's (1982) linear-by-linear, row, column, and row-column association models.

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). The probability of a "correct" answer is postulated to depend only on a person's ability and on the difficulty of the item concerned. A discrete variant of the Rasch model can be obtained by restricting

$$u_{wa}^{WA} = \theta_w \, \alpha \, x_a, \quad u_{wb}^{WB} = \theta_w \, \alpha \, x_b, \quad u_{wc}^{WC} = \theta_w \, \alpha \, x_c. \tag{6}$$

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 wishes to preserve the ANOVA-like restrictions, one may also score the two categories of A, B, and C as 1 and -1. The scores for the latent variable, sometimes called latent nodes, can be fixed or random quantities. The model with random nodes is equivalent to the semi-parametric Rasch model proposed by Lindsay, Clogg, and Grego (1991). When the latent scores are fixed and equidistant, the two-variable terms in (6) have the form of uniform associations. With random latent nodes, they have the form of row-association terms, where W operates as the row variable.

A discretized two-parameter logistic (2PL) model is obtained by the following set of restrictions:

$$u_{wa}^{WA} = \theta_w\,\alpha^A\,x_a\,,\ \ u_{wb}^{WB} = \theta_w\,\alpha^B\,x_b\,,\ \ u_{wc}^{WC} = \theta_w\,\alpha^C\,x_c\,. \tag{7}$$

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 2PL 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 uniform associations. On the other hand, with random scores for W, the two-variable terms are bi-linear as in Goodman (1979) and Clogg's (1982) row-column association models.

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

$$u_{wa}^{WA} = \theta_w\,\alpha_a^A\,,\ \ u_{wb}^{WB} = \theta_w\,\alpha_b^B\,,\ \ u_{wc}^{WC} = \theta_w\,\alpha_c^C\,.$$

Thus, the NR model contains one association parameter for each item category. With fixed and equidistant latent scores, these are column associations, where the latent variable acts as row variable. If the θ_w are treated as unknown parameters, we obtain a model with bi-linear row-column association terms.

A more parsimonious latent trait model for polytomous items is the partial credit model (Masters, 1982). Similarly to the Rasch model, it is obtained by the restrictions described in Equation (6). A less restricted model for polytomous items is obtained with the restrictions described in (7), yielding a variant of the partial credit model having item-specific discrimination parameters.

Combining a structural model with a measurement model

Several important extensions of the standard LCM have been proposed, such as models with several latent variables (Goodman, 1974; Haberman, 1979; Magidson and Vermunt, 2001), models with covariates (Clogg, 1981), and local dependence models (Hagenaars, 1988). These extension are all special cases of the log-linear path model with latent variables proposed by Hagenaars (1990, 1993) and extended by Vermunt

(1997b). This model combines a structural model with a measurement model for the latent variables. Because of the analogy with the LISREL model for continuous variables (Jöreskog and Sörbom, 1988), Hagenaars called it a modified LISREL approach. Here, we concentrate on its application to the analysis of longitudinal data.

Suppose that the endogenous variables in the log-linear path model given in Equation (1), W, Y, and Z, are now latent variables, and that each of them is measured indirectly by 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. Using the same structure as in Equation (1), but with the endogenous variables as latent variables, yields the following modified LISREL model:

$$\pi_{abcdefghirstwyz} = \left(\pi_{rst} \, \pi_{w|rst} \, \pi_{y|rstw} \, \pi_{z|rstwy} \right)$$
$$\left(\pi_{a|w} \, \pi_{b|w} \, \pi_{c|w} \, \pi_{d|y} \, \pi_{e|y} \, \pi_{f|y} \pi_{g|z} \, \pi_{h|z} \, \pi_{i|z} \right), \quad (8)$$

where the first part at the right-hand side is the structural model and the second part is the measurement model.

I should be noted that the model described in Equation (8) is more restricted that a standard latent class model. An important additional assumption is that the item responses at different time points are independent of one another given the latent variables. Another constraint is that the covariates have no direct effects of the indicators. Both assumptions can be tested and relaxed when applying the model to a data set.

As in models without latent variables, each of the conditional probabilities appearing at the right-hand side of Equation (8) can be restricted by a logit model. In other words, the various types of constraints discussed above in the context of log-linear path models and LC models can also be used here. Although 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.

The model described Equation (8) is an extension of the latent Markov model described by Wiggins (1973), Van de Pol and De Leeuw (1986), and Van de Pol and Lageheine (1990). On the one hand, it contains a more flexible way of including covariates (see also Vermunt, Langeheine and Böckenholt, 1999); on the other hand, it can be used to specify latent transition models with complicated measurement structures, like the discretized latent trait models described above.

The ℓEM program

The ℓEM program (Vermunt, 1997b) is especially developed for estimating log-linear path models with latent variables. Actually, in the model specification, latent and observed variables are treated in exactly the same way by the program. Maximum likelihood estimates of the parameters are obtained by means the EM algorithm (Dempster, Laird and Rubin, 1977), which is a general iterative algorithm to estimate models when there is missing data. Note that in the current application the latent variables are the missing data.

The EM algorithm consists of two separate steps per iteration cycle: an E(xpectation) step and a M(aximization) step. In the E step, estimates are obtained for the cell fre-

quencies in the completed table ($\hat{n}_{abcdefghirstwyz}$), conditional on the incompletely observed data and the parameter estimates from the last EM iteration; that is,

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

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

In the M step, as with completely observed data, ℓEM uses the IPF and unidimensional Newton algorithms treating the completed data as if it were observed data. In fact, the likelihood function in which $\hat{n}_{abcdefghirstwyz}$ appears as data, sometimes referred to as the completed data likelihood, is maximized. The improved parameter estimates are used again in the E step to obtain new estimates for the complete table. The EM cycles continue till the increase of log-likelihood is less than the specified convergence criterion.

Application

The approach to the analysis of longitudinal 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. Because of the German unification, the focus was on the comparison between the East and West German youth.

An important common theoretical construct in various Shell Surveys is the concept of youth-centrism (Georg, 1992). 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. For the example, we used the four items of the youth-centrism scale that proved to be scale in an earlier analysis (Georg, 1992). 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 in two categories (15-17 and 18-20). 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.

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
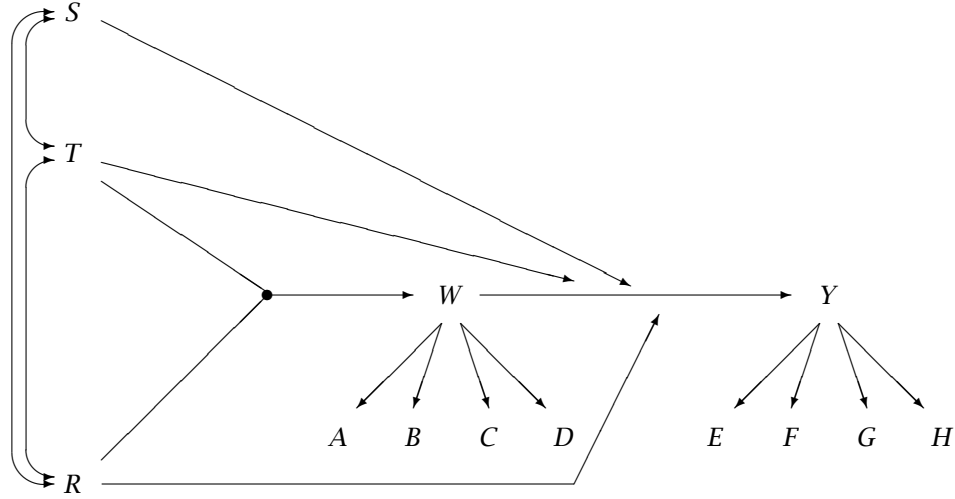
Figure 1: *Modified Lisrel model*

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 $n_{abcdefghrst}$.

The most general log-linear path model with latent variables we worked with is

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

The Figure 1 depicts a model that is obtained by imposing certain constraints on this general model. As can be seen, the latent variables W and Y have 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 depicted model 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,

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

Then, the stability of the measurement model over time was investigated by performing an analysis using information on both time points, but without covariates. 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

Table 1. Test results for the estimated structural models assuming a homogeneous two-class Rasch measurement model

Model		L^2	df	$p(L^2)$	BIC
1.	$\{W\}\{WY\}$	963.76	2032	1.000	-10543.38
2.	$\{RSTW\}\{RSTWY\}$	938.79	2011	1.000	-10449.43
3.	$\{RSTW\}\{WY\}$	953.93	2025	1.000	-10513.56
4.	$\{W\}\{RSTWY\}$	946.49	2018	1.000	-10481.36
5.	$\{RTW\}\{WY\}$	957.67	2029	1.000	-10532.47
6.	$\{RTW^1\}\{WY\}$	958.24	2031	1.000	-10543.24
7.	$\{W\}\{RWY,SWY,TWY\}$	949.82	2026	1.000	-10523.34
8.	$\{W\}\{RWY^2,SWY^3,TWY^4\}$	952.74	2029	1.000	-10537.41
9.	$\{RTW^1\}\{RWY^2,SWY^3,TWY^4\}$	948.76	2028	1.000	-10535.73
10.	$\{RTW^1\}\{RSTWY^5\}$	948.76	2030	1.000	-10547.05

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 $T = 1$

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 LCMs, Rasch models, and 2PL models. We used both the likelihood-ratio statistic L^2 and the Bayesian information criterion BIC for model selection. The L^2 follows asymptotically a chi-squared distribution. The BIC, which is defined as $2\,L^2 + \log(N)\,df$, weights model fit and parsimony. The model with the lower (most negative) BIC value is the model that should be preferred. This measure can also be used as an alternative to formal chi-squared tests when asymptotics do not hold, as, for example, with sparse tables.

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 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 2PL 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 2PL 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 2e) fits worse than the unrestricted model (Model 2c). This can be seen from the conditional tests between Models 2c and 2e ($\Delta L^2 = 12.87, df = 6, p = .045$), between Models 2c and 2d ($\Delta L^2 = 10.27, df = 5, p = .067$), and between Models 2c and 2f ($\Delta L^2 = 3.31, df = 3, p = .364$). 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 2PL 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 2PL is the best fitting three-class model.

The last column of Table 1 reports the *BIC* values for all models. As can be seen, according to the *BIC* criterion, the two-class Rasch model should be preferred for both time points. In other words, if we do not only take into account model fit but also parsimony, the most restricted measurement model comes out as the best.

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 low p values obtained with all the estimated models indicates that the assumption that item responses at different time points are independent given the latent variables does not hold perfectly. This problem could be tackled by introducing direct effects between the item responses at different occasions (Hagenaars, 1988), which is something that can be done within the framework introduced in this paper. For simplicity of exposition, we neglect this problem and concentrate on the stability of the measurement model.

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

Table 2. Test results for the estimated measurement models

Model		L^2	df	$p(L^2)$	BIC
separate models for time point one					
1a.	2 class	12.05	6	.061	-21.93
1b.	2 class Rasch	19.03	9	.025	-31.95
1c.	3 class (1 identifying restriction)	1.13	2	.288	-9.07
1d.	3 class Rasch (random nodes)	19.00	7	.008	-20.64
1e.	3 class Rasch	19.00	8	.015	-26.30
1f.	3 class 2PL	11.32	5	.045	-20.00
separate models for time point two					
2a.	2 class	9.40	6	.152	-24.58
2b.	2 class Rasch	16.29	9	.061	-34.68
2c.	3 class (1 identifying restriction)	2.71	2	.258	-5.91
2d.	3 class Rasch (random nodes)	12.98	7	.073	-26.61
2e.	3 class Rasch	15.49	8	.050	-29.81
2f.	3 class 2PL	6.02	5	.304	-22.31
simultaneous models for time points one and two					
3a.	2 class homogeneous	294.36	244	.015	-1087.40
3b.	2 class heterogeneous	278.27	236	.031	-1052.19
3c.	2 class $\pi_{a|w} \neq \pi_{e|y}$	292.07	242	.015	-1078.36
3d.	2 class $\pi_{b|w} \neq \pi_{f|y}$	291.21	242	.017	-1079.23
3e.	2 class $\pi_{c|w} \neq \pi_{g|y}$	294.10	242	.012	-1076.34
3f.	2 class $\pi_{d|w} \neq \pi_{h|y}$	284.10	242	.033	-1086.34
3g.	2 class Rasch homogeneous	298.62	247	.014	-1100.13
3h.	3 class homogeneous	271.30	235	.052	-1059.50
3i.	3 class Rasch homogeneous (random nodes)	281.39	241	.038	-1083.39
3j.	3 class Rasch homogeneous	281.91	242	.040	-1078.53
3k.	3 class 2PL homogeneous	277.24	239	.045	-1076.21

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.

As in the separate measurement models, the two-class Rasch model (Model 3g) does not fit worse than the unrestricted LCM ($\Delta L^2 = 4.26, df = 3, p = .235$). 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 2PL 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.

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

Table 3. Parameter estimates for the homogeneous
two-class Rasch model

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

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 almost equal sizes. The transition probabilities, $\pi_{y\mid w}$, 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. The result is that the proportion non-youth-centristic persons increases from .4699 to .7626 ($= .4699 \cdot .9596 + .5301 \cdot .5786$) between the two time points.

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.

Combining the measurement models with a structural model

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. As mentioned below, for the final model we estimated the p value by means of parametric bootstrapping.

Then, we tested several unsaturated hierarchical logit models for $\pi_{w|rst}$. The only two-variable effect that improved the model fit significantly was the effect of T on W. Although the two-variable term RW was not significant, including the three-variable term u_{rtw}^{RTW} in the model, improved the fit again. Model 5 includes these three terms. Inspection of the parameters of Model 5 showed that the three-variable 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 simple hierarchical logit models for the effects of R, S, and T on the transition from W to Y. Model 7, which contains the three-variable terms 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 ($\Delta L^2 = 11.02, df = 3, p = .012$). 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-centered ($W = 1$) toward youth-centered $Y = 2$.

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-centered to youth-centered. Apparently, the rather extreme logit parameters of Model 9 only describe this "outlier". Therefore, a 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, in order 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-centered 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-centered to youth-centered than do other persons.

In order to deal with the sparseness problem, for the final model, we estimated the p value by means of a parametric bootstrap. Using 1000 replications, we obtained a p value of slightly more than 6 percent. This indicates that Model 10 can not be rejected at a 5 percent significance level.

Discussion

This paper demonstrated how to the analyze categorical longitudinal data using log-linear path models with latent variables. In the application on youth-centrism, we used a restricted LCM to construct a semi-parametric Rasch model for the latent variable youth-centrism that 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 log-linear path 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-centered persons moved towards the non-youth-centered position. The possibility to detect such complicated interaction effects is specific for the presented approach.

The proposed model could be extended in several manners. Four interesting extensions are the possibility to deal with partially observed data (Vermunt, Langeheine, and Böckenholt, 1999), the possibility to have ordered-restricted latent class

measurement models similar to nonparametric item response models (Vermunt, 2001), the possibility to have dependent classification errors (Bassi et al., 2000), and the possibility to specify constraints on marginal distribution (Vermunt, Rodrigo, and Ato-Garcia, 2001).

References

Agresti, A. (1990). *Categorical Data Analysis.* Wiley, New York.

Bassi, F., Hagenaars, J.A., Croon, M. and Vermunt, J.K. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors. *Sociological Methods and Research, 29,* 230-268.

Clogg, C.C. (1981). New developments in latent structure analysis. In D.J. Jackson and E.F. Borgotta (Eds.), *Factor Analysis and Measurement in Sociological Research*, 215-246. Beverly Hills: Sage Publications.

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.

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 Jugendwerk der deutschen Shell (Ed.), *Jugend '92, Lebenslagen, Orientierungen und Entwicklungsperspectiven im vereinten Deutschland, Vol. 4: Methodenberichte, Tabellen, Fragebogen*, 15-16. Opladen: Leske und Budrich.

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.* New York: Academic Press.

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.* Newbury Park: Sage.

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

Heinen, A. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousand Oaks: Sage.

Jöreskog, K.G., and Sörbom, D. (1988). *Lisrel 7: A Guide to the Program and Applications.* Chicago: Scientific Software, Inc..

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

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.

Magidson, J. and Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots and tri-plots, *Sociological Methodology, 31*, 223-264.

Mare, R.D., and Winship, C. (1991). Loglinear models for reciprocal and other simultaneous effects. *Sociological Methodology 21*, 199-234.

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

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*, Nielson and Lydicke, Copenhagen.

Titterington, D.M., Smith, A.F., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Chichester, UK: John Wiley.

Van de Pol, F., and Langeheine, R. (1990). Mixed Markov latent class models, *Sociological Methodology, 20*, 213-247.

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

Vermunt, J.K. (1997a). *LEM: A General Program for the Analysis of Categorical Data.* Users' manual. Tilburg, NL: Tilburg University. (www.kub.nl\mto).

Vermunt, J.K. (1997b). *Log-linear Models for Event Histories.* Thousand Oakes: Sage Publications.

Vermunt, J.K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement, 25,* 283-294.

Vermunt, J.K., Langeheine, R. and Böckenholt, U. (1999). Latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24,* 178-205.

Vermunt, J.K., Rodrigo, M.F. and Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research, 30,* 170-196.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Chichester: Wiley.

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

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