MULTILEVEL LOGISTIC REGRESSION FOR POLYTOMOUS DATA AND RANKINGS

ANDERS SKRONDAL

DIVISION OF EPIDEMIOLOGY
NORWEGIAN INSTITUTE OF PUBLIC HEALTH, OSLO

SOPHIA RABE-HESKETH

DEPARTMENT OF BIOSTATISTICS AND COMPUTING
INSTITUTE OF PSYCHIATRY, LONDON

We propose a unifying framework for multilevel modeling of polytomous data and rankings, accommodating dependence induced by factor and/or random coefficient structures at different levels. The framework subsumes a wide range of models proposed in disparate methodological literatures. Partial and tied rankings, alternative specific explanatory variables and alternative sets varying across units are handled. The problem of identification is addressed. We develop an estimation and prediction methodology for the model framework which is implemented in the generally available gllamm software. The methodology is applied to party choice and rankings from the 1987–1992 panel of the British Election Study. Three levels are considered: elections, voters and constituencies.

Key words: multilevel models, generalized linear latent and mixed models, factor models, random coefficient models, polytomous data, rankings, first choice, discrete choice, permutations, nominal data, gllamm.

Introduction

Two kinds of nominal outcome variable are considered in this article; unordered polytomous variables and permutations. The outcome for a unit in the unordered polytomous case is one among several objects, whereas the outcome in the permutation case is a particular ordering of objects. The objects are nominal in the sense that they do not possess an inherent ordering shared by all units as is assumed for ordinal variables. The development in this paper is coaxed in the terminology of decision theory. Hence, objects are henceforth denoted *alternatives*, unordered polytomous variables denoted *first choices* and permutations denoted *rankings*. For instance, in election studies a central outcome variable is the first choice of a voter (say Conservatives) among a set of alternatives (say Labour, Conservatives and Liberals). Sometimes additional information is obtained in the form of rankings of the alternatives (say Liberals preferred to Labour preferred to Conservatives).

The standard statistical model for first choices and rankings is logistic regression. It has been pointed out in the econometric and psychometric literature that these models involve a questionable independence assumption known as "Independence from Irrelevant Alternatives" (IIA). Discussions of the ramifications of IIA have been confined to one-level designs. However, first choice and ranking data are often of a multilevel nature where units are nested within clusters. In the context of the election example, multilevel data could for instance arise from two-level designs where voters are nested within constituencies or election occasions nested within vot-

ers. The liability of the IIA assumption is compounded for multilevel designs because additional dependence, between units within clusters, is typically induced by unobserved heterogeneity between clusters.

Even when potential dependence is acknowledged, the dependence structures considered for polytomous data and rankings are often too restrictive. As far as we are aware a truly multi-level model with unobserved heterogeneity at more than one level has not yet been considered in the literature. For instance, Goldstein (1995) specifies a model for panel election data including unobserved heterogeneity between constituencies but not between voters in a constituency. Furthermore, first choice models considered in the multilevel literature have neither accommodated alternative specific explanatory variables, alternative sets varying across units nor factor structures (compare Goldstein, 1995; Raudenbush & Bryk, 2002). Contributions on rankings appear to be missing in the multilevel literature.

The main contribution of this paper is to provide a unifying framework for multilevel modeling of first choices and rankings, accommodating dependence at unit and cluster levels induced by different kinds of factor and random coefficient structures. We also consider partial and tied rankings, alternative specific explanatory variables as well as alternative sets varying across units. The framework subsumes a wide range of models proposed in the psychometric, econometric, (bio)statistical and marketing literatures. We discuss the problem of identification and develop an estimation and prediction methodology which is implemented in the generally available gllamm software. The methodology is finally applied to British election panel data. A three-level design is considered where elections are nested within voters and voters nested within constituencies.

Logistic Regression for First Choice and Ranking

In the following, we will use lower case letters for scalars, bold lower case letters for column vectors and bold upper case letters for matrices. Let $\mathcal{A}_i = \{a_i^1, \ldots, a_i^{A_i}\}$ be the set of A_i alternatives available for unit i and v_i^a a linear predictor for unit i and alternative a.

First Choice

Letting c_i be the first choice, a logistic model for first choices (e.g., Bock, 1969; Gurland, Lee & Dolan, 1960; Theil, 1969) is usually specified as

$$\Pr(c_i \mid \mathcal{A}_i) = \frac{\exp(v_i^{c_i})}{\sum_{s=1}^{A_i} \exp(v_i^s)}. \tag{1}$$

An alternative specification of the logistic regression model, based on random utility models (e.g., Block & Marchak, 1960; Luce & Suppes, 1965; Marchak, 1960), is often used in econometrics (e.g., Maddala, 1983). Here, the model is usually derived by introducing utilities u_i^a for each unit i and alternative a. The utilities are modeled as

$$u_i^a = v_i^a + \epsilon_i^a,$$

where v_i^a is a fixed linear predictor and ϵ_i^a is a random term, assumed to be independently distributed across both units and alternatives with density

$$g(\epsilon_i^a) = \exp\{-\epsilon_i^a - \exp(-\epsilon_i^a)\}. \tag{2}$$

This distribution has many names (e.g., Johnson, Kotz, & Balakrishnan, 1995), but will in the sequel be denoted as Gumbel.

The probability of a first choice can be construed as *utility maximization* and expressed in terms of $A_i - 1$ binary utility comparisons where the utility of the chosen alternative $u_i^{c_i}$ is larger than the utilities of all other alternatives. Under the Gumbel specification for the random utility this leads to the logistic regression model (1). The result that differences between Gumbel

distributed variables (with constant variances) have logistic distributions is attributed to Holman and Marley (see Luce & Suppes, 1965). The stronger result that only the Gumbel specification leads to the closed form logistic model is due to McFadden (1973) and Yellott (1977).

Ranking

Let r_i^ℓ be the alternative given rank ℓ and $\mathbf{R}_i \equiv (r_i^1, r_i^2, \ldots, r_i^{A_i})$ the ranking of unit i. The logistic model for rankings (e.g., Luce, 1959; Plackett, 1975) is then specified as

$$\Pr(\mathbf{R}_i \mid \mathcal{A}_i) = \prod_{\ell=1}^{A_i-1} \frac{\exp(v_i^{r_i^\ell})}{\sum_{s=\ell}^{A_i} \exp(v_i^{r_i^s})}. \tag{3}$$

The model is often denoted the "exploded logit" (Chapman & Staelin, 1982) since the ranking probability is written as a product of first choice probabilities for successively remaining alternatives.

The probability of a ranking can be construed as *utility ordering* and expressed in terms of $A_i - 1$ binary utility comparisons, where the utility of the alternative ranked first is larger than the utility of that ranked second, which is larger than that ranked third and so on. Under the Gumbel specification of random utility this leads to the closed form "exploded logit" specification (3). That such an explosion results was proven by Luce and Suppes (1965) and Beggs, Cardell and Hausman (1981). The explosion is not simply an application of conditional probabilities but follows from the so called Luce-decomposability of Gumbel models (e.g., Critchlow, Fligner, & Verducci, 1991). Importantly, an analogous explosion is not obtained under normally distributed utilities. The Gumbel model is not reversible in the sense that successive choices starting with the worst alternative would lead to a different ranking probability. It is also worth noting that the ranking probability given above has the same form as the partial likelihood contribution for a stratum in Cox regression with "surviving" alternatives corresponding to risk sets and choices corresponding to failures. Allison and Christakis (1994) spell out dualities between logistic models for rankings and Cox regression.

Partial rankings result when unit i only ranks a subset of the full set of alternatives; for example, when experimental designs are used in presenting specific subsets of alternatives to different units (e.g., Böckenholt, 1992; Durbin, 1951). Such designs are easily handled by the present methodology since the alternative sets $\mathcal{A}_i$ are permitted to vary over units i. Another kind of partial ranking is a top-ranking where not all alternatives are ranked but only the subset of the $P_i < A_i$ most preferred alternatives. The probability of a top-ranking is simply the product of the first P_i terms in equation (3). Note that the first choice probability is obtained as the special case of the ranking probability when $P_i = 1$ for all i.

Tied alternatives are given the same rank. Although the probability of tied rankings is theoretically zero since the utilities are continuous, ties are often observed in practice. Exploiting the duality between logistic models for rankings and Cox regression, we can utilize methods for handling ties previously suggested in the survival literature. We hence assume that the units have preference orderings for the tied alternatives, but the ordering is hidden from us. A unit i produces $P_i^* < A_i$ ranks when there are ties. We let t_i^ℓ denote the number of alternatives tied at rank ℓ for the unit, give the tied alternatives arbitrary labels $m = 1, \ldots, t_i^\ell$ and define $v_i^{r_i^{\ell[m]}}$ to be the linear predictors of the tied alternatives. The exact expression for the ranking probability is very complex when there are ties (e.g., Kalbfleisch & Prentice, 1980). Following Breslow (1974) we suggest using the following approximation:

$$\Pr(\mathbf{R}_i \mid \mathcal{A}_i) = \prod_{\ell=1}^{P_i^*-1} \prod_{m=1}^{t_i^\ell} \frac{\exp(v_i^{r_i^{\ell[m]}})}{\sum_{\ell'=\ell}^{P_i^*} \sum_{m'=1}^{t_i^{\ell'}} \exp(v_i^{r_i^{\ell'[m']}})}.$$

This approximation amounts to assuming that all tied alternatives are still available when any of the tied alternatives are chosen. Breslow's method appears to work well as long as the number of ties is moderate (Farewell & Prentice, 1980). Note that no approximation is required if there is only one set of tied alternatives and these alternatives have the lowest rank P_i^*, since this represents a top-ranking.

Heterogeneity and "Independence from Irrelevant Alternatives"

It follows from logistic first choice models that the odds for two alternatives a and b for unit i are

$$\frac{\Pr(c_i = a \mid \mathcal{A}_i)}{\Pr(c_i = b \mid \mathcal{A}_i)} = \exp(v_i^a - v_i^b),$$

which only depend on the linear predictors for the two alternatives. Hence, the odds do not depend on which other alternatives there are in the alternative set $\mathcal{A}_i$. In the ranking case it furthermore follows that the odds do not depend on which other alternatives have already been chosen, the number of alternatives already chosen or the order in which those alternatives were chosen. Luce (1959) denoted this property "Independence from Irrelevant Alternatives" (IIA).

The problem associated with the IIA property is best illustrated by adapting the red bus–blue bus example of McFadden (1973). Let there be three political parties, Labour I, Labour II and Conservatives. The first two parties are indistinguishable and have the same linear predictor v_i^a whereas the Conservative party has linear predictor v_i^b. The probability of voting for either Labour I or Labour II is

$$\frac{2\exp(v_i^a)}{2\exp(v_i^a) + \exp(v_i^b)} = \frac{1}{1 + \exp(v_i^b - v_i^a - \ln(2))}.$$

Consider now the scenario that Labour I and Labour II merge to form a single Labour party. According to the logistic model the probability of voting Labour now decreases to

$$\frac{\exp(v_i^a)}{\exp(v_i^a) + \exp(v_i^b)} = \frac{1}{1 + \exp(v_i^b - v_i^a)}.$$

This effect is clearly counterintuitive and the model is often regarded as unduly restrictive (e.g., Takane, 1987, and the references therein).

The ramifications of IIA are most pronounced for an indifferent voter with $v_i^b - v_i^a = 0$. In this case the merger reduces the probability of voting Labour from 0.67 to 0.50, consistent with an equiprobable choice among initially three and then two available parties. However, in reality almost all voters will have a party preference reducing the impact of IIA. For example, the probability decreases less substantially due to the merger, from 0.94 to 0.88, for a Labour supporter with $v_i^b - v_i^a = -2$.

More importantly, there will be heterogeneity in party preference among voters. This heterogeneity, observed and unobserved, ensures that the model does not imply a substantial reduction in the share of the Labour vote due to the merger. To illustrate the effect of observed heterogeneity, consider a fixed effect of gender giving $v_i^b - v_i^a = 2$ for men and $v_i^b - v_i^a = -2$ for women. For a population consisting of 50% men, the merger reduces the marginal probability for Labour from 0.57 to 0.50 (compared with 0.67 to 0.50 in a homogeneous, indifferent population) and, marginally to the observed covariates, IIA is therefore violated. In practice, observed covariates cannot explain all variability in individual party preferences. The remaining unobserved heterogeneity can be modeled by including a shared random effect for the two Labour parties in the linear predictor. For example, consider the situation where the random effect is normally distributed with variance 16 (yielding a correlation between the Labour utilities of 0.91) and the marginal (with respect to the random effect) probabilities of voting Labour remain the same as

before for men and women. In this case the merger only reduces the probability of voting Labour from 0.53 to 0.50. Increasing the magnitude of the fixed effects or the random effects variances decreases the difference even further.

A consequence of failing to include random effects would be that the unrealistic IIA property still holds conditional on the observed covariates, for example, among men and among women if gender is the only observed covariate. Even with voter specific random effects, IIA does, of course, still hold for a given voter. This conditional IIA property could be relaxed by including voting-occasion specific random effects if the resulting model is identified.

A model failing to include random effects is obviously misspecified if there are substantial correlations between the utilities within a cluster conditional on the observed covariates. It is well known that model-based standard errors for the parameter estimates will in this case be incorrect. Another consequence of failing to include random effects is that the fixed effects estimates will be attenuated towards zero. This dilution reflects a difference in interpretation; models without random effects provide marginal or population-averaged effect estimates whereas random effects models provide conditional or subject-specific effect estimates (e.g., Neuhaus, 1992). We argue that inference regarding etiology should be confined to conditional effects since these are "controlled" for unobserved as well as observed covariates. In fact, a major virtue of multilevel designs (for instance panel designs) is that unobserved heterogeneity can be modeled, improving the prospects of valid causal inferences (e.g., Hsiao, 1985).

A Model Framework for Multilevel Logistic Regression

For simplicity, we confine the explicit development of our multilevel framework to two-level designs. However, it is straightforward to extend our methodology to an arbitrary number of levels. This is illustrated in our application section where a three-level model is used for British election data.

Let there be J clusters with index $j = 1, \ldots, J$ at Level 2. At Level 1 there are N_j units nested within each cluster j indexed $i = 1, \ldots, N_j$. $\mathcal{A}_{ij} = \left\{a_{ij}^1, \ldots, a_{ij}^{A_{ij}}\right\}$ is the set of A_{ij} alternatives for unit i in cluster j, and the total number of alternatives is denoted A. For the election example, we let i index voters, j constituencies and let $\mathcal{A}_{ij}$ be the set of political parties eligible for voter i in constituency j.

Utility Components

We will now specify parametric models for the unobserved random utilities. Initially, the utilities are decomposed as

$$u_{ij}^a = f_{ij}^a + \delta_{ij}^a + \epsilon_{ij}^a.$$

f_{ij}^a is a deterministic fixed effects term, representing observed heterogeneity of alternatives, units and clusters. δ_{ij}^a are latent variables, representing unobserved heterogeneity that is *dependent* over alternatives, further decomposed as

$$\delta_{ij}^a = \delta_{ij}^{a(1)} + \delta_{ij}^{a(2)},$$

where $\delta_{ij}^{a(1)}$ and $\delta_{ij}^{a(2)}$ are latent variables inducing dependence at the unit and cluster levels, respectively. ϵ_{ij}^a is an alternative specific error term representing unobserved heterogeneity that is *independent* across alternatives, units and clusters. Finally, we define the linear predictor as

$$v_{ij}^a = f_{ij}^a + \delta_{ij}^a.$$

Fixed Effects

We let the fixed effects f_{ij}^a be structured as

$$f_{ij}^a = m^a + \mathbf{x}_{ij}^{a\prime} \mathbf{b} + \mathbf{x}_{ij}' \mathbf{g}^a. \tag{4}$$

Here, m^a is a constant, $\mathbf{x}_{ij}^a$ a vector which varies over alternatives and may also vary over units and/or clusters whereas the vector $\mathbf{x}_{ij}$ varies over units and/or clusters but not alternatives. The corresponding fixed coefficient vectors are $\mathbf{b}$ and $\mathbf{g}^a$, respectively. Note that the effect $\mathbf{b}$ is assumed to be the same on all utilities, so that this part of the model simply represents a linear relationship between the utilities and alternative (and possibly unit and/or cluster)-specific covariates. For some alternative and unit and/or cluster specific variables the effect may differ between alternatives. Such effects can be accommodated by including interactions between these variables and dummy variables for the alternatives in $\mathbf{x}_{ij}^a$. In the election example, the covariate x_{ij} could be the age of voter i in constituency j with different effects g^a on the utilities for different parties. The party-specific covariate x_{ij}^a could be a measure of the distance on the left-right political dimension between the ath party and the ith voter in constituency j.

Unit Level Latent Variables

The component $\delta_{ij}^{a(1)}$, inducing dependence among utilities within units, is structured as

$$\delta_{ij}^{a(1)} = \mathbf{z}_{ij}^{a(1)\prime} \boldsymbol{\beta}_{ij}^{(1)} + \boldsymbol{\lambda}^{a(1)\prime} \boldsymbol{\eta}_{ij}^{(1)}. \tag{5}$$

Here $\boldsymbol{\beta}_{ij}^{(1)}$ is a random vector with corresponding explanatory variables $\mathbf{z}_{ij}^{a(1)}$. If $\mathbf{z}_{ij}^{a(1)}$ is a subset of $\mathbf{x}_{ij}^a$, $\boldsymbol{\beta}_{ij}^{(1)}$ are random coefficients allowing the effects of alternative specific covariates to vary between units i. In the election example, letting $z_{ij}^{a(1)} = x_{ij}^{a(1)}$ represent the distance on the left-right political dimension, the random slope $\beta_{ij}^{(1)}$ allows the distance effect to vary between voters. Finally, $\boldsymbol{\eta}_{ij}^{(1)}$ are factors at the unit level, representing unobserved variables having effects $\boldsymbol{\lambda}^{a(1)}$ on u_{ij}^a. Alternatively, $\boldsymbol{\lambda}^{a(1)}$ can be interpreted as unobserved attributes of alternative a and $\boldsymbol{\eta}_{ij}^{(1)}$ as random effects on the utilities.

The expectations of the latent variables can be structured as

$$\mathrm{E}[\boldsymbol{\beta}_{ij}^{(1)} \mid \mathbf{w}_{ij}] = \mathbf{G}_{\boldsymbol{\beta}}^{(1)} \mathbf{w}_{ij},$$

$$\mathrm{E}[\boldsymbol{\eta}_{ij}^{(1)} \mid \mathbf{w}_{ij}] = \mathbf{G}_{\boldsymbol{\eta}}^{(1)} \mathbf{w}_{ij}, \tag{6}$$

where $\mathbf{G}$ are regression parameter matrices and $\mathbf{w}_{ij}$ exogenous covariates. Imposing a regression structure for the random coefficients $\boldsymbol{\beta}_{ij}^{(1)}$ is a convenient device for introducing interactions between alternative-specific and unit-specific variables in the expectation structure without making the fixed part of the model f_i^a appear too complex. For example, if the effect of the distance on the left-right political dimension depends on age, age can be used as a covariate in (6) instead of including an interaction term in (4). In contrast, the regression structure for $\boldsymbol{\eta}_{ij}^{(1)}$ allows the effect of unit-specific covariates $\mathbf{w}_{ij}$ to be modeled more parsimoniously than via f_{ij}^a. For a single factor, for example, one regression coefficient is estimated for each covariate, whereas inclusion of the covariates in f_{ij}^a would require estimation of $A - 1$ regression coefficients per covariate.

Cluster Level Latent Variables

Let $\delta_{ij}^{a(2)}$, inducing dependence among utilities within clusters, be structured as

$$\delta_{ij}^{a(2)} = \mathbf{z}_{ij}^{a(2)\prime} \boldsymbol{\beta}_j^{(2)} + \boldsymbol{\lambda}^{a(2)\prime} \boldsymbol{\eta}_j^{(2)} + \mathbf{z}_{ij}^{(2)\prime} \boldsymbol{\gamma}_j^{a(2)}, \tag{7}$$

where $\boldsymbol{\beta}_j^{(2)}$ and $\boldsymbol{\gamma}_j^{a(2)}$ are vectors of random coefficients with corresponding variable vectors $\mathbf{z}_{ij}^{a(2)}$ and $\mathbf{z}_{ij}^{(2)}$. $\boldsymbol{\beta}_j^{(2)}$ allows the effects of alternative specific covariates to vary between clusters j, whereas $\boldsymbol{\gamma}_j^{a(2)}$ represents the variability in the effect of unit specific covariates $\mathbf{z}_{ij}^{(2)}$ between clusters. Note that there is in this case a random coefficient for each alternative. In the context of the election example $\beta_j^{(2)}$ could represent heterogeneity in the distance effect between constituencies. $\boldsymbol{\gamma}_j^{a(2)}$ represent random intercepts when $\mathbf{z}_{ij}^{(2)} = \mathbf{1}$ and could more generally represent random coefficients (varying over constituencies) of a covariate that varies within constituencies (e.g., age). $\boldsymbol{\eta}_j^{(2)}$ are factors representing unobserved variables at the cluster level having effects $\boldsymbol{\lambda}^{a(2)}$ on u_{ij}^a.

The expectations of the latent variables can be structured as

$$\mathrm{E}[\boldsymbol{\beta}_j^{(2)} \mid \mathbf{w}_j] = \mathbf{G}_{\boldsymbol{\beta}}^{(2)}\mathbf{w}_j, \tag{8}$$

$$\mathrm{E}[\boldsymbol{\eta}_j^{(2)} \mid \mathbf{w}_j] = \mathbf{G}_{\boldsymbol{\eta}}^{(2)}\mathbf{w}_j, \tag{9}$$

$$\mathrm{E}[\boldsymbol{\gamma}_j^{a(2)} \mid \mathbf{w}_j] = \mathbf{G}_{\boldsymbol{\gamma}}^{a(2)}\mathbf{w}_j, \tag{10}$$

where $\mathbf{G}$ are regression parameter matrices and $\mathbf{w}_j$ exogenous covariates varying at the cluster level. Note that cluster and alternative specific covariates $\mathbf{w}_j^a$ could also be included in (10). For example, the mean of the cluster-level random effect $\gamma_j^{a(2)}$ of the voter's age $z_{ij}^{(2)}$ could depend on the age of the candidate of party a in constituency j, w_j^a. For notational simplicity we assume that such effects are specified in the fixed part (4).

The random terms at Level 1 in (5) and Level 2 in (7) are analogous except that there is no term corresponding to $\mathbf{z}_{ij}^{(2)\prime}\boldsymbol{\gamma}_j^{a(2)}$ at Level 1. If such a term were included at the unit level, the $\boldsymbol{\gamma}_{ij}^{a(1)}$ could not be interpreted as random effects of covariates since the covariates, being constant across alternatives, do not vary within units. Including a term $\mathbf{z}_{ij}^{(1)\prime}\boldsymbol{\gamma}_{ij}^{a(1)}$ in the model would result in a heteroscedastic unit and alternative specific error term. However, if one of the elements of $\mathbf{z}_{ij}^{(1)}$ is a constant so that the corresponding element of $\boldsymbol{\gamma}_{ij}^{a(1)}$ is a random intercept at level 1, model identification will be fragile (see Identification section).

The Framework in Vector Form

The utilities u_{ij}^a are stacked in the $A_{ij} \times 1$ column vector $\mathbf{u}_{ij}$. Analogously, we assemble $f_{ij}, \delta_{ij}^{a(1)}, \delta_{ij}^{a(2)}$ and ϵ_{ij}^a in $\mathbf{f}_{ij}, \boldsymbol{\delta}_{ij}^{(1)}, \boldsymbol{\delta}_{ij}^{(2)}$ and $\boldsymbol{\epsilon}_{ij}$, respectively. We can then write the model as

$$\mathbf{u}_{ij} = \mathbf{f}_{ij} + \boldsymbol{\delta}_{ij}^{(1)} + \boldsymbol{\delta}_{ij}^{(2)} + \boldsymbol{\epsilon}_{ij}.$$

The fixed effects vector $\mathbf{f}_{ij}$ can be expressed as

$$\mathbf{f}_{ij} = \mathbf{m} + \mathbf{X}_{ij}^{(1)}\mathbf{b} + (\mathbf{I}_{A_{ij}} \otimes \mathbf{x}_{ij}')\mathbf{g},$$

where the matrix $\mathbf{X}_{ij}^{(1)}$ contains the vectors $\mathbf{x}_{ij}^{a}{}'$ as rows. The constants m^a are placed in the column vector $\mathbf{m}$ and the parameter vectors $\mathbf{g}^a$ are stacked in the column supervector $\mathbf{g}$.

The unit level latent variables $\boldsymbol{\delta}_{ij}^{(1)}$ can be written as

$$\boldsymbol{\delta}_{ij}^{(1)} = \mathbf{Z}_{ij}^{(1)}\boldsymbol{\beta}_{ij}^{(1)} + \boldsymbol{\Lambda}_{[ij]}^{(1)}\boldsymbol{\eta}_{ij}^{(1)}.$$

Here, $\mathbf{z}_{ij}^{a(1)\prime}$ are the rows of the matrix $\mathbf{Z}_{ij}^{(1)}$ and $\boldsymbol{\lambda}^{a(1)\prime}$ are the rows of $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ where the index $[ij]$ signifies that the pertinent elements of the loading matrix may vary with ij since the alternative

set is indexed ij. However, the elements (parameters) for a given alternative remain constant across the ij. The cluster level latent variables $\boldsymbol{\delta}_{ij}^{(2)}$ can be written as

$$\boldsymbol{\delta}_{ij}^{(2)} = \mathbf{Z}_{ij}^{(2)}\boldsymbol{\beta}_j^{(2)} + \boldsymbol{\Lambda}_{[ij]}^{(2)}\boldsymbol{\eta}_j^{(2)} + \left(\mathbf{I}_{A_{ij}} \otimes \mathbf{z}_{ij}^{(2)\prime}\right)\boldsymbol{\gamma}_j^{(2)}.$$

$\mathbf{Z}_{ij}^{(2)}$ contains $\mathbf{z}_{ij}^{a(2)\prime}$ as rows, $\boldsymbol{\Lambda}_{[ij]}^{(2)}$ contains $\boldsymbol{\lambda}^{a(2)\prime}$ as rows and the stochastic vectors $\boldsymbol{\gamma}_j^{a(2)}$ are stacked in the column supervector $\boldsymbol{\gamma}_j^{(2)}$. The regression structure for $\boldsymbol{\gamma}_j^{a(2)}$, given in (10), must then be extended to:

$$\mathrm{E}[\boldsymbol{\gamma}_j^{(2)} \mid \mathbf{w}_j] = \mathbf{G}_{\boldsymbol{\gamma}}^{(2)}\mathbf{w}_j,$$

where the regression matrices $\mathbf{G}_{\boldsymbol{\gamma}}^{a(2)}$ are stacked in the supermatrix $\mathbf{G}_{\boldsymbol{\gamma}}^{(2)}$.

Distributional Specifications

The alternative specific errors ϵ_{ij}^a are specified as Gumbel without scale parameters as in (2). It can be shown that the error has expectation ϱ, where ϱ denotes Euler's constant, and variance $\pi^2/6$. We specify the errors to be independently distributed across a, i and j.

The unit level latent variables $\mathbf{L}_{ij}^{(1)} = (\boldsymbol{\beta}_{ij}^{(1)\prime}, \boldsymbol{\eta}_{ij}^{(1)\prime})\prime$ have a multivariate normal distribution with expectations $\overline{\mathbf{L}}_{ij}^{(1)} = ((\mathbf{G}_{\boldsymbol{\beta}}^{(1)}\mathbf{w}_{ij})\prime, (\mathbf{G}_{\boldsymbol{\eta}}^{(1)}\mathbf{w}_{ij})\prime)\prime$ and covariance matrix $\boldsymbol{\Psi}^{(1)}$ partitioned as

$$\boldsymbol{\Psi}^{(1)} = \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(1)} & \\ \mathbf{0} & \boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(1)} \end{pmatrix}.$$

The level two latent variables $\mathbf{L}_j^{(2)} = (\boldsymbol{\beta}_j^{(2)\prime}, \boldsymbol{\eta}_j^{(2)\prime}, \boldsymbol{\gamma}_j^{(2)\prime})\prime$ are multivariate normal with expectations $\overline{\mathbf{L}}_j^{(2)} = ((\mathbf{G}_{\boldsymbol{\beta}}^{(2)}\mathbf{w}_j)\prime, (\mathbf{G}_{\boldsymbol{\eta}}^{(2)}\mathbf{w}_j)\prime, (\mathbf{G}_{\boldsymbol{\gamma}}^{(2)}\mathbf{w}_j)\prime)\prime$ and covariance matrix partitioned as

$$\boldsymbol{\Psi}^{(2)} = \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(2)} & & \\ \mathbf{0} & \boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)} & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)} \end{pmatrix}.$$

We denote the submatrices of $\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$ in the following way: the covariance matrix for random coefficients $\boldsymbol{\gamma}^{a(2)}$ pertaining to a given alternative is denoted $\boldsymbol{\Psi}_{\boldsymbol{\gamma}^a}^{(2)}$ whereas the covariance matrix for random coefficients $\boldsymbol{\gamma}^{a(2)}$ and $\boldsymbol{\gamma}^{b(2)}$ pertaining to different alternatives is denoted $\boldsymbol{\Psi}_{\boldsymbol{\gamma}^a, \boldsymbol{\gamma}^b}^{(2)}$. The unit level and cluster level latent variables $\mathbf{L}_{ij}^{(1)}$ and $\mathbf{L}_j^{(2)}$ are assumed mutually independent, and independent of the alternative specific error vector $\boldsymbol{\epsilon}_{ij}$. This model framework can be denoted *Gumbel-Normal* since the random part of the utilities, $\boldsymbol{\delta}_{ij}^{(1)} + \boldsymbol{\delta}_{ij}^{(2)} + \boldsymbol{\epsilon}_{ij}$, comprises a mixture of a Gumbel and normal components.

Identification

For each unit ij we define a $(A_{ij} - 1) \times A_{ij}$ comparison matrix $\mathbf{H}_{ij}^F$, containing -1, 0 and 1 in appropriate positions, ensuring that a given choice can be compactly expressed in terms of positive utility differences $\mathbf{H}_{ij}^F\mathbf{u}_{ij} > \mathbf{0}$. For instance, if alternative 1 is chosen among three alternatives, the comparison matrix becomes

$$\mathbf{H}_{ij}^F = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Analogously, we define the $(A_{ij} - 1) \times A_{ij}$ comparison matrix $\mathbf{H}_{ij}^R$ for a ranking. For instance, if three alternatives are ranked in the order 1, 2, 3, the comparison matrix becomes

$$\mathbf{H}_{ij}^R = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

First choice and ranking probabilities may then be expressed as

$$\Pr\left(c_{ij} \mid \mathcal{A}_{ij}\right) = \Pr(\mathbf{H}_{ij}^F \mathbf{u}_{ij} > \mathbf{0}), \tag{11}$$

and

$$\Pr(\mathbf{R}_{ij} \mid \mathcal{A}_{ij}) = \Pr(\mathbf{H}_{ij}^R \mathbf{u}_{ij} > \mathbf{0}), \tag{12}$$

respectively.

These probabilities depend only on the utility differences and we can therefore investigate identification by considering the moments of the utility differences. The Logistic-Normal distribution for the utility differences cannot be completely characterized by the first and second order moments. However, due to the similarity between the logistic and the normal distributions, the third and higher order moments do not contain much information (e.g., Rabe-Hesketh & Skrondal, 2001). We therefore find it useful to consider sufficient conditions for local identification in terms of the first and second order moments of the utility differences.

Note that we cannot identify the moments of the utilities themselves. An important implication is that two models, one positing independence among utilities and another specifying equicorrelated utilities cannot be distinguished from first and second order information since they generate identical first and second order moments of the utility differences. Other examples are provided by Bunch (1991).

Letting $\mathbf{H}_{ij}$ be the comparison matrix for the first choice or ranking of unit ij, the expectation structure of the utility differences, given all the exogenous variables $\mathbf{E}_{ij}$, is

$$\mathrm{E}[\mathbf{H}_{ij}\mathbf{u}_{ij} \mid \mathbf{E}_{ij}] = \mathbf{H}_{ij}\mathbf{m} + \mathbf{H}_{ij}\mathbf{X}_{ij}\mathbf{b} + (\mathbf{H}_{ij} \otimes \mathbf{x}_{ij}')\mathbf{g} + \mathbf{H}_{ij}\mathbf{Z}_{ij}^{(1)}\mathbf{G}_{\boldsymbol{\beta}}^{(1)}\mathbf{w}_{ij} + \mathbf{H}_{ij}\boldsymbol{\Lambda}_{[ij]}^{(1)}\mathbf{G}_{\boldsymbol{\eta}}^{(1)}\mathbf{w}_{ij}$$
$$+ \mathbf{H}_{ij}\mathbf{Z}_{ij}^{(2)}\mathbf{G}_{\boldsymbol{\beta}}^{(2)}\mathbf{w}_j + \mathbf{H}_{ij}\boldsymbol{\Lambda}_{[ij]}^{(2)}\mathbf{G}_{\boldsymbol{\eta}}^{(2)}\mathbf{w}_j + (\mathbf{H}_{ij} \otimes \mathbf{z}_{ij}^{(2)'})\mathbf{G}_{\boldsymbol{\gamma}}^{(2)}\mathbf{w}_j,$$

since $\mathbf{H}_{ij}(\varrho\mathbf{1}_{A_{ij}}) = \mathbf{0}$. The conditional covariance structure of the differences is

$$\mathrm{Cov}[\mathbf{H}_{ij}\mathbf{u}_{ij} \mid \mathbf{E}_{ij}] = \mathbf{H}_{ij}\mathbf{Z}_{ij}^{(1)}\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(1)}\mathbf{Z}_{ij}^{(1)'}\mathbf{H}_{ij}' + \mathbf{H}_{ij}\boldsymbol{\Lambda}_{[ij]}^{(1)}\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(1)}\boldsymbol{\Lambda}_{[ij]}^{(1)'}\mathbf{H}_{ij}' + \mathbf{H}_{ij}\mathbf{Z}_{ij}^{(2)}\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(2)}\mathbf{Z}_{ij}^{(2)'}\mathbf{H}_{ij}'$$
$$+ \mathbf{H}_{ij}\boldsymbol{\Lambda}_{[ij]}^{(2)}\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)}\boldsymbol{\Lambda}_{[ij]}^{(2)'}\mathbf{H}_{ij}' + (\mathbf{H}_{ij} \otimes \mathbf{z}_{ij}^{(2)'})\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}(\mathbf{z}_{ij}^{(2)} \otimes \mathbf{H}_{ij}')$$
$$+ \mathbf{H}_{ij}\left(\frac{\pi^2}{6}\mathbf{I}\right)\mathbf{H}_{ij}'. \tag{13}$$

It follows from (11) and (12) that the first choice and ranking probabilities are invariant to changes in the scale of the utility differences. One implication is that freeing the scale of the utility differences by adding additional unit and alternative specific random effects $\gamma_{ij}^{a(1)}$ would lead to fragile identification (relying on scarce higher order information) unless extra restrictions are imposed. Note, however, that the scales of the fixed and random terms in the linear predictor are identified since the last term in (13) is constant. This is also evident by noting that the first choice and ranking probabilities in (1) and (3) are not invariant with respect to a rescaling of the linear predictor. Hence, the expectations, variances and covariances of differences in the linear predictor are identified. It is therefore useful to identify the parameters by designating a base alternative, here without loss of generality chosen as the first.

Consider first identification of the parameters in the fixed effects part. No restrictions need be imposed to identify $\mathbf{b}$. Imposing the restrictions $m^1 = 0$ and $\mathbf{g}^1 = \mathbf{0}$ it follows that $\mathbf{m}$ and $\mathbf{g}$ become identified. We cannot simultaneously identify m^a and a coefficient b of an x_{ij}^a that does not vary over ij. Even when the constant m^a is omitted, the number of purely alternative specific covariates $\mathbf{x}^a$ that can be included is limited by the size of the alternative set A. We then consider the regression parameters in the expectation structures for the latent variables. First of all, observe that we cannot simultaneously include $\mathbf{m}$ in the fixed part of the model and also include intercepts in the regressions by using covariates $\mathbf{w}_{ij}$ or $\mathbf{w}_j$ with one element equal to 1 for all i and j. Assuming that this problem is resolved, we see that $\mathbf{G}_{\boldsymbol{\beta}}^{(1)}$ and $\mathbf{G}_{\boldsymbol{\beta}}^{(2)}$ are directly identified. Regarding $\mathbf{G}_{\boldsymbol{\eta}}^{(1)}$ and $\mathbf{G}_{\boldsymbol{\eta}}^{(2)}$, we note that these parameters are identified if $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ and $\boldsymbol{\Lambda}_{[ij]}^{(2)}$ are identified, respectively. We identify $\mathbf{G}_{\boldsymbol{\gamma}}^{(2)}$ by imposing $\mathbf{G}_{\boldsymbol{\gamma}^1}^{(2)} = 0$.

Turning to the covariance structure, we first consider the random coefficients. We see that $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(1)}$ and $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(2)}$ are identified without imposing further restrictions. In contrast, we identify $\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$ by imposing the restrictions $\boldsymbol{\Psi}_{\boldsymbol{\gamma}^1}^{(2)} = \mathbf{0}$. Note that this implies that $\boldsymbol{\Psi}_{\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^a}^{(2)} = \mathbf{0}$ ($\forall a \neq 1$).

We then consider identification of the unit level factor model. Usually, fixing the variances in $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(1)}$ or one factor loading in each column of $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ leads to identification of $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(1)}$ and $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ in factor models. However, the situation is considerably more complicated for first choices and rankings. For rankings, at least one extra parameter must be fixed in each column of $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ to ensure identification since differences $\mathbf{H}_{ij}\boldsymbol{\Lambda}_{[ij]}^{(1)}$ are involved. Otherwise an arbitrary constant vector may be added to (or subtracted from) the columns of $\boldsymbol{\Lambda}_{[ij]}^{(1)}$ without changing the covariance structure. Analogous concerns arise in the identification of $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)}$ and $\boldsymbol{\Lambda}_{[ij]}^{(2)}$ at the cluster level.

Unit level factor models are identified for rankings subject to the above restrictions whether there are exogenous variables in the model or not. This is in sharp contrast to the case of first choices where unit level factor models are not identified unless exogenous variables are included (compare Heckman & Sedlacek, 1985). Unfortunately, models only including unit specific covariates are prone to empirical underidentification even in this case, but these problems appear to be reduced when unit and alternative specific covariates $\mathbf{x}_{ij}^a$ are included (e.g., Keane, 1992). Hence, the use of unit level factor models for first choice designs should proceed with caution.

Estimation and Prediction

Marginal Likelihood

It follows from (11) that one way of obtaining a first choice probability is through $A_{ij} - 1$ dimensional integration of a Logistic-Normal distribution. However, the dimension of integration becomes excessively high when the number of alternatives is moderately large. Instead we propose to take advantage of the fact that the Gumbel-Normal utility model reduces to the independent Gumbel utility model with constant utility variances when we condition on the latent variables (in addition to the exogenous). The conditional probabilities $\Pr(c_{ij} \mid \mathcal{A}_{ij}, \mathbf{L}_{ij}, \mathbf{E}_{ij})$ and $\Pr(\mathbf{R}_{ij} \mid \mathcal{A}_{ij}, \mathbf{L}_{ij}, \mathbf{E}_{ij})$, conditional on all latent variables $\mathbf{L}_{ij}$ and all exogenous variables $\mathbf{E}_{ij}$, then equal the simple closed form expressions in (1) and (3), respectively, keeping in mind that the linear predictor $v_{ij}^a = f_{ij}^a + \delta_{ij}^a$ now includes latent variables.

To obtain the likelihood contribution for cluster j, we need to integrate out the mutually independent multivariate normal latent variables $\mathbf{L}_{ij}^{(1)}$ and $\mathbf{L}_j^{(2)}$, with corresponding densities $\varphi(\mathbf{L}_{ij}^{(1)})$ and $\varphi(\mathbf{L}_j^{(2)})$, giving

$$\Pr\left(\mathbf{R}_{1j} \cap \mathbf{R}_{2j} \cap \cdots \cap \mathbf{R}_{N_j j} \mid \mathbf{E}_{ij}\right)$$

$$= \int \prod_{i=1}^{N_j} \left\{ \int \Pr(\mathbf{R}_{ij} \mid \mathcal{A}_{ij}, \mathbf{L}_{ij}, \mathbf{E}_{ij}) \varphi(\mathbf{L}_{ij}^{(1)}) \, d\mathbf{L}_{ij}^{(1)} \right\} \varphi(\mathbf{L}_j^{(2)}) \, d\mathbf{L}_j^{(2)}.$$

Hence, under this formulation the dimension of integration equals the number of latent variables which is often of a considerably smaller magnitude than $A_{ij} - 1$. Note that the conditional probabilities do not have closed forms if normally distributed utilities, giving a probit model, are specified instead of the Gumbel. $A_{ij} - 1$ dimensional integration would be required to obtain the marginal likelihood in this case.

The marginal log-likelihood for rankings becomes

$$\mathcal{L} = \sum_{j=1}^{J} \ln \Pr\left(\mathbf{R}_{1j} \cap \mathbf{R}_{2j} \cap \cdots \cap \mathbf{R}_{N_j j} \mid \mathbf{E}_{ij}\right),$$

whereas the marginal likelihood for first-choices is readily obtained as the special case where $P_{ij} = 1$ for all ij.

We suggest using parametric empirical Bayes methods (e.g., Morris, 1983) for predicting values of the latent variables for the individual units. Hence, the predictor is the expectation of the posterior distribution of the latent variables (given the units's ranking or first choice and exogenous variables), with parameter estimates plugged in.

Implementation: gllamm

All models in the general model framework can be estimated using the program gllamm (Rabe-Hesketh, Pickles & Skrondal, 2001). This program, written in Stata, implements maximum likelihood estimation and empirical Bayes prediction for many kinds of generalized linear mixed models with latent variables. Numerical integration by adaptive Gauss-Hermite quadrature (Rabe-Hesketh, Skrondal & Pickles, 2002, 2003) is used to obtain the marginal log-likelihood which is maximized by Newton-Raphson using numerical first and second derivatives. To the user, the only substantial difference between specifying models for rankings compared with other types of responses is that the data are first expanded as for exploded logits (Rabe-Hesketh, Pickles & Skrondal, 2001, chap. 9). This allows the fixed and random structure of the utilities for different alternatives to be specified. Different alternative sets per unit, partial rankings and ties (Breslow method) are all dealt with by simply expanding the data appropriately.

Special Cases of the Model Framework

We now consider special cases of our model framework which have been suggested in the psychometric, econometric, (bio)statistical and marketing literatures. Most developments have been restricted to first choices, and this is the case for the models presented below unless otherwise indicated. However, extension of the models and methods to rankings is straightforward as has been demonstrated for the general model framework. We will also consider probit versions of the models where the alternative specific errors $\boldsymbol{\epsilon}_{ij}$ are specified as normal instead of Gumbel. These models are not special cases of our framework but closely related.

Fixed Effects Only

Conventional logistic regression models only have a fixed part given by (4) in its most general form. Here it is implicitly specified that the utilities of the alternatives for a unit ij, u_{ij}^a, are stochastically independent, given the exogenous variables. Problems associated with these models were discussed in the section Heterogeneity and Independence from Irrelevant Alternatives.

First choice versions of the fixed effects model in (4) have been considered by for instance Train (1986), whereas ranking versions have been used by Chapman and Staelin (1982) and Allison and Christakis (1994). The *conditional logit model*, standard in econometrics, arises as the special case where $\mathbf{x}'_i\mathbf{g}^a = 0$ in (4). This model is very useful for prediction of choice probabilities when hypothetical values for $\mathbf{x}^a_i$ are stipulated for new alternatives (e.g., products). However, such predictions presuppose that the constants m^a are zero, that is, that the included attributes fully explain the choices. The model is used for both first choices (e.g., McFadden, 1973) and rankings (e.g., Hausman & Ruud, 1987). The *polytomous logistic regression model*, a standard model for first choices in for instance biostatistics (e.g., Hosmer & Lemeshow, 1989), results as the special case where $\mathbf{x}^a_{i'}\mathbf{b} = 0$ in (4).

Unobserved Heterogeneity between Units within Clusters

The fixed effects part for these models can have any of the structures outlined above. The dependence is confined to be within units by setting $\delta^a_{ij} = \delta^{a(1)}_{ij}$ in (5), a reasonable specification if there is no dependence induced by clusters.

1. Random coefficient models:

$$\delta^{a(1)}_{ij} = \mathbf{z}^{a(1)'}_{ij}\boldsymbol{\beta}^{(1)}_{ij}.$$

A probit version of this kind of model was considered by Hausman and Wise (1978).
2. Common factor models:

$$\delta^{a(1)}_{ij} = \boldsymbol{\lambda}^{a(1)'}\boldsymbol{\eta}^{(1)}_{ij}.$$

The alternative specific errors ϵ^a_{ij} can in this case be interpreted as unique factors, hence a common factor model arises. Bloxom (1972) and Arbuckle and Nugent (1973) are seminal contributions, introducing common factor models in the pair-comparisons literature. Probit versions for rankings, with the restriction $\boldsymbol{\Psi}^{(1)}_{\boldsymbol{\eta}} = \mathbf{I}$, have been discussed in a series of papers (Böckenholt, 1993; Brady, 1989; Chan & Bentler, 1998). Brady and Böckenholt note that these models are closely related to so-called ideal point models. A regression structure for $\boldsymbol{\eta}^{(1)}$ can be included as in (6). A probit special case suggested for pair-comparisons, the so-called wandering vector model (Carroll, 1980; De Soete & Carroll, 1983) specifies that the expectation only arises via the factors. This kind of structure is obtained by imposing the restrictions $\mathbf{m} = \mathbf{0}$ and $\mathbf{w}_i = \mathbf{1}$ in the regression structure. In the first choice case, regression structures for $\boldsymbol{\beta}^{(1)}_{ij}$ and $\boldsymbol{\eta}^{(1)}_{ij}$ have recently been suggested by Harris and Keane (1999). In contrast, the ranking models considered in the literature do in general not accommodate observed heterogeneity in terms of regression structures.
3. Random coefficient and common factor models:

$$\delta^{a(1)}_{ij} = \mathbf{z}^{a(1)'}_{ij}\boldsymbol{\beta}^{(1)}_{ij} + \boldsymbol{\lambda}^{a(1)'}\boldsymbol{\eta}^{(1)}_{ij}.$$

McFadden and Train (2000), among others, have developed a class of models denoted mixed logit models. The above model is the common special case where the mixing distribution is multivariate normal. The model is also essentially that suggested by Böckenholt (2001a) for rankings, except that Böckenholt also includes a finite mixture for the constants m^a.

Unobserved Heterogeneity between Clusters

In this case there is unobserved heterogeneity at the cluster level but not at the unit level so that $\delta^a_{ij} = \delta^{a(2)}_{ij}$ in (7).

1. *Random coefficient models.* Random coefficients of unit-specific covariates arise when

$$\delta_{ij}^{a(2)} = \mathbf{z}_{ij}^{(2)\prime} \boldsymbol{\gamma}_j^{a(2)}.$$

$\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$ is structured in different ways in the literature. Using the restrictions suggested in our section on identification, the elements $\Psi_{\boldsymbol{\gamma}_s^a, \boldsymbol{\gamma}_{s'}^b}^{(2)}$ of the matrix are identified, where $a, b = 2, \cdots, A$ and $s, s' = 1, \cdots, q$ (q is the number of elements in $\boldsymbol{\gamma}_j^{a(2)}$). Hedeker (1999) specifies $\Psi_{\boldsymbol{\gamma}_s^a, \boldsymbol{\gamma}_{s'}^b}^{(2)} = 0$ ($\forall a, b : a \neq b$), whereas Yang (2001) specifies $\Psi_{\boldsymbol{\gamma}_s^a, \boldsymbol{\gamma}_{s'}^b}^{(2)} = 0$ ($\forall a, b, s, s' : a \neq b$ and $s \neq s'$). Daniels & Gatsonis (1997) freely estimate all identified parameters. Hedeker, Yang and Daniels & Gatsonis all assume heterogeneity only between clusters. In the Daniels & Gatsonis paper, the fixed effect part is of the polytomous logistic regression type, not including alternative specific information, and the expectation structure (10) is imposed on $\boldsymbol{\gamma}_j^{(2)}$. Using our suggested restrictions on $\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$ and imposing $\mathbf{z}_{ij}^{(2)} = 1$ ($\forall i, j$) we obtain a 'correlated alternative specific random intercept model' of the kind specified by Goldstein (1995).

Random coefficients of alternative specific or alternative and unit specific covariates results when

$$\delta_{ij}^{a(2)} = \mathbf{z}_{ij}^{a(2)\prime} \boldsymbol{\beta}_j^{(2)}.$$

This kind of model was considered by Revelt & Train (1998) for repeated first choices of subjects (the clusters). In a conjoint choice analysis, Haaijer, Wedel, Vriens, and Wansbeek (1998) structure $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{(2)}$ using a single principal component.

2. *Factor models.* A common factor model arises if

$$\delta_{ij}^{a(2)} = \boldsymbol{\lambda}^{a(2)\prime} \boldsymbol{\eta}_j^{(2)} + \boldsymbol{\gamma}_j^{a(2)}.$$

Here, we have specified $\mathbf{z}_{ij}^{(2)} = 1$ ($\forall i, j$) and $\Psi_{\boldsymbol{\gamma}^a, \boldsymbol{\gamma}^b}^{(2)} = 0$ ($\forall a, b : a \neq b$). Importantly, $\boldsymbol{\gamma}_j^{a(2)}$ can in this case be interpreted as unique factors at the cluster level. The cluster level covariance structure becomes $\boldsymbol{\Lambda}_{[ij]}^{(2)} \boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)} \boldsymbol{\Lambda}_{[ij]}^{(2)}{}' + \boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$, where $\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)}$ is diagonal. Elrod and Keane (1995) considered a probit version for first choices, imposing $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)} = \mathbf{I}$ and $\boldsymbol{\Psi}_{\boldsymbol{\gamma}}^{(2)} = \kappa \mathbf{I}$, where κ is a scalar parameter. A regression structure (9) may also be included for $\boldsymbol{\eta}_j^{(2)}$ if desired. An unrestricted covariance matrix for the utilities can be obtained by using $A - 1$ factors with unrestricted $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)}$ but imposing the restriction $\boldsymbol{\Lambda}_{[ij]}^{(2)} = \mathbf{I}_{A_{ij}}$. A factor model without unique factors results if $\boldsymbol{\gamma}_j^{a(2)} = 0 \ \forall a, j$. Here, the cluster level covariance structure becomes $\boldsymbol{\Lambda}_{[ij]}^{(2)} \boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)} \boldsymbol{\Lambda}_{[ij]}^{(2)}{}'$. If we furthermore impose the restriction $\boldsymbol{\Psi}_{\boldsymbol{\eta}}^{(2)} = \mathbf{I}$, the covariances have a principal component structure $\boldsymbol{\Lambda}_{[ij]}^{(2)} \boldsymbol{\Lambda}_{[ij]}^{(2)}{}'$. This kind of lower rank approximation was used by Elrod (1988) in the first choice context.

3. *Random coefficient and factor models.* The random coefficient and factor model

$$\delta_{ij}^{a(2)} = \mathbf{z}_{ij}^{a(2)\prime} \boldsymbol{\beta}_j^{(2)} + \boldsymbol{\lambda}^{a(2)\prime} \boldsymbol{\eta}_j^{(2)} + \mathbf{z}_{ij}^{(2)\prime} \boldsymbol{\gamma}_j^{a(2)}, \tag{14}$$

has apparently not been discussed in the literature.

Unobserved Heterogeneity between Clusters and between Units within Clusters

Allenby and Lenk (1994) is the only contribution we are aware of modeling dependence at both unit and cluster levels although their model does not include explicit terms representing unit level heterogeneity. They considered a model where the fixed effects part is of the general

kind (4). The specification of latent variables at the cluster level entails a special case of (14) only including random coefficients for alternative-specific covariates

$$\delta_{ij}^{a(2)} = \mathbf{z}_{ij}^{a(2)\prime} \boldsymbol{\beta}_j^{(2)}.$$

Furthermore, we note that no expectation structure (8) is included for $\boldsymbol{\beta}_j^{(2)}$. In addition to the Gumbel error ϵ_{ij}^a, Allenby and Lenk introduce the error term ϵ_{ij}^{a*} in order to induce a serial correlation structure between occasions i in households j. A vector autoregressive lag-one structure

$$\boldsymbol{\epsilon}_{ij}^* = \boldsymbol{\Phi}\boldsymbol{\epsilon}_{i-1,j}^* + \boldsymbol{\zeta}_{ij}$$

is specified where $\boldsymbol{\Phi}$ is diagonal, $\boldsymbol{\epsilon}_{0j}^* \sim \text{MVN}(\mathbf{0}, \mathbf{C})$ and $\boldsymbol{\zeta}_{ij} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. Although not pointed out by Allenby and Lenk, ϵ_{ij}^{a*} induces dependence among alternatives at level 1 since $\mathbf{C}$ and $\boldsymbol{\Sigma}$ are nondiagonal.

Application: The British Election Panel 1987–1992

Data from the 1987–1992 panel of the British Election Study (Heath, Curtice, Jowell, Evans, Fields, & Witherspoon, 1991; Health, Jowell, & Curtice, 1994; Health, Jowell, Curtice, Brand, & Mitchell, 1993)[1] are used to illustrate multilevel modeling of polytomous data and rankings. 1608 respondents participated in the panel. We excluded voting occasions with missing covariates and where the voters did not vote for candidates from the major parties. The resulting data comprised 2458 voting occasions, 1344 voters and 249 constituencies. The alternatives Conservative, Labour, and Liberal (Alliance) are labeled as $a = 1, 2, 3$, respectively. The voters were not explicitly asked to rank order the alternatives, but the first choice clearly corresponds to rank 1. The voters also rated the parties on a five point scale from "strongly against" to "strongly in favour". We used these ratings to assign ranks to the remaining alternatives, ordering the parties in terms of their rating. Tied second and third choices were observed for 394 voting occasions yielding top-rankings.

We expect that there may be unobserved heterogeneity between elections i, between voters j and between constituencies k. Hence, we consider a three-level model with elections (Level 1) nested within voters (Level 2) and voters nested within constituencies (Level 3). The fixed part of all models considered includes the following election and/or voter specific covariates $\mathbf{x}_{ijk}$: [1987] and [1992] are dummy variables representing the elections, [Male] is a dummy for the voter being male, [Age] represents the age of the voters in 10 year units, [Manual] is a dummy for father of voter a manual worker and [Inflation] is a rating of perceived inflation since the last election on a five point scale. The fixed part also includes an election, voter and alternative specific covariate x_{ijk}^a; [LRdist]. This covariate represents the distance between a voter's position on the left-right political dimension and the mean position of the party voted for. The mean positions of the parties over voters were used to avoid rationalization problems (e.g., Brody & Page, 1972). The placements were constructed from four scales where respondents located themselves and each of the parties on a 11 point scale anchored by two contrasting statements (priority should be unemployment versus inflation, increase government services versus cut taxation, nationalisation versus privatisation, more effort to redistribute wealth versus less effort).

We consider three types of models for the random part

(a) A random coefficient model for political distance [LRdist], inducing dependence and allowing the effect of z_{ijk}^a to vary over elections: $z_{ijk}^a \beta_{ijk}^{(1)}$, over voters: $z_{ijk}^a \beta_{jk}^{(2)}$ and over constituencies: $z_{ijk}^a \beta_k^{(3)}$. Note that $z_{ijk}^{(a)} = x_{ijk}^{(a)}$ in this application.

---

[1]Data were supplied by the UK Data Archive. Neither the original data collectors nor the archive are responsible for the present analyses.

(b) A one-factor model, inducing dependence within elections: $\lambda^{a(1)}\eta^{(1)}_{ijk}$, within voters: $\lambda^{a(2)}\eta^{(2)}_{jk}$ and within constituencies: $\lambda^{a(3)}\eta^{(3)}_k$. A common factor model is only specified at the election level. Factor models without specific factors are used at the voter and constituency levels.

(c) A correlated alternative specific random intercept model, inducing dependence within voters: $\gamma^{a(2)}_{jk}$ and within constituencies: $\gamma^{a(3)}_k$. We do not consider correlated alternative specific random intercept models at the election level since they would be extremely fragile in terms of identification.

At a given level, for example, the voter level, the model in (b) with random terms $\lambda^{a(2)}\eta^{(2)}_{jk}$, $a = 2, 3$ is nested in the random coefficient model (c) with random terms $\gamma^{a(2)}_{jk}$, $a = 2, 3$ since the variances of the random terms are unconstrained in both cases whereas the correlation is fixed at one in the factor model and unconstrained in the random coefficient model.

When dependence is modeled at several levels, we use the same kind of model (e.g., (a), (b) or (c)) at the different levels in order to limit the set of models considered. Note that this is a practical consideration; combinations of models can be used at the same level and different models specified at different levels. Furthermore, no parameter restrictions are imposed across levels. The multilevel models are referred to by numbers indicating the levels followed by letters in parantheses for the model type, for instance $\mathcal{M}12$(b) for a one-factor model specified at the election and voter levels. To preserve space we confine the discussion to analyses based on rankings. Each model was fitted several times, comparing adaptive quadrature with different numbers of quadrature points per latent dimension to ensure reliable results. The sequence of estimated models, their number of parameters (# Par) and their maximized log likelihoods ($\mathcal{L}$) are reported in Table 1.

TABLE 1.
Estimated models

	Fixed Part	Random Part			# Par	$\mathcal{L}$
		Election	Voter	Constituency		
$\mathcal{M}0$	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$				13	-2963.68
$\mathcal{M}1$(a)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$	$z^a_{ijk}\beta^{(1)}_{ijk}$			14	-2945.82
$\mathcal{M}1$(b)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$	$\lambda^{a(1)}\eta^{(1)}_{ijk}$			15	-2842.59
$\mathcal{M}2$(a)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$z^a_{ijk}\beta^{(2)}_{jk}$		14	-2893.14
$\mathcal{M}2$(b)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$\lambda^{a(2)}\eta^{(2)}_{jk}$		15	-2693.72
$\mathcal{M}2$(c)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$\gamma^{a(2)}_{jk}$		16	-2646.48
$\mathcal{M}3$(a)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$			$z^a_{ijk}\beta^{(3)}_k$	14	-2948.44
$\mathcal{M}3$(b)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$			$\lambda^{a(3)}\eta^{(3)}_k$	15	-2846.16
$\mathcal{M}3$(c)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$			$\gamma^{a(3)}_k$	16	-2841.60
$\mathcal{M}12$(a)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$	$z^a_{ijk}\beta^{(1)}_{ijk}$	$z^a_{ijk}\beta^{(2)}_{jk}$		15	-2893.14
$\mathcal{M}12$(b)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$	$\lambda^{a(1)}\eta^{(1)}_{ijk}$	$\lambda^{a(2)}\eta^{(2)}_{jk}$		17	-2692.66
$\mathcal{M}23$(a)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$z^a_{ijk}\beta^{(2)}_{jk}$	$z^a_{ijk}\beta^{(3)}_k$	15	-2893.14
$\mathcal{M}23$(b)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$\lambda^{a(2)}\eta^{(2)}_{jk}$	$\lambda^{a(3)}\eta^{(3)}_k$	17	-2630.08
$\mathcal{M}23$(c)	$\mathbf{x}'_{ijk}\mathbf{g}^a + x^a_{ijk}b$		$\gamma^{a(2)}_{jk}$	$\gamma^{a(3)}_k$	19	-2600.90

$x^{(a)}_{ijk} = z^{(a)}_{ijk}$ is the distance on the left-right political dimension.

We first introduce latent variables only at the election level in $\mathcal{M}1(a)$ and $\mathcal{M}1(b)$ and see from the table that the fit is considerably improved compared to the conventional model $\mathcal{M}0$, indicating that there is cross-sectional dependence among the utilities at a given election (given the covariates in the fixed part). Second, latent variables are introduced only at the voter level in models $\mathcal{M}2(a)$, $\mathcal{M}2(b)$ and $\mathcal{M}2(c)$. The fit is much improved compared to the conventional model, indicating that there is unobserved heterogeneity at the voter level inducing longitudinal dependence within voters. We note in passing that models akin to $\mathcal{M}2(a)$ have attracted considerable interest in political science (e.g., Rivers, 1988). Third, we introduce latent variables only at the constituency level in $\mathcal{M}3(a)$, $\mathcal{M}3(b)$ and $\mathcal{M}3(c)$, and see that the fit is once more improved. Importantly, latent variables at a given level will not only induce dependence at that level but also at all lower levels. Hence, it is possible that latent variables at the election level are superfluous once latent variables are included at the voter level. To resolve this issue we include latent variables at both election and voter levels in $\mathcal{M}12(a)$ and $\mathcal{M}12(b)$. The improvement in fit achieved by including the additional election level latent variables is negligible suggesting that the cross-sectional dependence within elections is largely due to subject level heterogeneity. We therefore do not need to include latent variables at the election level in the subsequent analyses as long as latent variables for the voter level are included. It would be surprising if the latent variables at the voter level were not needed when latent variables are specified at the constituency level since this would imply conditional independence between a voter's utilities at the two elections given the constituency level effects and covariates. As expected, therefore, the $\mathcal{M}23$ models fit considerably better than the $\mathcal{M}3$ models. Furthermore, apart from (a), the $\mathcal{M}23$ models are superior to their $\mathcal{M}2$ counterparts suggesting that there is heterogeneity between constituencies after taking into account subject-level heterogeneity. The $\mathcal{M}2(a)$ model will not be considered further since its fit is considerably worse than that of the $\mathcal{M}23(b)$ and $\mathcal{M}23(c)$ models. Regarding the choice between the $\mathcal{M}23(b)$ and $\mathcal{M}23(c)$ models, which are nested, the correlated random intercepts model appears to be the preferred model.

The estimates for the conventional logistic model $\mathcal{M}0$, only including the fixed part, are reported in Table 2. Estimates for our retained model $\mathcal{M}23(c)$ based on 7-point adaptive quadrature are reported in Table 3 for rankings and first choices in the left and right panels, respectively. We see that the estimated effects of the election and/or voter specific covariates are in accordance with previous research on British elections. Being male and older increases the probability of voting Conservative, whereas a perceived high inflation since the last election harms the incumbent party (the Conservatives). The impact of social class is indicated by the higher probability of voting Labour among voters with a father who is/was a manual worker. Regarding our election, voter and alternative specific covariate [RLdist], the estimate also makes sense: the larger

TABLE 2.
Estimates for the conventional logistic model ($\mathcal{M}0$)

	Ranking		First Choice	
	Lab vs. Cons Est. (SE)	Lib vs. Cons Est. (SE)	Lab vs. Cons Est. (SE)	Lib vs. Cons Est. (SE)
FIXED PART:				
g_1^a [1987]	0.38 (0.20)	0.12 (0.17)	0.51 (0.23)	0.13 (0.22)
g_2^a [1992]	0.51 (0.20)	0.13 (0.18)	0.63 (0.24)	−0.13 (0.23)
g_3^a [Male]	−0.79 (0.11)	−0.53 (0.09)	−0.77 (0.13)	−0.67 (0.12)
g_4^a [Age]	−0.37 (0.04)	−0.18 (0.03)	−0.34 (0.04)	−0.20 (0.04)
g_5^a [Manual]	0.65 (0.11)	−0.05 (0.10)	0.69 (0.13)	−0.10 (0.12)
g_6^a [Inflation]	0.87 (0.09)	0.18 (0.03)	0.76 (0.10)	0.57 (0.09)
b [LRdist]	−0.62 (0.02)		−0.54 (0.02)	
$\mathcal{L}$	−2963.68		−1957.91	

TABLE 3.
Estimates for correlated alternative specific random intercepts model at voter and constituency levels ($\mathcal{M}23(c)$)

	Ranking		First Choice	
	Lab vs. Cons Est. (SE)	Lib vs. Cons Est. (SE)	Lab vs. Cons Est. (SE)	Lib vs. Cons Est. (SE)
FIXED PART:				
g_1^a [1987]	0.77 (0.56)	0.75 (0.37)	0.95 (0.52)	0.13 (0.52)
g_2^a [1992]	1.28 (0.59)	0.78 (0.39)	1.32 (0.54)	−0.30 (0.55)
g_3^a [Male]	−0.99 (0.31)	−0.71 (0.20)	−1.15 (0.28)	−0.96 (0.27)
g_4^a [Age]	−0.74 (0.11)	−0.37 (0.07)	−0.61 (0.10)	−0.36 (0.09)
g_5^a [Manual]	1.57 (0.34)	0.10 (0.22)	1.31 (0.30)	0.04 (0.29)
g_6^a [Inflation]	1.31 (0.18)	0.74 (0.13)	1.17 (0.19)	0.97 (0.18)
b [LRdist]	−0.79 (0.04)		−0.87 (0.05)	
RANDOM PART:				
Voter Level				
$\psi_{\gamma^a}^{(2)}$	16.13 (2.05)	6.03 (0.90)	7.43 (1.62)	9.11 (1.61)
$\psi_{\gamma^2,\gamma^3}^{(2)}$	8.53 (1.15)		5.90 (1.30)	
Constituency Level				
$\psi_{\gamma^a}^{(3)}$	4.91 (1.12)	0.60 (0.29)	3.12 (0.86)	1.74 (0.60)
$\psi_{\gamma^2,\gamma^3}^{(3)}$	1.21 (0.48)		1.11 (0.60)	
$\mathcal{L}$	−2600.90		−1748.95	

the political distance between voter and party the less likely it is that the voter will vote for the party. The random effects variances at the voter level are larger than at the consituency level consistent with a greater residual variability between voters within constituencies than between constituencues as would be expected. The variance of the random effect for Labour, representing residual variability in the utility differences between Labour and Conservatives, is particularly large implying that there are many voters with strong residual (unexplained) support for either the Labour or Conservative parties. There is a positive correlation between the random effects for the Labour and Liberal parties suggesting that those who prefer Labour to the Conservatives, after conditioning on the covariates, also tend to prefer the Liberal party to the Conservatives. This is consistent with the Liberal party being placed between the Labour and Conservative parties and suggests that the [LRdist] covariate has not fully captured this ordering. A useful way of further interpreting the estimates for the random part of multilevel regression models for polytomous data and rankings is to derive the model implied residual (conditional on the covariates) correlation matrices for the utility differences. For the retained model $\mathcal{M}23(c)$ the matrices are displayed in Table 4.

Here the signs of the utility differences are such that we expect positive correlations if the Liberal party is positioned between the Conservative and Labour parties conditional on the covariates. For example, those who prefer Labour to Conservative (positive $u^{\text{lab}} - u^{\text{con}}$) are likely to also prefer Liberal to Conservative (positive $u^{\text{lib}} - u^{\text{con}}$). As expected, the implied cross-sectional correlations at a given election (A) are larger than those implied from the fixed effects model (0.5 in column 1 and −0.5 in column 2). The longitudinal correlations across elections within voters (B) are all positive (the fixed effects model implies zero correlations), the difference in utilities between the Labour and Conservative parties being the most highly correlated across elections. As would be expected, the correlations between different voters in the same constituency (C) are much lower than between elections for the same voter. The correlation involving the Liberal-Conservative differences tend to be lower than the others, suggesting that

TABLE 4.
Model $\mathcal{M}23$(c): Predicted residual within-constituency correlation matrix for the utility differences (subscript k omitted)

Overview

		voter j		voter j'	
		1987	1992	1987	1992
voter j	1987	A			
	1992	B	A		
voter j'	1987	C	C	A	
	1992	C	C	B	A

A: Within voter and election

	$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{con}}$	$u_{ij}^{\mathrm{lib}} - u_{ij}^{\mathrm{con}}$	$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{lib}}$
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{con}}$	1		
$u_{ij}^{\mathrm{lib}} - u_{ij}^{\mathrm{con}}$	0.73	1	
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{lib}}$	0.77	0.14	1

B: Within voter between elections

	$u_{i'j}^{\mathrm{lab}} - u_{i'j}^{\mathrm{con}}$	$u_{i'j}^{\mathrm{lib}} - u_{i'j}^{\mathrm{con}}$	$u_{i'j}^{\mathrm{lab}} - u_{i'j}^{\mathrm{lib}}$
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{con}}$	0.86		
$u_{ij}^{\mathrm{lib}} - u_{ij}^{\mathrm{con}}$	0.63	0.67	
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{lib}}$	0.68	0.29	0.71

C: Between voters (within or between elections)

	$u_{i'j'}^{\mathrm{lab}} - u_{i'j'}^{\mathrm{con}}$	$u_{i'j'}^{\mathrm{lib}} - u_{i'j'}^{\mathrm{con}}$	$u_{i'j'}^{\mathrm{lab}} - u_{i'j'}^{\mathrm{lib}}$
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{con}}$	0.21		
$u_{ij}^{\mathrm{lib}} - u_{ij}^{\mathrm{con}}$	0.09	0.08	
$u_{ij}^{\mathrm{lab}} - u_{ij}^{\mathrm{lib}}$	0.22	0.06	0.27

these parties were less distinguishable from each other after adjusting for the covariates than the other pairs of parties.

Discussion

In addition to effects of observed covariates, there appears to be unobserved heterogeneity at both voter and constituency levels, but not at the election level, in our application. Therefore "Independence from Irrelevant Alternatives" (IIA) is not tenable here, neither conditionally on the fixed effects nor conditionally on the constituency level random effects. There is no evidence for residual correlations among the utilities for a given voter. However, for a given voter IIA only holds conditional on the left-right political distance and inflation rating, and not marginal to these election varying covariates. Furthermore, the impact of IIA will be small since the voters tend to have strong party preferences due to both the fixed and random parts of the model. The misspecification of the conventional model is reflected in the considerably higher absolute values of the fixed effect estimates in the model capturing unobserved heterogeneity $\mathcal{M}23(c)$ as compared to the conventional model $\mathcal{M}0$. A major virtue of multilevel designs is that unobserved heterogeneity can be modeled, improving the prospects of valid causal inferences.

We have employed the gllamm software of Rabe-Hesketh, Pickles and Skrondal (2001) in the analyses reported in this article. An important merit of this software is the generality of the corresponding Generalized Linear Latent and Mixed (GLLAMM) model framework (e.g., Rabe-Hesketh, Skrondal and Pickles, in press; Skrondal & Rabe-Hesketh, 2004), including the models considered here as special cases. This approach works well in a wide range of situations (Rabe-Hesketh, Skrondal & Pickles, 2002, 2003), but speed may become a problem when there are many latent variables. Other approaches to estimation include Markov Chain Monte Carlo methods (e.g., Yao & Böckenholt, 1999) and Maximum Simulated Likelihood (e.g., Hajivassil-iou & Ruud, 1994; Revelt & Train, 1998). A restricted class of models may also be fitted by linearizing the multinomial logistic link function (e.g., Goldstein, 1995), although this approach can be problematic as regards bias of estimated random part parameters (Yang, 1997).

The presented framework for multilevel logistic regression is general and has been shown to subsume numerous models suggested in previous research as special cases. It is nevertheless straightforward to further generalize the framework by for instance including higher order factors to structure the covariance matrices of the latent variables, covariates measured with error, several sets of first choices (e.g., Bock, 1972) or rankings, and pairwise comparisons (e.g., Böckenholt, 2001b). In fact, these extensions are all accommodated by the GLLAMM model framework and can be fitted in the gllamm software. We have not included finite mixtures for rankings like those proposed by Croon (1989) and Böckenholt (2001a) in this article. However, the Croon models are easily fitted in gllamm (Rabe-Hesketh et al., 2001, chap. 9). By using the maximum possible number of mixture components we can obtain the nonparametric maximum likelihood estimator (e.g., Rabe-Hesketh, Pickles, & Skrondal, in press).

Finally, some comments on the relative merits of first choice and ranking designs are in order. A drawback of rankings as compared to first choices in the context of stated preferences is that the decision task is often deemed more complex. For first choices only the most preferred alternative must be picked whereas rankings require the ordering of the alternatives. However, considerable simplification of the ranking task may result from using partial ranking designs. First choice data can sometimes be more reliable than rankings. This is because first choice data are more commonly based on revealed preferences (observed choices) whereas rankings tend to be based on stated preferences. On the other hand, it is well known that rankings lead to more efficient estimators than first choices (e.g., Chapman & Staelin, 1982; Hausman & Ruud, 1987), which is reasonable since rankings contain more information. The estimated standard errors in our application are smaller for the rankings than the first choices, but the increase in efficiency is not impressive. This could be because there are only three alternatives and because there were a considerable number of top-rankings. Rankings are also beneficial from the point of view of

identification. Some first choice models are not identified whereas the corresponding ranking models are identified. Even when the first choice models are formally identified they are known to be more fragile than their ranking counterparts.

References

Allenby, G.M., & Lenk, P.J. (1994). Modelling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association*, *89*, 1218–1231.

Allison, P.D., & Christakis, N.A. (1994). Logit models for sets of ranked items. In P.V. Marsden (Ed.), *Sociological Methodology 1994* (pp. 199–228). Oxford, U.K.: Blackwell.

Arbuckle, J., & Nugent, J.H. (1973). A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical and Statistical Psychology, 26*, 240–260.

Beggs, S., Cardell, S., & Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, *16*, 1–19.

Block, H.D., & Marchak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, W. Hoeffting, S. Ghurye, W.G. Madow & H.D. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 97–112). Stanford, CA: Stanford University Press.

Bloxom, B. (1972). The simplex in pair comparisons. *Psychometrika*, *37*, 119–136.

Bock, R.D. (1969). Estimating multinomial response relations. In R. Bose (Ed.), *Essays in probability and statistics* (pp. 111–132). Chapel Hill, NC: University of North Carolina Press.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51.

Böckenholt, U. (1992). Thurstonian representation for partial ranking data. *British Journal of Mathematical and Statistical Psychology*, *45*, 31–39.

Böckenholt, U. (1993). Application of Thurstonian models to ranking data. In M.A. Fligner & J.S. Verducci (Eds.), *Probability models and statistical analysis for ranking data* (pp. 157–172). New York, NY: Springer.

Böckenholt, U. (2001a). Mixed-effects analyses of rank-ordered data. *Psychometrika*, *66*, 45–62.

Böckenholt, U. (2001b). Hierarchical modeling of paired comparison data. *Psychological Methods*, *6*, 49–66.

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

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics*, *30*, 89–100.

Brody, R.A., & Page, B.I. (1972). Comment: The assessment of policy voting. *American Political Science Review*, *66*, 450–458.

Bunch, D.S. (1991). Estimability in the multinomial probit model. *Transportation Research, Part B: Methodological*, *25*, 1–12.

Carroll, J.D. (1980). Models and methods for multidimensional analysis of preference choice (or other dominance) data. In E.D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 234–289). Bern: Huber.

Chan, W., & Bentler, P.M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika*, *63*, 369–399.

Chapman, R.G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, *14*, 288–301.

Critchlow, D.E., Fligner, M.A., & Verducci, J.S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, *35*, 294–318.

Croon, M.A. (1989). Latent class models for the analysis of rankings. In G. De Soete & K.C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 99–121). Amsterdam: North-Holland.

Daniels, M.J., & Gatsonis, C. (1997). Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine*, *16*, 2311–2325.

De Soete, G., & Carroll, J.D. (1983). A maximum likelihood method for fitting the wandering vector model. *Psychometrika*, *48*, 553–566.

Durbin, J. (1951). Incomplete blocks in ranking experiments. *British Journal of Psychology*, *4*, 85–90.

Elrod, T. (1988). Choice map: Inferring a product-market map from panel data. *Marketing Science*, *7*, 21–40.

Elrod, T., & Keane, M.P. (1995). A factor analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, *32*, 1–16.

Farewell, V.T., & Prentice, R.L. (1980). The approximation of partial likelihood with emphasis on case-control studies. *Biometrika*, *67*, 273–278.

Goldstein, H. (1995). *Multilevel statistical models*. London: Arnold.

Gurland, J., Lee, I., & Dolan, P. (1960). Polychotomous quantal response in biological assay. *Biometrics*, *16*, 382–398.

Haaijer, M.E., Wedel, M., Vriens, M., & Wansbeek, T.J. (1998). Utility covariances and context effects in conjoint MNP models. *Marketing Science*, *17*, 236–252.

Hajivassiliou, V., & Ruud, P. (1994). Classical estimation methods for LDV models using simulation. In R. Engle & D. McFadden (Eds.), *Handbook of econometrics IV* (pp. 2384–2441). Amsterdam: North-Holland.

Harris, K.M., & Keane, M.P. (1999). A model for health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics*, *89*, 131–157.

Hausman, J.A., & Ruud, P.A. (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, *34*, 83–103.

Hausman, J.A., & Wise, D.A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *46*, 403–426.

Heath, A., Curtice, J., Jowell, R., Evans, G., Fields, J., & Witherspoon, S. (1991). *Understanding political change: The British voter 1964-1987*. Oxford, U.K.: Pergamon.

Heath, A., Jowell, R., & Curtice, J. (Eds.) (1994). *Labour's last chance? The 1992 election and beyond*. Aldershot: Dartmouth.

Heath, A., Jowell, R, Curtice, J.K., Brand, J.A., & Mitchell, J.C. (1993). *SN2983—British Election Panel Study, 1987–1992* [Computer file]. Colchester, Essex: The Data Archive [Distributor].

Heckman, J.J., & Sedlacek, G. (1985). Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy*, *93*, 1077–1125.

Hedeker, D. (1999). MIXNO: A computer program for mixed-effects logistic regression. *Journal of Statistical Software*, *4*, 1–92.

Hosmer, D.A., & Lemeshow, S.A. (1989). *Applied logistic regression*. London: Wiley.

Hsiao, C. (1985). Benefits and limitations of panel data. *Econometric Review*, *4*, 121–174.

Johnson, N.L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, Volume 2*. New York, NY: Wiley.

Kalbfleisch, J.D., & Prentice, R.L. (1980). *The statistical analysis of failure time data*. New York, NY: Wiley.

Keane, M.P. (1992). A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics*, *10*, 193–200.

Luce, R.D. (1959). *Individual choice behavior*. New York, NY: Wiley.

Luce, R.D., & Suppes, P. (1965). Preference, utility and subjective probability. In R.D. Luce, R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology III* (pp. 249–410). New York, NY: Wiley.

Maddala, G.S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge, MA: Cambridge University Press.

Marchak, J. (1960). Binary choice constraints on random utility indicators. In K.J. Arrow, S. Karlin & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959* (pp. 312–329). Stanford, CA: Stanford University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York, NY: Academic Press.

McFadden, D., & Train, K. (2000). Mixed MLN models for discrete choice. *Journal of Applied Econometrics*, *15*, 447–470.

Morris, C. (1983). Parametric empirical Bayes inference, theory and applications. *Journal of the American Statistical Association*, *78*, 47–65.

Neuhaus, J.M. (1992). Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, *1*, 249–273.

Plackett, R.L. (1975). The analysis of permutations. *Applied Statistics*, *24*, 193–202.

Rabe-Hesketh, S., & Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics*, *57*, 1256–1264.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM Manual* (Tech. Rep. 2001/01). London: King's College, University of London, Institute of Psychiatry, Department of Biostatistics and Computing. Computer software and documentation can be retreived at http://www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, *2*, 1–21.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2003). *Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects*. Submitted for publication.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (in press). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (in press). Generalized multilevel structural equation modeling. *Psychometrika*.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: household's choices of appliance efficiency level. *Review of Economics and Statistics*, *80*, 647–657.

Rivers, D. (1988). Heterogeneity in models of electoral choice. *American Journal of Political Science*, *32*, 737–757.

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall CRC.

Takane, Y. (1987). Analysis of covariance structures and probabilistic binary data. *Communication & Cognition*, *20*, 45–62.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review*, *10*, 251–259.

Train, K. (1986). *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.

Yang, M. (1997). Multilevel models for multiple category responses—A simulation. *Multilevel Modelling Newsletter*, *9*, 10–16.

Yang, M. (2001). Multinomial regression. In A.H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 107–123). Chichester, U.K.: John Wiley & Sons.

Yao, G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, *52*, 79–92.

Yellott, J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–44.