Categorical causal modeling
Hagenaars, J.A.P.

Publication date:
1997

Link to publication

Citation for published version (APA):
Categorical Causal Modeling: Directed Loglinear Models with Latent Variables

Jaques A. Hagenaars

WORC Paper 97.04.002/7
Categorical Causal Modeling:
Directed Loglinear Models with Latent Variables

JACQUES A. HAGENAARS
Tilburg University

Department of Methodology,
Faculty of Social Sciences,
Tilburg University,
P.O. Box 90153,
5000 LE Tilburg,
The Netherlands.

email: jacques.a.hagenaars@kub.nl
phone: +31.13.4662086

*This article is a modified version of a paper presented at the conference ‘Social Science and Statistics in honor of the late Clifford C. Clogg’ at Pennsylvania State University, State College, Penn, USA, 26-28 September 1996. The main ideas in this article were developed in close cooperation with other members of the Department of Methodology, Tilburg University especially, Marcel Croon, Ton Heinen, Ruud Luijkx, and Jeroen Vermunt. Specific references to their work will be found in the main text.
Latent Class Analysis (LCA) has matured into a very useful technique for the analysis of categorical data, as has been made eminently clear by Clogg in two comprehensive overviews (Clogg, 1981a, 1995). LCA is first of all a very general measurement model. It is a particular kind of latent structure model that can be used for the investigation of the reliability and validity of categorical measurements and has a lot in common with more traditional techniques such as factor analysis, Likert scales or Rasch models. Moreover, LCA has been shown to be useful for exploring the consequences of several kinds of frequently occurring systematic response errors, such as test-retest effects, item nonresponse, and missing data in general. LCA has also been fruitfully applied in longitudinal studies to discover systematic patterns of change and to separate true change from apparent changes caused by unreliability of the measurements. Further, it is a useful tool for carrying out comparative cross-cultural or cross-temporal research. And, finally, LCA offers a very general approach for dealing with the problems caused by unobserved heterogeneity. (For more detailed treatments and an extensive list of references with regard to these (and still other) topics, see Clogg, 1981a, 1995, McCutcheon and Hagenaars, forthcoming)

Several of the above features will be illustrated in this article, but from a consistent causal modeling point of view and with a special emphasis on LCA as a tool for taking response errors into account. It will be shown how LCA can be integrated into a general causal log-linear modeling approach and what the main advantages and problems connected with that integrated approach are. In the next Section, it will first be clarified what is meant by a causal model for categorical data. The discussion will be rather informal, emphasizing the essential features of the approach.
DIRECTED LOGLINEAR MODELS

CAUSAL MODELS FOR CATEGORICAL DATA

Figure 1 is an example of a traditional recursive causal model. It is a recursive model, because no direct or indirect causal loops are involved. It is assumed that all variables in Figure 1 are categorical, e.g., Age (A) is coded in ten years intervals; Gender (B) has obviously two categories Women, Men; Education (C) has 7 categories, a mixture of 'level of education' and 'general/vocational types of education'; Party Preference (D) at time 1 involves a choice among five main political parties six months before the elections; E has the same meaning, but now immediately before the elections take place.

[figure 1 about here]

The pointed arrows in Figure 1 represent asymmetrical (causal) direct partial relationships between the connected variables, controlling for the pertinent antecedent and intervening variables; the nodes connecting three or more variables represent higher order interaction effects among the connected three or more variables; the absence of an arrow indicates the absence of a direct effect and the absence of a node indicates the absence of the corresponding higher order interaction; symmetrical relationships in which the variables are not causally ordered will be represented by double-headed arrows.

Departing from the adage that what is 'causally posterior' cannot influence what is 'causally prior', essential to all (recursive) causal models such as Figure 1, whether the variables are continuous or not is that the relations between any two variables should be investigated while controlling for all antecedent and intervening variables, but ignoring the variables that appear later in the causal chain. For a set of categorical variables A through E, application of this principle leads to a series of submodels (a system of equations) for a corresponding series of (marginal) tables. These marginal tables are formed from the entries of the full cross-classification table ABCDE, that is, from the probabilities \( \pi_{ijklm}^{ABCDE} \), which is the probability that a randomly chosen element from the population belongs to cell \((ijklm)\) of table ABCDE. If the causal order of the variables from 'least' to 'most dependent' is alphabetical and if the causal system is recursive with variable A as the only exogenous variable, then the first equation or submodel concerns marginal 'table' A with entries \( \pi_A \) obtained by collapsing the
full cross-classification table ABCDE over B through E. In the next equation, the effects of A on B are investigated by defining a model for the conditional probabilities $\pi_{ij}^{AB}$, the probability that a randomly chosen element obtains score B=j, given A=i, obtained by means of marginal table AB. Then, the effects of A and B on C are determined, using conditional probabilities $\pi_{ij}^{ABC}$ in marginal table ABC. The remaining equations successively concern the effects of A, B, and C on D and the effects of A, B, C, and D on E, defining submodels for the conditional probabilities $\pi_{ijkl}^{ABCD}$ and $\pi_{ijkl}^{EABCD}$, respectively. Following this ‘causal logic’, the joint probabilities in the full cross-classification table ABCDE are seen as a function of an appropriate series of (marginal, conditional) probabilities:

$$
\pi_{ijklm}^{ABCDE} = \pi_{ij}^{AB} \pi_{ij}^{CAB} \pi_{ij}^{DiABC} \pi_{ijkl}^{EABCD}
$$

(1)

Because the causal model in Figure 1 contains two exogenous variables A (Age) and B (Gender) that do not depend on any of the other variables and are not ‘causally ordered’, for Figure 1, the first equation (submodel) concerns marginal table AB rather than marginal table A and the appropriate decomposition of the joint probabilities of the full table ABCDE becomes

$$
\pi_{ijklm}^{ABCDE} = \pi_{ij}^{AB} \pi_{ij}^{CAB} \pi_{ij}^{DiABC} \pi_{ijkl}^{EABCD}
$$

(2)

Symmetrical relationships among more than two exogenous variables can be handled in a similar manner. Symmetrical relationships among endogenous variables, leading to non-recursive models pose more serious problems. Mare and Winship have made some interesting proposals for a categorical data approach in which two variables have reciprocal causal effects or in which indirect causal loops occur, but, at least for the time being, their approach raises more questions than that it solved the problem (Mare and Winship, 1991). Easier to handle are block-recursive models. Block-recursive models are models in which all variables are divided into blocks, each variable belonging to one block and one block only and each block containing one or more variables and in which the blocks are causally ordered, as in an ordinary recursive system (Blalock, 1969, pp. 71-74, Whittaker, 1990, p. 79, Lauritzen, 1996, Section
4.6). So the causal order of variables belonging to different blocks is known (and recursive), but no causal order is specified for the variables belonging to the same block. Therefore, when setting up the decomposition of the joint probabilities of the full table, as in Equations (1) and (2), the variables within a particular block should always appear together as a joint dependent variable; otherwise, the variables are treated as in ordinary recursive systems. An example will be provided later on.

**MODIFIED PATH MODELS**

The decomposition of the joint probability in Equation (2) forms the starting point of the causal analysis of Figure 1. Given sample data with observed proportions $P_{ijklm}$ that follow a Poisson or (product)multinomial distribution and given substantive hypotheses about the nature of the causal relations involved, it is possible to find the appropriate maximum likelihood estimates $\pi$ for the parameters at the right hand side of Equation (2). The maximum likelihood estimates for the entries $\pi_{ijklm}$ in the full table (the quantity at the left hand side of Equation (2)) can then be found by substituting these parameter estimates $\pi$ in Equation (2). In this way, the validity of the complete causal model, i.e., the simultaneous validity of all hypotheses about the (marginal, conditional) probabilities at the right hand side of Equation (2) can be tested by multiplying the estimated entries $\pi_{ijklm}$ for the full table by sample size $N$ and compare them by means of the usual chi-square tests with the observed proportions $p_{ijklm}$ multiplied by $N$.

Goodman developed the Modified Path Analysis Approach for obtaining the maximum likelihood estimates (Goodman, 1973, Hagenaars, 1990). Essentially, he proposed to define for each of the (marginal) tables corresponding to the parameters at the right hand side of Equation (2), the appropriate logit or loglinear submodel that reflects the substantive hypotheses. Because in Figure 1, Age (A) and Gender (B) are assumed to be independent of each other, the appropriate loglinear submodel for marginal table AB is the independence model, model \{A,B\} in the usual shorthand notation for hierarchical loglinear models. In formula, where the $\lambda$'s are the loglinear parameters, constrained to sum to zero over each of their subscripts:

$$\ln \pi_{ij}^{AB} = \lambda + \lambda_i^A + \lambda_j^B$$

(3)
The maximum likelihood estimates $\pi_{ij}^{AB}$ for the submodel in Equation (3) are the maximum likelihood estimates for $\pi_{ij}^{AB}$ in Equation (2), given the validity of Figure 1.

The next marginal table to be considered is table ABC with conditional probabilities $\pi_{kij}^{C|AB}$. Age (A) and Gender (B) have both a direct partial effect on Education (C); moreover, there is a three-variable interaction effect: the gender differences with regard to education are different for the different age groups. The corresponding logit model is the saturated one, which, in its turn, corresponds with saturated loglinear model \{ABC\}:

$$\ln \pi_{ijk}^{ABC} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

(4)

The estimates $\pi_{ijk}^{ABC}$ obtained for the submodel in Equation (4) provide the estimates for $\pi_{kij}^{C|AB}$ in Equation (2).

In an analogous way, a logit model has to be defined for marginal table ABCD and the conditional probabilities $\pi_{ijk}^{D|ABC}$, in which A has a main effect on D, and B and C have main and interaction effects on D. This logit model is equivalent to loglinear model \{ABC,AD,-BCD\}. Finally, because E is only influenced by B and D, model \{ABCD,BE,DE\} has to be set up for table ABCDE, providing the estimates for $\pi_{mijkl}^{E|ABCD}$ in Equation (2).

A simultaneous test on the validity of all submodels and hypotheses involved can be carried out in the manner described above. Alternatively, with the same result, one may compute for each submodel the likelihood ratio chi-square test statistic $L^2$ with its corresponding degrees of freedom and then sum these separate test statistics and their degrees of freedom to arrive at the overall test statistic.

Goodman's procedure is a very elegant, easy to apply procedure that provides all flexibilities and possibilities of loglinear modeling. A possible drawback is that as we go further and further into the causal chain, the (marginal) tables contain more and more variables and the corresponding (sub)models more and more parameters to be estimated, ultimately, given space or time constraints, more than can be handled by existing software.

The modified path model can be seen as a special case of a general approach recently developed by Agresti and Lang in which they obtain the maximum likelihood estimates for, what they call, generalized loglinear models. In these models, loglinear models are defined simultaneously for the full cross-classification table and for marginal tables derived from the full
table (Lang and Agresti, 1994, Becker, 1994). Generalized loglinear models and extensions of it (‘marginal models’, Bergsma, 1997) will certainly lead to important new modifications of the modified path model.

**DIRECTED GRAPHS**

The ‘causal’ decomposition of the joint probability illustrated in Equations (1) and (2) leans heavily on work done in the area of graphical modeling, excellent overviews of which are provided by Kiiveri and Speed (1982), Whittaker (1990) and Lauritzen (1996). In graphical models, the relations among variables are defined in terms of the presence or absence of (conditional) statistical independence. In directed graphical models, the causal (asymmetrical) nature of the relations is explicitly taken into account. In principle, graphical models can handle (mixtures of) continuous and categorical data; here, only categorical data will be dealt with.

A (simplified) directed graphical analysis of the model in Figure 1 would also start with Equation (2). Then, as in the Goodman procedure, a more simplified parametrization is sought, this time not in terms of a more parsimonious loglinear representation, but in terms of (conditional) independence relations: the relevant ‘tables’ and parameters at the right hand side of Equation (2) are collapsed over those variables that do not have a direct influence on the dependent variable. Applying this principle to Figure 1 and Equation (2), the simplified equation becomes

\[ \pi_{ijklm}^{ABCDE} = \pi_i^A \pi_j^B \pi_{ijk}^{ABC} \pi_{ijkl}^{AB}. \]  

(5)

Comparing Equations (2) and (5), it is first noticed that \( \pi_{ij}^{AB} \) in (2) is replaced by \( \pi_i^A \pi_j^B \) in (5). Because, according to Figure 1, A and B are independent of each other within marginal table AB, the joint probability is equal to the product of the two marginal probabilities. Further, \( \pi_{ijkl}^{E|ABCD} \) in Equation (2) is replaced by \( \pi_{ijkl}^{E|BD} \) in Equation (5): according to Figure 1, when one controls for B and D, the variables A and C have no direct influence on E. So, the conditional probability of scoring m on E only depends on B and D, but is conditional independent of A and C, given B and D. If the causal model in Figure 1 is valid, the conditional probability of scoring m on E is the same whether determined within table ABCDE or within table BDE.
Maximum likelihood estimates for the parameters at the right hand side of Equation (5) can be obtained by substituting the observed proportions $p$ for the corresponding unknown parameters $\pi$ (given appropriate sampling schemes). The maximum likelihood estimates for the entries of the full table at the left hand side of Equation (5) are then easily calculated and so are the test statistics for the complete model.

The directed graphical modeling approach has certain advantages compared to the Goodman approach. For the simple recursive directed graphical models dealt with here, the parameter estimates and expected frequencies are very easy to calculate and no iterative procedures are required. Furthermore, the parameters being conditional probabilities have a very simple interpretation. Finally, the (marginal) tables to be dealt with may be much simpler, certainly at the end of the causal chains, and may contain much less variables than in the Goodman procedure.

A major disadvantage compared to the Goodman approach is that although the (marginal) tables to be used may be more parsimonious, the graphical (sub)models for these tables generally are not. Within the graphical approach, for each of the relevant marginal tables in Equation (5), viz. $A$, $B$, $ABC$, $ABCD$, and $BDE$, the pertinent loglinear (graphical) models are the saturated models, viz. $\{A\}$, $\{B\}$, $\{ABC\}$, $\{ABCD\}$, and $\{BDE\}$ respectively. Only if the ‘graphical’ (sub)models for the (marginal) tables are indeed saturated, the Goodman and the graphical approach yield the same results. Otherwise, the latter is needlessly complex and may not adequately reflect the hypotheses involved. For example, in Equation (5), the saturated model $\{ABCD\}$ for table $ABCD$ implies the existence of all three and four variable interaction effects of $A$, $B$, and $C$ on $D$, while the model in Figure 1 only implies direct main effects of $A$, $B$, and $C$ on $D$, plus one single three variable interaction effect ‘$BCD$’, but not higher order effects such as ‘$ACD$’ or even ‘$ABCD$’. The reason for the presence of these unwanted complex interactions in graphical models is that the absence of particular kinds of loglinear effects cannot be formulated in terms of conditional independence (the famous example being the no-three-variable-interaction model for a three-way table).

**DIRECTED LOGLINEAR MODELING (DLM)**

It is possible to combine modified path analysis with directed loglinear modeling, as Vermunt showed (Vermunt, 1996, 1997), leading to, what is called here, Directed Loglinear Modeling (DLM). Directed loglinear models define appropriate parsimonious loglinear
models for tables, that are as simple as possible in ‘graphical’ terms. Applied to the example of Figure 1, starting point is the graphical decomposition in Equation (5), resulting in a set of ‘simple’ (marginal) tables, for which the appropriate parsimonious loglinear models are defined. Because the model in Figure 1 does not imply any restrictions on the marginal distributions of A or B, the saturated model \{A\} is applied to marginal table A, and saturated model \{B\} to marginal table B. For observed marginal table ABC, saturated model \{ABC\} remains valid. For marginal table ABCD, model \{ABC, AD, BCD\} has to be defined. Finally, for the last (‘simple’) marginal table BDE in Equation (5), model \{BD, BE, DE\} is required by Figure 1. Each of these submodels provides the maximum likelihood estimates for the probabilities at the right hand side of Equation (5).

As follows from the above, DLM will lead to the same estimates as directed graphical models if and only if for all ‘simple’ (marginal) tables the saturated loglinear model is postulated. DLM always yields exactly the same parameter estimates, standard errors, and test statistics as the standard Goodman procedure, but using simpler (marginal) tables. Different from the Goodman approach is that in DLM the overall likelihood ratio chi-square test statistic \(L^2\) can no longer be obtained by summing the separate \(L^2\)'s for the subtables, but has to be computed in the usual way by employing the estimated probabilities at the right hand side of Equation (5) to estimate the joint probabilities at the left hand side (and comparing them with the observed probabilities).

Directed loglinear models provide an extremely flexible tool for the causal analysis of categorical data. The underlying logic of the approach mirrors the underlying logic of causal analysis, using tables and models that are as parsimonious as our complete causal model allows. Further, because all that is essentially done is fitting regular loglinear models to the appropriate (marginal) tables, everything that can be done with loglinear models in general can be done for each of the (sub)models and the causal model as a whole. Vermunt developed a very general, flexible and user friendly program \(\text{EM}\) that can actually and routinely carry out all necessary computations for rather complex models with a fair amount of variables (Vermunt, 1996, 1997).
DIRECTED LOGLINEAR MODELS WITH LATENT VARIABLES

One straightforward extension of DLM is the integration of the latent class model into it by introducing categorical latent variables, following the principles of the 'modified LISREL approach' (Hagenaars, 1990, 1993, Vermunt, 1996, 1997).

In the analysis of Figure 1 so far, it has been assumed that all variables have been perfectly validly and reliably measured. This assumption is relaxed in Figure 2 where, as in Figure 1, the variables A, B, and C represent Age, Gender, and Education respectively, but where Y now represents the categorical latent ('true score') variable Party Preference at time 1 and Z the categorical latent variable Party Preference at time 2. Variables D and E are now two observed (imperfect) categorical indicators of latent variable Y, and F and G are the (imperfect) categorical indicators of latent variable Z.

[figure 2 about here]

The 'modified LISREL model' in Figure 2 can be divided into two parts: the measurement part and the structural part. The measurement part represents the relations between the indicators D through G and the two latent variables Y and Z and essentially corresponds to a standard latent class model with two latent variables. The structural part represents the causal structure among the variables A, B, C, Y, and Z and is essentially identical to the causal structure in Figure 1.

Starting point for the analysis of the complete model in Figure 2 is (the partially unobserved) table YZABCDEFG with entries $\pi_{rsijklmn0}$. If all (conditional) independence relations among the variables implied by Figure 2 are taken into account, $\pi_{rsijklmn0}^{YZABCDEFG}$ can be (graphically) decomposed as follows:

$$\pi_{rsijklmn0}^{YZABCDEFG} = \pi^A_{ij} \pi^B_{kl} \pi^{ABC}_{ijkl} \pi^{ZBY}_{ijk} \pi_{rsj}$$

$$\pi_{lir} \pi_{nm} \pi_{ns} \pi_{os}$$

(6)

The first part at the right hand side of Equation (6), viz. $(\pi^A_{ij} \pi^B_{kl} \pi^{ABC}_{ijkl} \pi^{ZBY}_{ijk} \pi_{rsj})$ represents the structural part of Figure 2 and is analogous to Equation (5). The appropriate loglinear and logit models for each of the elements of the structural part also follow directly from the discussion above. Using the standard shorthand notation for hierarchical loglinear models
again, loglinear model \{A\} should be applied to marginal table A, that is, to \((\pi_i^A)\); model \{B\} to table B \((\pi_j^B)\); model \{ABC\} to table ABC \((\pi_k^{ABC})\); model \{ABC, AY, BCY\} to table ABCY \((\pi_r^{ABCY})\); and, finally, model \{BY, BZ, YZ\} to table BYZ \((\pi_s^{BYZ})\).

The second part of Equation (6), viz. \((\pi_i^{DY}, \pi_j^{EY}, \pi_k^{FZ}, \pi_r^{GZ})\) represents the measurement part of Figure 2. It can be written in this way and thus ‘added’ to the structural model because of the fact that D and E are only indicators of and are only ‘influenced’ by Y, and F and G are only determined by Z; no direct relationships exist among the indicators, nor are they influenced by any of the other variables in the model except Y and Z. In this case, without further restrictions, the relevant logit equations for the elements of the measurement part all correspond to the saturated loglinear model: for marginal table DY loglinear model \{DY\} is defined; for table EY, model \{EY\}; for table FZ, model \{FZ\}; and for table GZ, model \{GZ\}.

The maximum likelihood estimates for the parameters of all these loglinear models and the probabilities at the right hand side of Equation (6) can be obtained using observed table ABCDEFG with entries \(\pi_{ijklmn}^{ABCDEFG}\), employing the EM-algorithm or Newton-Raphson procedures or variants and combinations thereof. Complexities may arise because some or all parameters may be unidentified, because the maximum likelihood solution obtained may be a local rather than a global maximum or because parameter estimates may be boundary estimates, e.g., conditional probabilities equal to 1 or 0. All these possible estimation procedures and problems have been extensively discussed in the literature on latent class analysis (Goodman, 1974a,b, De Leeuw et al., 1990, Hagenaars, 1990, 1993, Vermunt, 1996, 1997). Vermunt’s program QEM mentioned above enables one to routinely estimate an enormous variety of directed loglinear models with latent variables.

In the next sections, it will be shown how these causal, categorical latent variable models can be used to account for several kinds of response error, first, in the next section, through modeling ‘random’ response error and, in later sections, through dealing with systematic response errors caused by ‘consistency’ effects and different survey conditions.\(^4\)
NOMINAL LCA

In the standard latent class model, i.e., in the ‘pure’ measurement part of DLM, the relationship between a latent variable and a particular manifest indicator is genuinely probabilistic, without any systematic interference of other variables or indicators. In that sense, LCA accounts for ‘random’ measurement error in the categorical indicators (Hagenaars, 1990, Section 4.4.1).

Given modern computers and programs, LCA can now easily be applied to rather large sets of indicators. Nevertheless, in this Section, a very simple model will be used with just one latent variable and only three indicators. Emphasis of the exposition will be on the flexibility of LCA as a ‘pure’ measurement model, discussing some similarities and differences with a few other widely used measurement models. The indicators are measures of Extrinsic Job Satisfaction and the data, analyzed before by Shockey, come from the Quality of Employment Survey 1977 Cross Section (Shockey, 1988). The data are presented in Table 1.

The standard latent class model for these data with one latent variable X and three indicators P, S, and B can be depicted as in Figure 3,

\[ \pi_{tijk}^{XPSB} = \pi_t^X \pi_i^P \pi_j^S \pi_k^B \]

(7)

Without further restrictions, the saturated model has to be defined for each of the elements at the right hand side of Equation (7). In this case, this is identical to directly applying loglinear model \{PX,SX,BX\} to Table 1 (Haberman, 1979, Chapter 10):
Because the indicators each have four response categories, the analysis of the data in Table 1 was started with four latent classes, treating all variables as categorical nominal level variables. The maximum likelihood estimates for this model, as for all models in this article, have been obtained by (a major new version of) EM. Model \{XP, XS, XB\} (Equation (7) or (8)), in which X has four categories fitted the data excellently: L^2 = 15.11 df = 24 p = .92 (and Pearson's X^2 = 14.31). The same model, but now with three latent classes did not fit the data at all: L^2 = 66.36 df = 34 p = .00 (X^2 = 72.25), while the model with five latent classes appeared not to be identified (as deduced from the rank of the estimated variance-covariance matrix of the parameter estimates).

From the parameters estimates of the four latent class model, viz. the (conditional) response probabilities in Equation (7) presented in Table 2a, and the excellent fit of the model, it is concluded that the three indicators may be regarded as measurements of one underlying latent variable X 'Extrinsic Job Satisfaction', where the four categories of X have similar meanings as the four categories of the observed variables. The indicators are not perfect measurements of this latent variable: the probabilities of a correct answer, an answer corresponding with the latent class a person belongs to, are far from 1. The estimates \( \pi^X_i \) provide the estimates for the distribution of the latent variable Extrinsic Job Satisfaction. By incorporating this latent class model into a general causal model in the manner described in the previous section, the causes and consequences of Extrinsic Job Satisfaction (X) can be determined, corrected for the unreliability of the indicators P, S, and B.

\[ \ln \pi^X_{t,j,k} = \lambda^X_i + \lambda^P_j + \lambda^S_k + \lambda^{XP}_{t,j} + \lambda^{XS}_{t,j} + \lambda^{XB}_{t,j} \] (8)

**OLCA**

The standard LCA in which all variables are treated as nominal level variables is a very flexible model. This flexibility certainly has many advantages, as shown, for example, by
Hagenaars and Halman (1989). At the same time, it is a very complex model in the sense that it contains a large number of parameters. Moreover, it may ignore some important available information about the possible nature of the relations between the manifest and latent variables. In this example, the scores of the three indicators P, S, and B are ordered and it seems useful to carry out a latent class analysis taking this ordered character of the categories into account.

Most proposals for dealing with the ordered character of the measurements either treat the variables essentially as nominal level measurements and just investigate whether or not the outcomes happen to turn out as ordered (as in the nominal LCA above) or they imply linear restrictions on the relations between the variables and thus implicitly assume that the variables have been measured at interval level. Few techniques exist for purely ordinal level variables and analyses (Agresti et al, 1987, Dykstra and Lemke, 1988).

Croon has developed some truly ordinal versions of LCA (Croon, 1990, 1993a, 1993b, Croon and Luijkx, 1992), in which he defines an ordinal relation between the latent variable X and an indicator A in terms of (weakly) monotonically in- or decreasing local odds, continuation odds or cumulative odds (for a definition of these terms, see also Agresti, 1990, Section 9.3). In his procedure OLCA (Ordinal Latent Class Analysis), that will be used here, ordinal (inequality) restrictions on particular cumulative odds are imposed.

The basic idea behind OLCA is that if the relationship between the indicator A, with ordered categories $i = 1,...,I$, and a latent variable X, with ordered categories $t = 1,...,T$ is truly ordinal, then, wherever the ordered scores of A are dichotomized, keeping their order, there should be a (weakly) monotonically in- or decreasing relationship between X and the dichotomized variable A. Assuming a positive relationship between A and X, the cumulative response probabilities and therefore the cumulative odds should obey the following (weak) inequality restrictions:

\[
\begin{align*}
\sum_{t=1}^{k} \pi_{i,t}^{A|X} & \geq \sum_{t=1}^{k} \pi_{i,t+1}^{A|X} \\
\frac{\sum_{i=k+1}^{l} \pi_{i,t}^{A|X}}{\sum_{i=1}^{k} \pi_{i,t}^{A|X}} & \leq \frac{\sum_{i=k+1}^{l} \pi_{i,t+1}^{A|X}}{\sum_{i=1}^{k} \pi_{i,t+1}^{A|X}}
\end{align*}
\] (9)
If the latent class model \{XP,XS, XB\} with four latent classes is estimated for the data in Table 1, subjected to the inequality restrictions in Equation (9), the outcomes in Table 2a hardly change. After rounding off to two decimals, all parameter estimates but two change less than .02 (in absolute values). The exceptions concern indicator P, as indicated in the note below Table 2a. The nominal outcomes violate the restriction in Equation (9): the cumulative probability of answering (1 or 2) on indicator P is smaller for X=1 than for X=2. In OLCA, these two cumulative conditional probabilities are set equal to each other: (.6103 + .1549) = (.1564 + .6088) = .7652. (And of course, a similar equality then applies to the cumulative probabilities of answering (3 or 4) on Indicator P for X=1 and X=2.) A further order violation occurs with regard to indicator S, where, as also indicated below Table 2a, OLCA imposes another equality.

The fit of the ordinal model is almost the same as the fit of the standard nominal model: $L^2 = 15.55$ ($X^2 = 14.77$) instead of $L^2 = 15.11$ ($X^2 = 14.31$) obtained before. Evaluation of the test statistics for the ordinal model is difficult, because the appropriate number of degrees of freedom is hard to determine. The nominal LCA had 24 degrees of freedom. Because the (weak) inequality restrictions in Equation (9) led to two independent strict equality restrictions (for P and S), one might think that OLCA in this case has 26 degrees of freedom. However, in another sample from the same population in which the ordinal model (Equation (9) holds, application of the same ordinal latent class model may result in a different number of strict equality restrictions and, consequently, in a different number of degrees of freedom. Obviously the number of degrees of freedom is not fixed for a particular ordinal model, but appears to have become a random variable. How to determine in that case the corresponding level of significance for the test statistics is an as of yet unsolved problem (see the Croon references above and Dijkstra 1992, Shapiro, 1985).

Another problem with OLCA is that the occurrence of local maxima seems to be a bit more of a problem than in standard (nominal) LCA. Nevertheless, given the fact that a large part of social science data must be considered as ordinal, rather than nominal or interval level data, developments such as OLCA are of the utmost importance.

**INTERVAL LEVEL LATENT VARIABLES**

One disadvantage of strictly ordinal analyses was not mentioned above: they do not diminish the number of parameters to be estimated. Models in which it is assumed that the variables
concerned are actually measured at interval level and have linear relationships with each other provide much more parsimonious descriptions and explanations of the data and, for that reason may better suit our purposes, if indeed the interval and linearity assumptions are valid.

If all categorical variables are treated as interval level variables, an interesting loglinear model is the linear-by-linear association model (Haberman, 1979, Goodman, 1984, Clogg and Shihadeh, 1994). For the latent class model in Equation (8), the linear-by-linear model implies the following restrictions on the two-variable association parameters

\[ \lambda_{xp}^{xt} = (\lambda_p)(X_i)(P_i) \]
\[ \lambda_{xs}^{xt} = (\lambda_s)(X_i)(S_j) \]
\[ \lambda_{xb}^{xt} = (\lambda_B)(X_i)(B_k) \]

where \( X_i \) denotes the ordered scores on the latent variable \( X \), restricted to sum to zero and with equal unit distance. \( P_i \), \( S_j \), and \( B_k \) have analogous meanings and are similarly restricted. The parameters \( \lambda_p \), \( \lambda_s \), and \( \lambda_B \) indicate the strength of the relationship between the latent variable and its indicators; they are comparable to the factor loadings in ordinary factor analysis. The first order effects \( \lambda_p \), \( \lambda_s \), and \( \lambda_B \) in Equation (8) are related to the 'difficulty' of the items, comparable to the item difficulties defined in Guttman or Rasch scaling models. As Heinen has extensively shown, latent class models with restrictions such as in Equation (10) have much in common with (discretized) latent trait models and categorical factor analysis models (Heinen, 1993, 1996; see also Formann, 1992, Rost, 1988, Mellenbergh, 1994, Van der Linden and Hambleton (1997), Rost and Langeheine (1997)).

For four latent classes, the linear-by-linear latent class model had to be rejected when applied to Table 1: \( L^2 = 115.79, df = 48, p = .00 (X^2 = 132.82) \). Increasing the number of latent classes to five did not really improve the fit of the model: \( L^2 = 107.01, df = 47, p = .00 (X^2 = 120.52) \). Moreover, one of the five latent classes was almost completely empty. If still more latent classes were added, only more (nearly) empty latent classes resulted, but not a better fitting model. Even a truly latent trait model in which an underlying, normally distributed, continuous variable is assumed to exist which has a similar linear-by-linear relationship with the indicators fits as badly as the ordinary linear-by-linear LCA model.

Although the four latent class model did not fit, its parameter estimates are presented in Table 2b. A comparison of the estimated response probabilities in Table 2a (LCA or OLCA)
and Table 2b (linear-by-linear LCA) clearly show why the latter model fails. Although the relations between the latent variable and its indicators are certainly ordinal, they are definitely not linear. Within the first three latent classes, the conditional distribution of the responses is much more ‘peaked’, more strongly ‘unimodal’ according to the nominal/ordinal LCA (Table 2a) than according to the linear LCA (Table 2b). Moreover, in the nominal/ordinal LCA, the modal category of the indicator is always the category that corresponds with the latent class, but not so in the linear LCA. The linear model smooths the conditional response probabilities too much. It seems, at least in this case, as if when answering the questions, perhaps conditioned by the precoded four-categories-answers, the respondent has only a few crude, ordered categories in mind rather than a continuous ‘scale’. Factor analysis and LISREL type models, using only the first and second moments of the distributions would never notice this phenomenon, unlike (the nominal or ordinal) LCA.

However, before drawing these far-reaching conclusions about the response process, an alternative explanation of the failure of the linear-by-linear model has to be explored. It is possible that the underlying variable X is an equally spaced interval level variable, but the indicators are not. The manifest scores are most probably ordered but possibly with unequal, unknown intervals between them. Within the loglinear framework, this situation can be dealt with by means of the Column association model (Goodman, 1984, Clogg and Shihadeh, 1994). Applied to the latent class model, it means that the latent variable X (the ‘row variable’) is still treated as an interval level variable, but each indicator (the ‘column variable’) is essentially treated as a nominal level variable. The ideas underlying this particular latent class model are strongly related to methods for dealing with manifest categorical variables in continuous latent variable models such as LISREL, that make use of thresholds and probit transformations rather than, as here, of a logistic one; see, among others, Muthén, 1988, Jöreskog and Sörbom, 1993, Arminger, 1995, Browne and Arminger, 1995, Kühnel, 1996, Long, 1997).

In this case, the column association model implies the following restrictions on the associations between the latent variable X and the indicators in Equation (8)
The restrictions in equation (11) have the same structure as the restrictions for the linear-by-linear model in Equation (10). The only difference is that now the category scores for the indicators S, P, and B are not fixed, but are parameters to be estimated, viz. \( \kappa_i \), \( \mu_j \), and \( \nu_k \). Two identification restrictions are needed for these parameters, often \( \sum \kappa_i = 0 \) and \( \sum \kappa_i^2 = 1 \), and similar restrictions for \( \mu_j \) and \( \nu_k \). The optimal spacing of the observed scores is estimated from the data for each indicator, assuming that the relationship between the latent variable \( X \) and the indicator is linear. The more the response categories are perceived by the respondents as expressing more or less the same intensity of the underlying (latent) attitude, the closer their estimated scores are expected to be to each other. Usually the estimated scores in cases like in our example will turn out to be ordered, but this is not necessary. For that reason this model is called an interval-nominal, rather than a interval-ordinal LCA.

Although it may seem that the restrictions in Equation (11) involve products of parameters and that therefore the resulting model is no longer an ordinary loglinear model, this is not true. The restrictions in Equation (11) maybe rewritten in a normal loglinear form as follows

\[
\begin{align*}
\lambda_{ij}^{XP} &= (\lambda_p)(X_i)(\kappa_j) \\
\lambda_{ij}^{XS} &= (\lambda_p)(X_i)(\mu_j) \\
\lambda_{ik}^{XB} &= (\lambda_p)(X_i)(\nu_k)
\end{align*}
\]

where \( \lambda_p \) in Equation (12) equals \( (\lambda_p \kappa_j) \) in Equation (11), etc. and the scores \( X_i \) are fixed.

Despite the fact that from a substantive point of view this particular latent class model is a very attractive one, and certainly more flexible than the linear-by-linear model, it does not fit the data when four latent classes are defined: \( L^2 = 105.86, \text{df} = 42, \ p = .00 \) (\( X^2 = 125.31 \)). And again, adding more latent classes did not improve the fit and only led to more (nearly) empty latent classes. This ‘weakly linear’ model has to be rejected too and much for the same reasons as the standard linear-by-linear model: the conditional response probabilities for the
Column association LCA show the same patterns as found for the linear-by-linear LCA in Table 2b.

Other variants of this approach could be and have been tried, variants in which the scores on the latent variable are estimated but the observed scores treated as fixed (see also Luijkx and Hagenaars, 1990) and models in which the scores on both latent and manifest variables are estimated.\(^6\) However, none of these models did fit; adding extra latent classes did not make a real difference and the pattern of the estimated conditional response probabilities was very similar to the outcomes for the linear-by-linear LCA in Table 2b. Moreover, it is debatable whether it is a good methodological practice to assume the existence of not directly observed latent variables, estimate the proportions of the people in the latent classes and then on top of that also estimate the latent scores.

It turns out to be possible to define restricted, parsimonious LCA models within the log-linear framework, but, at the same time, it is an important advantage that the assumptions needed to justify these parsimonious models can be tested in a straightforward way by comparing the results with the outcomes of less restricted latent class analyses.

**CORRELATED ERRORS OF MEASUREMENT**

Because of the 'random error' structure dealt with in the previous section, the latent class model amounted to a relatively straightforward loglinear latent variable model with a relatively simple causal structure that could be represented by one basic equation, viz. Equation (8). However, more complicated error structures may require more complicated causal models. A case in point is the analysis of longitudinal data where possible test-retest effects and autocorrelated response errors have to be reckoned with. Although the example chosen is a real world example, no data analyses will be presented in this Section. emphasis will be on the kinds of analyses one could carry out.

**THE SIPP PANEL**

The example that will be used is taken from a paper by Bassi et al. (1995). It concerns changes in the Labor Market Status. This central variable has three categories: Not in the labor force,
Unemployed, Employed. (For a similar example, using related models, see Van de Pol and Langeheine, 1997.) Bassi et al. use data from the Survey of Income and Program Participation (SIPP) panel, that was started in 1984 by the US Bureau of Census.

To understand the response errors that appear to occur in the SIPP panel, an elementary understanding of the SIPP interviewing scheme is needed. The respondents of the SIPP panel are divided into four rotation groups. In the beginning of every month (e.g., May), the members of one rotation group (e.g., R1) are interviewed and asked a number of questions about their labor status during the preceding four months (January through April), beginning with the month (actually, the week) closest to the moment of interviewing (April), then the next closest (March), etc. This rotation group R1 will be re-interviewed four months later (the beginning of September) and asked questions about the then previous four months (May through August), etc. Another rotation group (R2) starts a month later (June) and has an analogous schedule; in a similar way there is a third rotation group (R3-interviewed in the beginning of July) and a fourth one (R4-interviewed in the beginning of August). A particular rotation group stays active for eight consecutive interviewing periods and thus provides information over an uninterrupted period of 32 months. As a consequence of this rotation scheme, the information concerning the labor market status in one particular month is provided by four different rotation groups and gathered at the beginnings of four different consecutive months.

The main focus of this article will be the analysis of one group (R1). How to analyze several rotation groups at the same time is shown by Bassi et al. (1995) (and in the next section). Because in our example R1 has been interviewed in the beginning of May, the information on the labor market status gathered during this interview refers to the labor market states in January, February, March, and April and will be denoted as A, B, C, and D, respectively.

[figure 4 about here]

The gross turnover in labor market states is often analyzed by means of a markov model. The causal model corresponding with a first order markov chain is depicted in Figure 4a. The corresponding equation is
If for each of the ‘marginal tables’ at the right hand side of Equation (13) the saturated loglinear model is postulated, we have a first order, nonstationary markov chain in which the only direct effects are from month $t-1$ to month $t$. Stationary first order markov chains can be obtained by defining loglinear models in which the transition probabilities $A-B$, $B-C$, and $C-D$ are restricted to be equal to each other (Bishop et al., 1975, Chapter 7, Vermunt, 1996).

Second order markov chains result from adding extra direct effects from month $t-2$ to month $t$; still higher order chains are defined in an analogous way.

**RESPONSE ERRORS**

Using the data on labor market status from the SIPP panel, it is possible to arrive at well fitting causal loglinear models in the form of a markov chain, if necessary by relaxing the first-order or stationarity assumptions. However, as explained by Bassi et al., the outcomes show some peculiar regularities. First, the bivariate turnover tables for consecutive months determined for one rotation group within one particular interview (e.g., $A-B$, $B-C$, $C-D$ in Figure 4a) show too little gross change in labor status, less than in reality, that is, less than is expected from knowledge provided by reliable sources. Second, the turnover between two consecutive months that are close to the moment of interviewing ($C-D$) is larger than the turnover between two consecutive months that are farther away from the date the interview takes place ($A-B$). Further, when the analysis is extended to eight or more points in time, it generally appears that the turnover between two months for which the information is gathered within the same interview is much less than the turnover between two months for which the information is obtained from interviews at different points in time. Finally, there is the well documented Seam Effect. Analyzing the data for all four rotation groups, it follows from the interviewing scheme employed in the SIPP that the turnover between two particular months, e.g., April-May, is determined in one rotation group from the information obtained at two separate moments of interviewing, viz. in the beginning of May and in the beginning of September, but for each of the other three rotation groups, from the information gathered within one and the same interview at one moment in time, viz. in the beginning of June, July,
and August, respectively. The label Seam Effect is used for the phenomenon that the gross turnover in labor market status between two consecutive months is always much less for the three rotation groups where the information comes from one interview than for the one rotation group where two separate moments of interviewing have been used.

All these peculiarities forces one to conclude that the observed Labor Status is not a completely valid and reliable measurement of the respondents' true labor status and that particular kinds of response errors must have occurred. Latent markov chains have been proposed to correct for these response errors (Wiggins, 1955, 1973, Poulsen, 1982, Van de Pol, 1989, Van de Pol and Langeheine, 1990, 1997). The directed loglinear modeling proposed here makes their approach more general and flexible.\(^7\)

A straightforward translation of the first-order markov model of Figure 4a into a latent markov model is presented in Figure 4b and in Equation (14), where W, X, Y, and Z are the trichotomous latent Labor Market States in January, February, March, and April respectively and where the first part at the right hand side of Equation (14) represents the structural (markovian) component of the latent variable model and the second part the measurement model.

\[
\begin{align*}
\pi_{i w v j k l} & = (\pi_i \pi_w \pi_x \pi_y \pi_z) \\
&W, X, Y, \text{ and } Z \text{ are the trichotomous latent Labor Market States in January, February, March, and April respectively and where the first part at the right hand side of Equation (14) represents the structural (markovian) component of the latent variable model and the second part the measurement model.}
\end{align*}
\]

When the latent markov model in Equation (14) is applied to the SIPP data, using saturated loglinear models for each of the (conditional) probabilities at the right hand side of Equation (14) and extending the analysis to more points in time and more rotation groups, the results are not satisfactory. The turnover at the latent level turns out to be even smaller than at the manifest level (that already was too small) and also all the other peculiarities mentioned such as the seam effect do not disappear. Obviously, the latent class model in the form of the latent markov model in Figure 4b deals with the response error in the wrong way.

Crucial to the model in Figure 4b is that, as explained before, the response error is assumed to occur 'randomly', or, as it is often called in the literature on misclassification: the Independent Classification Error assumption (ICE) is made here. The observed scores on a particular indicator only depend probabilistically on the latent variable that indicator is supposed to measure and, consequently, the mismatches between the true states and the reported states at
time $t$ are not associated with the mismatches at time $t+k$, when the latent variable for time $t$ and/or $t+k$ are held constant.

However, it may be questioned whether the ICE assumption is a reasonable one, given the SIPP interviewing scheme, where information about the previous four months is collected cross-sectionally at one moment of time. During this one interview, due to all kinds of conditioning effects, memory effects, test-retest effects, tendencies to appear consistent, etc, the answers and response errors concerning the labor market status for the different months may be directly correlated with each other. If there are indeed 'conditioning effects' during the (cross-sectional) interview, then it is necessary to replace the strict 'random' ICE assumption by a milder one and to define local dependence LCA models (Hagenaars, 1988, 1990).

**UNOBSERVED HETEROGENEITY**

One way to account for possible direct correlations among the response errors is the introduction of an extra latent variable $V$ that is directly related to all indicators $A$ through $D$, but (at least here) uncorrelated with the proper latent variables $W$ through $Z$. The resulting model is presented in Equation (15) and (in a more parsimonious form) in Figure 4c:

$$
\pi_{r,s,t,u,v,i,j,k,l} = \frac{\pi_{r,s,t,u,v,i,j,k,l}}{\pi_{s,t,u,v,i,j,k,l}} \cdot \frac{\pi_{r,s,t,u,v,i,j,k,l}}{\pi_{r,s,t,u,v,i,j,k,l}}.
$$

Because an extra latent variable $V$ was added without additional observed indicators, Equation (15) has to be examined very carefully for identifiability problems. One possible way to achieve identifiability is to define more restrictive models, for example, by defining no-three-variable interaction models for the three-way tables involving latent variable $V$ (as shown in Figure 4c). Although this may not be sufficient for ensuring identifiability, in general, it certainly helps.

Because $V$ is treated here as a categorical latent variable, the number of categories of $V$ has to be determined. One usually starts with two categories and adds more and more categories till the model is no longer identified or a good fit with the data has been obtained. As there are seldom or never theoretical reasons for thinking that $V$ has a certain number of categories, this is the only way to proceed. Actually, its atheoretical nature is the most serious objection one might raise against the introduction of an extra latent variable such as $V$: it is hard to find a
compelling substantive interpretation of the latent variable. Essentially, V is just a variable modeling unobserved heterogeneity (Heckman and Singer, 1982, DeSarbo and Wedel, 1993, Vermunt, 1996), taking care of that part of the association among the indicators that is not explained by the other, proper latent variables in the model without providing a clue as to what V might stand for. V just accounts for the 'correlated error terms', but, from a substantive point of view, in an unknown way. Even when the model fits the data well, the results are usually not very informative, because almost any interpretation can be given to V. Therefore, preference should be given to models that somehow incorporate the presumed nature of the 'conditioning process' in a more theoretical way.

**MODELING SYSTEMATIC RESPONSE ERROR**

One possible reason for the presence of correlated error terms might be that during the cross-sectional interview, the respondents have the tendency to make all answers about their labor status during the previous four months consistent with their true position at the time of the interview. Because latent variable Z in Figure 4 represents as closely as possible this true position, this hypothesis implies that Z has a direct influence, not only on D, but also on A, B, and C. Equation (16) and (in more parsimonious form) Figure 4d represent this hypothesis:

\[
\begin{align*}
\pi_{wxyzabcd}^{tuvijkl} &= \pi_w^{w} \pi_y^{w} \pi_x^{w} \pi_z^{w} \pi_y^{w} \pi_z^{w} \\
\pi_{ijsv}^{wxyz} \pi_{jtv}^{wxyz} \pi_{kuv}^{wxyz} \pi_{lvi}^{wxyz} 
\end{align*}
\]  

Restricted (no-three-variable interaction) loglinear models may be applied to the conditional probabilities at the right hand side of Equation (16) and may be necessary to achieve identifiability.

Another reason for the correlation among the response errors might be that respondents have the tendency to adapt their answers not to their present latent positions, but to the first given answer. Given the fact that, during the cross-sectional interviews, the questions go back in time, the answer given first is represented by indicator D and thus D is supposed to have a direct effect on A, B, and C, as represented in Equation (17) and (parsimoniously) in Figure 4e.
Still another plausible assumption might be that a particular response is not made consistent with the first given answer, but with the immediately previous one, giving rise to the model in Equation (18) and (more parsimoniously) in Figure 4f.

\[
\pi_{stuvijkl}^{WXYZABCD} = \pi_{x}^{W} \pi_{z}^{X} \pi_{y}^{Y} \pi_{v}^{Z}.
\]

Which of the above models in Equations (13)-(18) should be preferred depends on what the investigator deems plausible and on the fit of the model when empirically tested. And, of course, many variations of the above models can be formulated. For example, Bassi et al., without having tried all models presented here, finally decided upon a model proposed by Hubble and Judkins (1987) in which essentially the ICE assumption is applied to all relationships among the measurements obtained in different interviews; however, for the measurements obtained within one and the same interview, it is assumed that, once a mistake is made and a mismatch occurs between the latent true and the manifest position, all next answers (going back in time) will be exactly in agreement with that mistake.

Once a particular (valid) measurement model for the correlated response errors has been obtained, the complete model provides the estimates for the parameters of the structural part, corrected for these response errors. The structural part itself may assume many forms, different from the one discussed above. For example, the assumptions of the latent Markov chain may be relaxed by defining higher order Markov chains or by allowing nonstationarity. The nature of the latent change may be further restricted by excluding certain transitions between particular latent states, by assuming no latent change between particular points of time or by assuming 'linear' latent change, etc. (Collins and Wugalter, 1992, Hagenaars, 1990, 1994, Jagodzinski et al., 1987).
Directed loglinear models with latent variables provide a very flexible tool to model simultaneously many kinds of latent changes and many types of response errors. At the same time, and this cannot be emphasized strongly enough, flexibility changes from a virtue into a vice when the models are not theoretically based and are only used as an arbitrary data fitting device. The introduction of correlated error terms, in one way or another, will almost always 'save' a model that otherwise would have been rejected (Hagenaars, 1988). However, correlated error terms often violate the very basic measurement assumptions the analysis started with and should not be applied unless for a very good theoretical reason.

CAUSAL MODELING, MULTIPLE GROUP COMPARISONS AND INTERVIEWING CONDITIONS

STANDARD MULTIPLE GROUP LCA

Response errors may assume different forms in different groups and in different research settings. Often, an explicit causal modeling approach is needed to account for such effects, as will be shown through a re-analysis of a data set on (again) labor market behavior, analyzed before by Shockey and coming from the June 1980 Current Population Survey (Shockey, 1988). The data are presented in Table 3.

The central dependent variable, Labor Force Status has again three categories: Not-in-the-labor-force, Unemployed, and Employed. The data in Table 3 have been provided by three different groups, viz. rotation group I, VI, and VII, where group I contains first-time respondents and groups VI and VII respondents that have been interviewed for the sixth and seventh time respectively. Given the sampling procedure employed in the Current Population Survey, each rotation group is in principle a representative sample of the US working-age population. Therefore, the unemployment rates in Table 3 should be the same for all three groups within the boundaries of sampling error. However, in Table 3, there exist statistically significant differences between the percentages Unemployed in the three groups: 5.84%, 4.83%, and
As Shockey notices, "...the unemployment rate reported in the first rotation group has consistently been higher than that reported for any of the other groups since the late 1950's. (Shockey, 1988, p. 304). A possible explanation of this phenomenon might be provided by the somewhat different interviewing conditions for the first and the other rotation groups. Most interviews in the first group are face-to-face interviews, while the interviews for the other groups are usually conducted over the telephone. Further, sometimes the information is provided by the respondent, sometimes by a 'proxy'. Although 'this (latter-JH) form of error is introduced evenly to all groups and should not directly lead to the rotation group problem', it 'may act to exacerbate it' (Shockey, 1988, p. 305).

If latent variable $X$ denotes the true Labor force Status in June 1980 with the same three categories as the corresponding manifest variable $L$ (Table 3), the purpose of the analysis of Table 3 is to find estimates for $\pi^X$, the true proportions not-in-the-labor-force, unemployed, and employed in the population, corrected for the extent to which the interviewing conditions $G$ (rotation groups), $F$ (interview format), and $R$ (respondent type) have led to false reports on one's labor force status. Shockey tries to accomplish this by making use of Goodman/Clogg's multiple group LCA (Clogg and Goodman, 1985). In terms of the variables in Table 3, treating $G$ as the grouping variable, he starts with the basic multiple group model

$$\pi^X_{ijkl} = \pi^X_i \pi^G_{ji} \pi^L_{ji} \pi^X_{ji} \pi^R_{ji} \pi^F_{ji}$$

(19)

which is identical to loglinear model \{XGL,XGR,XGF\}. Because he assumes that the true labor force status and response type $R$ are the same for all three rotation groups, the standard multiple group model is simplified to loglinear model \{X,G\} for table $XG$ and \{XGL,XR,XGF\} for table $XGLRF$, resulting in a simplified version of Equation (19):

$$\pi^X_{ijkl} = \pi^X_i \pi^G_{ji} \pi^L_{ji} \pi^R_{ji} \pi^F_{ji}$$

(20)

Partly for reasons of identifiability, Shockey imposes some additional equality restrictions on the parameters of Equation (20), but these will be ignored here as they are not relevant for the argument to be developed.

According to Shockey, the parameter estimates of $\pi^X_i$ in Equation (20) should provide the estimates of the true proportions of Unemployed, Employed and Not-in-the-labor-force; the
estimates of the conditional response probabilities $\pi_{jtk}^{LXG}$, $\pi_{kti}^{RX}$, and $\pi_{jti}^{FXG}$ should tell us how the interviewing conditions distorted the true labor force situation. However, for several reasons, his analysis is not convincing. One minor reason is that the model in Equation (20) may be needlessly complicated. Once the multiple-group LCA is recognized as a loglinear model with a latent variable, it is possible, as Hagenaars has shown, to test simpler, but still meaningful loglinear models in which, for example, the complicated three-variable interaction terms are left out and no-three-variable-interaction models are employed for marginal tables XGL and XGF (Hagenaars, 1985, Section, 3.5; 1990, Section 3.6, McCutcheon and Hagenaars, 1997). A more fundamental objection is that Shockey's analysis is not a logical one from a substantive point of view. To explain this point, it is helpful to make use of the causal diagram in Figure 5, that underlies the basic multiple group latent class model of Equation (20).

As appears clearly from Figure 5, in Shockey's analysis, the interviewing conditions R and F are treated as 'indicators' of X, together with the genuine indicator L. Consequently, in agreement with the assumption of local independence, for each latent category of X within each rotation group, the 'indicators' R, F, and L are independent of each other. There are no direct effects of R and F on L. This is strange as the whole analysis started from the hypothesis that the rotation groups reported different (un)employment rates L because of the direct influences of F and R on L.

**APPROPRIATE CAUSAL MODELS**

To test the validity of the hypotheses in hand, it is necessary to set up a causal loglinear model with a latent variable in which G and X (the presumed true labor status) are the exogenous variables, L is the ultimate dependent variable, and F and R are the intervening variables. Because there are no indications as to how the causal order between F and R might be, F and R are placed within the same (causal) block. The basic equation for the resulting block-recursive model is

$$\pi_{ijkl}^{XGRLF} = \pi_{ijkl}^{XG} \pi_{ijkl}^{RFXG} \pi_{ijkl}^{LXGFR}$$

(21)
Because it is assumed that the true (un)employment rates are the same for all three groups, \( \pi_{1tG}^{XG} \) in Equation (21) can be replaced by \( \pi_{1tG}^{XG} \); in other words, for marginal table \( XG \) independence model \( \{X, G\} \) is postulated. If it is further assumed that there are no three-variable interaction effects on the dependent variable, the appropriate model for \( \pi_{1tG}^{XG} \) (marginal table \( XGRF \)) is \( \{XG, RF, XR, XF, GR, GF\} \) and for \( \pi_{1tG}^{XG} \) (table \( XGLRF \)) it is model \( \{XG, RF, LR, RL, LF\} \). The causal diagram for Equation (21) with the above (log-linear) restrictions is depicted in Figure 6 (both solid and dotted lines). This restricted version of Equation (21) fits the data well: \( L^2 = 9.376, \text{df} = 6, p = .15 (X^2 = 9.358) \).

Further analyses made it clear that the effects represented by the dotted lines in Figure 6 were not significant and could be deleted. This resulted in the following (ordinary recursive) model

\[
\pi_{1tG}^{XGLRF} = \pi_{1tG}^{XG, RF, XR, XF, GR, GF, LR, RL, LF} \]

(22)

in which model \( \{X, G\} \) is defined for table \( XG \), model \( \{XR\} \) for table \( XR \), model \( \{GX, GF, XF\} \) for table \( GXF \) and model \( \{GXF, GL, XL, LF\} \) for table \( GXLF \). The test results for the thus restricted model (22) are: \( L^2 = 14.889, \text{df} = 11, p = .19 (X^2 = 14.792) \).

Given Shockey's central hypotheses, it is important that the direct effects of \( G \) on \( L \) cannot be deleted: the rotation group differences in reported unemployment cannot be fully explained through the mediating effects of \( F \) and \( R \). Further, it turns out that both \( G \) and \( F \) directly cause bias in the (un)employment reports. Respondent type \( R \) has no direct influence on \( L \) and is not an 'exacerbating' source of bias in the reports.

**MAKING SENSE OF THE OUTCOMES**

If the model in Figure 6 (solid lines only) is valid, the estimates \( \pi_{i}^{X} \) within this model should provide unbiased estimates for the proportion (un)employed in the population. The Total Row in Table 4a provides the relevant outcomes.
The outcomes are very strange. The smallest proportion, the one pertaining to latent class 2 amounts to .172, which would imply an unbelievable high estimate of the number of unemployed people at that moment in the USA, assuming that latent class 2 represents the Unemployed. But what does it represent? For this, given the intended meaning of the latent variable X, the most important clue is provided by the nature of the direct relationship between X and L. Tables 4a and 4b present the relevant estimates. In Table 4a, the conditional response probabilities for L, given X are given for those respondents who belong to the first rotation group and have been interviewed face-to-face. Choosing this particular conditional table XL (for GF = 12) is arbitrary, but the other conditional tables XL, for the other combinations of the categories of F and G showed essentially the same patterns. Further, in Table 4b, the direct loglinear effects X-L are presented. Because cell XL = 12 is (almost) empty, the size of the parameter estimates reported in Table 4b have to be interpreted with care. Replacing the zero in this cell by small figures like .001 or .0001 results in much less extreme estimates of the parameters, although always the same pattern of outcomes is seen. Now, no matter which figures are being used and how we look at the relationship between X and L, it is clear that the categories of X cannot be interpreted in terms of: Not-in-the-labor-force, Unemployed and Employed (see also note 8). Moreover, given the strengths of the relationships X-R and X-F (Tables 4c and 4d), it appears that the ‘meaning’ of X is determined as much by these relationships as by the relation of X with its single ‘genuine’ indicator L, resulting in a latent variable that cannot be given a coherent and intelligible substantive meaning.

NUMERICAL PROBLEMS

Essentially, the latent variable is not or too weakly identified. In close connection to this, a number of numerical problems occurred when trying to find the maximum likelihood solutions. Hundreds of runs with different starting values were needed to arrive at the above results, most of them leading to local maxima; small changes in the model specification resulted in very different outcomes; the number of iterations needed often ran into the hundreds of thousands; accuracy of the stopping criterion had to be set at an unbelievable small number; all kinds of variants of EM- and Newton/Raphson-algorithms had to be applied; even now, it would come as no surprise if somebody came up with a solution that yielded lower L²’s than the ones found above for the same models. Finally, it might well be that the above mentioned solutions are not identified or only identified given that the estimated limiting
values of 0 or 1 that were found for some of the estimated conditional probabilities are indeed valid in the population.

**OVERCOMING THE DIFFICULTIES**

Given that the crucial problem is the weak identification of the latent variable, the solution must be found in strengthening the relationship between L and X. Following Clogg, it is possible to introduce a latent variable X that has three ‘perfect’ latent classes with a conditional response probability of one to report their observed status as Not-in-the-labor-force, Unemployed, and Employed respectively (and a probability of zero to give the ‘wrong’ answer) and, in addition, one or more extra (restricted) latent classes that show the ‘normal’ probabilistic relations between X and L to account for response error (Goodman, 1975, Clogg, 1981b, Hagenaars, 1990, Section 4.4.2). This ‘extended’ latent variable X was introduced into the causal model of Figure 6 in several ways hoping to arrive at a latent variable with a more clear relationship with L and a more intelligible interpretation. However, to no avail: the extra ‘probabilistic’ latent classes were not at all interpretable in terms of labor force status.

When everything else fails, it might be hoped that it is at least possible to carry out some kind of sensitivity-for-response-error analysis. Even when it is impossible to obtain from the data direct estimates of the amount and nature of the response error, it may still be useful to set up a model with a latent variable X, supposedly indicating the true Labor Force Status, that is measured with varying degrees of a priori fixed unreliability, that is, with varying strength of the relation X-L. The purpose of the analysis is then to explore the extent to which the conclusions about the number of (un)employed people and about the strengths of the other relationships within the causal model are sensitive to the varying degrees of unreliability in the measurement of X. For the causal model in Figure 6, the relevant parameters that indicate the ‘reliability’ with which X is measured are $\lambda_{ij}$. By using the appropriate ‘weights’ for the initial estimates and the correct estimation procedures (incorporated in $\hat{\theta}$EM), it is possible to carry out analyses with fixed loglinear effects (Haberman, 1979, Chapter 9, Bishop *et al.*, 1975, Section 3.6, Rudas and Leimer, 1992). Several degrees of reliability were tried, roughly corresponding with (partial) conditional response probabilities of .70 through .95 of giving the correct answer, but again with negative results: either the models did not fit or they still
yielded parameter estimates than could not be given a meaningful interpretation in one way or another.

Although these latter methods might work well in other cases, here, the overall conclusion must be that it is impossible to correct for the response errors that are apparent in the data in Table 3. What is needed here, and has to be recommended in general is to use more and better data. Parsimonious models for the simultaneous analyses of the data for more points in time and for more rotation groups, preferably using more indicators for the central latent variable might provide the stable information for identifying the latent variable, that is missing here.

**EVALUATION**

Directed Loglinear Models provide a very flexible tool for the causal analysis of categorical data. They form a natural way of translating the theoretical causal hypotheses into a statistical model. Through the introduction of latent variables, they can take several important kinds of response error into account. Also, compared to attempts to incorporate categorical variables or nonlinear (interacting) relationships into ‘standard’ causal regression models as LISREL, no matter how useful these procedures are (see, e.g., Long, 1997, Arminger, 1995, Browne and Arminger, 1995, Jöreskog and Wan, 1995, Jaccard and Wan, 1996, Bollen, 1995), the approach advocated here, for many situations is less complex and less dependent upon crucial assumptions about underlying (normal) population distributions. Rather general and easy to use computer programs are available to carry out the necessary analyses in a routine manner (Vermunt, 1996).

Combinations of discrete and continuous variables can be handled by the directed loglinear approach in two ways. On the one hand, it is possible to discretize the continuous variables into a limited number of categories and treat them in a natural way within this approach as categorical variables, where necessary imposing linear restriction on the relationships among particular variables, as shown above. On the other hand, one may treat the continuous variables as truly continuous, assuming that they follow a (conditional) multivariate normal distribution (Whittaker, 1990, Chapter 11, Lauritzen, 1996, Chapter 6, Moustaki, 1996, Croon and Heinen, 1997, Van der Heijden and Dessens, 1996, Dayton and Macready, 1988,
Böckenholt, 1997). Discretizing makes less demanding assumptions about the underlying population distributions; moreover, the information loss involved in categorizing can usually be ignored if enough (five?) categories are being used.

Nevertheless, many problems remain, or rather, challenges to be taken up in the near future. The most important ones are directly related to the major plague of categorical data analysis: small sample size and sparse tables. Testing a particular model, even carrying out conditional tests comparing several models may become extremely difficult with sparse tables, as the asymptotics of the standard chi-square tests break down and the maximum likelihood estimates cannot be expected to follow a normal distribution. Exact testing and parametric bootstrapping procedures may provide an way out of these problems (Agresti, 1992, Van der Heijden, et al., 1997, Collins et al. 1993, Langeheine et al., 1996) although we still do not know enough of their behavior in the case of extremely sparse tables that occur often in applied research.

Observed sparse tables easily results in zero cells in the table of estimated expected frequencies. The interpretation and definition of the parameters may then become difficult, or, at least, nonstandard. Further, with zero estimated frequencies (and boundary parameters estimates), the appropriate number of degrees of freedom is hard to define. Sampling zeroes become a kind of aposteriori structural zeroes: the maximum likelihood estimates obtained are either terminal estimates or maximum likelihood estimates on the ‘apriori’ assumptions that the pertinent cells are indeed zeroes in the populations. As is true in the case of ordinal restrictions dealt with above, the degrees of freedom are not fixed by the model, but become dependent on the data and, in this way, become random variables.

Adding small constants to the observed frequencies will, at least in regular loglinear models without latent variables, prevent the occurrence of estimated zero cells, but the parameter estimates may be extremely unstable and very much dependent on the particular constants chosen.

When the relevant models are very parsimonious, generally sparse tables pose less of a problem, but for relatively complex models the testing and estimation problems are very serious, including numerical problems in finding the global maximum of the likelihood. Perhaps a Bayesian approach may solve some of these problems (Rubin and Stern, 1994, Gelman et al. 1995).
In a way, the introduction of latent variables contributes to the sparseness of the (complete) tables, because extra variables and parameters are added without increasing the number of independent observations. Having enough ‘genuine’ indicators for these latent variables will overcome many of the problems that have otherwise to be expected (as shown above). Further, more research is needed into two-step procedures in which, first, the measurement model is analyzed separately and, then, once a satisfactory measurement model for the latent variable(s) is obtained, ‘observed’ latent scores (‘factor’ scores) are computed, which, finally are introduced into the causal, structural part of the causal model instead of the latent variables and their indicators. In this way, it is not necessary to estimate simultaneously all relationships among all ‘causal’ factors and all indicators. As is well known, such ‘observed’ latent scores are usually not (completely) identified and, moreover, biased, but for several important models the bias is known and can be corrected for (Hagenaars, 1990, Section 3.3, Bolck, *et al.*, forthcoming, Croon and Bolck, forthcoming).

Finally, although the possibility to introduce latent variables is a very strong point of DLM, the meaning of these latent variables models is not always clearly understood. Latent variables denote the respondent’s ‘true’ position, corrected for response error by taking the probabilistic relationship between latent and observed variables (the indicators) into account. However, confining the exposition here to pure ‘random error’ (ICE) models: there are two kinds of ‘random error’ in the manifest data. On the one hand, there is a less-than-perfect relationship between a latent variable and its indicators because the respondent makes nonsystematic response errors and happens to misreport his/her true position. On the other hand, there is random error and a probabilistic relationship because there is truly ‘random behavior’. Figure 3b about changes in (un)employment can be used to elaborate this point. According to the model in Figure 3b, there is probabilistic relationship between the true (un)employment status and the observed one. This ‘imperfection’ is partly caused by the unsystematic response errors the respondents make, by nonsystematic recording errors made by interviewers or coders, etc. But another source of ‘random error’ is the fact that some persons are ‘truly’ (un)employed and are registered as such during most of the period, but at some point in time happen to loose their job for a short while or, on the contrary, happen to find employment for a while. Latent variable models as such cannot make the difference between somebody who is usually employed but at one point in time happens to be erroneously recorded as unemployed and a person who is usually employed but at one point in time happens to be without a job.
Whether this confusion is bad depends on the purposes of the investigation. If one wants to measure the true underlying attitude of a person or his/her weight, in the sense of the value that is not affected by daily or even hourly ‘random’ fluctuations (the operational true score, see note 9), latent variable models are the right models to employ. However, if one is interested in the real number of people that are actually unemployed at a certain moment of time (the platonic true score - see note 9), regardless whether they are ‘always’ unemployed or typically employed but accidentally at this moment unemployed, latent variable models will generally not be appropriate because (in this particular example) they will underestimate the real number of unemployed people and underestimate the actual changes that are going on the labor market.

At least part of the failure of latent variable models such as in Figure 3b that generally underestimate the true changes on the labor market may be caused by the presence of random, but real true change, that is not detected by these models. If this is true, the proposed models that take systematic response error into account may still not provide the correct estimates (or at least not for the right reasons).

Nevertheless, despite all these cautious and warning notes that forces one to recognize not only the potentialities, but also the pitfalls of the use of directed loglinear models with latent variables, and despite the many questions that remain unanswered and the many unresolved problems, ‘the questions are intriguing and the search for answers is fun (Nunnally, 1973, p. 109)."
REFERENCES

-Bolek A., Croon M.A., Hagenaars J.A. (*Forthcoming*),


- Lord F.M. and M.R. Novick (1968), *Statistical Theories of Mental Test Scores*, Reading: Addison-Wesley


- Suttcliffe J.P. (1966), A Probability Model for Errors of Classification; I: General Considerations, Psychometrika, vol. 30, pp. 73-96.


Figure 1
Figure 2
Figure 3
a) First-order markov model at the manifest level

\[ A \rightarrow B \rightarrow C \rightarrow D \]

b) First-order markov model at the latent level

\[ W \rightarrow X \rightarrow Y \rightarrow Z \]

\[ A \quad B \quad C \quad D \]

c) Latent markov model with unmeasured 'consistency-trait'

\[ W \rightarrow X \rightarrow Y \rightarrow Z \]

\[ A \quad B \quad C \quad D \]

\[ V \]

d) Latent markov model with consistency with true position at time of interview
e) Latent Markov model with consistency with first given answer

f) Latent Markov model with consistency with last given answer

figure 4
figure 5
figure 6
TABLE 1: Extrinsic Job Satisfaction

P - Pay is good (on present job)
S - The job security is good
B - The fringe benefits are good

1. not at all true  2. a little true  3. somewhat true  4. very true

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>S</th>
<th>B-1</th>
<th>B-2</th>
<th>B-3</th>
<th>B-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>42</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>5</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>S</th>
<th>B-1</th>
<th>B-2</th>
<th>B-3</th>
<th>B-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>11</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>29</td>
<td>18</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>10</td>
<td>21</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>S</th>
<th>B-1</th>
<th>B-2</th>
<th>B-3</th>
<th>B-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>14</td>
<td>24</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>22</td>
<td>65</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>6</td>
<td>45</td>
<td>172</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SOURCE Quality of Employment Survey 1977 Cross Section; Shockey (1988)
TABLE 2: Estimated Conditional Response Probabilities $\pi_i^{px}$, $\pi_j^{sx}$, $\pi_k^{bx}$ for LCA, applied to Table 1.

**a. Nominal level LCA**

<table>
<thead>
<tr>
<th></th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>.59</td>
<td>.16</td>
<td>.03</td>
<td>.03</td>
<td>.50</td>
<td>.09</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.66*</td>
<td>.15</td>
<td>.06</td>
<td>.32</td>
<td>.24</td>
<td>.05</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.20</td>
<td>.13**</td>
<td>.67</td>
<td>.25</td>
<td>.17</td>
<td>.50</td>
<td>.40</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>.06</td>
<td>.15</td>
<td>.66</td>
<td>.01</td>
<td>.17</td>
<td>.55</td>
<td>.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>.44</td>
<td>.02***</td>
<td>.05***</td>
<td>.02</td>
<td>.35</td>
<td>.06</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.44</td>
<td>.10</td>
<td>.07</td>
<td>.30</td>
<td>.16</td>
<td>.03</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.24</td>
<td>.35</td>
<td>.56</td>
<td>.09</td>
<td>.27</td>
<td>.42</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>.19</td>
<td>.30</td>
<td>.82</td>
<td>.08</td>
<td>.36</td>
<td>.69</td>
<td>.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
<th>X=1</th>
<th>X=2</th>
<th>X=3</th>
<th>X=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>.74</td>
<td>.21</td>
<td>.08</td>
<td>.02</td>
<td>.70</td>
<td>.10</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.13</td>
<td>.50</td>
<td>.16</td>
<td>.01</td>
<td>.22</td>
<td>.21</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>.12</td>
<td>.24</td>
<td>.56</td>
<td>.16</td>
<td>.08</td>
<td>.48</td>
<td>.26</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>.02</td>
<td>.05</td>
<td>.20</td>
<td>.82</td>
<td>.01</td>
<td>.21</td>
<td>.72</td>
<td>.95</td>
</tr>
</tbody>
</table>

$\hat{X}_i$ .17 .15 .39 .30 .19 .55 .13 .13

* in OLCA: .61 (see text)  
** in OLCA: .17 (see text)  
*** in OLCA: .04 (see text)

$\lambda_p = 1.408 \quad \lambda_s = 1.100 \quad \lambda_b = 1.852$
### TABLE 3: Labor Force Status by Interviewing Conditions and Rotation Group

<table>
<thead>
<tr>
<th>F R L</th>
<th>G = 1</th>
<th>G = 2</th>
<th>G = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1</td>
<td>47</td>
<td>1988</td>
<td>2116</td>
</tr>
<tr>
<td>2 1 1</td>
<td>2413</td>
<td>544</td>
<td>490</td>
</tr>
<tr>
<td>1 2 1</td>
<td>57</td>
<td>2829</td>
<td>3025</td>
</tr>
<tr>
<td>2 2 1</td>
<td>3310</td>
<td>725</td>
<td>560</td>
</tr>
<tr>
<td>1 1 2</td>
<td>8</td>
<td>372</td>
<td>423</td>
</tr>
<tr>
<td>2 1 2</td>
<td>579</td>
<td>131</td>
<td>104</td>
</tr>
<tr>
<td>1 2 2</td>
<td>6</td>
<td>225</td>
<td>209</td>
</tr>
<tr>
<td>2 2 2</td>
<td>409</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>1 1 3</td>
<td>115</td>
<td>4861</td>
<td>5009</td>
</tr>
<tr>
<td>2 1 3</td>
<td>5793</td>
<td>1086</td>
<td>847</td>
</tr>
<tr>
<td>1 2 3</td>
<td>164</td>
<td>3433</td>
<td>3577</td>
</tr>
<tr>
<td>2 2 3</td>
<td>4268</td>
<td>855</td>
<td>676</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17169</strong></td>
<td><strong>17149</strong></td>
<td><strong>17133</strong></td>
</tr>
</tbody>
</table>

**F** - Interview Format (1. telephone 2. face-to-face)

**R** - Respondent Type (1. proxy 2. self)

**L** - Labor Force Status (1. not-in-labor-force (NILF) 2. unemployed. 3. employed.)

**G** - Rotation Group (1. group I 2. group VI 3. group VII)

Table 4. Estimates for Causal Model in Figure 6 - solid lines

a. Manifest (L) by Latent (X) Labor Force Status; N-Not-in-the-labor-force; U-Unemployed; E-Employed; \( \tau_{ij}^{XY} \)

<table>
<thead>
<tr>
<th>L</th>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. N</td>
<td>.524</td>
<td>.000</td>
<td>.207</td>
<td></td>
</tr>
<tr>
<td>2. U</td>
<td>.021</td>
<td>.099</td>
<td>.105</td>
<td></td>
</tr>
<tr>
<td>3. E</td>
<td>.455</td>
<td>.901</td>
<td>.688</td>
<td></td>
</tr>
<tr>
<td>total ( \tau_{ij}^{LX} )</td>
<td>.546</td>
<td>.172</td>
<td>.282</td>
<td></td>
</tr>
</tbody>
</table>

b. Manifest (L) by Latent (X) Labor Force Status; N-Not-in-the-labor-force; U-Unemployed; E-Employed; \( \lambda_{ij}^{XL} \)

<table>
<thead>
<tr>
<th>L</th>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. N</td>
<td>6.44</td>
<td>-11.60</td>
<td>5.16</td>
<td></td>
</tr>
<tr>
<td>2. U</td>
<td>-3.56</td>
<td>5.88</td>
<td>-2.33</td>
<td></td>
</tr>
<tr>
<td>3. E</td>
<td>-2.88</td>
<td>5.71</td>
<td>-2.83</td>
<td></td>
</tr>
</tbody>
</table>

c. Respondent Type by Latent Labor Force Status (X); \( \lambda_{ik}^{XR} \)

<table>
<thead>
<tr>
<th>R</th>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. proxy</td>
<td>-2.35</td>
<td>4.63</td>
<td>-2.27</td>
<td></td>
</tr>
<tr>
<td>2. self</td>
<td>2.35</td>
<td>-4.63</td>
<td>2.27</td>
<td></td>
</tr>
</tbody>
</table>

d. Interview Format by Latent Labor Force Status (X); \( \lambda_{ij}^{XF} \)

<table>
<thead>
<tr>
<th>F</th>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. telephone</td>
<td>1.04</td>
<td>.00</td>
<td>-1.04</td>
<td></td>
</tr>
<tr>
<td>face-to-face</td>
<td>-1.04</td>
<td>-.00</td>
<td>1.04</td>
<td></td>
</tr>
</tbody>
</table>
NOTES

1. Throughout this article, the term 'causal' will be used in a rather loose sense to denote asymmetrical relationships among variables. For more precise definitions of causality and some opposing views, see, among others, Sobel (1995) and Glymour et al. (1987). Early 'causal' interpretations of LCA, although more limited in scope while staying within the classical LCA-approach have been offered by Clogg and by Goodman (Clogg, 1981a, p.229, Goodman, 1974a).

2. Sometimes it is possible to obtain the estimates \( \tau_{ijklm}^{ABCDE} \) for the full table directly by specifying one (loglinear) model, rather than a sequence of submodels. This is true for 'moral graphs', where the restrictions imposed on the data by the model with 'directed' (causal) arrows are the same as those of the corresponding model with 'undirected' arrows (Whittaker, 1990, Section 3.5).

3. One should note that in graphical diagrams, contrary to what is done here, 'knots' denoting higher order interaction effects are not used. If two (or more) variables are connected to a third one, in standard graphs, the three-variable-interaction (or higher order) term order term is 'automatically' included.

4. It may be interesting to note that almost all models proposed for misclassifications in categorical data, including those using external validation data (for a recent overview, see Kuha and Skinner, 1997) can be formulated in terms of categorical latent variable models, making these misclassification models more flexible than is usually the case now. Further, although not treated here, latent variable models are also excellently suited to handle nonresponse (Little and Rubin, 1987, Chapter 9, Hagenaars, 1990, Section 5.5, Vermunt, 1996, Section 3.2).

5. Computer programs such as \( \hat{EM} \) or PANMARK (Van de Pol, 1996) make efficiently use of the fact that the graphical decomposition of the joint table as in Equations (6) or (7) make it possible to estimate the parameters and expected frequencies for (marginal) tables of relatively low dimensionality without having to deal directly with the high dimensional full table. By storing further only the non-empty cells of the observed table or the individual records, handling a rather large amount of (15 to 50) categorical variables has now become practically feasible. However, it may still occur that the marginal tables to be used are forbiddingly large. A possible solution could then be, when using IPF, N-R procedures or variants thereof, never to store the (marginal) table with expected frequencies, but to use only the sufficient statistics and the estimated parameters and compute at each step the particular expected cell frequency needed to update the parameter estimates. Although in this way, in terms of memory capacity, there are virtually no limits to the number of variables that can be employed, having to compute each (marginal) cell frequency at each step may be very time consuming for very large tables.
6. In principle, one can define column, row, row-plus-column, and linear-by-linear association models with fixed scores for the ‘linear’ variable(s) (type-1 association models) or with scores to be estimated (type-2 association models). Several models of type-1 are identical to a type-2 model, especially for two-way tables. In multi-way tables, such equivalences also depend on the nature of the restrictions imposed on the scores (Clogg and Shihadeh, 1994).

7. Both Poulsen and Van de Pol/Langeheine defined not only Markov models at the latent level, but also mixed Markov models, dividing the respondents into different unobserved groups (latent classes) having different Markov parameters or into groups only some of which follow the Markov model, while for the others a different structural model is postulated. These and other forms of unobserved heterogeneity can also be handled by directed loglinear models as is illustrated below.

8. Except the objections that will be mentioned below, another problem is that the estimates of the parameters in Equation (20) provided by Shockey (1988, Table 5) (in a somewhat more restricted form) are wrong in the sense that the reported maximum likelihood estimates are local and not global maxima. Rather than his solution with \( L^2 = 73.53 \), estimates can be found for the same model and the same data with \( L^2 = 39.43 \). As will be mentioned below, it is extremely difficult for this particular model with this data set to avoid local maxima. Awkward is that the nice, substantively interpretable estimates that Shockey got for the relation between \( X \) and \( L \) disappears and that the correct, more optimal (best?) solution with \( L^2 = 39.43 \) showed the same strange outcomes that will be discussed below.

9. The distinction made here between ‘random response error’ and ‘random behavior’ is related to the distinction in ‘platonic true score’ and ‘operational true score’, where the platonic true score model assumes that there exists a real and actual true score and the operational true score model defines the true score as the score a person obtains on average in a series of independently conducted experiments (see, e.g., Sutcliffe, 1965, Lord and Novick, 1968, Section 2.9, Hagenaars 1990, Section 4.4.2 and Sobel, 1994). Whether this distinction matters depends on the purposes of the analysis. Here, it is argued that sometimes it does matter.