Regina Dittrich and Brian Francis and Reinhold Hatzinger and Walter Katzenbeisser

A Paired Comparison Approach for the Analysis of Sets of Likert Scale Responses

Paper

A Paired Comparison Approach for the Analysis of Sets of Likert Scale Responses

Regina Dittrich, Brian Francis, Reinhold Hatzinger, Walter Katzenbeisser

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 24
November 2005

http://statmath.wu-wien.ac.at/

A Paired Comparison Approach for the Analysis of sets of Likert scale Responses

Regina Dittrich†

Vienna University of Economics and Business Administration, Austria.

Brian Francis

Lancaster University, United Kingdom

Reinhold Hatzinger

Walter Katzenbeisser

Vienna University of Economics and Business Administration, Austria.

Summary. This paper provides an alternative methodology for the analysis of a set of Likert responses measured on a common attitudinal scale when the primary focus of interest is on the relative importance of items in the set. The method makes fewer assumptions about the distribution of the responses than the more usual approaches such as comparisons of means, MANOVA or ordinal data methods. The approach transforms the Likert responses into paired comparison responses between the items. The complete multivariate pattern of responses thus produced can be analysed by an appropriately reformulated paired comparison model. The dependency structure between item responses can also be modelled flexibly. The advantage of this approach is that sets of Likert responses can be analysed simultaneously within the Generalized Linear Model framework, providing standard likelihood based inference for model selection. This method is applied to a recent international survey on the importance of environmental problems.

Keywords: Paired comparisons, Bradley-Terry model, response patterns, log-linear model, generalized linear model, GLIM, R, Likert scale

1. Introduction

Likert scale items are commonly used to investigate the attitudes of respondents to a series of written or verbal statements (items). Typically, the statements form sets of questions, with respondents asked to represent their strength of feeling on a common categorical scale. Such response scales - often with five ordered categories labelled 1 to 5 - are typically defined by endpoints such as "not at all serious" to "very serious", "very unimportant" to "very important", or "strongly dislike" to "strongly like". The Likert scale is an essential tool in psychology and in social surveys, and is a ubiquitous method of collecting attitudinal data.

Two types of analysis are commonly carried out on sets of Likert responses. The first type relates to score building. Responses to items are treated as belonging on a numerical

†*Address for correspondence:* Regina Dittrich, Vienna University of Economics, Department of Statistics and Mathematics, Augasse 2-6, A-1090 Vienna, Austria. Tel.Nr.: +43 1 31336 4256, FAX: +43 1 31336 774
E-mail: Regina.Dittrich@wu-wien.ac.at

scale, and are either summed over the items, or a factor or latent variable analysis is carried out, and a weighted or unweighted score is produced, which is taken to measure a common characteristic of the item set for a respondent. For example, a set of attitudinal questions might aim to represent the degree of aggression of an individual through an aggression score. However, such approaches are not the concern of this paper.

The second type of analysis, which we focus on here, is concerned more with providing an ordering of the relative importance of a set of items, and how this relative importance might vary according to other characteristics of the individual. Commonly, simple methods are used to examine the relative importance of Likert items. Sometimes, Likert items are treated as categorical, where frequencies for each item are given to determine a ranking of the items; those items which are more often judged high are interpreted as more important than others (e.g. Denz, 2000). Other studies look at the percentage of responses in a particular combination of responses, such as the two highest categories of each item ignoring the rest of the information in the data (e.g. Witherspoon, 1994). Another common procedure is to treat each Likert scale as continuous, for example, for a five point Likert scale, the responses $\{1, 2, 3, 4, 5\}$ are treated as equally spaced points along a continuum. In such cases a mean and standard deviation is often reported for each of the Likert-scale questions and the items are ranked according to the means (e.g. Culp and Schwartz, 1999; Aby et al, 1995; O'Hara and Stagl, 2002). The effect of subject covariates for each item separately can also be investigated to account for differences between groups (e.g. Kemp and Burt, 2002). More sophisticated methods might use a multivariate approach, and simultaneously analyse the joint pattern of means for the set of items through MANOVA or multivariate regression.

Other classes of methods rely on models based on latent variable approaches; Tutz(1990) provides a good overview. The ordered categorical scale is assumed to be a manifestation of a latent quantitative variable. These models can be seen as a multivariate (standard normal) models that has been discretised using a set of threshholds. Maydeu-Olivares (2002) proposed a Thurstonian type model while Uesaka and Asano (1987a, 1987b) suggested latent scale linear models for ordinal responses within the Grizzle-Starmer-Koch approach (Grizzle, Starmer and Koch, 1969)

These approaches can be problematic for a variety of reasons. Simple categorical approaches either fail to utilise the complete information in the data, or have difficulty in determining a proper ranking of the items. In addition, much analysis is descriptive and lacks proper statistical analysis when comparing groups (e.g. Dalton and Rohrschneider, 1998). Methods analysing means (either univariate or multivariate) assume both that the distance between response categories are equal, and that the responses have an underlying normal or multivariate normal distribution. These assumptions made are often unrealistic in practice. The assumption of equidistant categories is not needed in the model proposed by Maydeu-Olivares but the normality assumption is still fundamental for the Thurstonian type model. The models proposed by Uesaka and Aanso are more general allowing various distributions for the latent variables to be specified.

Nevertheless, a common problem with all these methods is that item or Likert responses are treated as *absolute* measurements, and this can be a rather dubious assumption especially when dealing with subjective self assessments. In the psychometric literature (e.g. Fischer, 1974) it is a basic assumption that one requirement for defining measurements is that individuals giving the same answer to a Likert item (choosing the same category) do not only share the same response value but are equivalent or similar with respect to the attitudes, values, etc. to be measured. This implies that for example a Likert response of 5 (="very important") has the same meaning for all individuals which responded in this category. However, in social surveys and other attitudinal work this can be questioned. Brady (1989) addresses this problem in the context of factor and ideal point analysis for interpersonally incomparable data: statistical methods that rely on interpersonal comparability are prone to produce spurious results when this assumption is not met. This is a particular problem when comparing different countries or cultures, with various authors (e.g. Heine et al, 2002) coming to the conclusion that cross-cultural comparisons of means of Likert items are inaccurate. For example, the data analysed in the present paper is concerned with the importance of the perceived danger of various environmental issues. However, within a country, the perceived danger of an issue will depend on numerous factors - the safety standards of the country and the perceived danger of other issues not asked about. Thus, *absolute measures* of importance will assess local circumstances as much as the importance of issues. It these circumstances, it is often more meaningful to consider *relative* judgements, i.e. a Likert response value of 5 will only express a higher importance than a lower Likert response value for a given individual, and not more. Of course, with this approach the scale origin will be lost (Böckenholt, 2004) but following our arguments it can be doubted that the assumption of common origins in the decision process of item evaluations across all respondents is always reasonable.

The purpose of this paper is to present an alternative method for the analysis of Likert scale data. We suggest that attention should be focussed on the *relative importance* of items, rather than on their absolute importance: therefore we construct *comparative* judgements as an alternative to ratings usually treated as *absolute* judgements. Thus our approach is to interpret Likert responses as *rankings with ties* which can be further expanded to paired comparison data. Therefore we only use the ordinal information in the original Likert response data. This generated paired comparison data can then be analysed by appropriately defined paired comparison models. Finally a ranking of Likert items can be achieved. This is similar to the Mallows-Bradley-Terry approach for modelling rankings without ties (Critchlow and Fliegner, 1991).

The main advantage is that for this approach possible interpersonal incomparabilities do not matter anymore and an undesirable effect of individual idiosyncratic interpretations of category labels is avoided. There are further advantages: No assumption is made concerning underlying normality, or about the equidistance of response categories. The complete multivariate pattern of paired comparisons produced are modelled within the Generalised Linear Model framework and therefore standard software can be used for parameter estimation, assessments of goodness of fit and model selection. Parameters representing dependencies between items can also be introduced. Furthermore the effects of subject covariates can be assessed and the importance of these effects can be judged. Of course, a disadvantage is that the dimensionality of the problem will become very large, soon.

The structure of this paper is as follows: in Section 2 we describe a data set on environmental concerns to be analysed. This is followed in Section 3 by a brief description of paired comparison models. We then show how transformed Likert scale responses can be analysed using an appropriately reformulated version of a log-linear model for paired comparison response patterns. In Section 4 we analyse the environmental data set and give an interpretation of the results. The paper concludes with a discussion in Section 5.

2. Environmental problems in Europe

Environmental issues have increasingly become an important concern of modern society and cross-national surveys like the International Social Survey Programme (ISSP) have addressed this topic. The second, more recent, survey in the ISSP which related to environmental concerns was carried out in the year 2000. In all, 27 countries were surveyed with around 1000 respondents from each country, but a detailed analysis for all countries would exceed the purpose of this paper. We have instead chosen to focus on two countries - Austria and Great Britain - motivated by the fact that the authors come from these countries and therefore the results might better be understood by substantive knowledge about these countries.

The relevant part of the survey was related to major environmental concerns and respondents were asked about their perception of environmental dangers. The attitudinal questions used in the analysis were presented as five-point Likert items and the first question (`CAR`) had the form:

In general, do you think air pollution caused by cars is:

(1) extremely dangerous for the environment
(2) very dangerous for the environment
(3) somewhat dangerous for the environment
(4) not very dangerous for the environment
(5) not dangerous at all for the environment

The other questions in the set had identical response choices, but were related to other issues: *air pollution caused by industry* (`IND`), *pesticides and chemicals used in farming* (`FARM`), *pollution of country's rivers, lakes and streams* (`WATER`), *a rise in the world's temperature caused by the 'greenhouse effect'* (`TEMP`) and *modifying the genes of certain crops* (`GENE`).

We are interested in determining the relative ranking of these six items and how this ranking changes according to country. A common approach (cf. Witherspoon, 1994) is to examine the percentages of those responding 'extremely' or 'very' dangerous for each of the items.

Using this technique, Table 1 shows remarkable differences between Austria and Great Britain. We notice first of all that there are absolute differences in the level of concern - the percentages are higher in Austria. However, this paper is concerned with the relative ordering of the items, and we observe differences between the two countries in this respect as well. Genetic modification was ranked third highest in Austria, but was the lowest of all the issues in GB. In contrast, water pollution was the top issue for GB, but ranked in fifth place for Austria.

However, this rather simple descriptive analysis does not include the information from all available categories and does not provide inferential conclusions. One of the aims of this paper is to determine a ranking of the items by using a statistical model and to investigate whether these rankings vary according to characteristics of the respondents, i.e. subject covariates.

Table 1. Comparison of percentages in Austria and Great Britain

Percentage stating issue is "extremely" or "very" dangerous for environment

Issue	in Austria %	in GB %
air pollution caused by cars (`CAR`)	50.8	57.2
air pollution caused by industry (`IND`)	78.8	65.1
pesticides and chemicals used in farming (`FARM`)	69.4	52.5
pollution of country's rivers, lakes and streams (`WATER`)	63.2	65.6
a rise in the world's temperature (`TEMP`)	78.3	54.1
modifying the genes of certain crops (`GENE`)	73.9	44.9
number of respondents	782	813

The data used for the analysis in this paper consist of $N = 1595$ complete responses to the six environmental items on a Likert type scale and five covariate values each describing the subjects: country, age, gender, level of education attained and the locality of residence ($N = 782$ Austrian repondents, and $N = 813$ from Great Britain). The subject covariates country and gender are two-category variables (`country`: 1, Great Britain; 2, Austria; and `sex`: 1, male; 2, female). Age, originally a continuous variable, was recoded into a three-category variable (`age`: 1, < 40 years; 2, 41-59 years; 3, 60+ years). The level of education attained, which originally had seven potential valid categories, was recoded into a two category variable (`edu`: 1, below A-level/matrice; 2, A-level/matrice or higher). The location of residence was obtained by self-assessment and used in the analysis as it had been defined in the survey (`urb`: 1, urban area; 2, suburbs of large cities, small town, county seat; 3, rural area).

3. Modelling Likert Scale Data

In this section, we show how to analyse Likert scale data by means of paired comparison models.

The basic idea is to compare two selected Likert items A and B simply by observing whether one or the other is "preferred", that is, has a higher or more positive response on the Likert response scale. The transformed response has three values - that A is preferred to B, that B preferred to A, or that the two Likert responses are equal and there is no preference. By transforming all possible pairs of Likert items, we obtain a multivariate pattern of paired comparison responses. In order to model this data we use the correspondence between appropriately defined Likert-patterns and derived paired comparison patterns (PC-patterns). A wide selection of models for paired comparison data exist. We begin with a class of models known as Bradley-Terry models, which makes very few assumptions about the nature of the data. From these models one obtains a set of item worths which provide the relative importance of the Likert items.

We first describe a model for paired comparisons, then the data transformation process, and then formulate the statistical model for the multivariate response.

3.1. The Bradley-Terry model

Given a set of items $1, 2, \ldots, J$, the basic Bradley-Terry model (Bradley and Terry, 1952) can be written

$$P\{Y_{ij} = h|\pi_i, \pi_j\} = \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^h \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{1-h}, \; h = 0, 1\,,$$

where $\{Y_{ij} = 0\}$ ($\{Y_{ij} = 1\}$) represents the event that item i (item j) is chosen in the comparison of items i and j. The π's are unknown non-negative 'worth' parameters, describing the location of the items on the preference scale and we ensure identifiability by the requirement that $\sum_i \pi_i = 1$.

The Bradley-Terry model can be extended to deal with an ordinal response, representing the *degree* of preference between two items (Agresti, 1992). Thus, in the comparison of items i and j, the response scale could consist of five categories with the labels *strong preference for* i, *mild preference for* i, *no preference*, *mild preference for* j, and *strong preference for* j.

As a starting point we use the *Adjacent Categories* model (Böckenholt and Dillon, 1997; Dittrich et al, 2004) which postulates a power relationship between the response category and the probability of preferring item i over item j. For an ordinal response with $H + 1$ categories, let $\{Y_{ij} = h\}$, $h = 0, 1, \ldots, H$, denote the event that response category h is chosen in the comparison of item i and item j. $\{Y_{ij} = 0\}$ denotes the event that item i is strongly preferred over j, whereas $\{Y_{ij} = H\}$ represents the event that item j is strongly preferred over i. The AC-model is specified by

$$P\{Y_{ij} = h|\pi_i, \pi_j, c_h\} = a_{ij} c_h \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^h \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{H-h}, \; h = 0, 1, \ldots, H\,, \tag{1}$$

where all π_i's are positive, and where a_{ij} denotes a normalizing constant to let the probabilities in (1) sum to unity. c_h is interpreted as a parameter representing a possible response bias effect towards selecting category h.

Formula (1) can be rewritten by applying a substitution suggested by Sinclair (1982), which produces a simpler and more tractable form of (1):

$$P\{Y_{ij} = h|\pi_i, \pi_j, c_h\} = a^*_{ij} c_h \left(\frac{\sqrt{\pi}_i}{\sqrt{\pi}_j}\right)^{H-2h}, \quad h = 0, 1, \ldots, H\,, \tag{2}$$

where a^*_{ij} is again a normalizing constant.

In this paper we will restrict the number of response categories to three, i.e $H = 2$. This restriction will make fewer assumptions about the distances between the categories of the Likert item scales. For $H = 2$, the AC-model (2) has a particularly simple form:

$$P\{Y_{ij} = h|\pi_i, \pi_j, c_h\} = a^*_{ij}\, c_h \left(\frac{\pi_j}{\pi_i}\right)^{1-h}, \; h = 0, 1, 2\,. \tag{3}$$

With $c_0 = c_2 = 1$, this is equivalent to the model proposed by Davidson (1970), who extended the Bradley-Terry model to deal with tied preferences.

3.2. Transforming Likert scale data

The basic idea in transforming Likert scale data to paired comparisons is straightforward. For any two Likert items, if the response to the first item is greater on the numeric scale than the second, then we say that the first item is preferred to the second. If the responses are equal, this gives an undecided preference.

Formally, we suppose that an individual rates each of a set of J Likert items using a common κ-point scale, where higher values of the Likert responses correspond to higher preferences. Note that all judges have to rate all items and no missing values are allowed. For Likert item j, the response is denoted by l_j. These ratings produce an observed Likert-pattern vector

$$\ell = (l_1, l_2, \ldots, l_J),$$

which is one of all κ^J possible patterns, and $l_j \in \{1, 2, \ldots, \kappa\}$. For a given pair (i, j) of items consider the difference $l_j - l_i$ of the respective Likert response-patterns and define

$$w_{ij} = \begin{cases} -1 & \text{if } l_j < l_i, \quad (h = 0) \\ 0 & \text{if } l_j = l_i, \quad (h = 1) \\ 1 & \text{if } l_j > l_i. \quad (h = 2) \end{cases} \tag{4}$$

as derived PC-responses. A positive value for w_{ij} corresponds to a preference for item j. However, if the Likert scale has been set up such that lower numbers represent higher preferences, the definition would need to be changed by switching the 1 and -1 values to give the same result.

3.3. *Statistical Modelling of the transformed Likert scale data*

Let us consider the w_{ij}'s which are the derived PC-responses as realisations of random variables W_{ij}. We start by modelling the joint distribution of the random variables

$$(W_{12}, W_{13}, \ldots, W_{1J}, W_{23}, W_{24}, \ldots, W_{2J}, \ldots, W_{J-1:J})$$

through a multiplicative formulation. We first specify an independence model which is similar to the Mallows-Bradley-Terry ranking model (Mallows, 1957; Critchlow and Fligner, 1991). In their model the probability of each *ranking* of the items is taken to be proportional to the product of the probabilities of all pairwise comparisons that are consistent with the ranking. In our approach we use the correspondence between Likert-response patterns and the derived PC-patterns. This independence model will then further be generalised to include possible dependencies between the derived PC-responses (Dittrich et al, 2002).

Recall equation (4). We note that the event $\{W_{ij} = w_{ij}\}$ is equivalent to $\{Y_{ij} = w_{ij} + 1\}$ in our notation. We can therefore use a variant of equation (3) to construct the distribution of the random variable W_{ij}:

$$P(w_{ij}) = P\{W_{ij} = w_{ij}\} = P\{Y_{ij} = w_{ij} + 1\} = \begin{cases} a^*_{ij} \left(\frac{\pi_j}{\pi_i}\right)^{w_{ij}}, & \text{if } w_{ij} = -1, 1, \\ a^*_{ij}\, c_{ij}, & \text{if } w_{ij} = 0, \end{cases} \tag{5}$$

where a^*_{ij} is again a normalizing constant. The parameters c_0 and c_2 in (3) are set to unity as there can be no response preference bias for particular generated paired comparison responses when responses in fact occur on the Likert scales. The parameters c_{ij} allows for a different probability of equality of responses for each pair of items, and also allows for the fact that the number of cells in the $(\kappa \times \kappa)$ table of Likert responses which contribute towards each of the three PC-responses is not equal. For example in a 5×5 -point Likert scale comparison table, five cells contribute to $w_{ij} = 0$, ten cells contribute to $w_{ij} = 1$ and ten to $w_{ij} = -1$.

For the independence model we define the probability of a derived PC-response pattern $(w_{12}, w_{13}, \ldots, w_{J-1:J})$ by:

$$P(w_{12}, w_{13}, \ldots, w_{J-1:J}) = P(W_{12} = w_{12}, \ldots, W_{J-1:J} = w_{J-1:J}) = \alpha \prod_{i<j} P(w_{ij}), \qquad (6)$$

where α is a normalising constant to make the probabilities sum up to unity, and $P(w_{ij})$ is defined in (5).

3.4. Incorporating dependencies

As we are modelling the joint distribution of the W_{ij}, various types of dependencies can be considered between the derived PC-responses. Even if Likert responses are taken to be independent, dependencies between the derived PC-responses are likely to arise from the transformation process. For example, if for item 1 we form all $J - 1$ pairs where item 1 is included, then the derived PC-responses involving the common item 1 are partially determined by the Likert response of item 1. This is particularly true when the Likert response of the common item is judged to be on either of the extremes of the Likert scale *most favourable* (or *least favourable*), as all other items have to be either *equal* or *less* (*equal* or *more*) to that item. It is hypothesised that dependencies are introduced by repeated occurrence of identical objects in (two) pairs of derived PC-responses. These dependencies will be represented by further parameters $\theta_{ij,ik}$, i.e. certain two-way interactions between PC-responses, where i represents the common item (cf. Dittrich et al, 2002).

The joint distribution of the W_{ij} can then be written as:

$$P(w_{12}, w_{13}, \ldots, w_{J-1:J}) = \qquad (7)$$
$$\alpha^* \prod_{i<j} P(w_{ij}) \times \exp\{\theta_{12,13}\, w_{12}w_{13} + \theta_{12,14}\, w_{12}w_{14} + \cdots + \theta_{J-2:J,J-1:J}\, w_{J-2:J}w_{J-1:J}\},$$

with α^* as a normalising constant. Dependencies are introduced by the exponent term. If all $\theta = 0$ the model is reduced to that of independence. The inclusion of terms of this form is similar to the approach taken by Cox (Cox, 1972; Cox and Wermuth, 1994) when modelling dependence in multivariate binary data. The resulting distribution was termed the quadratic exponential distribution. Higher order dependencies (not considered here) can also be introduced by accounting for triple, quadruple etc. PC-responses where one item is in common in more than two pairs. An advantage of this specification is that the θ-parameters can be interpreted to be proportional to a log-odds ratio in the conditional distribution of two W's given all others.

3.5. Parameter estimation and the log-linear model

So far, there is no difference between a model based on real paired comparisons and a model based on derived PC-responses (derived from Likert responses). However, there is a main difference between these two types of responses concerning the set of different response patterns. A simple approach might be to take the transformed response w_{ij} for every (i, j) comparison and for every individual, as contributions to the counts of a contingency table with $3^{\binom{J}{2}}$ cells. In other words, we could assume that the transformed responses are generated by a paired comparison experiment which could be analysed using a standard paired comparison model. The assumption is then made that all possible PC-response patterns can occur. This approach, however, does not lead to correct parameter estimates in model fitting because the assumed paired comparison experiment also allows for inconsistent responses between all pairs of items, such as $A > B$, $B > C$ and $C > A$, responses that can not be generated from Likert scales.

Determining the number of such generated distinct PC-patterns is not straightforward. Since each J dimensional Likert-pattern vector

$$\ell = (l_1, l_2, \ldots, l_J)$$

is transformed into a $\binom{J}{2}$ dimensional paired comparison(PC)-pattern vector

$$\mathbf{w} = (w_{12}, w_{13}, \ldots, w_{1J}, w_{23}, w_{24}, \ldots, w_{2J}, \ldots, w_{J-1:J})\,, \tag{8}$$

where for every pair of items i and j, $w_{ij} \in \{-1, 0, 1\}$, but as we have already seen, not all combinations of paired comparison responses can be generated from Likert responses. The number must be less than $3^{\binom{J}{2}}$.

Moreover, the transformation causes different observed Likert-pattern vectors to be mapped into the same PC-pattern vector. For example, the Likert-patterns $(1, 1, \ldots, 1)$, $(2, 2, \ldots, 2), \ldots$ etc. will all be mapped into the same PC-pattern $(0, 0, \ldots, 0)$.

Let us assume that we have a set of κ^J Likert-pattern vectors. Then U, the number of unique PC-patterns is given by

$$U = \sum_{\nu=0}^{\kappa} \nu\,! \left\{ {J \atop \nu} \right\}, \tag{9}$$

where $\left\{ {J \atop \nu} \right\}$ stands for the number of ways to partition a set of J elements into ν nonempty subsets. In fact, $\left\{ {J \atop \nu} \right\}$ is a Stirling number of the second kind, which can be calculated by means of the following recurrence (Graham et al, 1989):

$$\left\{ {J \atop \nu} \right\} = \nu \left\{ {J-1 \atop \nu} \right\} + \left\{ {J-1 \atop \nu-1} \right\}, \text{integer } J > 0\,.$$

In the special case of $\kappa = J$ (that is, where the number of Likert categories is equal to the number of items) an ordered Bell number, $B(J) = \sum_{\nu=0}^{J} \nu\,! \left\{ {J \atop \nu} \right\}$, is obtained. This sequence runs: $1, 3, 13, 75, 541, 4683, \ldots$ (cf. Sloane's A000670 sequence, Sloane and Plouffe, 1995). This formula is also true for $\kappa > J$ because $\left\{ {J \atop \nu} \right\} = 0$ for all $\nu > J$. A more detailed explanation is given in Appendix A.

We base our parameter estimation on simple multinomial sampling over the derived PC-response pattern vectors. To estimate the parameters, we suppose each of the N subjects has responded completely to the J Likert items thus contributing a certain PC-response pattern. Let

N_u = number of times when the u-th unique PC-response pattern vector is observed,

where $u = 1, 2, \ldots, U$. Then the N_u's are multinomially distributed with $N = \sum_u N_u$ and with probabilities P_u . Thus, the likelihood function is

$$L = \Delta \prod_u P_u^{N_u}$$

where $P_u = P(w_{12,u}, w_{13,u}, \ldots, w_{J-1:J,u})$ given in (6) or (7). The expectation m_u of N_u can then be represented through a log-linear model:

$$\ln m_u = \ln E\{N_u\} = \ln(N\Delta) + \sum_{i<j} \ln P(w_{ij,u})$$

for the independence model, and

$$\begin{aligned} \ln m_u \quad &= \quad \ln(N\Delta^*) + \sum_{i<j} \ln P(w_{ij,u}) + \\ &\quad \theta_{12,13}\, w_{12,u} w_{13,u} + \theta_{12,14}\, w_{12,u} w_{14,u} + \cdots + \theta_{J-2:J,J-1:J}\, w_{J-2:J,u} w_{J-1:J,u} \end{aligned}$$

for the dependence model. Both Δ and Δ^* are normalising constants. The relation between the multinomial distribution and the Poisson distribution can then be used to fit the model as a Poisson log-linear model.

Using matrix notation, the model can be written as

$$\mathbf{y} = \mathbf{X}\,\beta = (\mathbf{1}, \mathbf{X}^*, \mathbf{C}, \mathbf{D})\,\beta \tag{10}$$

where $\mathbf{y}$ is the vector of the $\ln m_u$s, $\mathbf{X}$ is the design matrix and the vector β containing the model parameters is given by

$$\beta^T = (\delta, \lambda_1, \lambda_2, \ldots, \lambda_J, \gamma_{12}, \gamma_{13}, \ldots, \gamma_{J-1:J}, \theta_{12,13}, \theta_{12,23}, \ldots, \theta_{J-2:J,J-1:J})\,,$$

$\delta = \ln\{N\Delta^*\}$, $\lambda = \ln \pi$ and $\gamma_{ij} = \ln c_{ij}$. It is worth noting that there is a degree of over-parameterisation in this model. Specifically, there are too many λ parameters. In practical examples, we fix λ_J at zero to ensure estimability.

The design matrix $\mathbf{X}$ consists of the following submatrices. The matrix $\mathbf{X}^*$represents the λ parameters and can be generated by

$$\mathbf{X}^* = (-1)\mathbf{W}\mathbf{B}\,.$$

The rows of the matrix $\mathbf{W}$ are given by the unique derived PC-responses $\mathbf{w}$ as given in (8) and $\mathbf{B}$ is the paired comparison design matrix (Böckenholt and Dillon, 1997)

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 1 & 0 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix}.$$

Each column of this matrix corresponds to one of the Likert items and each row to one of the pairwise comparisons (i.e., the differences between two Likert items).

The coefficients associated with the parameters γ given in the matrix $\mathbf{C}$ can be generated by

$$\mathbf{C} = (\mathbf{J} - \mathbf{W} \odot \mathbf{W})\ ,$$

where $\mathbf{J}$ is a matrix of 1's with the same size as the matrix $\mathbf{W}$ and $\odot$ denotes the elementwise multiplication of the matrices.

The matrix $\mathbf{D}$ used to generate the coefficients of the dependency parameters θ can easily be calculated in the usual way as two-way interactions by multiplying the corresponding columns of the $\mathbf{W}$ matrix elementwise.

$$\mathbf{D} = (\mathbf{w}_{12} \odot \mathbf{w}_{13}, \mathbf{w}_{12} \odot \mathbf{w}_{23}, \ldots \mathbf{w}_{J-2:J} \odot \mathbf{w}_{J-1:J}) \ .$$

Finally, given the estimates of $\lambda_1, \lambda_2, \ldots, \lambda_J$, we can estimate the worth parameters $\pi_1, \pi_2, \ldots, \pi_J$ through the expression

$$\pi_i = \frac{\exp(\lambda_i)}{\sum_j \exp(\lambda_j)}$$

3.6. *An example*

To demonstrate the ideas consider the simple case of three items rated on a $\kappa = 2$ point Likert scale. The set of all different Likert-patterns consists of $2^3 = 8$ row vectors arranged in a matrix $\mathbf{L} = (\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)$, with columns $\mathbf{l}_i$. All possible pairwise differences of columns lead to the PC- response pattern matrix $\mathbf{W}^* = (\mathbf{w}^*_{12}, \mathbf{w}^*_{13}, \mathbf{w}^*_{23})$, where $\mathbf{w}^*_{ij} = \mathbf{l}_j - \mathbf{l}_i,\ j > i$ analogous to (4):

$$\mathbf{L} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \end{pmatrix}, \quad \mathbf{W}^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \\ -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \\ -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{pmatrix}.$$

It can be seen that many of the 3^3 possible PC-response patterns such as (1, 1, 1) or (1, 1, -1) do not appear in $\mathbf{W}$, as they cannot be generated by any Likert-pattern. It can also be seen that the Likert-patterns (1, 1, 1) and (2, 2, 2) will both be transformed into the same PC-response pattern (0, 0, 0). Removing the redundant pattern, here the last row in the $\mathbf{W}^*$ matrix, yields the unique (7×3) paired comparison pattern matrix $\mathbf{W} = (\mathbf{w}_{12}, \mathbf{w}_{13}, \mathbf{w}_{23})$, where the number of rows are given by $\left\{{3 \atop 0}\right\} + \left\{{3 \atop 1}\right\} + 2\left\{{3 \atop 2}\right\} = 0 + 1 + 2 \times 3$. Note that the matrix $\mathbf{W}^*$ can be generated from the matrix $\mathbf{L}$ by

$$\mathbf{W}^* = (-1)\mathbf{L}\,\mathbf{B}^T\,,$$

where the matrix $\mathbf{B}$ is the (3×3) paired comparison design matrix given by

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}.$$

To obtain the matrix $\mathbf{W}$ from $\mathbf{W}^*$ is not straightforward, since in general we do not know the positions (row numbers) of the redundant patterns in $\mathbf{W}^*$. For practical purposes we have to computationally remove these patterns (e.g. by using the *unique*-function in the R/S programming language). The same problem arises for the response variable, i.e. the

N_us. We need to aggregate over all related frequencies to get the total number of responses for each unique PC-pattern.

Once having obtained the properly dimensioned $\mathbf{W}$ (and the vector for the response variable) all other model structures can be set up according to the previous section. Using matrix notation we can write for all unique PC-response patterns:

$$\begin{pmatrix} \ln m_1 \\ \ln m_2 \\ \ln m_3 \\ \ln m_4 \\ \ln m_5 \\ \ln m_6 \\ \ln m_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 2 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -2 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & -2 & 1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & -2 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{23} \\ \theta_{12,13} \\ \theta_{12,23} \\ \theta_{13,23} \end{pmatrix} = \mathbf{X}\,\beta\,,$$

where $\mathbf{X} = (\mathbf{1}, \mathbf{X}^*, \mathbf{C}, \mathbf{D})$ and the vector $\beta^T = (\delta, \lambda_1, \lambda_2, \lambda_3, \gamma_{12}, \gamma_{13}, \gamma_{23}, \theta_{12,13}, \theta_{12,23}, \theta_{13,23})$. Note that in this example there are three pairs of paired comparisons, each pair with one item in common, i.e. the pairs (12,13), (12,23) and (13,23).

To be more specific, for $m_3 = E\{N_3\} = N\,P(1, 0, -1)$ we get

$$m_3 = E\{N_3\} = N\Delta\frac{\pi_2}{\pi_1} \times c_{13} \times \frac{\pi_2}{\pi_3} \times \exp\{-\theta_{12,23}\}$$

which can be written in log-linear form

$$\ln m_3 = \ln E\{N_3\} = \delta - \lambda_1 + 2\lambda_2 - \lambda_3 + \gamma_{13} - \theta_{12,23}\,.$$

□

3.7. *Subject-specific and item-specific covariates*

The model can easily be extended to allow for subject-specific covariates. This will involve replacing the parameters λ_i by λ_{is}, where $s = 1, \ldots, S$ represents the effect of subject s. We can then model the λ_{is} through a suitable model involving the covariates. This will necessitate duplicating the design matrix S times, and the observed pattern counts m_u are replaced by m_{us}, which is now a binary indicator indicating whether the pattern u is the pattern observed for subject s. For large S the problem quickly becomes intractable. However, if all covariates are categorical, we can consider instead all possible combinations T of the levels of the covariates and index these combinations by t. The level of duplication needed is then much reduced, as T is usually very much smaller than S. Estimability is ensured by fixing all λ_{Jt} to be zero. Worths are calculated for each covariate combination t through the expression

$$\pi_{it} = \frac{\exp(\lambda_{it})}{\sum_j \exp(\lambda_{jt})}$$

In certain applications it might be desirable to reparameterise the items, e.g. to combine items with common characteristics or to investigate some common properties. This extension of the model can be achieved by replacing the parameters λ_i (or λ_{is} as above) by the linear predictor

$$\lambda_i = \sum_{\nu=1}^{P} z_{i\nu}\beta_\nu^Z ,$$

where the $z_{i\nu}$'s denote the values of the covariates describing the ν'th property of item i and the β^Zs are unknown regression parameters. For these extensions see Francis et al (2002) and Dittrich et al (1998).

4. Analysis and interpretation of the environmental data set

We use the dependence model proposed in Section 3 to analyse the dataset on the subjective assessment of environmental dangers. The standard Likert paired comparison model without covariates will have a dataset with as many rows as there are non-redundant patterns. The actual number of rows can be calculated by formula (9), which for our data is 3963 (6 items, 5 categories). The number of columns is 82 and is given as follows: by 1 column for the grand mean, 6 columns for the item parameters `CAR`, `IND`, `FARM`, `WATER`, `TEMP`, `GENE`, 15 for the γ_{ij}-parameters (denoted by $G1$ to $G15$), one for each comparison for the case of equal Likert responses and 60 for the dependency parameters $\theta_{ij,ik}$'s (denoted by $T1$ to $T60$).

However, this covariate-free model is too simple for our needs and we extend it to include the effect of the five categorical covariates `country`, `age`, gender (`sex`), education (`edu`) and location (`urb`). These covariates have $2, 3, 2, 2$ and 3 levels respectively, giving 72 possible covariate combinations. This expands the dataset to $3963 \times 72 = 285336$ rows. Each observed row count represents the number of times a particular PC-pattern occurs for a particular combination of covariate categories.

4.1. Model selection

In the Wilkinson and Rogers notation (Wilkinson and Rogers, 1973), the basic model for the expanded data but with no covariate effects on the items can now be expressed as:

$$\begin{aligned}\text{Basic model} = & \texttt{country} * \texttt{sex} * \texttt{age} * \texttt{urb} * \texttt{edu} + \texttt{G1} + \texttt{G2} + \cdots + \texttt{G15} + \\ & \texttt{T1} + \texttt{T2} + \texttt{T3} + \cdots + \texttt{T60} + \texttt{CAR} + \texttt{IND} + \texttt{FARM} + \texttt{WATER} + \texttt{TEMP} + \texttt{GENE}\end{aligned}$$

For the analysis the highest possible interaction between all subject covariates

```
country*sex*age*urb*edu
```

now has to be included into the model to ensure that the fitted count totals for each covariate combination are equal to the observed count totals, thus properly fixing the marginal distribution. The main effect of a covariate on the item parameters are represented in our model by a set of interactions of a covariate with each of the six items. Thus the effect of `age` for example on the item parameters is represented by the terms

```
age*(CAR+IND+FARM+WATER+TEMP+GENE).
```

We started with a complex model (model 1) which included all possible interactions between three subject covariates and the items `CAR,IND,FARM,WATER,TEMP,GENE`. For example the three-way interactions of `sex, urb, age` and the items can be written as

`sex*urb*age*(CAR+IND+FARM+WATER+TEMP+GENE)`

which we simplify in our model notation to `sex*urb*age*ITEMS`. This notation implies that all lower interactions (that is, between two subject covariates and the items, and between single covariates and the items) are also included. The model can be written as a combination of four-way interaction groups as follows:

$$\begin{aligned}
&\text{Basic model}\\
&+\texttt{sex}*\texttt{urb}*\texttt{age}*\texttt{ITEMS}+\texttt{sex}*\texttt{urb}*\texttt{country}*\texttt{ITEMS}\\
&+\texttt{sex}*\texttt{age}*\texttt{country}*\texttt{ITEMS}+\texttt{urb}*\texttt{age}*\texttt{country}*\texttt{ITEMS}\\
&+\texttt{sex}*\texttt{urb}*\texttt{edu}*\texttt{ITEMS}+\texttt{sex}*\texttt{age}*\texttt{edu}*\texttt{ITEMS}\\
&+\texttt{urb}*\texttt{age}*\texttt{edu}*\texttt{ITEMS}+\texttt{sex}*\texttt{edu}*\texttt{country}*\texttt{ITEMS}\\
&+\texttt{urb}*\texttt{edu}*\texttt{country}*\texttt{ITEMS}+\texttt{age}*\texttt{edu}*\texttt{country}*\texttt{ITEMS}
\end{aligned}$$

We used a backward elimination procedure aiming for a parsimonious model. Each group of four-way interactions (between three subject covariates and the items, such as `sex.urb.age.ITEMS`) was removed from model 1 in turn (and added afterwards). No interaction group showed a significant deviance change as can be seen in table 2. Removing all possible four-way interactions between each triplet of three subject covariates and the items led to a general change in deviance of 112.9 on 125 degrees of freedom (p-value is 0.773), giving no evidence for the inclusion of any four-way interaction terms . After this step we obtained a model without four-way interactions which we call model 2. The same procedure was applied to model 2. Removing all possible three-way interaction groups (between two subject covariates each and the items) gave a general change in deviance of 130.3 on 95 degrees of freedom (p-value is 0.009) - and provided evidence that some of the three-way interaction groups were needed. Examination of the effect of the removal from model 2 of each of the three-way interaction groups in turn gave changes in deviance which are presented in table 2. Only the interaction group `sex·age·items` showed a significant change in deviance at this step of the analysis.

We then started again with model 2 removing the interaction groups step by step in the order of the p-values (starting with the least significant term `age·edu` and so on) and leaving them out if the change in deviance was not significant. The three-way interaction `urb ·edu· items` became significant in this analysis with a deviance change of 26.34 on 10 degrees of freedom (p-value of 0.0033) and the interaction term `sex·age·items` had a deviance change of 24.66 on 10 degrees of freedom (p-value of 0.006). At this stage we have model 3 which is:

$$\text{Basic model}+\texttt{sex}*\texttt{age}*\texttt{ITEMS}+\texttt{urb}*\texttt{edu}*\texttt{ITEMS}+\texttt{country}*\texttt{ITEMS}$$

As the subject covariates `sex`, `age`, `urb` and `edu` are all included in significant three-way interaction terms the two-way interactions of these covariates and the items can not be removed from the model. Leaving out the two-way interaction `country` $*$ `ITEMS` gives a change in deviance of 165.3 on 5 degrees of freedom and therefore it can not be left out. A

Table 2. Differences of deviances for models when omitting four-way and three-way interaction groups

omitting four-way interactions	deviance difference to model 1	df	p-values
sex · urb · age · ITEMS	17.79	20	0.60
sex · urb · country · ITEMS	14.69	10	0.14
sex · age · country · ITEMS	4.67	10	0.91
urb · age · country · ITEMS	20.18	20	0.45
sex · urb · edu · ITEMS	12.55	10	0.25
sex · age · edu · ITEMS	5.51	10	0.85
urb · age · edu · ITEMS	15.49	20	0.75
sex · edu · country · ITEMS	5.23	5	0.39
urb · edu · country · ITEMS	9.36	10	0.50
age · edu · country · ITEMS	8.58	10	0.57
omitting three-way interactions	deviance difference to model 2	df	p-values
sex · urb · ITEMS	14.05	10	0.17
sex · age · ITEMS	22.75	10	0.01
sex · country · ITEMS	3.15	5	0.68
sex · edu · ITEMS	4.26	5	0.51
urb · age · ITEMS	20.79	20	0.41
urb · country · ITEMS	13.97	10	0.17
urb · edu · ITEMS	13.24	10	0.21
age · country · ITEMS	8.24	10	0.61
age · edu · ITEMS	5.66	10	0.84
edu · country · ITEMS	3.76	5	0.58

final model is therefore model 3, showing that all five covariates are affecting the ordering of the items, and showing that the effect of age is different for males and females, and the effect of education differs according to residential location.

4.2. Interpretation

We first need to remind ourselves that we are looking at the differences in relative rather than absolute values of the Likert responses. Taking this into account, we focus on three main areas of difference in response which were identified in the analysis - country differences, age and sex differences and education and location differences. We interpret the results through examination of the parameter estimates λ_{it} (Tables 3 and 4), and also by visual inspection of the item worths π_{it}. Figures 1 and 2 display the worth parameters of the items for various groups defined by combinations of gender and age with location of residence and educational level. Figure 1 gives the parameter estimates for Austria and Figure 2 for Great Britain.

Country-level effects: One main issue of this analysis was the question if there are differences between Great Britain and Austria concerning the relative perception of environmental dangers. ; in order of the effect size they are: `country.WATER`, `country.CAR` and `country.IND`. Taking Great Britain as the reference country, in all these interactions, the effect is negative. Austrians rate `WATER`, `CAR` to be of lesser relative importance than respondents from Great Britain. The plots in Figures 1 and 2 tell a similar story. The top scoring items in Austria for most groups is industry (labelled `I`) competing with "a rise in

Table 3. Final model: parameters of interest; part 1

parameter	estimate	s.e.	z-value	p-value
CAR	0.113	0.047	2.40	0.0163
IND	0.396	0.051	7.84	0.0000
FARM	0.052	0.052	1.00	0.3163
WATER	0.321	0.050	6.45	0.0000
TEMP	0.232	0.051	4.52	0.0000
GENE	0.000	–	–	–
sex(fem).CAR	-0.126	0.039	-3.24	0.0012
sex(fem).IND	-0.107	0.042	-2.55	0.0106
sex(fem).FARM	-0.128	0.043	-2.97	0.0030
sex(fem).WATER	-0.122	0.041	-2.99	0.0028
sex(fem).TEMP	-0.217	0.043	-5.05	0.0000
sex(fem).GENE	0.000	–	–	–
urb(suburb).CAR	0.039	0.041	0.95	0.3418
urb(rural).CAR	-0.044	0.035	-1.26	0.2068
urb(suburb).IND	-0.082	0.044	-1.88	0.0607
urb(rural).IND	-0.120	0.036	-3.32	0.0009
urb(suburb).FARM	0.049	0.046	1.07	0.2867
urb(rural).FARM	-0.049	0.038	-1.27	0.2032
urb(suburb).WATER	-0.013	0.043	-0.31	0.7569
urb(rural).WATER	-0.068	0.036	-1.89	0.0593
urb(suburb).TEMP	-0.049	0.045	-1.10	0.2728
urb(rural).TEMP	-0.067	0.038	-1.79	0.0733
urb(suburb).GENE	0.000	–	–	–
urb(rural).GENE	0.000	–	–	–
age(2).CAR	-0.064	0.042	-1.52	0.1274
age(3).CAR	-0.069	0.043	-1.59	0.1114
age(2).IND	-0.157	0.045	-3.51	0.0004
age(3).IND	-0.163	0.046	-3.54	0.0004
age(2).FARM	-0.036	0.047	-0.77	0.4388
age(3).FARM	-0.052	0.048	-1.08	0.2789
age(2).WATER	-0.148	0.044	-3.35	0.0008
age(3).WATER	-0.153	0.045	-3.36	0.0008
age(2).TEMP	-0.184	0.046	-3.97	0.0001
age(3).TEMP	-0.215	0.048	-4.51	0.0000
age(2).GENE	0.000	–	–	–
age(3).GENE	0.000	–	–	–
country(Austria).CAR	-0.269	0.032	-8.28	0.0000
country(Austria).IND	-0.089	0.034	-2.62	0.0089
country(Austria).FARM	-0.057	0.036	-1.61	0.1085
country(Austria).WATER	-0.326	0.034	-9.48	0.0000
country(Austria).TEMP	0.006	0.035	0.16	0.8741
country(Austria).GENE	0.000	–	–	–
edu(A-level+).CAR	0.129	0.044	2.92	0.0035
edu(A-level+).IND	0.040	0.046	0.86	0.3920
edu(A-level+).FARM	0.144	0.049	2.95	0.0032
edu(A-level+).WATER	0.118	0.046	2.59	0.0096
edu(A-level+).TEMP	0.115	0.048	2.40	0.0165
edu(A-level+).GENE	0.000	–	–	–

the world's temperature" (labelled `T`) and "modifying genes of certain crops" (labelled `G`) in some groups. "Pollution caused by cars" (labelled `C`) is considered the least dangerous item in all groups and "pollution of rivers, lakes and streams" (`W`) is in general the second lowest item in nearly all groups.

In Great Britain industry (`I`) is also considered to be a more dangerous item in nearly all groups but competing with "pollution of rivers, lakes and streams" (`W`) and with "a rise in the world's temperature" (`T`) on top of the list of perceived environmental dangers. In contrast, in Great Britain cars (`C`) are towards the middle of the relative scale of dangers while "modifying genes of certain crops" (`G`) are in general at the lower end. Examination of the parameter estimates in Table 3 shows that there are three strongly significant interaction terms in the `country` $*$ `items` group: `country(austria).CAR` and `country(austria).WATER` and a less strong effect of `country(austria).INDUSTRY`. All effects are negative which means that for Austria the subjective impression of danger for the items `CAR`, `WATER` and `INDUSTRY` is significantly decreased compared to Great Britain.

For the item "pollution of rivers, lakes and streams", its relative dangerousness is likely to relate to the perceived quality of the water in lakes, rivers and streams in the two countries. In Austria lake and river quality has been improved substantially in the last decade and most of them have "drink water quality". In contrast, the British press were concentrating on stories of river pollution - one story had the headline "Rivers hit by wave of pollution accidents" (The Guardian, August 31 2000), followed a week later by "How pollution is making river fish change sex" (Daily Mail, September 7, 2000). The car and industry effects are perhaps less easily explained, but in Great Britain, there was a nationwide blockade of fuel depots in September 2000 by lorry hauliers protesting against the high "green" taxes on fuel, prompting a lively debate on the issue of taxing fuel. In Austria we speculate that there might be a more differentiated view concerning motor vehicles. Whereas most public attention is concentrated on problems with unrestricted alpine transit by lorries, environmental issues caused by (smaller) cars may be perceived as less dangerous in contrast. Industrial pollution was also to the forefront of the public consciousness in Great Britain, with a toxic leak from an ICI plant in Runcorn, Cheshire in 1999 and 2000 caused over 200 homes to be permanently evacuated. Both events may have contributed towards a raising of the public consciousness of this item. The higher relative danger of "modifying genes of certain crops" in Austria may be due to a campaign in 1997 supported by many environmental groups and the major Austrian newspapers, when 21% of the eligible voters signed a petition for a referendum against the cultivation and growing of genetically modified organisms and food. Accordingly, legal regulations were established thereafter.

Age and sex differences: Differences in perception of relative dangerousness over age groups and between sexes, in contrast, are unlikely to be caused by national events. In fact, the picture is complex. From the "main effect" parameter estimates in Table 3, we see that young females are more likely than young males to perceive genetic modification as a danger, and less likely to perceive rising temperatures as dangerous. The main age effects show little difference between the two older male age groups; between the youngest male group and the two older male groups we see that the younger generation are more likely to perceive rising temperatures, industrial pollution and water pollution as dangerous, and less likely to perceive genetic modification as dangerous.

Finally, we identify from the interaction parameters in Table 4 that the major effect is that older females are more likely than younger females to perceive rising temperatures as dangerous.

Table 4. Final model: parameters of interest; part 2

parameter	estimate	s.e.	z-value	p-value
sex(fem).age(2).CAR	0.085	0.055	1.54	0.1231
sex(fem).age(3).CAR	0.099	0.057	1.74	0.0810
sex(fem).age(2).IND	0.062	0.058	1.06	0.2872
sex(fem).age(3).IND	0.010	0.060	0.17	0.8623
sex(fem).age(2).FARM	0.135	0.061	2.22	0.0266
sex(fem).age(3).FARM	0.128	0.063	2.04	0.0418
sex(fem).age(2).WATER	0.108	0.057	1.88	0.0596
sex(fem).age(3).WATER	0.186	0.059	3.13	0.0017
sex(fem).age(2).TEMP	0.228	0.060	3.80	0.0001
sex(fem).age(3).TEMP	0.198	0.062	3.21	0.0013
sex(fem).age(2).GENE	0.000	–	–	–
sex(fem).age(3).GENE	0.000	–	–	–
urb(suburb).edu(A-level+).CAR	-0.120	0.058	-2.08	0.0377
urb(rural).edu(A-level+).CAR	-0.095	0.067	-1.41	0.1581
urb(suburb).edu(A-level+).IND	0.005	0.061	0.08	0.9350
urb(rural).edu(A-level+).IND	0.009	0.071	-0.12	0.9031
urb(suburb).edu(A-level+).FARM	-0.191	0.064	-2.99	0.0028
urb(rural).edu(A-level+).FARM	-0.146	0.074	-1.96	0.0505
urb(suburb).edu(A-level+).WATER	-0.123	0.060	-2.04	0.0410
urb(rural).edu(A-level+).WATER	-0.131	0.070	-1.87	0.0616
urb(suburb).edu(A-level+).TEMP	0.038	0.063	0.60	0.5502
urb(rural).edu(A-level+).TEMP	-0.135	0.073	-1.84	0.0654
urb(suburb).edu(A-level+).GENE	0.000	–	–	–
urb(rural).edu(A-level+).GENE	0.000	–	–	–

In summary, therefore, we can identify that the two older age groups show little differences between themselves, and few differences between males and females. Younger females however differ from younger males, who in turn differ from older people of either sex.

It is hard to explain this difference in perception between young males and young females. Perhaps the gender bias of science education could be a partial reason; young females are more likely to question science, both in the desirability of genetic modification , and in the prediction of future disaster from rising temperatures. What is of interest, however, if this is true, is why the male-female dichotomy disappears to a great extent in later life.

Education and location differences: As so many of the interaction terms are significant, interpretation is best considered by careful examination of Figures 1 and 2. The more highly educated group appears to rate relatively lower the dangerousness of genetic modification and to rate relatively higher the dangerousness of rising temperatures apart from in rural areas, where rising temperatures are seen as not so much of a problem. The location effect is again focused on genetic modification - in both countries there is an increase in relative dangerousness as as we move from city to suburban to rural. The main interaction effects can be identified in the educated suburban group, who are relatively more likely to rank rising temperatures as high compared to other groups.

A partial explanation for this might be that the rural population are more concerned with both the quality and safety of food production compared with the more urban residents, although this concern is less amongst the better educated. In contrast, the concerns of the more urban and educated group relate more to global concerns, and in particular global warming.

Fig. 1. Item worths for **Austria**: Groups are defined by age and gender, for combinations of location of residence and educational level (a) – (f)

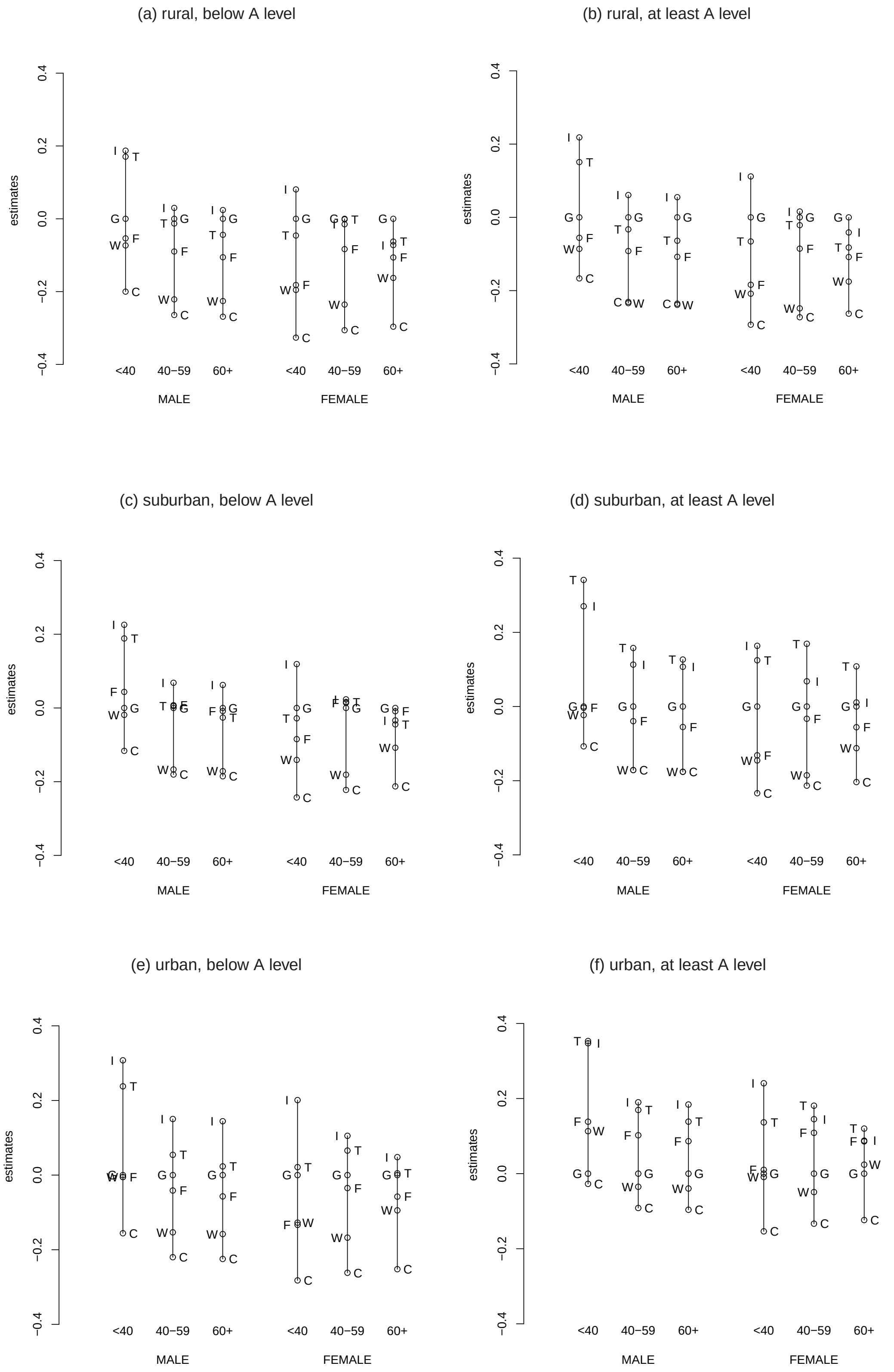
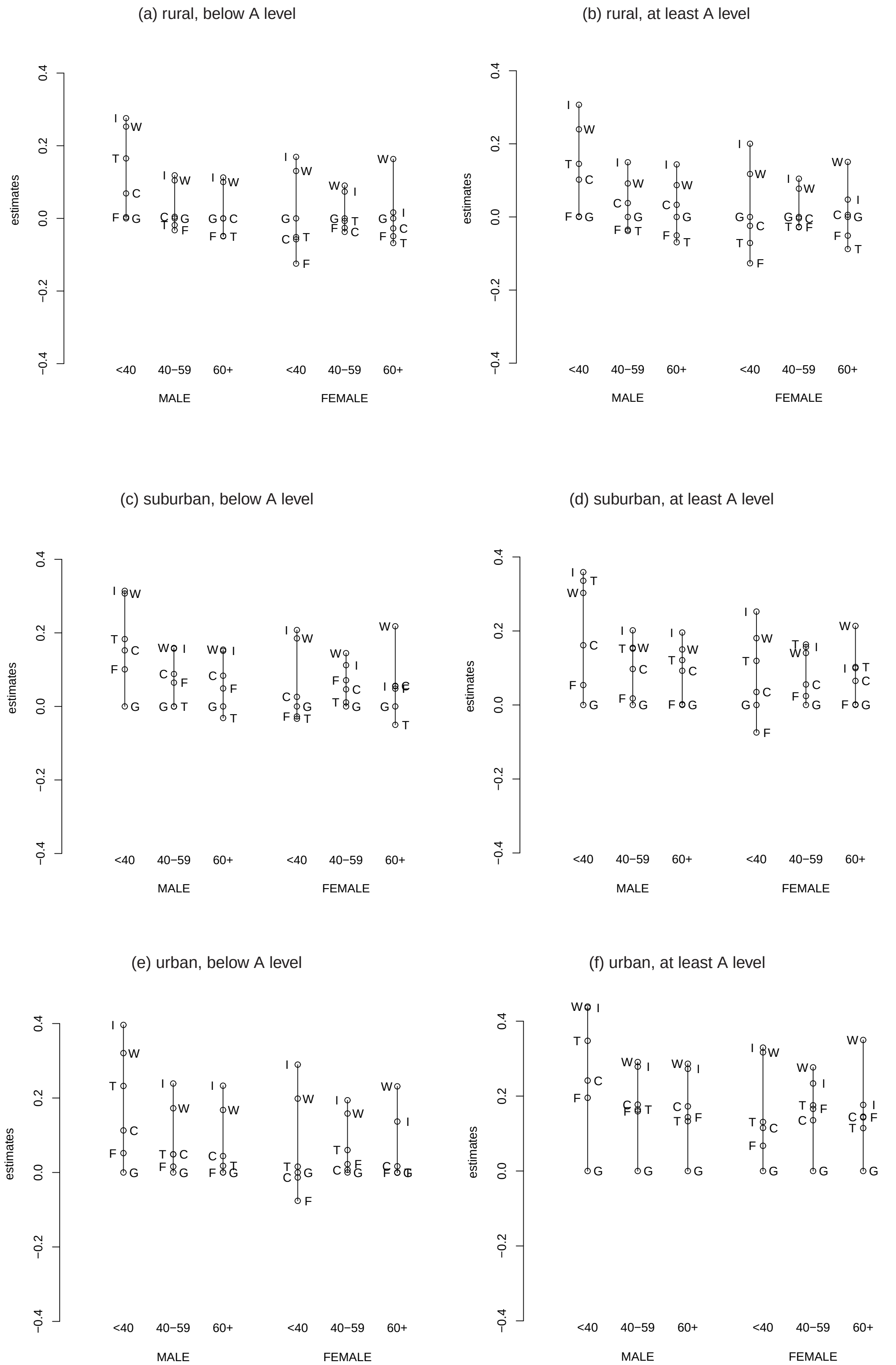


Fig. 2. Item worths for **Great Britain**: Groups are defined by age and gender, for combinations of location of residence and educational level (a) – (f)



5. Discussion

This paper has demonstrated that alternative models for sets of Likert items can be constructed which relax the assumptions of Normality. The paired comparison approach has a number of advantages. First, the model belongs to the class of Generalised Linear Models, and the parameters can therefore be estimated by standard software using the Poisson-multinomial equivalence to fit models using a Poisson distribution and log link. Covariates can be included in the model at the cost of some increase in the size of the problem. Usual model-bulding strategies based on the deviance or AIC can be used. Moreover, the nature of the model means that low-order dependencies can easily be included. Higher-order dependencies can as well be considered in the usual way as higher order interactions between the columns of the $\mathbf{W}$ matrix. Moreover it can be checked if these dependencies are needed in the model by comparing deviance differences.

The model can be extended in several ways. Firstly, we could use the full form of the AC-model and analyse ordered paired comparisons rather than restricting the paired comparison response to three categories. This will use more information in the data, at the cost of increasing the complexity of the model and losing the simplicity of the more basic paired comparison model. The ordered paired comparison model when used for likert data has an underlying assumption of equal distances between categories and this may be an undesirable characteristic. Secondly, it would be possible to relax the assumption of the need for complete responses by developing models which include a missing value indicator. However, this again would increase the size of the problem to be analysed.

Practically, the model produces useful results which can be interpreted by political scientists and sociologists. In theorising reasons for differences in environmental attitudes, various competing hypotheses have been put forward. Thus, Guha(2000) summarised two reasons for differences in environmental attitudes between countries. One theme relates these to country differences in post-materialist values (Inglehart, 1995)- that is, those values concerned with ethics and quality of life; and paired comparison methodology provides a way forward in analysing this concept (Francis et al, 2002). The other theme suggests that differences are related to poverty and real experiences of the environment. In this second model, it is the poorer countries which have greater concern for the environment, and citizens are concerned about existing natural resource-based problems rather than future potential problems which are not observable. Brechin(1999) considered this debate as too simplistic, and viewed environmental attitudes as a complex social phenomenon, with views formed by local environmental perceptions. It is interesting to observe in our analysis that in Great Britain it is concerns about industrial pollution and water quality -natural resource issues- which are mostly to the forefront, whereas in Austria, concern about industrial pollution is as much as important as concerns about the future (rising temperatures and genetic modification). This comparison of two western European countries tends to support the Brechin view of environmental attitudes.

In terms of covariates, the concentration has been on absolute differences. Hayes(2001), for example, analysed the 1993 ISSP data, and found few differences between men and women in their attitudes to the environment. Carriere and Scruggs (2001), in contrast, found strong differences in level of attitude to environmental risks between gender, income and urban/rural groups. The method proposed here, with its focus on the relative ordering of items, provides an alternative insight into the differences of emphasis of different environmental problems, and gives an alternative viewpoint on attitudinal data in this area.

More generally, the method provides a relatively assumption-free analysis which is useful for social surveys, particularly cross-national studies, where absolute differences in attitude may not be meaningful but where the assessment of relative differences are of real interest.

6. References

Aby, J. S. , Pheley, A. M. and Steinberg, P. (1996). Motivation for participation in clinical trials of drugs for the treatment of asthma, seasonal allergic rhinitis, and perennial nonallergic rhinitis. *Annals of Allergy, Asthma, and Immunology*, **76**, 348-354.

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

Agresti, A. (1992). Analysis of ordinal paired comparison data. *Applied Statistics*, **41**, 287-297.

Böckenholt, U. and W. Dillon (1997). Modelling within-subject dependencies in ordinal paired comparison data. *Psychometrika*, **62**, 411-434.

Böckenholt, U. (2004). Comparative judgements as an alternative to ratings: Identifying the scale orgigin. *Psychological Methods*, **9**, 453-465.

Bradley, R. A. and M. E. Terry (1952). Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika*, **39**, 324-345.

Brady, H. E. (1989). Factor and ideal point analysis for interpersonally incomparable data. *Psychometrika*, **54**, 181-202.

Brechin, S. R. (1999) Objective problems, subjective values and global environmentalism: evaluating the postmaterialist argument and challenging a new explanation. *Social Science Quarterly* **80(4)** 793-809.

Carriere, E. and L. Scruggs (2001) A Cross-national Study of the Cultural Sources of Environmental Attitudes: Evidence from the 2000 ISSP. Unpublished.
`http://www.sp.uconn.edu/~scruggs/csenvatt.pdf`

Cox, D. R. (1972) The analysis of multivariate binary data. *Applied Statistics*, **21**, 113-120

Cox, D. R. and Wermuth, Nanny (1994) A note on the quadratic exponential binary distribution. *Biometrika*, **81**, 403-408

Critchlow, D. E. and M. A. Fligner (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, **56**, 513-533

Culp, K. and Schwartz, V. J. (1999). Motivating Adult Volunteer 4-H Leaders. *Journal of Extension*, **37** (1) `http://joe.org/joe/1999february/rb5.html`.

Dalton, R. and R. Rohrschneider (1998). The greening of Europe. In: Jowell, R., Curtice, J., Park, A., Brook, L., Thomson, K. and Bryson, C. (eds.) *British and European Social Attitudes: How Britain Differs. The 15th BSA Report*. Ashgate Publishing, Aldershot.

Davidson, R. R. (1970). Extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *J.Am.Stat.Ass.*, **65**, 317-328.

Denz, H. (2000). Materialismus - Postmaterialismus. Ranking oder Rating. *Österreichische Zeitschrift für Soziologie*, **25**, 79-94.

Dittrich, R., R. Hatzinger and W. Katzenbeisser (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Applied Statistics*, **47**, 511-525.

Dittrich, R., R. Hatzinger and W. Katzenbeisser (2002). Modelling dependencies in paired comparison data. A log linear approach. *Computational Statistics and Data Analysis*, **40**, 39-57.

Dittrich, R., R. Hatzinger and W. Katzenbeisser (2004). A log-linear approach for modelling ordinal paired comparison data on motives to start a PhD programme. *Statistical Modelling*, **4**, 181-193

Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests.* Hans Huber, Bern Stuttgart Wien.

Francis, B., R. Dittrich, R. Hatzinger and R. Penn (2002). Analysing partial ranks by using paired comparison methods: an investigation of value orientation in Europe. *Applied Statistics*, **51(3)**, 319-336.

Graham, R. L., D. E. Knuth, and O. Patashnik (1989). *Concrete Mathematics.* Addison-Wesley, Reading, Mass.

Grizzle, J. F., C. F. Starmer, and G. G. Koch (1969). Analysis of categorical data by linear models *Biometrics*, **25**, 489-504.

Guha, R. (2000) *Environmentalism: a global history.* Longman: New York

Hayes, B.C. (2001) Gender, scientific knowledge and attitudes towards the environment: a cross-national analysis. *Political Research Quarterly*, **54(3)**, 657-671.

Inglehart, R. (1995) Public support for environmental protection: objective problems and subjective values in 43 societies *Political Science and Politics*, **28(1)**, 57-72

Kemp, S. and C. D. B. Burt (2002). Altruism in valuing government and market-supplied goods. Journal of Socio-Economics, **31**, 167-169.

Mallows, C. L. (1957). Non-null ranking models: I. *Biometrika*, **44**, 114-130.

Maydeu-Olivares, A. (2002). Limited information estimation and testing of Thurstonian models for preference data. *Mathematical Social Sciences*, **43**, 467-483.

O'Hara, S. U. and S. Stagl (2002). Endogenous preferences and sustainable development. *Journal of Socio-Economics*, **31**, 511-527.

Sinclair, C. D. (1982). GLIM for preference. In: Gilchrist, R. (ed.): *GLIM 82. Proceedings of the International Conference on Generalised Linear Models*, Springer Lecture Notes in Statistics, **14**, 164-178.

Sloane, N.J.A. and S. Plouffe (1995) *The Encyclopedia of Integer Sequences*, Academic Press.

Tutz, G. (1990) *Modelle für kategoriale Daten mit ordinalem Skalenniveau*, Vandenhoeck & Ruprecht .

Uesaka, H., and C. Asano (1987). Latent scale linear models for multivariate ordinal responses and analysis by the method of weighted least squares *Ann. Inst. Statist. Math.*, **39**, 191-210.

Uesaka, H., and C. Asano (1987). Analysis of ordered categorical data from repeated measurements assuming a quantitative latent variable *Ann. Inst. Statist. Math.*, **39**, 513-531.

Wilkinson, G. N. and C. E. Rogers (1973) Symbolic description of factorial model for analysis of variance. *Applied Statistics*, **22**, 392-399.

Witherspoon, S. (1994). The Greening of Britain: Romance and Rationality. In: Jowell, R., Cutice, J., Brook, L., Ahrendt, D. and Park, A. (eds.) *British Social Attitudes: the 11th Report* Dartmouth Publishing, Aldershot.

7. Appendix A

Consider the simple example of three items $i_1\, i_2\, i_3$ rated on a 3-point Likert scale. There are 3^3 possible Likert-patterns given in the first column of Table 5 but many of these Likert-patterns are mapped into the same PC-pattern. For example the PC-pattern (0,1,1) is the result of forming pairwise differences from Likert responses with the property that the first two items are rated equally (denoted by $\underbrace{i_1\, i_2}$ in column 4) and better than the third item (denoted by $\underbrace{\ i_3\ }$). This can be interpreted that the set of three items is partitioned into two nonempty subsets, where the first subset contains the (equally but) better rated item(s) and the second subset contains the item(s) rated worse (but equally). There are $\left\{ {3 \atop 2} \right\}$ ways to partition a set of three items into two subsets. Because the order of the subsets is essential we have to permute all possible partitions.

In general the number of unique PC-patterns can be explained by the following considerations: The set of J items is partitioned into $\nu = 0, 1, 2, \ldots, \kappa$ subsets and the number of those partitions is given by $\left\{ {J \atop \nu} \right\}$, where κ denote the number of points on the Likert scale. Within each subset the items are rated equally but the subsets are ordered in the sense that the items in the first subset are rated better than the items in the second subset and these are rated better than items contained in the third subset and so on. Because the order of the partitions is essential, we have to permute those, which can be done in $\nu!$ ways. Summation over all possibilities yields the number of unique PC-responses which in this case is an ordered Bell number.

Table 5. Illustrating the number of unique PC-patterns for $J = 3$ and $\kappa = 3$

Likert-pattern	number of Likert-patterns with same PC-pattern	unique PC-pattern	ordered partition	number of PC-patterns with certain partition
(111), (222), (333)	3	0 0 0	$\underbrace{i_1\, i_2\, i_3}$	${3 \brace 1} = 1$
(112), (113), (223)	3	0 1 1	$\underbrace{i_1\, i_2}\ \underbrace{i_3}$	
(221), (331), (332)	3	0 -1 -1	$\underbrace{i_3}\ \underbrace{i_1\, i_2}$	
(121), (131), (232)	3	1 0 -1	$\underbrace{i_1\, i_3}\ \underbrace{i_2}$	
(212), (313), (323)	3	-1 0 1	$\underbrace{i_2}\ \underbrace{i_1\, i_3}$	${3 \brace 2} \times 2 = 6$
(211), (311), (322)	3	-1 -1 0	$\underbrace{i_2\, i_3}\ \underbrace{i_1}$	
(122), (133) (233)	3	-1 -1 0	$\underbrace{i_1}\ \underbrace{i_2\, i_3}$	
(123)	1	1 1 1	$\underbrace{i_1}\ \underbrace{i_2}\ \underbrace{i_3}$	
(132)	1	1 1 -1	$\underbrace{i_1}\ \underbrace{i_3}\ \underbrace{i_2}$	${3 \brace 3} \times 3! = 6$
⋮	⋮	⋮	⋮	
Total	27			13 = B(3)