Causal log-linear modeling with latent variables and missing data

Jeroen K. Vermunt

Appeared in "Analysis of Change: Anvanced Techniques in Panel Data Analysis", Uwe Engel and Jost Reincke (Editors), 1996.

1 Introduction

The log-linear model has become a widely used method for the analysis of multivariate frequency tables. A general approach for analyzing categorical data which combines three important extensions of the standard log-linear model will be presented. Modified path models, latent class models and models for nonresponse are integrated within one general model.

Log-linear models are used to describe the observed frequencies or proportions in a multi-way cross-tabulation by means of a limited number of parameters. In the standard log-linear model, no distinction is made between dependent and independent variables. However, if one is interested in the effects of a set of independent variables on a dependent variable, one can use a 'regression' variant of the standard log-linear model, the well known logit model (Goodman, 1972; Agresti, 1990). When the dependent variable has more than two categories it sometimes also called a multinomial response model (Haberman, 1979; Agresti, 1990).

Goodman (1973) introduced a 'path analytic' extension of the logit model. He proposed a log-linear model which takes a priori information on the causal ordering of the variables into account. The so-called 'modified path analysis approach' consists of specifying a 'recursive' system of logit models in which a variable appearing as the dependent variable in a particular logit equation may appear as one of the independent variables in one of the next equations.

Often we want to study phenomena which are difficult to observe directly. This has given rise to a family of measurement models for identifying unobserved or latent variables from a set of observed variables. The latent class model is a 'factor analytic' model for categorical data (Lazarsfeld and Henry, 1968; Goodman 1974). Haberman (1979) demonstrated that the latent class model is equivalent to a log-linear model in which one or more of the variables are unobserved.

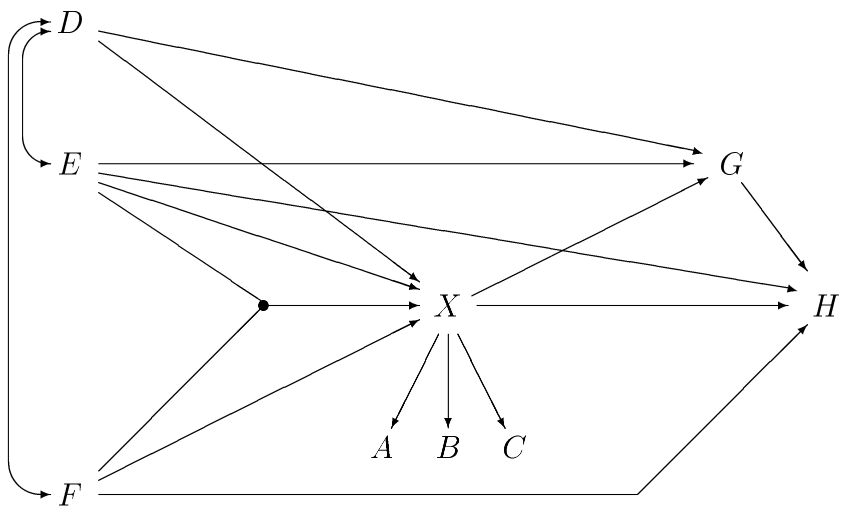
For interval level data, the combination of factor analysis and path analysis led to the famous Lisrel model (Jöreskog and Sörbom, 1988). Hagenaars (1990, 1993) developed a 'Lisrel' model for categorical data by combining the modified path model and the latent class model. He implemented this so-called 'modified Lisrel approach' into his latent class analysis program LCAG (Hagenaars en Luijkx, 1990).

As mentioned above, latent class models are log-linear models in which one or more variables are completely unobserved. However, in social research, and especially in panel studies, we often are confronted with another type of missing data, i.e., with variables which are unobserved for a part of the sample due to panel attrition, item nonresponse or the data collection design. Fuchs (1982) proposed a method which makes it possible to use partially observed data when estimating the parameters of a log-linear model. Fay (1986) extended Fuchs' work by making it possible to specify and test explicitly ones assumptions with regard to the mechanism causing the missing data. He proposed to model the response mechanism via log-linear path models in which so-called response indicators are included.

Both Hagenaars' modified Lisrel models and Fay's models for nonresponse are modified path models in which some information is missing on particular variables. By combining these two approaches, one obtains a more general modified path model in which unobserved variables, partially observed variables, completely observed variables and response indicators can be included (Vermunt, 1988, 1995; Hagenaars, 1990). A program called ℓEM (Log-linear and event history analysis with missing data using the EM algorithm) has been developed which can be used to estimate this rather general log-linear model by means of the EM algorithm (Vermunt, 1993).

The three next sections present log-linear path models, log-linear models with latent variables, and log-linear models for nonresponse, respectively. The models are illustrated by means of an application in Section 5. The data for the example is taken from a long-term Dutch panel study on educational and occupational careers of persons who were in the last grade of primary school in 1952, the so-called 'extended Mathijssen-Sonnemans cohort' (Hartog, 1986; Vermunt, 1988). The data are collected via a very specific data collection design which resulted in a lot of missing data in addition to the usual panel attrition. The other relevant aspects of the data will be introduced when discussing the different kinds of models.

Figure 1: Modified Lisrel model (Model 6)



2 Log-linear path models

Suppose we want to investigate the causal relationships among father's educational level (D), father's occupation (E), sex (F), school ability at the end of the primary school (X), educational level (G) and occupation (H) by means of log-linear analysis. The data required for such an analysis is the six-way cross-tabulation $DEFXGH$ of the above-mentioned variables. An observed cell frequency in this cross-table will be denoted by n_{defxgh}, where the lower case subscripts denote the categories of the variables D, E, F, X, G and H. In contrast to standard log-linear models, we do not only want to estimate the strength of the association among these variables, but we also want to use a priori information on the causal order among the variables. Figure 1 shows the assumed causal order among the variables used in the example. The variables D, E and F will be treated as exogenous variables. The other ones are endogenous, where X is assumed to be posterior to G, and G is assumed to be posterior to H. For the moment, it will be assumed that all variables are observed directly. In the next section, the variables A, B and C appearing Figure 1 will be used as indicators for the latent variable X.

2.1 Probability structure

Let π_{defxgh} denote the probability that $D = d$, $E = e$, $F = f$, $X = x$, $G = g$ and $H = h$. Using the a priori information on the causal order among the variables, π_{defxgh} can be written as (Goodman, 1973)

$$\pi_{defxgh} \;=\; \pi_{def}\, \pi_{x|def}\, \pi_{g|defx}\, \pi_{h|defxg} \;. \tag{1}$$

So imposing a causal ordering can be simply accomplished by decomposing the joint probability into a product of marginal and conditional probabilities. This is a straightforward way to express that the value of a particular variable can only depend on the preceding variables and not on the posterior ones. For instance, if the causal order is true, G can only depend on the preceding variables D, E, F and X, but not on the posterior variable H. Therefore, the probability that $G = g$ depends only on the values of D, E, F and X, and not on the value of H.

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

The simplest way to specify more parsimonious models is to restrict directly the conditional probabilities appearing in Equation (1). Suppose that, as depicted in Figure 1, G depends on D, E and X, but not on F. This assumption can be incorporated in the model by replacing $\pi_{g|defx}$ by $\pi_{g|dex}$ since in that case $\pi_{g|defx} = \pi_{g|dex}$. This is the easiest procedure for restricting the number of parameters. It is also applied in, for instance, discrete Markov models (Van de Pol and Langeheine, 1990). When the same kinds of restrictions can be imposed on the other elements appearing at the right hand side of Equation (1), the number of parameters can be reduced a lot. For instance, on the basis of the relationships depicted in Figure 1, a more restricted version of the general Equation (1) would be

$$\pi_{defxgh} \;=\; \pi_{def}\, \pi_{x|def}\, \pi_{g|dex}\, \pi_{h|efxg} \;. \tag{2}$$

Thus, in addition to the already mentioned restriction on $\pi_{g|defx}$, H is assumed not to depend directly on D, that is, $\pi_{h|defxg} = \pi_{h|efxg}$.

This rather simple procedure for obtaining more restricted models has, however, one important disadvantage. The dependent variable must always be related to the joint independent variable. For instance, in Equation 2, G depends on the joint variable DEX. Thus, if a particular variable is thought to influence the dependent variable concerned, all interactions with the other independent variables must be included in the model as well. As a result, the model will generally still contain more parameters than necessary.

2.2 Logit models for probabilities

By using a log-linear or logit parameterization of the marginal and conditional probabilities, it is possible to specify and test more parsimonious causal models for categorical data. This leads to what Goodman called a 'modified path analysis approach' (Goodman, 1973). This approach consist of specifying a 'recursive' system of logit models. As in path analysis, a particular variable which appears as dependent variables in one equation can be used as independent variables in one of the next equations. The relationships among the exogenous variables can be restricted by means of a log-linear model. A model for the relationships among the variables used in the example would consist of four so-called modified path steps or submodels: one model for the exogenous variables D, E and F, and three logit models in which X, G and H appear as dependent variables. Because of simplicity of exposition, here only simple hierarchical log-linear models will be used, but the results can easily be generalized to log-linear models which include more sophisticated restrictions on the parameters, such as symmetric relationships, linear-by-linear interactions and log-multiplicative row and column effects.

Suppose that G depends on D, E and X, and that there exist no three variables interactions between G and the independent variables (see Figure 1). In that case, the following logit parameterization of the conditional probability concerned would apply,

$$\pi_{g|defx} = \pi_{g|dex} = \frac{\exp\left(u_g^G + u_{dg}^{DG} + u_{eg}^{EG} + u_{xg}^{XG}\right)}{\sum_g \exp\left(u_g^G + u_{dg}^{DG} + u_{eg}^{EG} + u_{xg}^{XG}\right)} ,$$

where the u's denote log-linear parameters which fulfill the well known ANOVA-like constraints. Specifying this logit model for $\pi_{g|defx}$ is equivalent to specifying log-linear model $\{DEFX, DG, EG, XG\}$ for marginal table $DEFXG$, or

$$\log m_{defxg} \;=\; \alpha^{DEFX}_{defx} + u^{G}_{g} + u^{DG}_{dg} + u^{EG}_{eg} + u^{XG}_{xg} \;,$$

where m_{defxg} denote the expected frequencies in the marginal table concerned. Moreover, α^{DEFX}_{defx} denotes the effect which fixes the marginal distribution of the dependent variables. Including this effect makes a log-linear model equivalent to a logit model (Goodman, 1972; Agresti, 1990).

So specifying a causal model for a set of categorical variables can simply be accomplished by specifying separate log-linear or logit models for different marginal tables, or subtables. The marginal tables are formed by the variables used in the previous marginal table and the variable which appears as dependent variable. In this case, one must specify log-linear models for tables DEF, $DEFX$, $DEFXG$ and $DEFXGH$, where the margin formed by the variables of the previous marginal table must always be fixed. Goodman (1973) presented his 'modified path analysis approach' showing how to specify separate log-linear models for different marginal tables. And next, he showed how to combine the expected frequencies of the separate submodels by an equation similar to Equation (1). Note that the probabilities in Equation (1) can be obtained by means of the expected frequencies via

$$\pi_{g|defx} \;=\; \frac{m_{defxg}}{\sum_g m_{defxg}} \;. \tag{3}$$

One additional remark has to made with regard to the modified path models. When a particular variable does not depend on all its preceding variables, the procedure proposed by Goodman can be modified somewhat. As already mentioned above, G does not depend on F and therefore $\pi_{g|defx} = \pi_{g|dex}$. Therefore, the log-linear restrictions which were imposed on $\pi_{g|defx}$ can also be imposed directly on $\pi_{g|dex}$, namely by means of log-linear model $\{DEX, DG, EG, XG\}$ for marginal table $DEXG$, or

$$\log m_{dexg} \;=\; \alpha^{DEX}_{dex} + u^{G}_{g} + u^{DG}_{dg} + u^{EG}_{eg} + u^{XG}_{xg} \;.$$

Estimating parameters in a marginal table which includes only the dependent variables which are really used has two important advantages: it is computationally more efficient, and, moreover, it prevents fitted zeroes when the fixed margin $DEFX$ contains observed zeroes cell which do not appear in the margin DEX.

2.3 Estimation

Maximum likelihood estimates for the log-linear parameters and the expected frequencies for the various subtables can be obtained using standard programs for log-linear analysis. In that case, the models for the various subtables must be estimated separately. The estimated cell probabilities for the overall model can be computed via Equations (3) and (1). Model testing can be performed, for instance, by means of the log-likelihood ratio statistic L^2. The various submodels can be tested separately. A test of the overall fit can simply be obtained by adding both the L^2-values and the degrees of freedom of the various submodels.

A program called ℓEM has been developed to estimate log-linear path models without the necessity to set up the different marginal tables (Vermunt, 1993). In ℓEM, specifying a log-linear path model is the standard way of modeling an observed frequency table. The current available version of the program (ℓEM 0.11) is based on the original procedure of Goodman, but the most recent working version uses the more efficient procedure in which the subtables contain only the independent variables which are really used. The procedure implemented in ℓEM to estimate hierarchical log-linear models is the iterative proportional fitting algorithm (Agresti, 1990). But ℓEM can also be used to estimate more complex log-linear models in which the parameters are linearly restricted in some way (Haberman, 1979; Agresti, 1990). This is accomplished by allowing the user to specify his own design matrix for particular log-linear effects. In ℓEM, it is also possible to us log-multiplicative effects, such as the type II association models developed by Goodman (Goodman, 1979; Clogg, 1982; Xie, 1992). These non-hierarchical log-linear models are estimated by means of the one-dimensional Newton algorithm (Goodman, 1979; Vermunt, 1995). This algorithm differs from the well known Newton-Raphson algorithm in that only one parameter is updated at once instead of updating all parameters simultaneously. It is a very simple and fast algorithm per iteration which, as we will see in the next section, fits very well into the EM algorithm used to estimate models with missing data.

3 Log-linear models with latent variables

In the previous section, it was assumed that all variables used in the causal log-linear model can be directly observed. However, often one encounters problems in which several indicators are used to measure a concept which itself cannot be measured directly. In the example, the variable school ability (X) is such a latent variable. Three different school ability tests, denoted by A, B and C, are used as indicators for X.

3.1 Latent class models

Latent class analysis is a variant of factor analysis which is especially suited for analyzing categorical latent and manifest variables. The latent class model was first proposed by Lazarsfeld (Lazarsfeld and Henry, 1968). Goodman (1974) and Haberman (1979) made the model practically applicable by introducing estimation and test procedures. As factor analysis, the latent class model can be used to identify the latent construct X using the indicators A, B, C. Moreover, just as factor analysis, the latent class model is based on he assumption of local independence. Using the classical parameterization proposed by Lazarsfeld, the latent class model for one latent variable X and three indicators A, B, C can be written as

$$\pi_{xabc} \; = \; \pi_x \, \pi_{a|x} \, \pi_{b|x} \, \pi_{c|x} \; , \tag{4}$$

where π_{xabc} denotes the joint probability of the latent variable and its three indicators, π_x denotes the probability of belonging to particular latent class, and $\pi_{a|x}$ denotes the probability that $A = a$ given $X = x$. The latent distribution is assumed to be formed by X^* mutually exclusive and exhaustive categories, that is, $\sum_{x=1}^{X^*} \pi_x = 1$. From Equation 4, it can easily be seen that, given a particular value of X, the variables of A, B and C are assumed to be independent.

Haberman (1979) demonstrated that the unrestricted latent class model can also be parametrized as a log-linear model in which one or more variables are unobserved. Using the log-linear parameterization, the latent class model of Equation (4) can be written as

$$m_{xabc} \; = \; u + u_x^X + u_a^A + u_b^B + u_c^C + u_{xa}^{XA} + u_{xb}^{XB} + u_{xc}^{XC} \; .$$

This is equivalent to writing the separate conditional response probabilities in terms of log-linear parameters (Haberman, 1979; Formann, 1992; Heinen, 1993). For instance, the probability that $A = a$ given $X = x$ can also be written as

$$\pi_{a|x} \;=\; \frac{\exp\left(u_a^A + u_{xa}^{XA}\right)}{\sum_a \exp\left(u_a^A + u_{xa}^{XA}\right)} \,.$$

Formann (1992) used this parametrization of the latent class model for the formulation of his linear logistic latent class model. Heinen (1993) used this parameterization to demonstrate the equivalence between latent class models and latent trait models in which the latent variable is discretized (see also Vermunt and Georg, 1995).

3.2 Modified Lisrel models

Several extensions of the standard latent class model have been proposed, such as models for more than one latent variable (Goodman,1974a, 1974b; Haberman 1979), models with so-called external variables (Clogg, 1981) and models for multiple-group analysis (Clogg and Goodman, 1984; McCutcheon, 1987). A limitation of these extensions is, however, that they are all developed within the framework of either the classical or the log-linear latent class model. Therefore, it is not always possible to postulate the wanted a priori causal order among the structural variables incorporated the model.

Hagenaars (1990, 1993) solved this problem by combining the modified path model discussed in the previous section with the latent class model. More precisely, he showed how to specify a modified path model for the joint distribution of the external and the latent variables in a latent class model. Not surprisingly, he called this extension which he implemented in his program LCAG (Hagenaars and Luijkx, 1990) a modified Lisrel approach.

If X is a latent variable with indicators A, B and C, and the same causal order among D, E, F, X, G and H is assumed as in the previous section (see Figure 1), the joint probability of all variables can be written as

$$\pi_{defxghabc} \;=\; \pi_{def}\, \pi_{x|def}\, \pi_{g|defx}\, \pi_{h|defxg}\, \pi_{abc|defxgh} \;, \tag{5}$$

where

$$\pi_{abc|defxgh} \;=\; \pi_{abc|x} \;=\; \pi_{a|x}\, \pi_{b|x}\, \pi_{c|x} \;.$$

Thus, including latent variables in a modified path model involves specifying one or more additional modified path steps in which the relationships among the latent variables and their indicators are specified. These additional steps form the measurement part of the model, while the other steps form the structural part of a modified Lisrel model.

3.3 Estimation

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

In the E step of the EM algorithm, the missing data is estimated. In our case, we must obtain estimates for the unobserved frequencies of the complete table $DEFXGHABC$, the $\hat{n}_{defxghabc}$'s, conditional on the observed data and the parameters estimates from the last EM iteration. This is accomplished using the observed incomplete data and the parameter estimates from the last iteration by

$$\hat{n}_{defxghabc} = n_{defghabc}\,\hat{\pi}_{x|defghabc}\,. \tag{6}$$

Here, $n_{defghabc}$ denotes an observed frequency, and $\hat{\pi}_{x|defghabc}$ denotes the probability that $X = x$ given the observed variables.

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

Hagenaars' latent class analysis program LCAG (Hagenaars and Luijkx, 1990) includes an option to specify a modified path model for the joint latent distribution. Observed variables can be included in the modified path model by means of a trick, namely by making them quasi-latent via particular restrictions on the conditional probabilities. This, however, can become a laborious operation, especially if, as in the example, many observed variables appear in the structural part of the model.

The program ℓEM is especially developed for estimating modified path models with latent variables. Latent and manifest variables are treated in exactly the same way by the program. That is why ℓEM is more user friendly and more efficient than LCAG for estimating modified Lisrel models. Moreover, in LCAG only hierarchical log-linear models can be specified, while in ℓEM, as we already saw in the previous section, all kinds of linear restrictions can be imposed on the log-linear parameters. Although in ℓEM 0.11 the number of cells of the cross-tabulation of all variables is still limited, the current working version of ℓEM can handle much bigger problems because the size of an application does not depend on the total number of cells in the complete table anymore.

The algorithm used in ℓEM is a modified version of the original EM algorithm because the M step always consists of only one iteration. So generally the complete data likelihood is not maximized but only improved within a particular M step. This is a special case of the so-called GEM algorithm which states that every increase in the complete data likelihood also leads to an increase of the incomplete data likelihood we actually want to maximize (Dempster, Laird and Rubin, 1977; Little and Rubin, 1987). In fact, the algorithm which is used in ℓEM is also a version of the ECM algorithm (Meng and Rubin, 1993). In the ECM algorithm, the M step is replaced by a conditional maximization (CM) step. Conditional maximization implies that instead of improving all the parameters simultaneously, subsets of parameters are updated fixing the other ones at their previous values. This is just what is done by IPF and by the one-dimensional Newton algorithm. Meng and Rubin (1993) state that such simple and stable linear convergence methods are often more suitable for the M (or CM) step of the EM (or ECM) algorithm than superlinear converging but less stable algorithms, such as Newton-Raphson. This GEM or ECM algorithm converges in nearly the same amount of iterations as the true EM algorithm. This makes it much faster than the true EM algorithm, especially in applications where a real M step would need many iterations to converge.

4 Log-linear models for nonresponse

In survey research, it almost always occurs that information on some variables is missing for a part of the sample. This can be caused, for instance, by item-nonresponse, by panel attrition or by the data collection design. The easiest way to deal with partially observed variables is to perform the analysis using only complete cases. This, however, may lead to biased parameter estimates in case the nonresponse is selective in some way. Another possibility is to impute the missing data. The model based approach to partially observed data that will be discussed in this section is strongly related to imputation. The main difference between imputation and the approach to be presented below is that the data are imputed during the estimation of the model parameters and not beforehand. Moreover, the approach presented here allows to specify and test models for the mechanism causing the missing data.

The data of the 'extended Mathijssen-Sonnemans cohort' were collected via a very specific data collection design which resulted in a lot of missing data in addition to the usual panel attrition. In 1952, 5823 pupils from the last grade of all primary schools in the Dutch province Brabant were tested with regard to their school ability (A, B and C). For 5387 of these pupils, the head master was able to supply information on the occupation of their father (E). In 1957, information on the educational level of the father (D) was collected for the children who had a school ability score above the mean. In 1958 and 1959, the same information was collected for sons of farmers and workers. Altogether, information on the father's educational level was collected for 2740 persons. In 1983, the complete group was approached again to obtain information about their finished education level and their current occupation (G and H). This resulted in useful information for 2587 person. This group contains more males because more effort was made to get their information, that is, they were approached more often than females in case of nonresponse.

Summarizing, starting with the 5387 persons for which information is available on A, B, C, E and F, there are four subgroups of persons with exactly the same kind of information. Subgroup $DEFGHABC$ consists of 1279 persons, subgroup $EFGHABC$ of 506, subgroup $DEFABC$ of 1344 and subgroup $EFABC$ of 1358, where the capital letters indicate which variables are observed for the subgroup concerned.

4.1 Fuchs' approach

Fuchs (1982) proposed to apply the EM algorithm to incorporate missing data when estimating the parameters of a log-linear model. Hagenaars (1990) showed how to adapt Fuchs' approach when the log-linear model concerned contains latent variables, such as the modified Lisrel models discussed above. In that case, one has a double missing data problem, namely partially observed variables and latent variables. Let $n_{defghabc}$, $n_{efghabc}$, n_{defabc}, n_{efabc} denote the observed frequencies in the subgroups $DEFGHABC$, $EFGHABC$, $DEFABC$ and $EFABC$, respectively. Applying Fuchs' method equates replacing the E step of Equation (6) by

$$\begin{aligned} \hat{n}_{defxghabc} &= n_{defghabc}\,\hat{\pi}_{x|defghabc} + n_{efghabc}\,\hat{\pi}_{dx|efghabc} \\ &\quad + n_{defabc}\,\hat{\pi}_{xgh|defabc} + n_{efabc}\,\hat{\pi}_{dxgh|efabc} \,. \end{aligned}$$

The M step of the EM algorithm is equivalent to the one discussed in the previous section. This version of the EM algorithm is implemented in Hagenaars' program LCAG (Hagenaars and Luijkx, 1990).

Although perhaps it might not be immediately clear, the method proposed by Fuchs (1982) is based on the assumption that the response mechanism is ignorable for likelihood based inference (Rubin, 1976; Little, 1982; Little and Rubin, 1987). The response mechanism is called ignorable if for every sampled unit the response distribution is independent of the missing data. In other words, for every subgroup it is assumed that the probability of belonging to that particular subgroup, or having missing information on just those variables, depends only on the observed variables. If the response mechanism can be assumed to be ignorable, cases with missing data can be used to estimate the parameters without the need to specify the exact ignorable response mechanism. This results from the fact that under the ignorable condition the likelihood can be partioned into a part which depends only on the model parameters and a part which depends only on the response mechanism. Both parts can be maximized separately. So also if the response mechanism is specified, every ignorable response mechanism leads to the same parameters estimates. That is the reason why when applying Fuchs' approach, the least restrictive ignorable response mechanism can be assumed to be valid: missing at random (=MAR). However, Fuchs actually tests the most restrictive form of ignorable nonresponse mechanism, namely missing completely at random (=MCAR) (Vermunt, 1995). After presenting

Fay's approach, I will discuss more in detail the various kinds of response mechanism.

4.2 Fay's approach

Although ignorability of the response mechanism is often a reasonable assumption, Fuchs' method has particular disadvantages which can be overcome by making use of an extension which was independently proposed by Fay (1986) and by Baker and Laird (1988). Using Fuchs' method it is neither possible to test a priori assumptions about the response mechanism, nor to specify nonignorable response mechanisms.

Fay (1986) proposed to include response indicators into a log-linear path model in which the relationships among the survey variables and the mechanism causing nonresponse are specified together. A response indicator is a variable which indicates whether a particular set of variables is observed or not. In the example, two response indicators are needed, one indicating whether D is observed or not and another one indicating whether G and H are observed or not. They will be denoted by the letters R and S, where $R = 1$ means that D is observed and $R = 2$ that D is missing. And, $S = 1$ means that G and H are observed and $R = 2$ that they are not observed. It will be clear that the levels of R and S identify the four above-mentioned subgroups.

The two response indicators can be used in a modified path model in the same way as the other variables. That makes it possible to relate, for instance, the probability of responding on D (=R) to the variables used in the analysis. There is, however, one restriction with regard to the use of the response indicators. They can only be used either as dependent variables or as independent variables in a logit equation in which another response indicator is explained. Because the response indicators are not allowed to influence the other variables, the modified Lisrel model from the previous section can simply be extended by including two additional modified path steps,

$$\pi_{defxghabcrs} = \pi_{def}\pi_{x|def}\pi_{g|defx}\pi_{h|defxg}\pi_{abc|x}\pi_{r|defxghabc}\pi_{s|defxghabcr} \,.$$

Of course, the conditional probabilities for R and S have to be restricted in some way to get a model which is identifiable.

As mentioned above, for the example data set, the mechanism causing the missing data is at least partially known. The probability of observing D depends on A, B and C, or on X, and on the interaction between E and F. The probability of observing G and H depends on F. Thus, according to the available information on the response mechanism, plausible logit models for R and S are

$$\begin{aligned} \pi_{r|defxghabc} &= \pi_{r|efx} = \frac{\exp\left(u_r^R + u_{xr}^{XR} + u_{er}^{ER} +_{fr}^{FR} + u_{efr}^{EFR}\right)}{\sum_r \exp\left(u_r^R + u_{xr}^{XR} + u_{er}^{ER} +_{fr}^{FR} + u_{efr}^{EFR}\right)} , \\ \pi_{s|defxghabcr} &= \pi_{s|fr} = \frac{\exp\left(u_s^S + u_{fs}^{FS} + u_{rs}^{RS}\right)}{\sum_s \exp\left(u_s^S + u_{fs}^{FS} + u_{rs}^{RS}\right)} . \end{aligned} \tag{7}$$

This is equivalent to assuming log-linear model $\{EFX, XR, EFR\}$ for marginal table $EFXR$ and model $\{FR, FS, RS\}$ for table FRS. The effect RS is included in the last model to reproduce exactly the number of persons in every particular subgroup.

4.3 Ignorable versus nonignorable response mechanisms

Because much of the theoretical work on nonresponse is based on the distinction between ignorable and nonignorable response mechanisms, I will pay some attention to the link between the approach presented here and these two types of response mechanisms (see also Vermunt, 1995).

As we saw above, the nonresponse is MAR if the probability of belonging to a particular subgroup depends only on the observed variables for every sample unit. So in terms of the response indicators R and S, the response mechanism is MAR if $\pi_{rs|defxghabc}$ equals to either $\pi_{11|defghabc}$, $\pi_{21|efghabc}$, $\pi_{12|defabc}$ or $\pi_{22|efabc}$. This is the least restrictive assumption about the distribution of R and S under which the response mechanism is ignorable for likelihood based inference. Note that the values on R and S depend on different variables for the different subgroups. More precisely, in every subgroup, the probability of observing just those variables depends only on the observed variables. In other words, the probability of belonging to a particular subgroup is independent of the missing variables for the subgroup concerned, including the latent variable X. These rather strange restrictions on the probability on R and S can, however, not be imposed by means of the approach presented here. So there is no direct link between the log-linear models for nonresponse and the distinction between ignorable and nonignorable nonresponse.

Using the models for nonresponse, the least restrictive model which fulfills the requisites of an ignorable response model is a model in which the values of the response indicators depend on all variables which are observed for all persons, that is, $\pi_{rs|defxghabc} = \pi_{rs|efabc}$. On the other hand, in the most restrictive ignorable model, the MCAR model, $\pi_{rs|defxghabc} = \pi_{rs}$, that is, the response indicators are assumed to be independent of all other variables. The MCAR random model is the response model which is actually tested by Fuchs (1982).

The only situation in which there exist a log-linear path model which is equivalent to a MAR response mechanism occurs in case of nested or monotone patterns of nonresponse (Vermunt, 1995). A pattern of nonresponse is nested when particular variables are missing more often than other ones, and for all persons with a particular missing variable all variables which are missing equally or more often are missing as well. Nested patterns of nonresponse occur often in panel studies: nonparticipation in one panel wave generally leads to nonparticipation in the subsequent waves too. In case of a nested pattern of nonresponse, a MAR model can be obtained by specifying a log-linear path model in which every response indicator is assumed to depend on the variables which are observed more often and on the response indicators belonging to these variables.

All response models which do not fulfill the above-mentioned conditions for ignorability are nonignorable. If R depends on D, it is clear that the response mechanism is nonignorable since the variable with missing data is directly related to its own response indicator. In other words, the probability of nonresponse depends on the variable with missing data. But also when S depends on D, the response mechanism is nonignorable. Although S does not indicate missingness on D, the mechanism is nonignorable because D is missing for some persons. The response model proposed in Equations (7) is nonignorable as well because both R and S depend on X which is missing for all persons.

It will be clear that, although the distinction between ignorable and nonignorable response mechanisms is valuable, it is just a theoretical distinction based on the fact whether it is necessary or not to specify the response mechanism for likelihood inference about the model parameters. Ignorability has no substantive meaning like the log-linear models for nonresponse discussed in this section. Therefore, one must be cautious when labeling a particular log-linear response model as ignorable or nonignorable. In the context of log-linear models for nonresponse, to my opinion, it has more sense to use another type of classification of response mechanism: the probability of responding on a particular variable depends also on the variable concerned or it does only depend on other variables. In the former case, the response mechanism is always nonignorable. Baker and Laird (1988) gave a nice example of an application in which the probability of nonresponse is allowed to depend on the variable with nonresponse. In the latter case, the response mechanism can be either ignorable or nonignorable.

4.4 Estimation

Fay (1986) proposed to estimate his causal models for patterns of nonresponse using the EM algorithm. However, he did not consider log-linear models with latent variables. Vermunt (1988, 1995) demonstrated how to adapt the E step of the EM algorithm proposed by Fay to situations in which also latent variables are included in the modified path model (see also Hagenaars, 1990). In fact, it is the same kind of solution as Hagenaars applied to generalize Fuchs' approach (Hagenaars, 1990). In the E step, the unobserved frequencies of the table $DEFXGHABCRS$ are computed via

$$\begin{aligned}
\hat{n}_{defxghabc11} &= n_{defghabc}\,\hat{\pi}_{x|defghabc11}\ , \\
\hat{n}_{defxghabc12} &= n_{efghabc}\,\hat{\pi}_{dx|efghabc12}\ , \\
\hat{n}_{defxghabc21} &= n_{defabc}\,\hat{\pi}_{xgh|defabc21}\ , \\
\hat{n}_{defxghabc22} &= n_{efabc}\,\hat{\pi}_{dxgh|efabc22}\ .
\end{aligned}$$

It can be seen that in every subgroup, or in every level of the joint response indicators, the expectation of the complete data given the observed data and the parameters estimates from the last iteration is obtained in a slightly different manner.

The above-mentioned procedure is the standard way of handling partially observed data in the program ℓEM (Vermunt, 1993). So estimation of log-linear models for nonresponse, including models with latent variables, can easily be performed using ℓEM. After specifying which variables are manifest variables, latent variables and response indicators, all variables can be used in the same way when specifying the various submodels of a modified path model.

5 Application

In this section, an application of the models discussed in the previous sections will be presented. For that purpose, the already mentioned 'Mathijssen-Sonnemans data' will be used. The observed variables are again: three ability tests (A, B and C), father's educational level (D), father's occupation (E), sex (F), educational level (G) and occupation (H). All variables are dichotomized, except for father's education, which has the following 4 categories: employees (1), independents (2), workers (3) and others (4). The dichotomous variables all have the categories low (1) and high (2), except for the variable sex (F), which has the categories male (1) and female (2). The reason why most variables are dichotomized is that when preparing the frequency tables, the models had to be estimated by means of LCAG (Vermunt, 1988). But still days of computer time were needed to estimate the models presented in this section. Using ℓEM 0.11, estimation of each of the models to be presented below toke less than two minutes.

First, I will present the analysis performed using only complete cases. Then, incomplete cases will be used in the analysis by means of Fuchs' (1982) procedure, that is, assuming MCAR nonresponse. And finally, Fay's models for nonresponse will be used to specify and test different kinds of models for the response mechanism.

5.1 Using only complete cases

The model which serves as starting point is the modified Lisrel model given in Equation (5), where the aim is, of course, to restrict the marginal and conditional probabilities using a logit parameterization.

As recommended by Hagenaars (1993), when specifying a modified Lisrel model, one can best start by investigating the measurement part of the model, or in other words, the relationships among the latent variables and their indicators. In this case, a latent class model has to be specified for the relationships among the latent ability variable X and the three ability test A, B, and C. The latent ability variable X is assumed to have two categories.

	Model	Model fit L^2	df	p	Conditional tests models	L^2	df	p
1	$\{DEFXGH, XA, XB, XC\}$	363.6	378	.70				
2	(1) + $\{DE, DF\}$	374.2	385	.64	(2)-(1)	10.6	7	.18
3	(1) + $\{DX, EFX\}$	366.6	385	.74	(3)-(1)	3.0	7	.89
4	(1) + $\{DG, EG, XG\}$	388.5	404	.70	(4)-(1)	24.9	26	.52
5	(1) + $\{EH, FH, XH, GH\}$	417.7	435	.72	(5)-(1)	54.1	57	.58
6	(1) + (2) + (3) + (4) + (5)	458.8	474	.68	(6)-(1)	95.2	96	.50

Table 1: Test results for some models using only complete cases

To test the fit of the measurement model, one can start assuming the relationships among the other variables to be saturated. So no restrictions need to be imposed yet on the relationships among the structural variables D, E, F, X, G and H. The simplest way to accomplish this is by specifying log-linear model $\{DEFXGH, XA, XB, XC\}$ for the complete table $DEFXGHABC$. Note that, given X, the variables A, B and C are not only assumed to be mutually independent, but also to be independent of the joint variable $DEFGH$. As can be seen from the test results given in Table 1, the measurement model (Model 1) fits very well ($L^2 = 363.6, df = 378, p = .70$).

Next, it was tried to find a more parsimonious specification for the structural relationships among the variables D, E, F, X, G and H. For that purpose, a step-wise model selection procedure was used per subtable, leaving the other submodels saturated. By testing these models against each other and against Model 1, it could be seen whether the more parsimonious specification for the marginal or conditional probability concerned deteriorated the fit or not. Here, only the test results for the best fitting subtable specific models will be presented.

For the relationships among the exogenous variables father's education (D), father's occupation's (E) and sex (F), model $\{DE, DF\}$ (Model 2) provides a good description of the data. Model 2 does not fit worse than Model 1 ($L^2 = 10.6, df = 7, p = .18$). Since it is not plausible that in the population sex is related to the father's educational level, the significance of the effect DF is almost certainly the result of the selectivity of the nonresponse.

The latent variable ability (X) was found to depend on D, E, F on the interaction of E and F. Restricting $\pi_{x|def}$ via log-linear model $\{DEF, DFX, EX\}$ for marginal table $DEFX$ (Model 3) does not deteriorate the fit compared to Model 1 ($L^2 = 3.0, df = 7, p = .89$).

A good fitting parsimonious model for $\pi_{g|defx}$ (educational level) is obtained via model $\{DEFX, DG, EG, XG\}$ for subtable $DEFXG$. So G depends on D, E and X. The conditional test of this model (Model 4) against Model 1 gives a nonsignificant result ($L^2 = 24.9, df = 26, p = .52$).

Variable H (occupational level) seemed to depend only on E, F, X and G. Model $\{DEFXG, EH, FH, XH, GH\}$ for table $DEFXGH$ (Model 5) does no fit worse than Model 1 ($L^2 = 54.1, df = 57, p = .58$).

The final model in which all best fitting submodels were combined (Model 6) fits the data very well ($L^2 = 458.8, df = 474, p = .68$. Moreover, like all submodels presented above, it does not significantly fit worse than Model 1 ($L^2 = 95.2, df = 96, p = .50$. The final model which is depicted in Figure 1 is very parsimonious. It contains only 34 independent log-linear parameters.

Table 2 presents the estimates for the two-variable and three-variable interaction terms according to Model 6. The log-linear parameters for the measurement model show that the indicators are strongly related to the latent variable X: $u_{11}^{XA} = .9225$, $u_{11}^{XB} = .5307$ and $u_{11}^{XC} = .6829$.

The parameters for the relationships among the exogenous variables indicate that children of employees have higher educated father's that other children ($u_{11}^{DE} = -.8441$). The value .1233 for u_{11}^{DF} indicates that males have lower educated father's than females. It can be expected that this artificial effect disappears when the partially observed data is used in the analysis.

Children of lower educated father's have a much lower school ability than children of higher educated father's ($u_{11}^{DX} = .4125$), males have a lower school ability than females ($u_{11}^{FX} = .1089$), and children of employees have a much higher school ability than other children ($u_{11}^{EX} = -.4970$). Moreover, the three-variable parameters u_{efx}^{EFX} show that the relationship between father's occupation and school ability is stronger for males than for females.

The educational level of the father has a positive effect on the educational level of the respondent ($u_{11}^{DG} = .1848$). Moreover, children of employees are relatively high educated ($u_{11}^{EG} = -.3988$) and children of workers are relatively low educated ($u_{11}^{EG} = .3893$). School ability has a strong positive effect on the final educational level ($u_{11}^{XG} = .3893$).

The most important determinant of the occupational status of the respondent is the respondent's own educational level ($u_{11}^{GH} = .5426$). Moreover, holding constant other factors, males have more often an occupation with a high status than females ($u_{11}^{FH} = -.3387$). Also, there exists a small positive

Parameter	Model 6	Model 11a
u_{11}^{DE}	-0.8441	-0.8461
u_{12}^{DE}	0.3724	0.2908
u_{13}^{DE}	0.5061	0.4940
u_{11}^{DF}	0.1233	
u_{11}^{DX}	0.4125	0.3047
u_{11}^{EX}	-0.4970	-0.5370
u_{21}^{EX}	0.2137	0.1041
u_{31}^{EX}	0.2708	0.1478
u_{11}^{FX}	0.1089	-0.0680
u_{111}^{EFX}	-0.2039	
u_{211}^{EFX}	0.2346	
u_{311}^{EFX}	0.2944	
u_{11}^{DG}	0.1848	0.2062
u_{11}^{EG}	-0.3988	-0.3587
u_{21}^{EG}	-0.0576	-0.0813
u_{31}^{EG}	0.3893	0.3929
u_{11}^{XG}	0.3724	0.4216
u_{11}^{EH}	-0.1846	-0.2460
u_{21}^{EH}	-0.1045	-0.1166
u_{31}^{EH}	0.1605	0.1541
u_{11}^{FH}	-0.3387	-0.3353
u_{11}^{XH}	0.1316	0.1736
u_{11}^{GH}	0.5426	0.5484
u_{111}^{EXH}		-0.0837
u_{211}^{EXH}		-0.1204
u_{311}^{EXH}		0.0311
u_{11}^{XA}	0.9225	1.0489
u_{11}^{XB}	0.5307	0.5289
u_{11}^{XC}	0.6829	0.6975

Table 2: Log-linear effects among the survey variables according to Model 6 and Model 11a

effect of X on H ($u_{11}^{XH} = .1316$). And finally, children of workers have a higher probability of having an occupation with a low status ($u_{31}^{EH} = .1605$) than others.

5.2 Fuchs' procedure for nonresponse

Fuchs' procedure to handle partially observed tables (Fuchs, 1982) can be implemented in ℓEM by specifying a model for the response mechanism which is equivalent to the MCAR assumption. In other words, the probability of $R = r$ and $S = s$ is assumed to independent of the other variables included in the model.

Table 3 presents the test results for some models which were estimated using all the available data. The best way to start the analysis when some data is missing, is to test the MCAR assumption itself. This can be accomplished by specifying model $\{DEFGHABC, RS\}$ for the table $DEFGHABCRS$ (Model 7). By assuming a saturated model for the relationships among the variables and, moreover, R and S to be independent of the other variables, one has a direct test for the MCAR assumption. As could be expected on the basis of the available information on the response mechanism, this response model must be rejected ($L^2 = 2419.6, df = 445, p = .00$). Nevertheless, more parsimonious models for the relationships among the research variables may be specified, that is, for the structural model and the measurement model. The fit of such models can be tested by comparing them with Model 7. Because any ignorable response model gives the same parameters estimates for the structural and the measurement model, a conditional test against Model 7 is a test of the model concerned, given that the response mechanism is MAR (Fuchs, 1982).

Model 8 is equivalent to Model 1, that is, it can be used to test separately the measurement part of the model. On the basis of the conditional test against Model 7, it must be concluded that the measurement model fit less well when all available data is used ($L^2 = 438.6, df = 378, p = .00$). This may be the result of the increased power of the significance test. However, it may also be an indication that the measurement model is only appropriate for the selective group without missing data. Although the fit of Model 8 can be improved by supplying it with some additional parameters, here the measurement model will be left just as it is. In other words, prevalence will be given to the principle of parsimony. Like Model 1, Model 8 will be used

	Model	Model fit L^2	df	p	Conditional tests models	L^2	df	p
7	MCAR	2419.6	445	.00				
8	(1) + MCAR	2858.2	823	.00	(8)-(7)	438.6	378	.02
9	(6) + MCAR	2950.9	919	.00	(9)-(8)	92.7	96	.58
9a	(9) - DF	2951.7	920	.00	(9a)-(9)	0.8	1	.37
9b	(9a) - EFX	2957.8	923	.00	(9b)-(9a)	6.9	3	.08
9c	(9b) + EXH	2949.8	920	.00	(9b)-(9c)	8.0	3	.05
				.00	(9c)-(7)	530.2	475	.04

Table 3: Test results for some models using Fuchs' approach

as a reference point for the more restricted structural models.

Model 9, which is equivalent to the final model for the complete data (Model 6), does not fit worse than Model 8 ($L^2 = 92.7, df = 96, p = .58$). In Model 9a, the effect DF was excluded from the model to test whether the significance of the effect DF was caused by the selectivity of the nonresponse. Since leaving out this effect does not lead to a worse fit ($L^2 = 0.8, df = 1, p = .37$), it can be concluded that the significance of the effect DF resulted from analyzing complete cases only.

Starting with Model 9a, it was tried to find more parsimonious models for the other submodels as well. The only effect which was not significant anymore was the three-variable interaction among E, F and X. In Model 9b, this effect is set equal to zero ($L^2 = 6.9, df = 3, p = .08$). It is plausible that this effect is also caused by the nonresponse because particular combinations of X, E and F have higher probabilities of responding on D.

Because of the greater power of the significance tests when using all available data, particular effects which where not significant when using only the complete data can become significant now. This was checked for effects with significance levels between 5 and 10 percent in the analysis performed using only complete cases. There is only one effect which becomes significant now: the interaction between E and X with respect to their effect on H. Model 9c containing the effect EXH fits better than Model 9b ($L^2 = 8, df = 3, p < .05$).

	Model	Model fit L^2	df	p	Conditional tests models	L^2	df	p
10	(9c) + $\{EFXR\}$							
	+ $\{DEFXS, RS\}$	1032.0	874	.00	(9c)-(10)	1917.8	46	.00
10a	(9c) + $\{EFABCR\}$							
	+ $\{DEFABCS, RS\}$	735.4	730	.44	(9c)-(10a)	2214.4	190	.00
11	(9c) + $\{EFR, XR\}$							
	+ $\{RS, FS\}$	1150.5	911	.00	(9c)-(11)	1799.3	9	.00
					(11)-(10)	118.5	37	.00
11a	(11) + FXS + FXR	1097.5	908	.00	(11)-(11a)	53.0	3	.00
					(11a)-(10)	65.5	34	.00
12	(11a) + DR	1097.3	907	.00	(11a)-(12)	0.2	1	.65
13	(11a) + GS + HS	1093.5	906	.00	(11a)-(13)	4.0	2	.14

Table 4: Test results for some models for nonresponse

5.3 Models for patterns of nonresponse

When modeling the response mechanism, Model 9c will be taken as starting point. It will be tried to make this model better fitting by adding parameters describing the response mechanism. From Model 7, it is known that at most 2419.6 in L^2 can be gained by using all 445 degrees of freedom for the specification of the response model. In that case, the L^2-value of the model would be 530.2 with $df = 475$ (see (9c)-(7) in Table 3), which is the test result for Model 9c in case the nonresponse is MAR. The MAR model can be seen as a saturated ignorable response model. Of course, here we are interested in much more parsimonious specifications of the response mechanism.

First, log-linear models $\{EFXR\}$ and $\{DEFXS, RS\}$ were specified for the probability that $R = r$ and $S = s$ respectively (Model 10). The other part of the model is equivalent to Model 9c. Actually, Model 10 is the most extended plausible model in which the response indicators are not influenced by the variables which response probability they indicate. In Model 10, R is assumed to depend on E, F and X, including all their interaction terms. Of course, it not possible that R depends on the variables G and H which are measured many years later. Furthermore, S is assumed to depend on all variables, except for G and H, the variables which missingness it indicates. The effect RS is included in the model to fix the sample sizes of the four subgroups. Comparison of Model 10 with Model 9c, shows that most of the information on the response mechanism is captured by Model 10 (see Table 4): The L^2-value improves with 1917.8, using only 46 degrees of freedom. However, there are still 399 degrees of freedom left to gain 502.2 in L^2. The reason why Model 10 does not fit as perfect as could be expected is that R and S are regressed on the latent variable X instead of the indicators A, B and C. This can be seen from the fact that Model 10a, where X is replaced by A, B and C, does not fit significantly worse than the saturated MAR model ($L^2 = 735.4 - 530.2 = 205.2, df = 730 - 475 = 245, p = .97$). Apparently, the assumption that A, B, C and R are mutually independent given X is a bit too strong. This is the same kind of problem that was encountered when testing the measurement model (Model 8). On the basis of the same arguments as we used above, we will continue assuming that given X the indicators are conditional independent of all other variables, including the response indicators. Therefore, X will be used as regressor when specifying the response mechanism and not the indicators A, B and C.

Model 11 is the response model of Equations (7), the model that was formulated on the basis of the a priori knowledge about the response mechanism. Using only 9 additional parameters compared to Model 9c, Model 11 captures a very large part of the mechanism causing the nonresponse: $L^2 = 1799.3$. So the a priori information on the mechanism causing the nonresponse is confirmed by the analysis. Omitting any of the parameters describing the response mechanism of Model 11 deteriorates the fit a lot. However, in terms of fit, Model 11 is inferior to Model 10 ($L^2 = 118.5, df = 37, p = .00$). Therefore, it was tried to improve the fit of Model 11 by adding some extra parameters. This resulted in Model 11a which contains 3 additional parameters, namely: the interaction of F and X with respect to their effect on R, the effect of X on S and the interaction effect of F and X on S. Although Model 11a still differs significantly from Model 10 ($L^2 = 65.5, df = 34, p = .00$), no other single parameter could considerably improve the fit anymore.

The parameters estimates for Model 6 and Model 11a are given in Table 2. It can be seen that apart from the fact that particular parameters are not significant anymore and that one effect becomes significant, the parameter estimates for the relationships among the research variables do not change very when partially observed cases are used in the analysis. Perhaps the most interesting difference between the two models is that according to Model 6, males have a bit lower school ability than females ($u_{11}^{FX} = .1089$), while according to Model 11a, males have a bit higher school ability than females ($u_{11}^{FX} = -.0680$).

Table 5 contains the parameter estimates for the response model according to Model 11a. The parameter estimates show that school ability deter-

Parameter	Model 11a
u_{11}^{ER}	0.0929
u_{21}^{ER}	0.2356
u_{31}^{ER}	0.3747
u_{11}^{FR}	0.2037
u_{11}^{XR}	-0.7070
u_{111}^{EFR}	-0.0968
u_{211}^{EFR}	0.1427
u_{311}^{EFR}	0.1133
u_{111}^{FXR}	0.1814
u_{11}^{FS}	0.2182
u_{11}^{XS}	-0.0672
u_{11}^{RS}	0.1424
u_{111}^{FXS}	0.0603

Table 5: Log-linear effects for the response model according to Model 11a

mines very strongly the probability of observing D: $u_{11}^{XR} = -.7070$. Moreover, children of independents and workers have higher probabilities of D being observed: $u_{21}^{ER} = .2356$ and $u_{31}^{ER} = .3747$. The three-variable interaction term shows that this effect is stronger for males than for females. And, for males D is observed more often than for females ($u_{11}^{FR} = .2037$). The parameter $u_{11}^{FS} = .2182$ shows that for males there is a higher probability of observing G and H than for females. Moreover, females with a low school ability have a lower probability of responding on G and H ($u_{11}^{XS} - u_{211}^{FXS} = -.0672 - .0603 = -.1275$), while males with a high school ability have a higher probability of responding on G and H.

Finally, two models were estimated in which a direct effect between the variables with missing variables and their response indicators was included. Model 12 is equivalent to Model 11a, except for that it contains a direct effect of D on R. This effect is clearly not significant. The same applies to Model 13: Neither the effect of G on S nor the effect of H of S is significant.

Summarizing, for this application, the modified Lisrel model extended with Fay's approach to partially observed data gave a very parsimonious description of both the causal relationships among the research variables and the response mechanism. The 960 observed frequencies were described using only 52 parameters: 35 for the causal log-linear model and 17 for the response model. Using partially observed data made it possible to detect some artificial effects. Moreover, because of the greater power of the tests, one effect became significant. The models for nonresponse confirmed the a priori knowledge about the response mechanism. However, the overall fit of the final model is not perfect. This is caused by the fact that the measurement model does not fit very well when all available data is used. By means of a more elaborate analysis of this data set, it would, of course, be possible to find the additional parameters to get an even better description of the observed data.

6 Discussion

A general approach for specifying and testing causal models for categorical variables was presented. It can be used to specify log-linear models with observed, partially observed and unobserved variables. The general log-linear model combines a structural model in which the causal relations among the structural variables are specified, a measurement model in which the relations between the latent variables and their indicators are specified, and a model for the response mechanism. The application demonstrated very well the value of this model. The relationships among the research variables could be decribed using a small number of log-linear parameters. When using the partially observed data, particular effects became significant as a result of the increased power of the statistical tests, and artificial effects were discovered which resulted from selective nonresponse. Although not demonstrated by the example, when the probability of nonresponse on a particular variable depends strongly on the value of the variable concerned, the parameter estimates of the model for the survey variables may change a lot as well (Baker and Laird, 1988).

In the application, only hierarchical log-linear model were used. However, it is also possible to impose all kinds of linear restriction on the log-linear parameters, that is, to specify non-hierarchical log-linear models per subtable. This becomes more important, when the model contains several polytomous variables with ordered or equidistant categories. The possibility to restrict the log-linear parameters can, for instance, be used to specify discrete approximations of latent trait models (Heinen, 1992, Vermunt and Georg, 1995). Because of the equivalence of the more general log-linear or logit model to the logistic regression model, it can also be used to incorporate continuous exogenous variables into a modified Lisrel model (Vermunt, 1995).

The way of specifying the causal order among the structural variables in modified Lisrel models is similar to the specification of (latent) Markov models (Van de Pol and Langeheine, 1990). Actually, the (latent) Markov model is a special case of the modified Lisrel model. The approach presented here is more general since the more flexible way of specifying the conditional probability structure makes it possible to relax the basic assumptions of the (latent) Markov model. The modified Lisrel model can, for instance, be used to specify multivariate Markov models (Vermunt, 1995) and latent Markov models with correlated errors (Bassi, Croon, Hagenaars and Vermunt, 1995). Furthermore, the possibility to parametrize the conditional probabilities by means of a logit model can, among other things, be used to specify regression models for the transition probabilities which are similar to discrete-time event history models (Vermunt, 1995).

Although in the example the latent class model was used as a measurement model, it can also be used for unmixing populations having different structural parameters (Titterington, Smith and Makov, 1985). Examples of the use of the latent class model as a finite mixture or discrete compound model are the mixed logit model (Formann, 1992), the mixed Markov model (Van de Pol and Langeheine, 1990), and the mixed Rasch model (Rost, 1990). Specifyng these kinds of mixed models within the modified Lisrel approach involves including a latent variable without indicators into the model.

And finally, the causal log-linear models that were presented are recursive models, that is, models in which the causal relationships among the variables are uni-directional. Recently Mare and Winship (1991) proposed 'non-recursive' log-linear models with reciprocal effects. Although not demonstrated here, the rather complicated 'non-recursive' log-linear models proposed by Mare and Winship can also be handled within the modified Lisrel approach.

References

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

Baker, S.G., N.M. Laird, 1988, Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Sta-*

tistical Association, 83, 62-69.

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

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

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

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

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Ser. B.*, 39, 1-38.

Fay, R.E., 1986, Causal models for patterns of nonresponse. *Journal of the American Statistical Association*, 81, 354-365.

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

Fuchs, C., 1982, Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, 77, 270-278.

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

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

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

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

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

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

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

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

Hartog, J., *Survey nonresponse in relation to ability and family background: structure and effects on estimated earning functions.* Research Memorandum no. 8620, University of Amsterdam.

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

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

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

Little, R.J., 1982, Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.

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

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

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

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

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

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

Rubin, D.B., 1976, Inference and missing data. *Biometrika*, 63, 581-592.

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

Vermunt, J.K., 1988, *Loglineaire modellen met latente variabelen en missing data.* Thesis, Tilburg.

Vermunt, J.K. (1993). *LEM: log-linear and event history analysis with missing data using the EM algorithm.* WORC PAPER 93.09.015/7, Tilburg University.

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

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

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