

Tilburg University

A general class of nonparametric models for ordinal categorical data

Vermunt, J.K.

Published in:
Sociological Methodology

Publication date:
1999

Document Version
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Vermunt, J. K. (1999). A general class of nonparametric models for ordinal categorical data. *Sociological Methodology*, 29, 197-221.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A general non-parametric approach to the analysis of ordinal categorical data

Jeroen K. Vermunt

Department of Methodology

Tilburg University

Abstract

This paper presents a general class of models for ordinal categorical data which can be specified by means of linear and/or log-linear equality and/or inequality restrictions on the (conditional) probabilities of a multi-way contingency table. Some special cases are models with ordered local odds ratios, models with ordered cumulative response probabilities, order-restricted row association and column association models, and models for stochastically ordered marginal distributions. A simple uni-dimensional Newton algorithm is proposed for obtaining the restricted maximum likelihood estimates. In situations in which there is some kind of missing data, this algorithm can be implemented in the M step of an EM algorithm. Computation of p-values of testing statistics is performed by means of parametric bootstrapping.

A general class of non-parametric models for ordinal categorical data

Jeroen K. Vermunt

Department of Methodology

Tilburg University

1 Introduction

Although the variables and the relationships that are studied in the social sciences are often of an ordinal nature, truly ordinal models are rarely used. Researchers confronted with ordinal data generally use nominal, interval, or quasi-ordinal methods. When using nominal analyses methods, such as standard hierarchical log-linear models, the ordinal variables are treated as nominal variables, which means that the information on the order of their categories is ignored. Interval analyses are based on assigning scores to the categories of the ordinal variables, as in linear-by-linear association models (Goodman, 1979; Haberman, 1979). The assumption of known category scores implies that the ordinal variables are actually treated as interval level variables. And finally, quasi-ordinal analyses involve estimating category scores for the ordinal variables, such as in log bi-linear association and correspondence models (Goodman, 1979, 1986; Clogg, 1982; Gilula and Haberman, 1988; Clogg and Shihadeh, 1994). Although the latter type of methods yield easily interpretable results when the estimated scores have the assumed order, there is no guarantee that the estimated ordering of the categories will be the expected one.

This paper follows a different modeling strategy for ordinal categorical variables. A non-parametric approach is presented which is based on imposing linear or log-linear inequality restrictions on (conditional) probabilities. This approach is truly ordinal in the sense that the estimated probabilities satisfy the specified order restrictions without the necessity of assuming the variables to be measured on interval level. Although not very well-known among social scientist, linear and log-linear inequality restrictions have been advocated by several authors for the specification of relationships between ordinal categorical variables (Grove, 1980; Agresti and Chuang, 1986; Agresti, Chuang and Kezouh, 1987; Dykstra and Lemke, 1988; Robertson, Wright, and Dykstra 1988; Croon, 1990, 1991; Ritov and Gilula, 1993; Agresti and Coull, 1996; Evans, Gilula, Guttman, and Swartz, 1997; Hoijtink and Molenaar, 1997).

The general form in which the inequality restrictions are presented in this paper makes it possible to formulate non-parametric variants of log-linear models for cell probabilities, of logit models, and of linear models for cumulative and mean responses. It is also shown that the combination of inequality restrictions with equality restrictions makes it possible to specify hybrid models having both parametric and non-parametric features. An example is a row association model with ordered row scores.

Estimation of the order-restricted probabilities is performed by means of

maximum likelihood using the method of activated constraints. A simple uni-dimensional Newton procedure for solving the corresponding Lagrange likelihood equations is presented. It is also demonstrated that the same procedure can be used in conjunction with the EM algorithm, which makes it possible to apply the proposed inequality restrictions in situations in which some of the variables are partially or completely missing (latent). In addition, attention is paid to likelihood-ratio tests based on asymptotic distribution functions and bootstrapping.

First, a small empirical example is presented to illustrate the differences between parametric and non-parametric analyses of ordinal categorical data. Then, the general class of restrictions yielding the ordinal models of interest is described. Next, attention is paid to maximum likelihood estimation with and without missing data and to model testing. And finally, the use of the proposed ordinal models is exemplified by means of a number of empirical examples.

2 An example

This section illustrates the possible benefits of using non-parametric models for ordinal data by means of a small example. The example concerns the analysis of the two-way cross-classification reported in table 1. This table, which is taken from table 2 of Clogg's 1982 Journal of the American Statistical Association paper on ordinal log-linear models, describes the relationship between number of siblings (S) and happiness (H). The original table is a three-way cross-classification of the ordinal variables years of schooling, number of siblings, and happiness. For this example, the original table is collapsed over education yielding the 5-by-3 table formed by S and H , in which S serves as row variable and H as column variable.

[INSERT TABLE 1 ABOUT HERE]

The use of parametric or non-parametric ordinal approaches to the analysis of categorical data makes, of course, only sense if there is some reason to assume that the relationship between the variables of interest is of an ordinal nature. Let us assume that we want to test whether there is a positive relationship between number of siblings and happiness, or, worded differently, whether individuals having more siblings are happier than individuals having less siblings.

One way of defining such a positive relationship is on the basis of the cumulative conditional responses on happiness given a person's number of siblings. In that case, we treat happiness as dependent variable and number of siblings as independent. Let $\pi_{h|s}$ denote the conditional probability that $H = h$ given that $S = s$. In addition, let $F_{h|s}$ denote the cumulative conditional probability that $H \leq h$ given that $S = s$, which is defined as

$$F_{h|s} = \sum_{p=1}^h \pi_{p|s}.$$

A positive relationship between S and H implies that

$$F_{h|s} \geq F_{h|s+1}, \quad (1)$$

or that the cumulative conditional probability that $H = h$ decreases or remains equal as S increases. We may also say that the cumulative probabilities are monotonically non-increasing. Note that even if this assumption holds for the population, as result of sampling error, this may not hold for the data. Table 2 reports the cumulative conditional probabilities calculated from the observed cell entries reported in table 1. As can be seen, there are several order violations in the data.

[INSERT TABLE 2 ABOUT HERE]

Another way of defining a positive relationship is on the basis of the local odds ratios. Let π_{sh} denote the probability that $H = h$ and $S = s$. In addition, let θ_{sh} denote a local odds ratio, which is defined as

$$\theta_{sh} = \frac{\pi_{sh}\pi_{s+1h+1}}{\pi_{sh+1}\pi_{s+1h}}.$$

Using this measure, a positive relationship involves

$$\theta_{sh} \geq 1, \quad (2)$$

or that each local odds ratio in the two-way table is larger than or equal to 1. As can be seen from table 3, the pattern of observed odds ratios is not in agreement with the definition of a positive relationship since some of them are smaller than 1.

[INSERT TABLE 3 ABOUT HERE]

The fact that the data are not fully in agreement with the assumption of an ordinal relationship may be the result of sampling error. One way of testing whether the observed order violations are the result of sampling error is by imposing restrictions on the cumulative conditional probabilities or the local odds ratios using some kind of parametric model.

[INSERT TABLE 4 ABOUT HERE]

Table 4 reports the test results for the estimated parametric models. As can be seen, the independence model does not fit the data ($L_{A.1}^2 = 26.27, df = 8, p < .01$), which indicates that there is an association between H and S . The next model (A.2) is a logit model for the cumulative response probabilities, i.e.,

$$\ln \frac{F_{h|s}}{1 - F_{h|s}} = \alpha_h + \beta_s.$$

It should be noted that this parametric model fulfils the conditions specified in equation 1 only if $\beta_s \geq \beta_{s+1}$. The model could be characterized as ordinal-nominal because it treats the dependent variable as ordinal and the independent as nominal. The cumulative logit model does not fit the data:

$L_{A,2}^2 = 18.52, df = 4, p = .00$. Apparently, its underlying assumption of proportional odds does not hold for this data set. In addition, the estimated β_s 's are out of order, which means that also the assumption of monotonically non-increasing $F_{h|s}$'s is not satisfied.

A parametric model that can be used to restrict local odds ratios is the row-column (RC) association model (Goodman, 1979; Clogg, 1982) which is defined by

$$\ln \pi_{sh} = u + u_s^S + u_h^H + \nu_s^S \nu_h^H .$$

Here, ν_s^S and ν_h^H are unknown 'scores' for the levels of S and H . The RC model satisfies the conditions described in equation 2 if $\nu_s^S \leq \nu_{s+1}^S$ and $\nu_h^H \leq \nu_{h+1}^H$, i.e., if the row and column scores are monotonically non-decreasing. Because the RC model does not restrict the row and column scores to be ordered, it can be labeled as nominal-nominal, or quasi ordinal. The RC model fits the data quite well: $L_{A,3}^2 = 7.33, df = 3, p = .06$. A problem is, however, that both the row and the column scores are out of order. More precisely, the order of the scores for rows 3 and 4 and for columns 1 and 2 is incorrect.

One way to prevent the occurrence of solutions that are out of order is to assign a priori scores to the levels of S and H . Note that this amounts to assuming that the variable concerned is of interval measurement level. Although for simplicity of exposition here I will work only with equal-interval scores, any set of scores which is in agreement with the assumed order may be used. Three restricted variants of the above RC model can be obtained, depending on whether we use a priori (equal-interval) scores for the column variable (H), the row variable (S), or both. The resulting models are known as row (R), column (C), and uniform (U) association model, respectively. These models could be labeled as nominal-interval, interval-nominal, and interval-interval.

The test results reported in table 4 show that the R model does not fit the data ($L_{A,4}^2 = 17.52, df = 4, p < .01$), which indicates that H may not be treated as interval level variable. In addition, the estimated scores for S are not ordered: the score for row 4 is slightly higher than for row 5. The C model fits very well ($L_{A,5}^2 = 8.36, df = 6, p = .21$), but again the category scores, in this case for $H = 1$ and $H = 2$, have an incorrect order. The uniform association model (Model A.6) does not fit the data at all ($L_{A,6}^2 = 20.21, df = 7, p = .01$), which indicates that the assumption that H and S are interval level variables is too strong. Nevertheless, the uniform association parameter is significant and has the 'expected' positive sign.

The above parametric ordinal approach illustrates that on the one hand the specified models make too strong assumptions, such as proportional odds or constant local odds ratios. On the other hand, they may not be restrictive enough in the sense that they do not force the solution to be ordered in one of the ways defined above. This is the main reason for proposing a non-parametric approach for the kinds of problems we are dealing with here.

The next two models reported in table 4 are non-parametric. More precisely, Model A.7 is defined by the inequality restrictions described in equation 1 and Model A.8 by the restrictions described in equation 2. The model obtained by

imposing inequality restrictions on the cumulative conditional probabilities fits quite well: $L_{A,7}^2 = 5.50, p \approx .10$. The fit of the model restricting the local odds ratios to be at least 1 is somewhat worse: $L_{A,8}^2 = 8.32, p \approx .04$.¹

The estimated cumulative probabilities for Model A.7 and the estimated odds ratios for Model A.8 are reported in tables 2 and 3, respectively. These ‘parameter’ estimates show very well the nature of an order-restricted maximum likelihood (ML) solution: as long as an order restriction is not violated, nothing happens, but if an order is violated, the corresponding estimate gets a boundary value. In the current situation, this involves equating adjacent cumulative conditional probabilities or equating local odds ratios to 1. It should be noted that although such a procedure seems to be simple to implement, from that data it cannot always be determined which restrictions have to be imposed. This can be seen from the ML solution for Model A.8 which contains 3 odds ratios equal to 1 while in the observed table there were 4 odds ratios smaller than 1.

A disadvantage of using the non-parametric approach is that there are no real parameters to report. The interpretation of the results has to be based on the fit statistics and on the estimated values of the probabilities or the functions of probabilities for which order restrictions were specified: in the above examples, these were the cumulative probabilities and the local odds ratios. To deal with this problem, we will also present models which combine parametric and non-parametric features, such as row association models with order-restricted row scores.

Another disadvantage of using non-parametric models is that estimation and testing is much more complicated than for parametric models. One of the objectives of this paper is to show that a quite general class of non-parametric models can be estimated with a very simple algorithm. In addition, the availability of fast computers, makes goodness-of-fit testing and computation of standard errors of the relevant measures, such as local odds ratios and cumulative probabilities, using computationally intensive resampling methods, feasible.

3 A general class of (in)equality restrictions

This section discusses linear and log-linear restrictions on (conditional) probabilities that can be used for specifying ordinal models for categorical data. Although the specification of ordinal models is based on imposing inequality constraints, we also discuss equality constraints of the same form. This is because of didactic reasons – equality constraints are somewhat easier to understand and most inequality constraints are variants or extensions of the simpler equality constraints – and because in some situations it may be relevant to combine the two types of restrictions. For each of the four types of constraints – linear equalities, linear inequalities, log-linear equalities, and log-linear inequalities –

¹As is explained in section 5, the p -values for the order-restricted models are estimated by means of bootstrapping. The reported p -values are point estimates based on 1000 bootstrap samples. The number of degrees of freedom is not defined in models with inequality constraints.

a number of possible applications is presented.

Let n_{ij} denote an observed cell count in an I -by- J table, where i serves as index for the (possibly composite, possibly degenerate) independent variable X and j for the (possibly composite) dependent variable Y . For example, Y might be a bivariate random vector (Y_1, Y_2) , in which case $j=(j_1, j_2)$ would index the possible level combinations of Y_1 and Y_2 . In situations in which no distinction is made between dependent and independent variables, X has only one level, which makes the index i redundant. The conditional probability that $Y = j$ given that $X = i$ is denoted by $\pi_{j|i}$.

Linear equality restrictions

The first type of restrictions are linear equality restrictions on the (conditional) probabilities $\pi_{j|i}$. The p th restriction of this form is defined by

$$\sum_{ij} z_{1ijp} \pi_{j|i} - c_{1p} = 0. \quad (3)$$

As can be seen, a linear combination of $\pi_{j|i}$'s defined by the z_{1ijp} 's minus some constant c_{1p} is postulated to be equal to zero. In most situations, c_{1p} will be 0. It should be noted that, in fact, we have a linear model for (conditional) probabilities that is well-known from the GSK framework (Grizzle, Starmer, and Koch, 1977).

These linear equality restrictions can be used to test several types of assumptions on the relationships between categorical variables. Some examples are independence, equal means, conditional independence, marginal homogeneity, equal marginal means, and symmetry.

Suppose that we are studying the relationship between two categorical variables A and B , with category indexes a and b , respectively. Using the above linear equality restrictions, independence between A and B can be specified as

$$\pi_{a|b} - \pi_{a|b+1} = 0.$$

Let $F_{a|b}$ denote that (cumulative) probability that $A \leq a$, given $B = b$: $F_{a|b} = \sum_{p=1}^a \pi_{p|b}$. An alternative formulation of the independence assumption is in terms of these cumulative probabilities:

$$F_{a|b} - F_{a|b+1} = \sum_{p=1}^a \pi_{p|b} - \sum_{p=1}^a \pi_{p|b+1} = 0.$$

Although working with cumulative probabilities seems to be unnecessary complicated, it will prove very useful in the context of inequality restrictions.

A less restrictive assumption than independence, which makes sense only if A is an interval level variable, is that the mean of A is the same for each category of B . Using ν_a^A as category scores for A , such an assumption can be formulated as

$$\mu_b^A - \mu_{b+1}^A = \sum_a \nu_a^A \pi_{a|b} - \sum_a \nu_a^A \pi_{a|b+1} = 0,$$

where μ_b^A is the mean of A for $B = b$.

The generalization of the independence assumption to a multivariate context yields the conditional independence assumption. Suppose that A and B are independent of one another within the levels of a third variable C with index c . Such a conditional independence model can be specified as

$$\pi_{a|bc} - \pi_{a|b+1c} = 0.$$

Of course, as in the bivariate case, we may also specify this hypothesis using the cumulative probabilities $F_{a|bc}$.

Rather than testing assumptions on the level of conditional probabilities as in the above examples, it is also possible to formulate hypotheses which have to be specified in the form of linear restrictions on the joint probability distribution of a set of categorical variables. An example is the marginal homogeneity assumption for a two-way square table formed by the variables A and B , which can be defined as:

$$\pi_{a.} - \pi_{.a} = \sum_b \pi_{ab} - \sum_b \pi_{ba} = 0.$$

Here, π_{ab} denotes the probability that $A = a$ and $B = b$, and a dot indicates that the corresponding probability is obtained by summation over the subscript concerned. For instance, $\pi_{a.} = \sum_b \pi_{ab}$.

Let $F_{a.}$ denote the cumulative marginal probability that $A \leq a$: $F_{a.} = \sum_{p=1}^a \sum_b \pi_{pb}$. In a similar way, we can define the cumulative marginal probability that $B \leq b$, $F_{.b}$. The marginal homogeneity model can also be specified in the form of constraints on these cumulative marginal probabilities, i.e.,

$$F_{a.} - F_{.a} = \sum_{p=1}^a \sum_b \pi_{pb} - \sum_{p=1}^a \sum_b \pi_{bp} = 0. \quad (4)$$

Another marginal hypothesis, which may be relevant in situations in which A and B are interval level categorical variables, is the assumption of equal marginal means for A and B (for an example, see Haber and Brown, 1986). This is obtained by

$$\mu^A - \mu^B = \sum_a \sum_b \nu_a^A \pi_{ab} - \sum_a \sum_b \nu_b^B \pi_{ab} = 0, \quad (5)$$

where the ν 's denote category scores assigned to the levels of A and B .

Another interesting model for squared tables is the well-known symmetry model. Using linear equality restrictions, such a model can be specified as

$$\pi_{ab} - \pi_{ba} = 0. \quad (6)$$

It should be noted that the independence, conditional independence, and symmetry models can also be formulated as log-linear models. The other examples of linear restrictions cannot be specified as standard log-linear models.

Linear inequality restrictions

The linear inequality restrictions of interest are of the form

$$\sum_{ij} z_{2ijq} \pi_{j|i} - c_{2q} \geq 0. \quad (7)$$

Here, the z_{2ijq} are used to define the q th linear combination of probabilities $\pi_{j|i}$. This linear combination minus a constant (c_{2q}) is assumed to be at least 0. Restrictions of this form can be used to specify ordered variants of the equality restrictions discussed above, such as ordered conditional distributions, ordered conditional means, and ordered marginal distributions.

Rather than assuming independence between A and B , it is possible to postulate a positive relationship. This can be specified in the form of monotonically non-increasing cumulative conditional probabilities

$$F_{a|b} - F_{a|b+1} = \sum_{p=1}^a \pi_{p|b} - \sum_{p=1}^a \pi_{p|b+1} \geq 0. \quad (8)$$

It should be noted that this formulation of a positive relationship, which has been used by several authors (see, for instance, Grove, 1980; Croon, 1990; Evans, Gilula, Guttman, and Swartz, 1997), yields an asymmetrical ordinal hypothesis. The same type of assumption but with B as dependent yields a different model. In other words, stochastically ordered $F_{a|b}$'s do not imply stochastically ordered $F_{b|a}$'s.

Linear inequalities can also be used to specify hypotheses about the conditional relationship between A and B , given an individual's score on a third variable C . Suppose that A and B are positively related given $C = c$. This can be specified as follows:

$$F_{a|bc} - F_{a|b+1c} = \sum_{p=1}^a \pi_{p|bc} - \sum_{p=1}^a \pi_{p|b+1c} \geq 0,$$

in other words, in the form of monotonically non-increasing cumulative probabilities. If C is also ordinal, we may also wish to assume that

$$F_{a|bc} - F_{a|bc+1} = \sum_{p=1}^a \pi_{p|bc} - \sum_{p=1}^a \pi_{p|bc+1} \geq 0.$$

This model, in which the cumulative distribution of A is assumed to be stochastically ordered in two directions, was described by Robertson, Wright, and Dykstra (1988:32-33).

An ordinal variant of the marginal homogeneity model (see equation 4) is obtained by assuming that the cumulative marginal probabilities for A ($F_{a.}$) are at least as large as those of B ($F_{.a}$):

$$F_{a.} - F_{.a} = \sum_{p=1}^a \sum_b \pi_{pb} - \sum_{p=1}^a \sum_b \pi_{bp} \geq 0.$$

This model of stochastically ordered cumulative marginal distributions was described by Robertson, Wright, and Dykstra (1988:290-292). In a similar way, we could formulate order-restricted variants of the equal marginal means model described in equation 5 and the symmetry model described in equation 6 by replacing the ‘=’ sign by a ‘ \geq ’ sign.

Log-linear equality restrictions

The r th log-linear equality restriction on the probabilities $\pi_{j|i}$ is defined as

$$\sum_{ij} z_{3ijr} \ln \pi_{j|i} - c_{3r} = 0, \quad (9)$$

where the z_{3ijr} define the r th linear combination of logs of cell probabilities which minus a constant (c_{3r}) is postulated to be equal to zero. Restrictions of this form can be used to specify any kind of log-linear model, such as independence, row association, linear-by-linear association, conditional independence, and no-three-variable interaction models. In addition, the term c_{3r} makes it possible to impose fixed-value restrictions on the log-linear parameters. It should be noted that this is, actually, the orthogonal complement notation of the standard log-linear model. Such a reformulation is also used by Lang and Agresti (1994) and Bergsma (1997) for specifying extended log-linear models. This orthogonal complement formulation is very appealing in many situations because, as is demonstrated below, assumptions about relationships between variables are specified directly in terms of restrictions on (local) odds ratios.

Let θ_{ab} denote a local odds ratio in the two-way table formed by the variables A and B . It is defined as

$$\theta_{ab} = \frac{\pi_{ab}\pi_{a+1b+1}}{\pi_{ab+1}\pi_{a+1b}}.$$

In an independence model, it is assumed that each θ_{ab} equals 1, or, equivalently, that each $\ln \theta_{ab}$ equals zero. Using the above log-linear equality restrictions, such a model can be specified as

$$\ln \theta_{ab} = \ln \pi_{ab} - \ln \pi_{ab+1} - \ln \pi_{a+1b} + \ln \pi_{a+1b+1} = 0.$$

In a similar way, other types of non-saturated log-linear models can be defined for the same two-way table. A row association model, for example, assumes that the local odds ratios are independent of the columns. This can be specified as

$$\begin{aligned} \ln \theta_{ab} - \ln \theta_{ab+1} &= \ln \pi_{ab} - 2 \ln \pi_{ab+1} - \ln \pi_{a+1b} + 2 \ln \pi_{a+1b+1} \\ &\quad + \ln \pi_{ab+2} - \ln \pi_{a+1b+2} = 0. \end{aligned} \quad (10)$$

Note that this is a standard row association model with equal-interval scores for the levels of the column variable. However, it is also possible to use other scoring schemes for the column variable. The general row association model specified in the form of log-linear restrictions on θ_{ab} (and on π_{ab}) is

$$\frac{\ln \theta_{ab}}{\nu_{b+1}^B - \nu_b^B} - \frac{\ln \theta_{ab+1}}{\nu_{b+2}^B - \nu_{b+1}^B} = 0, \quad (11)$$

where ν_b^B denotes the score assigned to level b of B . As can be seen, the logs of local odds ratio are weighted by the inverse of the distance between the corresponding column scores. In a similar way, it is possible to specify column association models and linear-by-linear association with any type of category scoring.

To illustrate the use of the constant c_{3r} , it is also possible to test the assumption that the local odds ratios are equal to a specific value. By

$$\ln \theta_{ab} - c = \ln \pi_{ab} - \ln \pi_{ab+1} - \ln \pi_{a+1b} + \ln \pi_{a+1b+1} - c = 0,$$

we obtain an uniform association model in which the local odds ratios are fixed to be equal to $\exp(c)$.

As in the case of linear restrictions, it is also possible to constrain the relationships between more than two variables. For instance, restrictions of independence, row association, column association, linear-by-linear association, and fixed uniform association could be applied conditionally on C . Such restricted conditional association models can be specified by replacing π_{ab} by π_{abc} or $\pi_{ab|c}$ in the corresponding log-linear restrictions.

Another interesting assumption in a three-way table is the no-three-variable interaction model, which implies that the local odds ratios are independent of the third variable. This can be specified as follows:

$$\begin{aligned} \ln \theta_{ab|c} - \ln \theta_{ab|c+1} &= \ln \pi_{abc} - \ln \pi_{ab+1c} - \ln \pi_{a+1bc} + \ln \pi_{a+1b+1c} \\ &\quad - \ln \pi_{abc+1} + \ln \pi_{ab+1c+1} + \ln \pi_{a+1bc+1} + \ln \pi_{a+1b+1c+1} = 0. \end{aligned}$$

Here, $\theta_{ab|c}$ denotes a conditional local odds ratio for variables A and B within level c of variable C . The specification of log-linear models using these types of contrasts of log odds ratios can easily be generalized to higher-way tables.

Log-linear inequality restrictions

The fourth and last type of restriction presented here are log-linear inequality restrictions. The s th restriction of this form is

$$\sum_{ij} z_{4ijs} \ln \pi_{j|i} - c_{4s} \geq 0. \quad (12)$$

Log-linear inequality restrictions can be used to specify ordinal variants of the log-linear models discussed above. We may, for instance, define models with a positive bivariate relationship in the form of non-negative local log odds ratios, row or column association models with monotonically non-decreasing scores, or models assuming a bivariate association to be stronger for one group than for another group.

With the linear inequality restrictions, a postulated positive relationship between two ordinal variables was defined in the form of non-increasing cumulative conditional probabilities (see equation 8). A natural definition of a positive relationship between A and B in log-linear terms is that all local odds

ratios are at least 1 (Dykstra and Lemke, 1988). This yields the following set of log-linear inequality restrictions on the θ_{ab} 's or the π_{ab} 's:

$$\ln \theta_{ab} = \ln \pi_{ab} - \ln \pi_{ab+1} - \ln \pi_{a+1b} + \ln \pi_{a+1b+1} \geq 0. \quad (13)$$

It should be noted that, contrary to the definition in terms of cumulative conditional probabilities, this definition of a positive relationship is a symmetric one since reversing A and B yields the same model.

A positive association could also be specified by means of a row association model with monotonically non-decreasing row scores (see Agresti, Chuang, and Kezouh, 1987). Such a model, which assumes that the column variable is an interval level variable and that the row variable is ordinal, can be specified by combining the restriction of column independent local odds ratios (equation 10) with the restriction of non-negative local log odds ratios (equation 13). The more general order-restricted row association proposed by Agresti, Chuang, and Kezouh (1987) is obtained by using equation 11 instead of 10. In a similar way, we can specify ordered variants of the column and linear-by-linear association models.

Unfortunately, the log-linear inequality restrictions cannot be used to define row-column association models with ordered row and column scores as proposed by Ritov and Gilula (1991) since these models are not log-linear but log bi-linear (see also Vermunt, 1998). The log-linear inequality restrictions can, however, be used to specify correspondence or correlation models with ordered row and column scores. As was demonstrated by Ritov and Gilula (1993), this can be accomplished by specifying the row-column correlation model as a latent class model with log-linear inequality restrictions of the form described in equation 13.

As in the case of log-linear equality restrictions, the above examples of log-linear inequality restrictions can also be used in a multivariate setting. We may, for instance, assume a positive association, a row association with ordered scores, or a correlation model with ordered scores for A and B within levels of a third variable, say C . An example is the binary logit model with ordered-restricted parameters for one of the two regressors proposed by Agresti and Coull (1996).

Another interesting ordinal hypothesis for a three-way table is that a bivariate relationship is stronger in one subgroup than in another subgroup. Suppose that we assume a non-negative association between A and B within levels of C . In addition, we want that the association increases with C . The latter assumption can be specified by the following additional set of log-linear constraints on the conditional local odds ratio

$$\begin{aligned} \ln \theta_{ab|c+1} - \ln \theta_{ab|c} &= -\ln \pi_{abc+1} + \ln \pi_{ab+1c+1} + \ln \pi_{a+1bc+1} - \ln \pi_{a+1b+1c+1} \\ &\quad + \ln \pi_{abc} - \ln \pi_{ab+1c} - \ln \pi_{a+1bc} + \ln \pi_{a+1b+1c} \geq 0. \end{aligned}$$

Note that this set of order restrictions concerns the three-variable interaction term between A , B and C .

4 Maximum likelihood estimation

Maximum likelihood estimation of cell probabilities under ordinal restrictions is an optimization problem under inequality constraints. One of the methods for solving such a problem is the Lagrangian method with activated constraints (see Gill and Murray, 1974; Gill, Murray, and Wright, 1981). The Lagrangian method, which is well-known in maximum likelihood estimation with equality constraints, involves augmenting the object function to be maximized with one Lagrange term for each of the constraints. If a constraint has the form of an inequality constraint, the corresponding equality constraint is activated or deactivated during the optimization process, depending on whether the corresponding inequality constraint is violated or not. Appendix I describes some of the basic principles of optimization under equality and inequality constraints.

Assuming a (product-)multinomial sampling scheme, maximum likelihood estimation of the $\pi_{j|i}$ parameters under the restrictions described in equations 3, 7, 9, and 12 involves finding the saddle point of the following (Lagrange) function

$$\begin{aligned} \mathcal{L} = & \sum_{ij} n_{ij} \ln \pi_{j|i} + \sum_i \alpha_i \left(\sum_j \pi_{j|i} - 1 \right) \\ & + \sum_p \beta_{1p} \left(\sum_{ij} z_{1ijp} \pi_{j|i} - c_{1p} \right) + \sum_q \beta_{2q} \left(\sum_{ij} z_{2ijq} \pi_{j|i} - c_{2q} \right) \\ & + \sum_r \beta_{3r} \left(\sum_{ij} z_{3ijr} \ln \pi_{j|i} - c_{3r} \right) + \sum_s \beta_{4s} \left(\sum_{ij} z_{4ijs} \ln \pi_{j|i} - c_{4s} \right), \end{aligned} \quad (14)$$

with

$$\begin{aligned} \beta_{2q} & \geq 0 \\ \beta_{4s} & \geq 0, \end{aligned}$$

where the α and the β parameters are Lagrange multipliers. As can be seen, the first term at the right-hand side of equation 14 is the well-known kernel of the (product-)multinomial log-likelihood function. The second component specifies a set of Lagrange terms which guarantee that the probabilities $\pi_{j|i}$ sum to 1 within each level of the independent variable X . The other four terms belong to the linear equality, linear inequality, log-linear equality, and log-linear inequality restrictions, respectively.

Because the second and the fourth set of constraints are inequality constraints, the β_{2q} and β_{4s} parameters must be greater than or equal to zero, which implies that the corresponding equality constraints are only activated if the inequality constraints concerned are violated. More precisely, an active constraint corresponds with a β_{2q} or β_{4s} which is larger than 0, while an inactive constraint corresponds with a β_{2q} or β_{4s} which equals 0.

Taking the first derivative with respect to $\pi_{j|i}$ and setting the result equal

to zero yields the following expression for the ML estimate of $\pi_{j|i}$:

$$\pi_{j|i} = \frac{n_{ij} + \sum_r z_{3ijr}\beta_{3r} + \sum_s z_{4ijs}\beta_{4s}}{-\alpha_i - \sum_p z_{1ijp}\beta_{1p} - \sum_q z_{2ijq}\beta_{2q}}. \quad (15)$$

Thus, given the Lagrange multipliers, there is a closed form solution for the $\pi_{j|i}$'s. What is needed is a method for finding the Lagrange multipliers. This can, for instance, be accomplished by means of the uni-dimensional Newton method. This method involves updating one parameter at a time, fixing all the other parameters at their current values.² For α_i , an uni-dimensional Newton update is of the form

$$\alpha'_i = \alpha_i - \text{step} \frac{\sum_j \pi_{j|i} - 1}{\sum_j \pi_{j|i} / \left(-\alpha_i - \sum_p z_{1ijp}\beta_{1p} - \sum_q z_{2ijq}\beta_{2q} \right)}, \quad (16)$$

for β_{1p} ,

$$\beta'_{1p} = \beta_{1p} - \text{step} \frac{\sum_{ij} z_{1ijp}\pi_{j|i} - c_{1p}}{\sum_{ij} z_{1ijp}^2 \pi_{j|i} / \left(-\alpha_i - \sum_p z_{1ijp}\beta_{1p} - \sum_q z_{2ijq}\beta_{2q} \right)}, \quad (17)$$

and for β_{3r} ,

$$\beta'_{3r} = \beta_{3r} - \text{step} \frac{\sum_{ij} z_{3ijr} \ln \pi_{j|i} - c_{3r}}{\sum_{ij} z_{3ijr}^2 / (n_{ij} + \sum_r z_{3ijr}\beta_{3r} + \sum_s z_{4ijs}\beta_{4s})}. \quad (18)$$

In each of these updating equations, the numerator is the function that must become zero and the denominator its first derivative with respect to the parameter concerned.

The updating equations for β_{2q} and β_{4s} have the same form as for β_{1p} and β_{3r} , respectively. As already indicated above, the Lagrange parameters pertaining to the inequality restrictions must be greater than or equal to zero, which implies that β'_{2q} and β'_{4s} must be set equal to zero if they become negative. This amounts to not activating or deactivating the equality constraint corresponding to an inequality constraint.

With *step* it is possible to change the step size of the adjustments. This may be necessary if $\pi_{j|i}$ takes on an inadmissible value, or, more precisely, a value smaller than zero. In addition, *step* may be used to start with somewhat smaller step sizes in the first iterations.

The exact iteration scheme is as follows:

1. set $\alpha_i = -n_{i\cdot}$, $\beta_{1p} = \beta_{2q} = \beta_{3r} = \beta_{4s} = 0$, and $\text{step} = 1/4$, and compute $\pi_{j|i}$'s using equation 15
2. save current α 's, β 's, and $\pi_{j|i}$'s
3. for each Lagrange parameter,
 - (a) update parameter using equation 16, 17, or 18

²Vermunt (1997:312-315) applied uni-dimensional Newton for a similar problem, that is, for the estimation of (conditional) probabilities under simple equality and fixed-value restrictions.

- (b) only for β_{2q} and β_{4s} : if smaller than 0, set parameter equal to 0
 - (c) compute new $\pi_{j|i}$'s using equation 15
 - (d) if one or more $\pi_{j|i} < 0$: half *step*, restore saved α , β 's, and $\pi_{j|i}$'s from 2, and restart with 3(a).
4. if no convergence is reached: double *step* if $step < 1$ and restart with 2, that is, is go to next iteration.

As can be seen from step 1, the starting values for $\pi_{j|i}$ are $n_{ij}/n_{i.}$, that is, the unrestricted observed probability of $Y = j$ given $X = i$. Step 3(b) shows how the algorithm deals with inequality constraints: if an update of β_{2q} or β_{4s} yields a value smaller than zero, the parameter concerned is set to zero. In this way, an inactive constraint may remain inactive or an active constraint may become inactive, depending on whether its previous value was zero or positive. An inactive inequality constraint is activated if the value of the corresponding Lagrange multiplier changes from zero into a positive value. The convergence mentioned in step 4 can be defined either in terms of a maximum change of the Lagrange parameters or a minimum change of the log-likelihood function.

The above uni-dimensional Newton method will converge to the ML solution if the restrictions do not contradict with one another, if all observed cell entries are larger than zero, and if the model does not combine linear with log-linear restrictions.

The first condition states that the algorithm will not converge if contradictory restrictions are imposed, such as $a = 0$, $b = 3$, and $a \geq b$. This is, of course, not specific for the current algorithm. It should be noted that, contrary to multi-dimensional methods like Newton-Raphson and Fisher-scoring, the uni-dimensional method does not have problems with redundant restrictions, such as $a \geq b$, $b \geq c$, and $a \geq c$.

The problem associated with the second condition is well-known in the analysis of categorical data, and is therefore not specific for this algorithm. As in standard log-linear models, some parameters may be undefined because some observed cells are equal to zero. A simple way to overcome this problem is to add a small number to each cell entry. To solve the numerical problems associated with zero cells a very small number, say 10^{-10} , already suffices.³

A third problem is that the algorithm may fail to converge to the ML solution if a model combines linear and log-linear restrictions. This problem was noted by Bergsma (1998) in the context of the algorithm proposed by Haber and Brown (1986) for log-linear models with linear (equality) restrictions on the expected cell entries. Haber and Brown proposed an algorithm in which at each iteration cycle first the log-linear parameters and then the parameters associated with the linear restrictions are updated. Bergsma showed that their proof of convergence contains an incorrect assumption, namely, that the term belonging to the linear part of the model, the denominator of equation 15, is

³Adding somewhat larger numbers to the observed cell entries, can very well be defended from a Bayesian point of view (Clogg and Eliason, 1987). With an informative (Dirichlet) prior, the estimated cell entries can, for instance, be smoothed to the independence model. For an excellent overview of this topic, see Schafer(1997).

positive for each cell entry. In the ML solution, both the numerator and denominator may be negative for some cells. A problem is, however, that an algorithm that does not simultaneously update the terms belonging to the linear and to the log-linear restrictions may not converge because of the requisite that the probabilities should remain positive after each update. The results by Haber and Brown hold asymptotically, which means that if the model holds and the sample size is large enough this problem will not occur. Thus, in practice, this problem is more likely to occur if the model of interest fits badly or if the sample size is small.

Bergsma (1998) proposed estimating models that combine linear and log-linear equality restrictions with a Fisher-scoring algorithm developed for the estimation of extended log-linear models (Lang and Agresti, 1994; Bergsma, 1997). This multi-dimensional saddle point method for finding ML estimates under a general class of equality constraints can easily be modified into an activate set procedure to allow for inequalities (see Appendix I). Another advantage of applying this more advanced method is that an even more general class of inequality constraints can be formulated, such as log-linear inequality restrictions on marginal probabilities. This may, for instance, yield a non-parametric variant of the cumulative logit model. Nevertheless, the procedure described above remains very attractive because of its simplicity. It can easily be implemented using macro languages of packages as SAS, GLIM, and S-plus. For the examples presented in the next section, we used both the simple uni-dimensional Newton algorithm and an adaptation of Bergsma's (1997) algorithm to inequalities. In all estimated models, both procedures yielded the same results.

Robertson, Wright, and Dykstra (1988: Chapter 1) described an alternative procedure for obtaining order-restricted maximum likelihood estimates. They showed that some order-restricted maximum likelihood problems can be transformed into isotonic regression problems. One of the algorithms they proposed for solving these isotonic regression problems is the pooling adjacent violators algorithm (PAVA), which is a simple IPF-like algorithm that can be used to solve models with simple order restrictions. Another method for finding ML estimates under equality and inequality constraints is to transform the constrained ML estimation problem into a quadratic programming problem (see, for instance, Fahrmeir and Klinger, 1994, and Schoenberg, 1997).

Latent variables and other types of missing data

The proposed non-parametric ordinal modeling approach can also be applied in situations in which there is some type of missing data, such as in latent class models and in models for panel data subject to partial nonresponse. However, to be able to deal with missing data, we have to adapt the estimation algorithm described in section 4.

The simplest option is to use the EM algorithm (Dempster, Laird, and Rubin, 1977). The main advantage of using this iterative method is that it is obtained with minor modifications of the estimation procedure for complete data. In the E step of the EM algorithm, we estimate the complete data on the basis of the incomplete data and the current parameter estimates. The M

step of the algorithm involves estimating the model parameters as if all data were observed. Croon (1990), for instance, implemented the pooling adjacent violators algorithm (PAVA) in the M step of the version of the EM algorithm that he used for estimating his ordinal latent class model. Here, we will use an EM algorithm which implements the simple uni-dimensional Newton method in the M step. Appendix II discusses the EM algorithm for a marginal model with partially missing data and for an order-restricted latent class model.

5 Model testing

Suppose that H_1 denotes the hypothesized order-restricted model, H_0 is a more restrictive alternative obtained by transforming the inequality restrictions into equality restrictions, and H_2 is less restrictive alternative which is obtained by omitting the inequality restrictions. This could, for instance, be non-negative local odds-ratios (H_1), independence (H_0), and the saturated model (H_2). The two tests of interest are between H_0 and H_1 and between H_1 and H_2 . Such tests can be performed using standard likelihood-ratio statistics. The corresponding statistics, $L_{1|0}^2$ and $L_{2|1}^2$, are defined as

$$\begin{aligned} L_{1|0}^2 &= 2 \sum_{ij} n_{ij} \ln \left(\frac{\hat{\pi}_{j|i(1)}}{\hat{\pi}_{j|i(0)}} \right) \\ L_{2|1}^2 &= 2 \sum_{ij} n_{ij} \ln \left(\frac{\hat{\pi}_{j|i(2)}}{\hat{\pi}_{j|i(1)}} \right), \end{aligned}$$

where $\hat{\pi}_{j|i(0)}$, $\hat{\pi}_{j|i(1)}$, and $\hat{\pi}_{j|i(2)}$ denote the estimated probabilities under H_0 , H_1 , and H_2 , respectively.

A complication in using these test statistics is, however, that they are not asymptotically χ^2 distributed. Wollan (1985) has shown that the above two test statistics follow chi-bar-squared distributions, which are weighted sums of chi-squared distributions, when H_0 holds (see Robertson, Wright, and Dykstra, 1988:321).

Let l_{max} denote the number of inequality constraints, or the maximum number activated constraints, and df_0 the number of degrees of freedom under H_0 . The p -values for $L_{1|0}^2$ and $L_{2|1}^2$ are approximated as follows

$$\begin{aligned} P(L_{1|0}^2 \geq c) &\approx \sum_{l=0}^{l_{max}} P(l) P(\chi_{(df_0-l)}^2 \geq c) \\ P(L_{2|1}^2 \geq c) &\approx \sum_{l=0}^{l_{max}} P(l) P(\chi_{(l)}^2 \geq c), \end{aligned}$$

that is, as the sum over the all possible numbers of active constraints of the probability of the corresponding number of constraints times the asymptotic p -value concerned.

A problem is, however, the computation of the $P(l)$'s, which depend on the maximum number of constraints, a vector of weights \mathbf{w} , in our case the

observed frequencies, and the type of order restrictions that is used. For simple order restrictions, the $P(l)$'s can be computed analytically up to $l_{max} = 5$. Robertson, Wright, and Dykstra (1988) reported $P(l)$ tables for $1 \leq l_{max} \leq 19$, assuming uniform weights \mathbf{w} and simple order restrictions. Simulation studies by Grove (1980) and Robertson, Wright, and Dykstra (1988) showed that the uniform weights assumption does not seriously distort the results when testing whether multinomials are stochastically ordered.

Rather than combining asymptotic results with an approximation of the $P(l)$'s, it is also possible to determine the p -values for the test statistics using parametric bootstrapping methods, which are also known as Monte Carlo studies. This very simple method, which is based on an empirical reconstruction of the sampling distributions of the test statistics, is the one followed here. Ritov and Gilula (1993) proposed such a procedure in ML correspondence analysis with ordered category scores. Schoenberg (1997) advocated using bootstrap testing methods in a general class of constrained maximum likelihood problems. Langeheine, Pannekoek, and Van de Pol (1996) proposed using bootstrapping in categorical data analysis for dealing with sparse tables, which is another situation in which we cannot rely on asymptotic theory for the test statistics. Agresti and Coull (1996) used Monte Carlo studies in combination with exact tests to determine the goodness-of-fit of order-restricted binary logit models which were estimated with a small sample.

In the $L_{1|0}^2$ case, T frequency tables with the same number of observations as the original observed table are simulated from the estimated probabilities under H_0 . For each of these tables, we estimate the models defined by H_0 and H_1 and compute the value of $L_{1|0}^2$. This yields an empirical approximation of the distribution of $L_{1|0}^2$. The estimated p -value is the proportion of simulated tables with an $L_{1|0}^2$ at least as large as for the original table. The standard error of the estimated p -value equals $\sqrt{p(1-p)/T}$. The bootstrap procedure for $L_{2|1}^2$ differs only from the above one in that frequency tables have to be simulated from the estimated probabilities of the order-restricted maximum likelihood solution, i.e., H_1 .⁴

A simulation study by Ritov and Gilula (1993) showed that parametric bootstrapping yields reliable results when applied in order-restricted correlation models, which are special cases of the models presented in this paper. To further assess the performance of bootstrapping, the examples for which Grove (1980) and Robertson, Wright, and Dykstra (1988:234-239) reported multinomial likelihood-ratio tests based on asymptotic chi-bar-squared distribution were replicated. For these examples, the bootstrapped p -values were very close to the reported asymptotic p -values. It should be noted that although bootstrapping seems to work well in these situations, it is not clear at all how the method performs when applied to sparse tables.

⁴As was noted by one of the reviewers, in the $L_{2|1}^2$ case, the bootstrap is not estimating the p -value corresponding to the chi-bar-squared distribution. The chi-bar-squared approximation of $P(L_{2|1}^2 \geq c)$ requires that H_0 holds, which means that it yields what could be called the least favourable p -value. On the other hand, the empirical bootstrap approximation of the distribution of $L_{2|1}^2$ holds under H_1 , which is more in agreement with standard tests.

6 Examples

This section discusses four situations in which the non-parametric ordinal models presented in this paper may be useful. The first example is a continuation of the bivariate example presented in section 2. The second example illustrates the use of inequality restrictions in logit regression models for ordinal dependent and independent variables. The third example is on marginal models for longitudinal data subject to partial nonresponse. The last example deals with latent class models for ordinal items.

Association between two ordinal variables

In section 2, some parametric and non-parametric models were presented for the 5-by-3 cross-classification of number of siblings (S) and happiness (H) reported in table 1. More precisely, we specified independence (A.1), cumulative logit (A.2), row-column association (A.3), row association (A.4), column association (A.5), and uniform association (A.6) models, as well as a model assuming non-increasing cumulative probabilities (A.7) and a model assuming local odds ratios of at least 1 (A.8).

The models presented so far for this two-way table are either parametric or non-parametric. There is, however, another interesting class of models for this type of data, that is, models that combine parametric with non-parametric features, such as order-restricted variants of the row-column, row, and column association models. According to the assumed measurement level of the row and the column variables, these three models could be labeled as ordinal-ordinal, ordinal-interval, and interval-ordinal, respectively.

The order-restricted RC model⁵ fits the data quite well: $L_{A.9}^2 = 8.36, p \approx .08$. While in the unrestricted RC model the scores for rows 3 and 4 and for columns 1 and 2 were out of order, in the order-restricted ML solution, only one equality restriction is imposed: the score for column 1 is equated to the score for column 2. This demonstrates again that it is dangerous to specify ordinal models by post hoc equality constraints.

Since the unrestricted row association model (A.4) fits badly, it is not surprising that the order-restricted R model fits very badly too ($L_{A.10}^2 = 18.60, p \approx .00$). In the ML solution for this model, the estimated scores for rows 5 and 6 are equated. On the other hand, the ordinal C model fits very well ($L_{A.11}^2 = 8.84, p \approx .22$). The ML solution for this model contains one activated constraint: the parameters belonging to the first two columns are equated.

On the basis of these results, it can be concluded that the relationship between number of siblings and happiness can be described by means of a (partially) non-parametric ordinal model. The two non-parametric models, as well as the order-restricted RC and C models, fit the data quite well. The most parsimonious model that fits the data is the order-restricted C model. This

⁵It should be noted that the order-restricted RC model cannot be specified with the linear or log-linear constraints presented in this paper. It was estimated with a modified version of the PAVA-like procedure proposed by Ritov and Gilula (1991) which is described in detail in an accompanying paper (Vermunt, 1998).

indicates that the row variable, number of siblings (S), may be treated as an interval level variable with equal-interval scored categories, while the column variable, happiness (H), should be treated as ordinal.

Logit regression models for ordinal variables

This example uses the data reported in table 2 of Clogg's 1982 JASA paper on ordinal log-linear models. The table concerns a 4-by-5-by-3 cross-classification of the ordinal variables years of schooling (Y), number of siblings (S), and happiness (H).⁶ We treat happiness (H) as a dependent variable and years of schooling (Y) and number of siblings (S) as independent variables. In fact, we are interested in modeling the probability that $H = h$ given that $Y = y$ and $S = s$, denoted by $\pi_{h|ys}$. With this example we want to illustrate the use of order restrictions in the context of a logit model with ordinal dependent and independent variables.

The test results for the estimated logit models are reported in table 4. A standard multinomial logit analysis treating each of the three variables as nominal shows that the three-variable interaction is not significant ($L_{B.1}^2 = 24.88, df = 24, p = .41$). In addition, the test results for Models B.2 and B.3 indicate that both independent variables have a significant effect on happiness.

The usual way of dealing with the fact that Y , S , and H are ordinal variables within the framework of logit analysis is the assignment of a priori scores to the levels of Y , S and H . This yields linear-by-linear (partial) associations, or, in the case equal-interval scoring, uniform associations for the YH and SH interactions. Note that such an approach actually assumes that we are dealing with interval level variables. The model that further restricts Model B.1 by assuming uniform two-way interactions does not fit at all: $L_{B.4}^2 = 54.58, df = 36, p = .02$.

An alternative way of specifying an ordinal logit model is by means of inequality restrictions on the conditional local odd ratios, that is, $\theta_{hy|s} \leq 1$ and $\theta_{hs|y} \geq 1$. This is a way of formulating that Y has a negative effect on H within each level of S and that S has a positive effect on H within each level of Y . If we also want to exclude the three-variable interaction term, we need the additional constraint $\theta_{hy|s} = \theta_{hy|s+1}$. The model, which combines these log-linear equality and inequality constraints, fits well ($L_{B.5}^2 = 35.30, p \approx .35$). The conclusion could be that the partial effects of Y and S on H are ordinal and equal across levels of the other explanatory variable.

Marginal models for partially missing longitudinal data

This example illustrates the use of marginal models with linear (in)equality constraints in the context of longitudinal data. In addition, it demonstrates the possibility to deal with partially missing data. The data, which are taken from the 1986-1987 SIPP panel, concern measurements of a person's employment status at four time points, where each time point is separated by three

⁶The original table of Clogg (1982) is a 3-by-4-by-5 table. For convenience, here, another order between the variables is used.

months.⁷ Employment status is classified into two categories: employed and not employed. A complication in the analysis of this data is that for many subjects in the sample there is missing information. More precisely, for 28 percent of the 6754 cases, information on the employment status is missing for one or more time points. In addition, except for observing only the first and the last time point, all possible missing data pattern are present in the sample. The non-zero observed frequencies are reported in table 5.

[INSERT TABLE 5 ABOUT HERE]

We are interested in studying the trend in the employment rate over the four periods. Suppose that because of macro-economic conditions one expects a monotonically increasing employment rate during the observation period. As will be shown below, the data are not fully agreement with such a trend, which may, however, be the result of sampling error.

Since this paper does not deal with missing data mechanisms, we will just assume an ignorable, missing at random (MAR), missing data mechanism (Little and Rubin, 1987). For the partially observed SIPP data, four marginal models were estimated: a saturated, a marginal homogeneity, a non-decreasing marginal, and a linearly changing marginal model. The test results are presented in table 4.

According to the saturated model, which, of course, fits perfectly, the estimated marginal probabilities of being employed at each of the four time points are .587, .607, .599, and .605, respectively. This indicates that there is a small increase in the number of employed individuals during the observation period. The increase is, however, not monotone. The marginal homogeneity model tests whether the observed differences between the time points are significant. The bad fit of this model ($L_{C.1}^2 = 22.36, df = 3, p < .01$) shows that this is the case. The third model assumes that the marginal probability of being employed is non-decreasing between consecutive time points. This model, which has one activated constraints, fits quite well ($L_{C.2}^2 = 3.17, p \approx .07$). As could be expected on the basis of the marginal distribution from the saturated model, the inequality constraint concerning the second and third time point is activated, which means that in the ML solution the marginal distributions of these time points are equated. And finally, a model was estimated with a linear change in the number of employed. As can be seen, this model is too restrictive: $L_{C.3}^2 = 13.73, df = 2, p < .01$.

Latent class models for ordinal items

The last example illustrates the use of the non-parametric approach in the context of latent class models for ordinal items. For this purpose, we use a 4-by-4-by-4 cross-tabulation of three extrinsic job satisfaction items used by Shockey (1988) in a paper on latent class analysis (see also Hagenaars, 1998). The three ordinal items measure an individual's satisfaction with job security (S), pay (P), and fringe benefits (B). The levels of the items are: (1) not at

⁷For more information about this data set, see Vermunt (1997:216 and 286-287). Here, we use only the information on the first four of the six panel waves.

all true, (2) a little true, (3), somewhat true, and (4) very true. The latent variable will be denoted by W .

For the three-way classification, different types of latent class models are specified, each having the general form

$$\pi_{wspb} = \pi_w \pi_{s|w} \pi_{p|w} \pi_{b|w}. \quad (19)$$

The models, which are all four class models, differ with respect to the restrictions that are imposed on the conditional response probabilities $\pi_{s|w}$, $\pi_{p|w}$, and $\pi_{b|w}$. The test results are reported in table 4.

As reported by Shockey (1988), the unrestricted latent class model with four latent classes fits the job satisfaction data very well: $L_{D.1}^2 = 15.11, df = 24, p = .92$. When using such a standard latent class model, there is, however, no guarantee that the latent classes are ordered. By ordered we mean that the higher the latent class the more satisfied one becomes with each of the job items. In this context, it also means that the latent variable is uni-dimensional. The linear and log-linear equality and inequality constraints proposed in this paper can be used to impose such an ordinal structure on the relationships between the latent variable and the indicators. More precisely, they can be used to further restrict the conditional response probabilities $\pi_{s|w}$, $\pi_{p|w}$, and $\pi_{b|w}$.

The most restricted model that is used is a four-class model in which the WS , WP , and WB interactions are assumed to be uniform. This model does not provide a good description of the data: $L_{D.2}^2 = 115.79, df = 48, p < .01$. A less restrictive model is obtained by using column associations for the WS , WP , and WB interactions, with the items as column variables. This means that the latent variable is treated as interval level and the items as nominal. Although the category scores for each of the indicators have the expected order, the model fits badly: $L_{D.3}^2 = 105.86, df = 42, p < .01$.

It is also possible to use the non-parametric ordinal specifications in the context of latent class analysis. One interesting type of assumption is that each of the cumulative response probabilities, $F_{s|w}$, $F_{p|w}$, and $F_{b|w}$, is stochastically ordered, which means that they have to be restricted as described in equation 8. This yields the ordinal latent class model proposed by Croon (1990). Another option is to use log-linear inequality constraints on the local odds ratios θ_{ws} , θ_{wp} , and θ_{wb} (see equation 13). The former specifications yields a well fitting model ($L_{D.4}^2 = 15.55, p \approx .96$). Actually, the unrestricted four-class model was already very close to this solution. This can be seen from the fact that the L^2 values of the nominal and the ordinal model are almost identical and that, in addition, only 2 of the 36 inequality constraints need to be activated in the ordinal model. Although the four-class model with non-negative local log odds ratios does not perform as well as the other ordinal model, it also fits the data quite well: $L_{D.5}^2 = 39.20, p \approx .34$. The ML solution for this ordinal latent class model contains 16 activated constraints, which means that 16 estimated local odds ratios are equal to one.

[INSERT TABLE 6 ABOUT HERE]

Table 6 reports the estimated latent class probabilities π_w , as well as the estimated cumulative conditional probabilities according to Model D.4. As can be seen, the restriction imposed is that the probability that an individual selects a particular item category or lower decreases or remains equal as w increases. This is one way of expressing a positive relationship between the latent variable and the items. Only two inequality constraints are activated in the reported ML solution.

This example showed that non-parametric ordinal restrictions may yield well fitting and easy to interpret latent class models. The log-linear latent class models with uniform and column association structures were much too restrictive, while the results obtained by unrestricted latent class analyses may be difficult to interpret.

7 Discussion

This paper described a general non-parametric approach for dealing with ordinal categorical data, which is based on specifying linear or log-linear inequality constraints on (conditional) probabilities. Several types of ordinal models can be defined with the proposed inequality constraints. In addition, inequality constraints can be combined with equality constraints, which makes it possible to define models which combine non-parametric with parametric features, such as order-restricted row association models, ordinal row-column correlation models, and ordinal regression models in which higher-order interaction terms are omitted.

A simple estimation method was proposed which performs very well in most situations. Implementation of this uni-dimensional Newton method in the M step of the EM algorithm makes it possible to use the ordinal restrictions when there is partially missing data or when the model contains one or more latent variables. The difficulties associated with goodness-of-fit testing in models with inequality constraints were overcome by using bootstrap or Monte Carlo methods rather than relying on asymptotic distribution functions. The proposed estimation algorithm and testing procedure perform well in the analysis of tables which are not too sparse.

The examples showed that, in most situations, truly ordinal models fit much better than models in which a priori scores are assigned to the categories of the ordinal variables. In addition, these ordinal models do not have the interpretation problems associated with quasi-ordinal models, in which estimated category scores may be out of order.

A possible extension of the approach proposed here is the application of inequality constraints in extended log-linear models (Lang and Agresti, 1994; Bergsma, 1997). This would yield new types of ordinal models, such as cumulative logit models for ordinal independent variables, ordinal models for global odds ratios, and ordinal models for a general class of association measures. For this purpose, the saddle point algorithm proposed by Bergsma (1997), which is a generalization of the algorithm proposed by Lang and Agresti (1994), should be transformed into an active set method.

Another interesting direction for future research is the use of Bayesian approaches for estimating parameters and assessing fit of non-parametric ordinal models for categorical data. Some work has already been done on this subject by Agresti and Chuang (1986), Evans, Gilula, Guttman, and Swartz (1997), Hoijtink and Molenaar (1997), McDonald and Prevost (1997).

Appendix I: Optimization under (in)equality constraints

Suppose we have to find the value of a set of parameters γ that maximizes a function $f(\gamma)$ under the following r equality constraints:

$$h_1(\gamma) = 0, h_2(\gamma) = 0, \dots, h_r(\gamma) = 0. \quad (20)$$

This is a standard constraint optimization problem that can be solved by finding the saddle point of the Lagrange function

$$k(\gamma, \lambda) = f(\gamma) + \sum_{i=1}^r \lambda_i h_i(\gamma), \quad (21)$$

where the λ_i 's are called Lagrange parameters. This objective function contains, besides the γ parameters of interest, a set of parameters corresponding to the constraints. It should be noted that the saddle point of $k(\gamma, \lambda)$ is the maximum of $f(\gamma)$ under the above equality constraints.

The saddle point of the Lagrange function is the point in the parameter space at which the first derivatives to all parameters are equal to zero, in this case,

$$\frac{\partial k(\gamma, \lambda)}{\partial \gamma_j} = \frac{\partial f(\gamma)}{\partial \gamma_j} + \sum_{i=1}^r \lambda_i \frac{\partial h_i(\gamma)}{\partial \gamma_j} = 0 \quad (22)$$

$$\frac{\partial k(\gamma, \lambda)}{\partial \lambda_i} = h_i(\gamma) = 0. \quad (23)$$

As can be seen, the second set of conditions correspond to the constraints that we want to impose. The first set of conditions are the modification of the standard condition $\frac{\partial f(\gamma)}{\partial \gamma_j} = 0$ resulting from the imposed constraints. The solution to these equations can be found using standard algorithms, such as Fisher scoring, Newton-Raphson, or uni-dimensional Newton.

When (some of) the constraints have the form of inequalities, $h_i(\gamma) \geq 0$, the situation is slightly different. In that case, we have to formulate the additional condition that $\lambda_i \geq 0$. Actually, this condition guarantees that the constraint $h_i(\gamma) = 0$ is imposed only if the unrestricted $h_i(\gamma)$ is smaller than 0. In other words, the inequality restriction concerned is activated, which means that the corresponding equality restriction is imposed only if it is violated.

In optimization under inequality constraints, one may also refer to the Kuhn-Tucker conditions. These state that an optimum of $f(\gamma)$ under the inequality constraints $h_i(\gamma) \geq 0$ satisfies the following four conditions:

1. $h_i(\gamma) \geq 0$
2. $\frac{\partial f(\gamma)}{\partial \gamma_j} + \sum_{i=0}^r \lambda_i \frac{\partial h_i(\gamma)}{\partial \gamma_j} = 0$
3. $\lambda_i \geq 0$
4. $\lambda_i h_i(\gamma) = 0$.

The first condition states that inequality restrictions should be fulfilled. The second corresponds to setting the first derivative of the Lagrange function to zero for all γ_j 's. The third is the above-mentioned condition with respect to the sign of the Lagrange parameters. The fourth condition is automatically fulfilled because, depending on whether a constraint is inactive or active, either λ_i or $h_i(\gamma)$ will be equal to zero.

As in the case of equalities, standard algorithms can be used for finding the optimum of $f(\gamma)$ under the specified inequality constraints. The only necessary modification is that at each iteration cycle it must be checked which inequalities should be activated and which should be deactivated. This is exactly what was done by so-called active set methods. A possible implementation is the following. Start with all λ_i 's equal to zero. Each iteration cycle consists of two steps: 1] determine the active set of constraints, and 2] update the γ_j 's, as well as the λ_i 's belonging to the active set of constraints. Step 1 involves deactivating the constraints that are no longer necessary, which correspond with λ_i 's smaller than zero, and activating constraints that are violated, which correspond with gradients indicating that the λ_i 's will become larger than zero. Note that, actually, we are checking the first and third Kuhn-Tucker conditions.⁸

Appendix II: The EM algorithm for models with (in)equality constraints

The linear and log-linear (in)equality constraints described in this paper can also be applied when there are missing data or latent variables. This can be accomplished by implementing the active set variant of the uni-dimensional Newton method in an EM algorithm.

In the E step of the EM algorithm, we have to calculate the expectation of the complete data, given the observed data and the current parameter estimates. The M step involves estimating the 'parameters' of interest, treating the expectation of the observed data as if it were the observed data. This means that a single M step has the same form as ML estimation with fully observed data. The EM algorithm cycles between the E step and the M step until convergence.

Suppose we are interested in the estimation of a model with stochastically ordered marginal distributions for three-wave panel data. The variable of interest at the three points in time is denoted by A , B , and C . What we are interested in is obtaining estimates for the probabilities π_{abc} under the linear inequality constraint $F_{a..} \geq F_{.a.} \geq F_{..a}$. Suppose that respondents may have missing values on B , on C , or on both B and C . In other words, there is a subgroup for which we observe A , B , and C , a subgroup for which we observe A and B , a subgroup for which we observe A and C , and a subgroup for which we observe A . The cell entries in the frequency tables for these four subgroups are denoted by n_{abc} , n_{ab} , n_{ac} , and n_a , respectively.

⁸McDonald and Diamond (1983) gave an overview of methods that can be used to determine the active set in the estimation on generalized linear models with linear inequality constraints.

The E step of the t th iteration cycle involves computing the expected value of the complete data, \hat{n}_{abc} , in the following way:

$$\hat{n}_{abc}^{(t)} = n_{abc} + n_{ab}\hat{\pi}_{c|ab}^{(t-1)} + n_{ac}\hat{\pi}_{b|ac}^{(t-1)} + n_a\hat{\pi}_{bc|a}^{(t-1)}. \quad (24)$$

Note that the $\hat{\pi}$'s are computed from the estimated probabilities from the previous iteration ($t - 1$). In the M step, new $\hat{\pi}_{abc}^{(t)}$ are obtained with the active set method described in section 4 using $\hat{n}_{abc}^{(t)}$ as observed frequencies.

Another example of the implementation of the EM algorithm concerns an order-restricted latent class model. Suppose we have a latent class model with a single latent variable X and three indicators A , B , and C . The model has the form

$$\pi_{xabc} = \pi_x\pi_{a|x}\pi_{b|x}\pi_{c|x}, \quad (25)$$

in which the probabilities $\pi_{a|x}$, $\pi_{b|x}$, and $\pi_{c|x}$ are assumed to fulfil some kind of order restriction, for instance, that all local odds are at least 1.

The E step of t th EM cycle involves obtaining the expectation of the complete data, $\hat{n}_{xabc}^{(t)}$, by

$$\hat{n}_{xabc}^{(t)} = n_{abc}\hat{\pi}_{x|abc}^{(t-1)}. \quad (26)$$

In the M step, new order-restricted estimates $\hat{\pi}_{a|x}^{(t)}$, $\hat{\pi}_{b|x}^{(t)}$, and $\hat{\pi}_{c|x}^{(t)}$ can be obtained by using $\hat{n}_{x.a.}^{(t)}$, $\hat{n}_{x.b.}^{(t)}$, and $\hat{n}_{x..c}^{(t)}$ as data in the standard restricted ML procedure described in section 4.

References

- Agresti, A., and Chuang, C. (1986). Bayesian and maximum likelihood approaches to order restricted inference for models from ordinal categorical data. R.Dykstra and T.Robertson (eds.), *Advances in ordered statistical inference*, 6-27. Berlin: Springer-Verlag.
- Agresti, A., Chuang, C., and Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association*, 82, 619-623.
- Agresti, A., Coull, B.A. (1996). Order-restricted tests for stratified comparisons of binomial proportions. *Biometrics*, 52, 1103-1111.
- Bergsma, W.P. (1997). *Marginal models for categorical data*. Tilburg University: Tilburg University Press.
- Bergsma, W.P. (1998). *A note on maximum likelihood methods for log-linear models when expected cell frequencies are subject to linear constraints*. Tilburg University: Research note.
- Clogg, C.C. (1982). Some models for the analysis of association in multi-way cross-classifications having ordered categories. *Journal of the American Statistical Association*, 77, 803-815.
- Clogg, C.C., and Eliason, S.R. (1987). Some common problems in log-linear analysis. *Sociological Methods and Research*, 16, 8-14.
- Clogg, C.C., and Shihadeh, E.S. (1994). *Statistical models for ordinal data*. Thousand Oakes, CA: Sage Publications.

- Croon, M.A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology*, 43, 171-192.
- Croon, M.A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology*, 44, 315-331.
- De Leeuw, J., and Van der Heijden, P.G.M. (1991). Reduced rank models for contingency tables. *Biometrika*, 8, 229-232.
- 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.
- Dykstra, R.L., and Lemke, J.H. (1988). Duality of I projections and maximum likelihood estimation for log-linear models under cone constraints. *Journal of the American Statistical Association*, 83, 546-554.
- Evans, M., Gilula, Z., Guttman, I., and Swartz, T. (1997). Bayesian analysis of stochastically ordered distributions of categorical variables. *Journal of the American Statistical Association*, 92, 208-214.
- Fahrmeir, L., and Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers*, 35, 211-229.
- Gill, P.E., and Murray, W. (1974). *Numerical methods for constrained optimization*. London: Academic Press Inc..
- Gill, P.E., Murray, W., and Wright, M.H. (1981). *Practical optimization*. London: Academic Press.
- Gilula, Z., and Haberman, S.J. (1988). The analysis of contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83, 760-771.
- 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.
- Goodman, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear approach in the analysis of contingency tables. *International Statistical Review*, 54, 243-309.
- Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics*, 25, 489-504.
- Grove, D.M. (1980). A test of independence against a class of ordered alternatives in a 2 X C contingency table. *Journal of the American Statistical Association*, 75, 454-459.
- Haber, M., and Brown, M.B. (1986). Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. *Journal of the American Statistical Association*, 81, 477-482.
- Hagenaars, J.A. (1998). Categorical causal modeling: directed log-linear models with latent variables. *Sociological Methods and Research*.
- Haberman, S.J. (1979). *Analysis of qualitative data, Vol. 2, New developments*. New York, San Francisco, London: Academic Press.
- Hojtink, H., and Molenaar, I.W. (1997). A multidimensional item response model: constrained latent class analysis using the Gibbs sampler and posterior predictives. *Psychometrika*, 62, 171-190.
- Lang, J.B., and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, 89, 625-632.

- Langeheine, R., Pannekoek, J., and Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, 24, 492-516.
- Little, R.J., and Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- McDonald, J.W., and Diamond, I. (1983). Fitting generalised linear models with linear inequality parameter constraints. *Glim Newsletter*, 29-36.
- McDonald, J.W., and Prevost, A.T. (1997). The fitting of parameter-constrained demographic models. *Mathematical Computation and Modelling*, 26, 79-88.
- Ritov, Y., and Gilula, Z. (1991). The order-restricted RC model for ordered contingency tables: estimation and testing for fit. *The Annals of Statistics*, 19, 2090-2101.
- Ritov, Y., and Gilula, Z. (1993). Analysis of contingency tables by correspondence models subject to order constraints. *Journal of the American Statistical Association*, 88, 1380-1387.
- Robertson, T., Wright, F.T., and Dykstra, R.L (1988). *Order restricted statistical inference*. Chichester: Wiley.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.
- Schoenberg, R. (1997). CML: constrained maximum likelihood estimation. *The Sociological Methodologist*, 1-8.
- Shockey, J. (1988). Latent class analysis: an introduction to discrete data models with unobserved variables. *J.S. Long (red.), Common problems in quantitative research*. Newbury Park, CA: Sage.
- Vermunt, J.K. (1997). *Log-linear models for event histories*. Thousand Oakes: Sage Publications.
- Vermunt, J.K. (1998). *RC association models with ordered row and column scores: estimation and testing*. Paper presented at the SMABS'98 conference, Leuven, July 13-15 1998.
- Wollan, P.C. (1985). *Estimation and hypothesis testing under inequality constraints*. University of Iowa: Ph.D. thesis.

Table 1: Cross-classification of number of siblings and happiness: observed frequencies

number of siblings	Happiness		
	not too happy	pretty happy	very happy
0-1	99	155	19
2-3	153	238	43
4-5	115	163	40
6-7	63	133	32
8+	99	118	47

Table 2: Cross-classification of number of siblings and happiness: observed and estimated cumulative probabilities

number of siblings	Happiness		
	not too happy	pretty happy	very happy
0-1	0.363/0.363	0.930/0.930	1.000/1.000
2-3	0.353/0.356	0.901/0.912	1.000/1.000
4-5	0.362/0.356	0.874/0.873	1.000/1.000
6-7	0.276/0.329	0.860/0.870	1.000/1.000
8+	0.375/0.329	0.822/0.809	1.000/1.000

Table 3: Cross-classification of number of siblings and happiness: observed and estimated odds ratios

number of siblings	Happiness	
	not too happy/pretty happy	pretty happy/very happy
0-1/2-3	0.994/1.000	1.474/1.471
2-3/4-5	0.911/1.000	1.358/1.308
4-5/6-7	1.489/1.023	0.980/1.125
6-7/8+	0.565/1.000	1.655/1.327

Table 4: Test results for the four examples

Model	L^2 value	df^1	p value ²
A. Ordinal association models			
1. independence	26.27	8	.00
2. cumulative logit	18.52	4	.00
3. row-column association	7.33	3	.06
4. row association	17.52	4	.00
5. column association	8.36	6	.21
6. uniform association	20.21	7	.01
7. non-increasing $F_{h s}$	5.50	0+2	.10
8. non-negative $\log \theta_{sh}$	8.32	0+3	.04
9. ordered row-column association	8.36	3+1	.08
10. ordered row association	18.60	4+1	.00
11. ordered column association	8.84	6+1	.22
B. Ordinal regression models			
1. no three-variable interaction	24.88	24	.41
2. only effect of S	53.42	30	.01
3. only effect of E	39.62	32	.17
4. 1 + uniform associations	54.58	36	.02
5. 1 + order-restricted local odds ratios	35.30	24+6	.23
C. Marginal models with missing data			
1. marginal homogeneity	22.36	3	.00
2. non-decreasing marginals	3.17	0+1	.07
3. linearly changing marginals	13.73	2	.00
D. Ordinal latent class models			
1. unrestricted four-class model	15.11	24	.92
2. uniform associations	115.79	48	.00
3. column associations	105.86	42	.00
4. non-increasing cumulative probabilities	15.55	24+2	.96
5. non-negative local log odds ratios	39.20	24+16	.34

1. The reported number of degrees of freedom for the order-restricted models is the df of the model without constraints plus the number of activated constraints.

2. The p values of the models with inequality constraints are estimated on the basis of 1000 bootstrap samples. The standard errors of these estimates are less than .01 for $p \leq .11$ and $p \geq .89$, and at most .02 for other p values.

Table 5: Observed response patterns with frequencies from SIPP panel

pattern	freq.	pattern	freq.	pattern	freq.
1 1 1 1	2447	1 1 1 2	114	1 1 2 1	75
1 1 2 2	79	1 2 1 1	87	1 2 1 2	9
1 2 2 1	22	1 2 2 2	84	2 1 1 1	147
2 1 1 2	31	2 1 2 1	41	2 1 2 2	80
2 2 1 1	103	2 2 1 2	36	2 2 2 1	61
2 2 2 2	1450	1 1 1 0	106	1 1 2 0	9
1 2 1 0	8	1 2 2 0	5	2 1 1 0	8
2 1 2 0	5	2 2 1 0	3	2 2 2 0	75
1 1 0 1	38	1 1 0 2	3	1 2 0 1	5
1 2 0 2	3	2 1 0 1	4	2 2 0 1	7
2 2 0 2	23	1 1 0 0	103	1 2 0 0	15
2 1 0 0	24	2 2 0 0	99	1 0 1 1	14
1 0 1 2	4	1 0 2 1	1	1 0 2 2	1
2 0 1 1	3	2 0 2 1	1	2 0 2 2	18
1 0 1 0	5	2 0 2 0	7	1 0 0 0	183
2 0 0 0	155	0 1 1 1	70	0 1 2 1	7
0 1 2 2	7	0 2 1 1	8	0 2 1 2	7
0 2 2 1	2	0 2 2 2	40	0 1 1 0	19
0 1 2 0	3	0 2 1 0	3	0 2 2 0	13
0 1 0 1	3	0 1 0 0	39	0 2 0 0	28
0 0 1 1	65	0 0 1 2	2	0 0 2 1	17
0 0 2 2	56	0 0 1 0	26	0 0 2 0	16
0 0 0 1	89	0 0 0 2	64	0 0 0 0	369

0 = missing; 1 = employed; 2 = not employed

Table 6: Parameter estimates for Model D.4 (order-restricted latent class model)

π_w	$W = 1$	$W = 2$	$W = 3$	$W = 4$
	0.16	0.17	0.37	0.30
$F_{b w}$	$S = 1$	$S = 2$	$S = 3$	$S = 4$
$W = 1$	0.61	*0.76	0.95	1.00
$W = 2$	0.16	*0.76	0.94	1.00
$W = 3$	0.03	0.17	0.85	1.00
$W = 4$	0.03	0.09	0.34	1.00
$F_{c w}$	$P = 1$	$P = 2$	$P = 3$	$P = 4$
$W = 1$	0.45	0.62	0.86	1.00
$W = 2$	*0.04	0.46	0.81	1.00
$W = 3$	*0.04	0.13	0.70	1.00
$W = 4$	0.02	0.08	0.18	1.00
$F_{b w}$	$B = 1$	$B = 2$	$B = 3$	$B = 4$
$W = 1$	0.75	0.87	0.98	1.00
$W = 2$	0.22	0.70	0.95	1.00
$W = 3$	0.08	0.22	0.79	1.00
$W = 4$	0.02	0.03	0.19	1.00

A ^(*) indicates an activated constraint