PSYCHOMETRIKA—VOL. 72, NO. 4, 489–504
DECEMBER 2007
DOI: 10.1007/S11336-007-9020-5

DISCRETE CHOICE MODELS FOR ORDINAL RESPONSE VARIABLES: A GENERALIZATION OF THE STEREOTYPE MODEL

TIMOTHY R. JOHNSON

UNIVERSITY OF IDAHO

In this paper I present a class of discrete choice models for ordinal response variables based on a generalization of the stereotype model. The stereotype model can be derived and generalized as a random utility model for ordered alternatives. Random utility models can be specified to account for heteroscedastic and correlated utilities. In the case of the generalized stereotype model this includes category-specific random effects due to individual differences in response style. But unlike standard random utility models the generalized stereotype model is better suited for ordinal response variables and can be interpreted as a kind of unidimensional unfolding model. This paper discusses the specification, interpretation, identification, and estimation of generalized stereotype models. Two applications are provided for illustration.

Key words: stereotype model, ordinal regression, response style.

1. Introduction

Let $Y_i \in \{1, 2, \ldots, p\}$ denote an ordinal response variable where the integer-valued realizations imply only an assumed ordering. Ordinal regression models the probability mass of Y_i, conditional on a vector of covariates $\mathbf{x}_i$, while taking into account the order of the realizations. Agresti (1990) discusses a typology of ordinal regression models based on different types of log-odds ratios or "logits," and Mellenbergh (1995) discusses this typology in the context of item response theory. The typology includes models based on the cumulative logit, $\log[P(Y_i \le y|\mathbf{x}_i)/P(Y_i > y|\mathbf{x}_i)]$, the continuation-ratio or sequential logit, $\log[P(Y_i = y|\mathbf{x}_i)/P(Y_i < y|\mathbf{x}_i)]$, and the adjacent-categories logit, $\log[P(Y_i = y + 1|\mathbf{x}_i)/P(Y_i = y|\mathbf{x}_i)]$. Anderson (1984) proposed an adjacent-categories logit ordinal regression model called the "stereotype model." To date the stereotype model has not been widely used, particularly in comparison to models based on the cumulative logit, and it is rarely discussed in the methodological literature except in some surveys of regression models for ordinal response variables (e.g., Clogg & Shihaden, 1994; Greenland, 1994; Holtbrügge & Schumacher, 1991; Liu & Agresti, 2005). This oversight on the part of behavioral statistics and psychometrics is surprising because Anderson argued that the stereotype model is particularly well suited for *subjective* ordinal response variables.[1] The intention of this paper is to extend the usefulness of the stereotype model by introducing a *generalized* stereotype model as a discrete choice random utility model for ordered alternatives. The generalized model can account for individual differences in category usage, and is a more parsimonious model for ordinal response variables than standard random utility models. It also leads to a mixed effects generalization of the stereotype model.

The organization of this paper is as follows. After briefly reviewing the stereotype model, I present the generalized stereotype model as a random utility model. I show that the generalized

This paper benefited significantly from the comments and suggestions of the editor, associate editor, and three anonymous reviewers. It is dedicated to my late colleague, peer, and friend Bradley D. Crouch.

Requests for reprints should be sent to Timothy R. Johnson, Department of Statistics, University of Idaho, Moscow, ID 83844-1104, USA. E-mail: trjohns@uidaho.edu

[1] As noted by Liu and Agresti (2005), oversight of the stereotype model may be due in part to the death of its author and would-be primary advocate prior to the publication of Anderson (1984).

stereotype model shares some of the advantages of multinomial probit and mixed logit models by accounting for heteroscedastic and correlated utilities, but is better suited for ordinal response variables. The remaining sections concern the identification of and inference for generalized stereotype models with more technical details given in the Appendix. The paper concludes with two illustrative applications.

2. The Stereotype Model

The stereotype model specifies the log-linear conditional probability mass function

$$P(Y_i = y|\mathbf{x}_i) \propto \exp(\alpha_y + \gamma_y \mathbf{x}_i' \boldsymbol{\beta}). \tag{1}$$

This is an ordinal adjacent-categories logit model where

$$\log\bigl[P(Y_i = y+1|\mathbf{x}_i)/P(Y_i = y|\mathbf{x}_i)\bigr] = (\alpha_{y+1} - \alpha_y) + (\gamma_{y+1} - \gamma_y)\mathbf{x}_i'\boldsymbol{\beta} \tag{2}$$

is nondecreasing in $\mathbf{x}_i'\boldsymbol{\beta}$ provided that $\gamma_y \leq \gamma_{y+1}$. The motivation and interpretation of this model is facilitated by several additional ordinal properties as discussed in Anderson (1984). In cases where (1) does not provide adequate fit, the model can be expanded into the T-dimensional stereotype model

$$P(Y_i = y|\mathbf{x}_i) \propto \exp\left(\alpha_y + \sum_{t=1}^{T} \gamma_y^{(t)} \mathbf{x}_i' \boldsymbol{\beta}^{(t)}\right).$$

The multidimensional model is a compromise between the unidimensional model in (1) and the nominal multinomial logit model

$$P(Y_i = y|\mathbf{x}_i) \propto \exp(\alpha_y + \mathbf{x}_i'\boldsymbol{\beta}_y),$$

equivalent to a "saturated" p-dimensional stereotype model where $\gamma_y^{(t)} = I(y = t)$. The unidimensional stereotype model is a special case of the nominal model where $\boldsymbol{\beta}_k \propto \boldsymbol{\beta}_{k'}$.

3. A Generalized Stereotype Model

Let Y_{ij} denote the jth response ($j = 1, 2, \ldots, m$) from the ith respondent.[2] Let U_{ijk} denote the "utility" of the kth response category for the jth observation from the ith respondent. The generalized stereotype model proposed here is defined by the random utility model

$$U_{ijk} = \alpha_k + \xi_{ik} + \gamma_k(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i) + \epsilon_{ijk}, \tag{3}$$

where $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{ip})'$ and $\boldsymbol{\zeta}_i$ are p- and q-dimensional random vectors, respectively, jointly distributed as $N(\mathbf{0}, \boldsymbol{\Phi})$, and independent of all ϵ_{ijk} where $E(\epsilon_{ijk}|\mathbf{x}_{ij}, \mathbf{z}_{ij}) = 0$. The observed response is determined by the largest latent response such that $Y_{ij} = y \Leftrightarrow \max_k(U_{ijk}) = U_{ijy}$. The original stereotype model is a special case where $m = 1$, $\xi_{ik} = \mathbf{z}_{ij}'\boldsymbol{\zeta}_i = 0$, and all ϵ_{ijk}

[2] For simplicity and without loss of generality, it is assumed here that the number of responses per respondent and the number of categories per response are equal over respondents and responses, respectively.

are independently and identically distributed Type I extreme value random variables. The generalized stereotype model belongs to a broad family of linear random utility models of the general form

$$U_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta}_k + \varepsilon_{ijk}. \tag{4}$$

Models vary in terms of the decomposition of the systematic component $\mathbf{x}_{ijk}\boldsymbol{\beta}_k$ and the random error component ε_{ijk}. This family includes the mixed multinomial logit model (McFadden and Train, 2000) and the multiperiod/panel multinomial logit model (Geweke, Keane, & Runkle, 1997) where $\boldsymbol{\beta}_k = \boldsymbol{\beta}_{k'}$ and $\varepsilon_{ijk} = \mathbf{z}'_{ijk}\boldsymbol{\zeta}_i + \epsilon_{ijk}$, and where ϵ_{ijk} is a Type I extreme value (logit) or normal (probit) random variable. Similarly the generalized stereotype model can be specified as a multinomial probit and mixed logit model. But unlike these discrete choice models the generalized stereotype model does not have (observed) category-specific regressors, but allows the regression coefficients to be proportional over categories. As argued by Anderson (1984), this parametrization makes the stereotype model more suitable for *ordered* response variables. The random parameters in the generalized stereotype model make it suitable for accounting for both respondent-specific effects represented by the term $\mathbf{z}'_{ij}\boldsymbol{\zeta}_i$, and category-specific effects represented by the term ξ_{ik}. The random category-specific effects are included to capture individual differences in category usage.

3.1. Individual Differences in Category Usage

The original stereotype model is closely related to the conditional logit choice model (McFadden, 1974). Both have simple and tractable forms, but imply the restrictive assumption of independence from irrelevant alternatives (IIA), meaning that the ratio of the choice probabilities of any two alternatives does not depend on the remaining $p-2$ alternatives (Luce, 1959). In terms of the utilities the implicit assumption is that they are uncorrelated and have constant variance. With the stereotype model these assumptions can be violated due to individual differences in response category usage. It is well known that the analysis of ordinal rating scales can be problematic due to individual differences in response style (e.g., Alwin & Krosnick, 1985; Baumgartner & Steenkamp, 2001; Messick, 1991), such as individual differences in the "extreme response style" (i.e., the tendency for respondents to use extreme versus moderate responses; Hamilton, 1968), or simply individual differences in the interpretation of the response categories (Cronbach, 1946, 1950). Individual differences in these and other response styles may manifest as individual differences in the tendency to use certain response categories inducing heteroscedasticity in and correlations among the utilities.

The mixed multinomial logit and multinomial probit models overcome violations of IIA by allowing heteroscedastic and correlated utilities by decomposing the errors in (4) into the sum of random effects and independent errors as shown earlier. The generalized stereotype model has two distinct kinds of random effects: the respondent-specific effects represented by $\mathbf{z}'_{ij}\boldsymbol{\zeta}_i$, making the generalized stereotype model a mixed effects generalization of the original, and the random effects represented by ξ_{ik} that are both respondent- and category-specific, included to account for individual differences in response category usage.[3] Mixed effects models for individual differences in category usage have been proposed assuming random respondent-specific thresholds in *cumulative* logit and probit models (e.g., Böckenholt, 2001a; Fielding, Yang, & Goldstein, 2003; Johnson, 2003; Lenk, Wedel, & Böckenholt, 2006; Tutz & Hennevogl, 1996; Wolfe & Firth, 2002). The generalized stereotype model takes a different approach to this problem based on a discrete choice model for ordered alternatives that can be viewed as an unfolding model.

[3]This mixed effects generalization of the stereotype model includes factor-analytic and item response models where the "covariate" vector $\mathbf{z}_{ij} = \boldsymbol{\lambda}_j$ is an unknown vector of loadings (see Rabe-Hesketh, Skrondal, & Pickles, 2004.)

3.2. The Stereotype Unfolding Model

Unlike other choice models, the stereotype model assumes an underlying ordering to the alternatives. Anderson (1984) motivated the stereotype model in terms of ordered distributions of the covariates conditional on the responses, and in terms of the ordering of the category response functions conditional on the covariates. The now generalized stereotype model can also be motivated and interpreted as an unfolding model. Suppose that rather than the linear model in (3) that the utility of each category is a quadratic function of $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_i$ such that

$$U_{ijk} = \theta_{0ik} + \theta_{1k}(\eta_{ij} - \theta_{2k})^2 + \epsilon_{ijk}, \tag{5}$$

where $\theta_{1k} < 0$ and $E(\epsilon_{ijk}|\eta_{ij}) = 0$ so that the expected utility of the kth category achieves a maximum of θ_{0ik} when $\eta_{ij} = \theta_{2k}$. This model postulates a kind of unidimensional unfolding model where the location of the kth category is θ_{2k}, and where the expected utility of each category is proportional to the square of the distance between the respondent's ideal point, η_{ij}, and the category point, θ_{2k}. The quadratic model appears to be a more natural random utility model for ordered choices in comparison to the linear model in (3) where the utility of each response category is a linear function of η_{ij}. However, the (generalized) stereotype model is consistent with (5) because the response probabilities depend only on the *differences* between the utilities since $Y_{ij} = y \Leftrightarrow U_{ijy} - U_{ijk} > 0$ for all $k \neq y$. If $\theta_{1k} = \theta_{1k'} = \theta_1 < 0$, then

$$U_{ijk} - U_{ijk'} = (\alpha_k + \xi_{ik}) - (\alpha_{k'} + \xi_{ik'}) + (\gamma_k - \gamma_{k'})\eta_{ij} + \epsilon_{ijk} - \epsilon_{ijk'}, \tag{6}$$

where $\alpha_k + \xi_{ik} = \theta_{0ik} + \theta_1\theta_{2k}^2$ and $\gamma_k = -2\theta_1\theta_{2k}$. But note that (6) is also obtained by assuming the linear utility model in (3). The relationship between the quadratic and linear utility models is shown in Figure 1 which depicts (a) $E(U_{ijk}|\eta_{ij})$ and (b) $E(U_{ijk} - U_{ij1}|\eta_{ij})$ as a function of η_{ij} for $p = 3$ categories where $\theta_{1k} = -1$ and $k = 1, 2, 3$.

The generalized stereotype model is consistent with a unidimensional random coordinate unfolding model with category points proportional to $\gamma_1, \gamma_2, \ldots, \gamma_p$. The category-specific random parameters $\xi_{i1}, \xi_{i2}, \ldots, \xi_{ip}$ can be interpreted as reflecting individual differences in the heights of the expected utility curves defined in (5) to account for individual differences in the baseline preference for each category independent of the position of the respondent's ideal point.

The equivalence of quadratic/unfolding and linear/vector utility models has also been noted in the context of paired comparisons (Tsai & Böckenholt, 2001) and ideal point discriminant analysis (Takane, Bozdogan, & Shibayama, 1987). The original stereotype model is equivalent to a unidimensional ideal point discriminant analysis.

3.3. Symmetry Constraints

Ordinal response scales often have a symmetric structure in the sense that the kth and $(p - k + 1)$th categories are perceived by respondents to have the same "intensity" relative to a (possibly imaginary) middle or neutral response such as when categories denote gradations between opposing responses (e.g., *strongly disagree*, *disagree*, *neither agree nor disagree*, *agree*, *strongly agree*). In such cases it may be reasonable to assume symmetry in the utility curves described by (5) such that

$$\theta_{0ik} = \theta_{0i,p-k+1} \quad \text{and} \quad \theta_{2,k+1} - \theta_{2k} = \theta_{2,p-k+1} - \theta_{2,p-k}$$

for $k < p$, implying symmetry in expected utility curves defined in (5). In terms of the generalized stereotype model defined in (3), these constraints imply

$$\alpha_k - \alpha_{k+1} = \alpha_{p-k+1} - \alpha_{p-k}, \qquad \gamma_{k+1} - \gamma_k = \gamma_{p-k+1} - \gamma_{p-k},$$

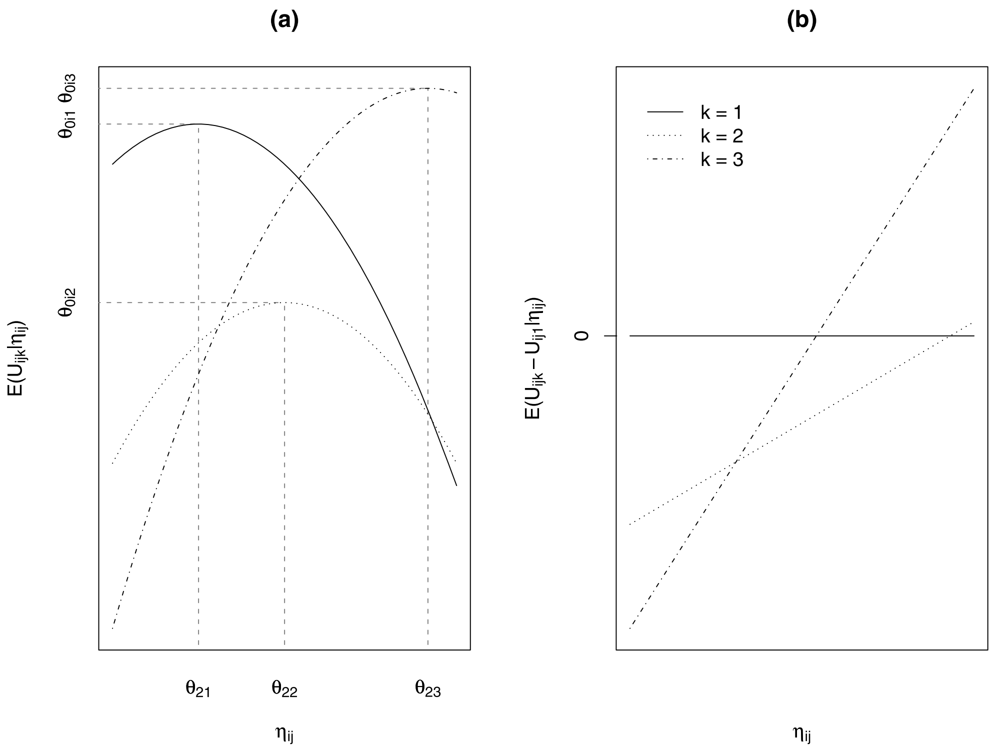

FIGURE 1.
(a) Quadratic expected utility functions and (b) linear expected utility function differences for an ordinal three-category response scale.

and

$$\xi_{i,k} - \xi_{i,k+1} = \xi_{i,p-k+1} - \xi_{i,p-k}$$

for $k < p$. Figure 2 shows a symmetric structure for an example with $p = 5$ categories.

For the multinomial probit version of the stereotype model the symmetry constraints on ξ_{ik} can alternatively be imposed directly on the covariance structure of the utilities (see the Appendix). The advantage of symmetry constraints is the reduction in the number of unknown category-specific parameters.

3.4. *Identification*

Careful analysis and specification of the generalized stereotype model is necessary to ensure that it is meaningfully identified. Some identification issues can be addressed by writing the generalized stereotype model as a bilinear mixed model for the differences among the utilities (see the Appendix). There are indeterminacies in the category-specific parameters because the observed responses depend only on the differences among the utilities as shown by (6). It is necessary to fix, say, $\alpha_{k^*} = \gamma_{k^*} = \xi_{ik^*} = 0$ for some specified reference category k^*. The choice of reference category is arbitrary, but if a symmetric structure is specified then it is convenient to set a zero midpoint for the scale so that the symmetry constraints become $\alpha_k = \alpha_{p-k+1}$, $\gamma_k + \gamma_{p-k+1} = 0$, and $\xi_{ik} = \xi_{i,p-k+1}$. Further discussion of the identification of the multinomial probit stereotype model is given in the Appendix. The scale of the parameters must also be fixed. $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Phi}$ are not jointly identified but this can be resolved by fixing, say, $\gamma_1 = 1$. The scale of the utilities must also be fixed (e.g., $\text{Var}(\epsilon_{ijk}) = 1$).

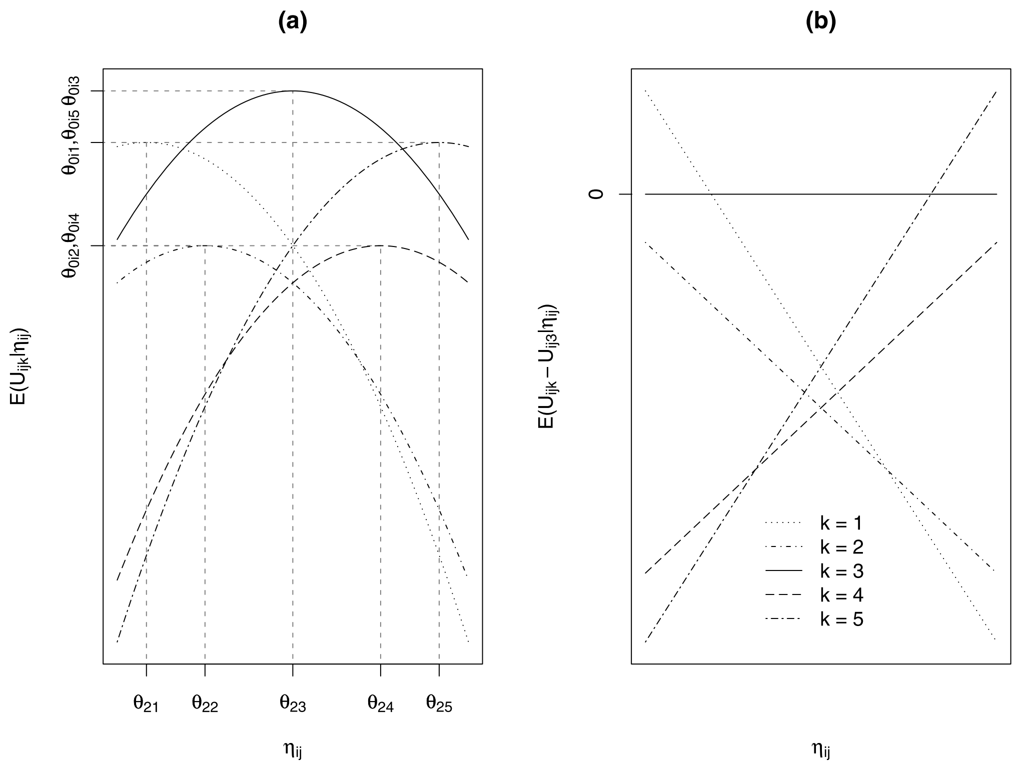

FIGURE 2.
(a) Quadratic expected utility functions and (b) linear expected utility function differences for a symmetric ordinal five-category response scale.

Although a given model with the appropriate constraints may be formally identified, it may not be well identified *empirically* resulting in estimates that are too unreliable to be useful in practice. Keane (1992; see also Weeks, 1997) showed that multinomial probit models for a single p-category response can be unstable in the absence of category-specific covariates. The stereotype model does not have (observed) category-specific covariates, and so has this potential instability. But this problem can be remedied by hierarchical data because the joint distribution of the observations provides additional information. In the author's experience hierarchical multinomial probit models with a sufficiently constrained covariance structure are not necessarily fragile provided that the covariance structure induced by the random effects is of small rank relative to m. Using the logit version of the model does not necessarily avoid the problem of fragile models because of the close similarity of the logit and probit models (see Rabe-Hesketh & Skrondal, 2001, for a similar issue in the context of multivariate logit and probit models).

3.5. *Inference*

The generalized stereotype model can be viewed as a nonlinear mixed model for multinomial responses. Likelihood-based inferential methods based on the "marginal" likelihood function require integration over the distribution of $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$. This can be done numerically using quadrature provided that the dimension of integration—i.e., the number of free random effects—is not too high. The integration can be reduced to a manageable level in some cases by using a logit version of the model. But in cases where the dimension of integration is prohibitively high for standard quadrature methods a more efficient approach is to use simulation-based methods, to which the multinomial probit version of the model is often more amenable. Computational details concerning likelihood-based inference for the generalized stereotype model are given in the Appendix.

4. Application—A Mixed Model for Verbal Aggression

This first application illustrates the use of the generalized stereotype model with data from Vansteelandt (2000) that were analyzed in depth in De Boeck and Wilson (2004). That study elicited responses on a 3-category scale indicating whether respondents would agree to give an aggressive verbal response ("no," "perhaps," or "yes") in a variety of scenarios. The characteristics of the scenarios comprised a 2 by 2 by 3 fully-factorial design. The factors and their coding are *behavior mode* (wanting to do, $x_{ij1} = -1$, versus actually doing, $x_{ij1} = 1$), *situation type* (blaming another, $x_{ij2} = 1$, versus blaming self, $x_{ij2} = -1$), and *behavior* type (curse, scold, or shout). The last factor is coded using two contrasts representing *blaming* ($x_{ij3} = 1/2$ if the behavior type is curse or scold, $x_{ij3} = -1$ otherwise), and *expressing* ($x_{ij4} = 1/2$ if the behavior type is curse or shout, $x_{ij4} = -1$ otherwise). The same coding was used in the analyses reported in De Boeck and Wilson. Respondents in a sample of $n = 316$ each gave responses to 24 scenarios (within each cell of the design there were two pseudo-replications).

With only $p = 3$ categories, the likelihood of either the multinomial probit or mixed multinomial logit stereotype models can be approximated relatively quickly and accurately using quadrature. A mixed logit model was used here. The reference category was specified as $k^* = 1$ so that $D_{ijk} = U_{ijk} - U_{ij1}$ with $\alpha_1 = \gamma_1 = 0$ and $\gamma_2 = 1$ for identification. The models discussed below were estimated using PROC NLMIXED from the SAS/STAT software package. The fit of each model is summarized in Table 1.

Consider first two stereotype models, one with only a random "intercept" such that $U_{ijk} = \alpha_k + \gamma_k(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_i) + \epsilon_{ijk}$, and another with category-specific random effects to account for individual differences in category usage such that $U_{ijk} = \alpha_k + \xi_{ik} + \gamma_k(\mathbf{x}'_{ij}\boldsymbol{\beta}) + \epsilon_{ijk}$ (inspection of the design matrix for the random effects, as defined in the Appendix, reveals that category-specific random effects and a random intercept are not jointly identified, so here ζ_i is "absorbed" into $\boldsymbol{\xi}_i$). The model allowing for individual differences in response style has a superior fit to the data in comparison to the intercept-only model. For comparison, mixed models with cumulative link functions, with a random intercept, and with and without random thresholds were also fit, where the random thresholds are intended to account for individual differences in response style as described by Johnson (2003). Here again a case can be made for the necessity of accounting for individual differences in response style because the random-thresholds model has a better fit than the fixed-thresholds model. Interestingly, both the stereotype and cumulative link models that account for individual differences in response style have almost the same fit (the AIC for

TABLE 1.
Goodness of fit measures for models for the verbal aggression data.

Model	Random parameters	Log-likelihood	Parameters	AIC
Stereotype	intercept	−6431	8	12,877
Stereotype	category	−6265	10	12,549
Cumulative	intercept	−6431	7	12,876
Cumulative	thresholds	−6265	9	12,547
Nominal	category	−6248	13	12,522
Stereotype (2d)	category	−6249	11	12,519

TABLE 2.
Maximum likelihood parameter estimates and standard errors for the stereotype models for the verbal aggression data.

	Model and random effect(s) structure			
	Stereotype (intercept)	Stereotype (category)	Multinomial (category)	Stereotype (2D) (category)
α_1	0.00	0.00	0.00	0.00
α_2	−0.53 (0.07)	−0.65 (0.08)	−0.65 (0.08)	−0.65 (0.08)
α_3	−1.40 (0.11)	−1.49 (0.12)	−1.49 (0.12)	−1.49 (0.12)
$\gamma_1^{(1)}$	0.00	0.00	0.00	0.00
$\gamma_2^{(1)}$	1.00	1.00	1.00	1.00
$\gamma_3^{(1)}$	1.67 (0.06)	1.70 (0.07)	0.00	1.49 (0.06)
$\gamma_1^{(2)}$			0.00	0.00
$\gamma_2^{(2)}$			0.00	0.00
$\gamma_3^{(2)}$			1.00	1.00
$\beta_1^{(1)}$	−0.27 (0.02)	−0.26 (0.02)	−0.28 (0.03)	−0.29 (0.02)
$\beta_2^{(1)}$	0.46 (0.02)	0.46 (0.03)	0.35 (0.03)	0.35 (0.03)
$\beta_3^{(1)}$	1.08 (0.05)	1.06 (0.05)	1.17 (0.05)	1.17 (0.05)
$\beta_4^{(1)}$	0.53 (0.04)	0.53 (0.04)	0.61 (0.05)	0.59 (0.04)
$\beta_1^{(2)}$			−0.44 (0.04)	
$\beta_2^{(2)}$			0.84 (0.04)	0.31 (0.05)
$\beta_3^{(2)}$			1.74 (0.07)	
$\beta_4^{(2)}$			0.88 (0.06)	
σ_{22}	1.18 (0.13)	1.55 (0.17)	1.56 (0.17)	1.55 (0.17)
σ_{23}	1.97 (0.20)	1.68 (0.21)	1.68 (0.21)	1.68 (0.21)
σ_{33}	3.28 (0.34)	3.35 (0.36)	3.35 (0.36)	3.35 (0.36)

Note: Standard errors are in parentheses.

the latter is superior only because of a difference of one parameter). Now consider whether the unidimensional ordinal structure is sufficient to account for the data by considering a multidimensional stereotype model. Here the saturated multidimensional stereotype model

$$U_{ijk} = \alpha_k + \xi_{ik} + \gamma_k^{(1)}\mathbf{x}'_{ij}\boldsymbol{\beta}^{(1)} + \gamma_k^{(2)}\mathbf{x}'_{ij}\boldsymbol{\beta}^{(2)} + \epsilon_{ijk}$$

is equivalent to a nominal mixed multinomial logit model when $\gamma_k^{(t)} = I(k = t)$. The fit of this model is superior to all other models that have been considered so far, which might be a reason to question the assumption that the response scale is ordinal. But inspection of $\hat{\boldsymbol{\beta}}^{(1)}$ and $\hat{\boldsymbol{\beta}}^{(2)}$ shows that with the exception of $\beta_2^{(1)}$ and $\beta_2^{(2)}$ (situation type) these vectors are nearly proportional such that it might be assumed that $\beta_1^{(1)}/\beta_1^{(2)} = \beta_3^{(1)}/\beta_3^{(2)} = \beta_4^{(1)}/\beta_4^{(2)}$, which is consistent with a two-dimensional stereotype model where $\boldsymbol{\beta}^{(2)} = (0, \beta_2^{(2)}, 0, 0)'$ with $\gamma_1^{(2)} = \gamma_2^{(2)} = 0$ and $\gamma_3^{(2)} = 1$ for identification. For this model the second dimension is necessary only for the explanatory variable situation type. The fit of this model is comparable to that of the nominal model.

The parameter estimates and their standard errors for the four stereotype models are shown in Table 2.

First consider the fixed category-specific effects. In the unidimensional models the ordering of the estimates of γ_1, γ_2, and γ_3 is consistent with the response scale, but their spacing does not suggest a symmetric structure. To interpret the random category-specific effects consider the

TABLE 3.
Estimated odds ratios for the verbal aggression data.

	Model and random effect(s) structure			
	Stereotype (intercept)	Stereotype (category)	Multinomial (category)	Stereotype (2D) (category)
"yes" vs. "perhaps"				
Beh. mode	0.70 (0.02)	0.69 (0.03)	0.71 (0.05)	0.75 (0.03)
Sit. type	1.85 (0.09)	1.89 (0.11)	2.63 (0.21)	2.63 (0.21)
Blaming	2.92 (0.21)	3.03 (0.25)	2.34 (0.24)	2.37 (0.22)
Expressing	1.71 (0.08)	1.74 (0.08)	1.48 (0.13)	1.55 (0.08)
"perhaps" vs. "no"				
Beh. mode	0.58 (0.03)	0.56 (0.03)	0.57 (0.04)	0.55 (0.03)
Sit. type	2.50 (0.12)	2.50 (0.12)	2.02 (0.13)	2.02 (0.12)
Blaming	4.98 (0.36)	4.93 (0.38)	5.81 (0.47)	5.76 (0.46)
Expressing	2.32 (0.13)	2.21 (0.13)	2.51 (0.19)	2.44 (0.15)

Note: Standard errors are in parentheses.

(co)variances σ_{22}, σ_{23}, and σ_{33} in Table 2 which refer to $\mathrm{Var}(\xi_{i2} - \xi_{i1})$, $\mathrm{Cov}(\xi_{i2} - \xi_{i1}, \xi_{i3} - \xi_{i1})$, and $\mathrm{Var}(\xi_{i3} - \xi_{i1})$, respectively. For comparison the stereotype model with only a random intercept can be reparametrized as $U_{ijk} = \alpha_k + \xi_{ik} + \gamma_k \mathbf{x}'_{ij}\boldsymbol{\beta} + \epsilon_{ijk}$ where $\xi_{ik} = \gamma_k \zeta_i$ so that $\sigma_{22} = \gamma_2^2 \,\mathrm{Var}(\zeta_i)$, $\sigma_{23} = \gamma_2\gamma_3 \,\mathrm{Var}(\zeta_i)$, and $\sigma_{33} = \gamma_3^2 \,\mathrm{Var}(\zeta_i)$. These (co)variances characterize the (co)variation of the usage of the "perhaps" and "yes" categories *relative* to that of the "no" category (the reference category).[4] The model without the category-specific effects constrains the variances of the ξ_{ik} to depend only on the category-specific parameters γ_k and the variance of ζ_i, while the correlations among the ξ_{ik} (and their differences) are constrained to be 1, reflecting a complete absence of category-specific random effects and individual differences in category usage. By allowing for individual differences in response category usage the other models do not constrain the covariance structure of $\boldsymbol{\xi}_i$. In all four models the variance of $\xi_{i3} - \xi_{i1}$ is significantly higher than that of $\xi_{i2} - \xi_{i1}$. For the intercept-only model this difference is due only to the fact that $\gamma_3^2 > \gamma_2^2$ since there $\sigma_{22} = \gamma_2^2 \,\mathrm{Var}(\zeta_i)$ and $\sigma_{33} = \gamma_3^2 \,\mathrm{Var}(\zeta_i)$, but in the other models the difference between σ_{22} and σ_{33} is somewhat smaller due presumably to individual differences in category usage not captured by the first model. In the three models with category-specific random effects the individual differences in the usage of the "yes" and "perhaps" categories relative to the "no" category are positively correlated ($\sigma_{23}/\sqrt{\sigma_{22}\sigma_{33}} \approx 0.73$), but not perfectly as implied by the random intercept model without category-specific random effects, suggesting random effects that are correlated but category-specific.

To consider the effects of the covariates it is helpful to examine the estimated adjacent odds ratios comparing categories $y+1$ and y conditional on the random effects. Generalizing from (2) these are defined as $\exp[c_j \sum_t (\gamma_{y+1}^{(t)} - \gamma_y^{(t)})\beta_j^{(t)}]$ where c_j is a scaling factor to standardize the coding of indicator variables ($c_1 = c_2 = 2$, $c_3 = c_4 = 3/2$). The estimated odds ratios for each covariate and each pair of response categories are given in Table 3.

Taking into account the coding of the covariates, the estimated odds ratios indicate that respondents are more inclined to agree to want to give a verbally aggressive response rather than

[4]It is important to note that because one of the ξ_{ik} must be fixed for identification, the covariance structure of $\boldsymbol{\xi}_i$ can only be interpreted in terms of the covariance structure of the *differences* among the elements of $\boldsymbol{\xi}_i$. See Böckenholt (2001b) for a discussion of this issue within the context of models for paired comparisons. The interpretation of the covariance structure of the differences among the elements of $\boldsymbol{\xi}_i$ is further complicated here because ζ_i is confounded with $\boldsymbol{\xi}_i$.

doing it (behavior mode), to agree to direct it at another rather than oneself (situation type), and to agree to a verbal response that involves blaming or expressing (behavior type). These results are statistically significant at traditional significance levels, regardless of which of the four stereotype models is considered. But note that there are some differences in the *magnitude* of the estimated effects among the four models. The largest differences in the odds ratios are between the unidimensional and multidimensional models—particularly with respect to the effects of situation type, blaming, and to a lesser extent expressing. For example, the odds of responding "perhaps" versus "no" increases by a factor of 2.5 when the situation is "blaming another" instead of "blaming self" under the unidimensional models, whereas the odds increase by a factor of 2 under the multidimensional models. Differences in these estimates are attributable to differences in the assumed covariance structure of the utilities, and in the assumed dimensionality of the model.

5. Application—An Item Response Model

This second application illustrates the use of a generalized stereotype model as an item response model for a 8-item personality inventory for vertical individualism. The items were originally developed by Singelis, Triandis, Bhawuk, and Gelfand (1995) and later refined by Triandis and Gelfand (1998) as part of a battery of four subscales designed to measure horizontal and vertical individualism and collectivism. The original scales used a 9-category ordinal response scale, but the data used here consist of $n = 589$ responses to a 7-category ordinal scale with the response categories "strongly agree/disagree," "agree/disagree," "slightly agree/disagree," and "neither agree nor disagree."[5] Consider a unidimensional item response model for these data where

$$U_{ijk} = \alpha_k + \xi_{ik} + \gamma_k(\beta_j + \lambda_j \zeta_i) + \epsilon_{ijk}.$$

The item-specific parameters β_j and λ_j then determine the location of the ideal point of a respondent as a function of ζ_i. Along the latent unidimensional continuum of the unfolding model this is $\beta_j + \lambda_j \zeta_i$ so that β_j and λ_j can be interpreted in much the same way as they are in test theory and factor analysis. Here β_j is a location parameter for the jth item (i.e., the ideal point of a respondent for the jth item when $\zeta_i = 0$), while λ_j captures the discrimination of the jth item (i.e., the change in the ideal point per unit change in ζ_i). The category-specific parameters α_k, ξ_{ik}, and γ_k capture the effects of the response categories, with ξ_{ik} representing respondent-specific usage of each response category. Because of the natural symmetry in the response scale it is reasonable to consider here a symmetric structure on the category specific parameters such that $\alpha_{jk} = \alpha_{j,7-k+1}$, $\gamma_{jk} - \gamma_{j,7-k+1} = 0$, and $\xi_{jk} = \xi_{j,7-k+1}$ where $k^* = 4$ is the reference category.

Although there are no questions regarding the structure of the scale itself, it can be interesting to consider various covariance structures for $\mathbf{\Phi}$. Let $\mathbf{\Phi}_{11}$ and $\mathbf{\Phi}_{12}$ denote the covariances of the category-specific random effects, and the covariances between the category-specific random effects and ζ_i, respectively. I fit five different models by considering three structures for the category-specific random effects (none so that $\mathbf{\Phi}_{11} = \mathbf{0}$, symmetric random category-specific effects, and asymmetric random category-specific effects), and whether or not the category-specific random effects are correlated with $\boldsymbol{\zeta}_i$ (i.e., $\mathbf{\Phi}_{12} = \mathbf{0}$ or $\mathbf{\Phi}_{12} \neq \mathbf{0}$). The dimensionality of the random effects structure due to the latent trait and the category-specific effects is four with the symmetry constraint, and seven without, so direct quadrature approximation of the likelihood

[5] The data were made available by Kristine M. Kuhn of the Department of Management and Operations at Washington State University.

TABLE 4.
Maximum likelihood estimates and standard errors for the category and item parameters for the stereotype item response models for the vertical individualism inventory.

	Constraints		
	$\mathbf{\Phi}_{11} = \mathbf{0}$	$\mathbf{\Phi}_{12} = \mathbf{0}$	None
$\alpha_1 = \alpha_7$	−1.16 (0.01)	−1.51 (0.04)	−1.56 (0.05)
$\alpha_2 = \alpha_6$	−0.26 (0.01)	−0.20 (0.03)	−0.23 (0.03)
$\alpha_3 = \alpha_5$	−0.02 (0.01)	0.09 (0.02)	0.04 (0.03)
α_4	0.00	0.00	0.00
γ_4	0.00	0.00	0.00
$\gamma_5 = -\gamma_3$	0.28 (0.01)	0.26 (0.01)	0.25 (0.01)
$\gamma_6 = -\gamma_2$	0.71 (0.01)	0.67 (0.01)	0.68 (0.01)
$\gamma_7 = -\gamma_1$	1.00	1.00	1.00
β_1	0.00	0.00	0.00
λ_1	1.30 (0.02)	1.45 (0.03)	1.45 (0.03)
β_2	1.47 (0.02)	1.68 (0.02)	1.68 (0.02)
λ_2	0.29 (0.02)	0.36 (0.02)	0.34 (0.02)
β_3	−1.25 (0.03)	−1.40 (0.03)	−1.40 (0.03)
λ_3	1.42 (0.03)	1.58 (0.02)	1.58 (0.04)
β_4	−0.41 (0.02)	−0.41 (0.02)	−0.42 (0.02)
λ_4	0.86 (0.03)	0.87 (0.04)	0.88 (0.03)
β_5	0.01 (0.02)	0.03 (0.02)	0.03 (0.02)
λ_5	0.67 (0.02)	0.68 (0.03)	0.67 (0.03)
β_6	−0.54 (0.02)	−0.58 (0.02)	−0.59 (0.02)
λ_6	0.90 (0.03)	0.95 (0.03)	0.97 (0.03)
β_7	−0.24 (0.03)	−0.26 (0.03)	−0.25 (0.03)
λ_7	1.22 (0.03)	1.31 (0.03)	1.30 (0.03)
β_8	0.80 (0.02)	0.89 (0.02)	0.89 (0.02)
λ_8	0.92 (0.03)	0.99 (0.03)	0.97 (0.03)
$\ell(\hat{\boldsymbol{\theta}}; \mathbf{y})$	−8062	−7863	−7861
AIC	16,164	15,779	15,780

Note: Standard errors are in parentheses.

is numerically cumbersome. The multinomial probit generalized stereotype model was implemented here since it is more amenable to simulation-based methods (see the Appendix). Models without the symmetry of the category-specific random effects proved to be too fragile. The maximum likelihood parameter estimates, standard errors, log-likelihoods, and the AIC for the other models are given in Table 4.

Estimates from all three models are in relatively close agreement regarding the ordering of the categories and items. Estimates of the γ_k parameters indicate that the categories are ordered as would be expected given their labels, and these estimates are quite similar between the three models. Estimates from all three models are in agreement regarding the ordering of the items with respect to their location and discrimination parameters, although the estimates of item location and discrimination parameters are somewhat attenuated for items 1, 2, 3, 7, and 8 based on the model without the category-specific random effects. Based on the relative fit of these models in terms of the AIC, a case can be made for the need for category-specific random effects to capture individual differences in category usage. Inferences are largely unchanged depending on whether or not correlations are allowed between the latent trait and the category-specific effects, but failing to specify category-specific random effects appears to attenuate the item parameters of several of the items.

TABLE 5.
The estimated identified random effects covariance/correlation matrix for the category-symmetric stereotype model.

	$\xi_{i1}-\xi_{i4}$	$\xi_{i2}-\xi_{i4}$	$\xi_{i3}-\xi_{i4}$	$\xi_{i5}-\xi_{i4}$	$\xi_{i6}-\xi_{i4}$	$\xi_{i7}-\xi_{i4}$	ζ_{i1}
$\xi_{i1}-\xi_{i4}$	1.24	0.68	0.21	0.21	0.68	1.00	0.09
$\xi_{i2}-\xi_{i4}$	0.51	0.45	0.70	0.70	1.00	0.68	0.07
$\xi_{i3}-\xi_{i4}$	0.12	0.24	0.26	1.00	0.70	0.21	0.24
$\xi_{i5}-\xi_{i4}$	0.12	0.25	0.26	0.26	0.70	0.21	0.24
$\xi_{i6}-\xi_{i4}$	0.51	0.45	0.24	0.25	0.45	0.68	0.07
$\xi_{i7}-\xi_{i4}$	1.24	0.51	0.12	0.12	0.51	1.24	0.09
ζ_{i1}	0.10	0.05	0.12	0.12	0.05	0.10	1.00

Note: Estimated correlations are above the main diagonal.

Table 5 shows the estimated (co)variances and correlations for the *identified* covariance matrix (see the Appendix) for the stereotype model with symmetric category parameters but no other constraints (the nonzero estimates for the stereotype model with $\mathbf{\Phi}_{12}$ are quite similar).

There is greater variability between respondents in the usage of the extreme response categories relative to that of the middle category than in the usage of the moderate response categories relative to that of the middle category. These individual differences are positively correlated, particularly so between adjacent response categories. This pattern is similar to that found by Johnson (2003) with mixed-effects cumulative probit models for individual differences in the extreme response style. The correlations between the category-specific random effects, relative to the middle category, and ζ_i are all small, indicating only perhaps a slight increased tendency to prefer the moderate response categories for respondents with higher vertical individualism. The relatively low absolute value of these correlations explains why the two models with and without the constraint $\mathbf{\Phi}_{12}=\mathbf{0}$ had very similar fit.

6. Conclusion

Anderson (1984) motivated the stereotype model by assuming an underlying subjective discrimination process between response categories. An alternative motivation for the model can be made by assuming a random utility maximization discrete choice process among ordered alternatives. The advantage of a random utility model over Anderson's original model is that it can account for individual differences in category usage. Its advantage over standard random utility models for discrete choice data is that it is more parsimonious because it assumes an ordering of the alternatives like that of an unfolding model. The generalized stereotype model is a mixed effects multinomial mixed model for ordinal response variables that has broad applicability.

Further methodological research concerning the generalized stereotype model might focus on the modeling of individual differences in category usage. One question is the robustness of inferences to failure to properly take into account such individual differences. For example, although there was strong evidence in both applications for the improvement of model fit when individual differences were taken into account, only in the second application did it seem to impact the parameters of interest. Another issue is the modeling of the structure of the category-specific effects. Alternative mean and covariance structures for $\boldsymbol{\xi}_i$ could be considered such as mixtures or adding a systematic component that depends on observed covariates. Substantive researchers might now (re)consider the (generalized) stereotype model as a useful means of modeling ordinal response variables. In its logistic form it falls under the generalized latent variable modeling framework of Rabe-Hesketh et al. (2004) as a special case of the multilevel polytomous logistic regression model (Skrondal & Rabe-Hesketh, 2003), and the item response modeling framework of Rijmen et al. (2003). The now generalized stereotype model provides researchers with a useful ordinal regression model that recognizes the inherent subjectivity in the response process—a motivation that is very much in the spirit of Anderson's original model.

Appendix: Computational Details

Bilinear Mixed Model Parametrization

To discuss the computational details of the generalized stereotype model it is useful to reparametrize it as a bilinear mixed model for the differences among the category utilities. Let $\mathbf{u}_i$ denote the mp-dimensional vector of utilities. The $m(p-1)$-dimensional vector of differences among the utilities can be written as $\mathbf{d}_i = (\mathbf{I}_m \otimes \mathbf{C}_{k^*})\mathbf{u}_i$, where $\mathbf{C}_{k^*}$ denotes a matrix which is a column-wise permutation of $[\mathbf{I}_{p-1}|-\mathbf{1}_{p-1}]$ such that the k^*th column of $\mathbf{C}_{k^*}$ is $-\mathbf{1}_{p-1}$.[6] It can be shown that

$$\mathbf{d}_i = \ddot{\mathbf{X}}_i(\boldsymbol{\gamma})\ddot{\boldsymbol{\beta}} + \ddot{\mathbf{Z}}_i(\boldsymbol{\gamma})\ddot{\boldsymbol{\xi}}_i + \ddot{\boldsymbol{\epsilon}}_i, \tag{7}$$

where

$$\ddot{\mathbf{X}}_i(\boldsymbol{\gamma}) = \left[(\mathbf{1}_m \otimes \mathbf{C}_{k^*}\mathbf{A})|(\mathbf{X}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)\right], \qquad \ddot{\boldsymbol{\beta}} = \left(\boldsymbol{\alpha}^{*'}, \boldsymbol{\beta}'\right)',$$

$$\ddot{\mathbf{Z}}_i(\boldsymbol{\gamma}) = \left[(\mathbf{1}_m \otimes \mathbf{C}_{k^*}\mathbf{A})|(\mathbf{Z}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)\right], \quad \ddot{\boldsymbol{\xi}}_i = \left(\boldsymbol{\xi}_i^{*'}, \boldsymbol{\zeta}_i'\right)', \quad \text{and} \quad \ddot{\boldsymbol{\epsilon}}_i = (\mathbf{I}_m \otimes \mathbf{C}_{k^*})\boldsymbol{\epsilon}_i.$$

The parametrization in (7) is like that of a classical linear mixed-effects model (e.g., Laird & Ware, 1982), although it is a *bilinear* model because the design matrices are functions of $\boldsymbol{\gamma}$. Symmetry constraints have been imposed via reparametrization by substituting $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\alpha}^*$, $\boldsymbol{\xi}_i = \mathbf{A}\boldsymbol{\xi}_i^*$, and $\boldsymbol{\gamma} = \mathbf{G}\boldsymbol{\gamma}^*$ where $\boldsymbol{\alpha}^*$, $\boldsymbol{\xi}_i^*$, and $\boldsymbol{\gamma}^*$ determine the distinct/free elements of $\boldsymbol{\alpha}$, $\boldsymbol{\xi}_i$, and $\boldsymbol{\gamma}$, respectively, and $\mathbf{A}$ and $\mathbf{G}$ are appropriately specified design matrices. For a 5-category response variable with symmetric categories, for example, symmetry is enforced by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Without symmetry constraints $\mathbf{A} = \mathbf{T} = \mathbf{I}_p$.

Identification of the Covariance Structure of the Multinomial Probit Stereotype Model

Methods for estimating multinomial probit models, as discussed in a later section, are usually based on the *marginal* distribution of $\mathbf{d}_i$. To identify the multinomial probit stereotype model constraints must be imposed on the covariance structure directly rather than on $\boldsymbol{\xi}_i$. If $\epsilon_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}_{mp})$, then from (7) it follows that the covariance matrix of $\mathbf{d}_i$ is

$$\begin{aligned}\mathrm{Cov}(\mathbf{d}_i) = {} & \mathbf{1}_m\mathbf{1}_m' \otimes \mathbf{C}_{k^*}\mathbf{A}\boldsymbol{\Phi}_{11}\mathbf{A}'\mathbf{C}_{k^*}' \\ & + (\mathbf{1}_m \otimes \mathbf{C}_{k^*}\mathbf{A}\boldsymbol{\Phi}_{12})(\mathbf{Z}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)' + (\mathbf{Z}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)(\mathbf{1}_m \otimes \mathbf{C}_{k^*}\mathbf{A}\boldsymbol{\Phi}_{12})' \\ & + (\mathbf{Z}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)\boldsymbol{\Phi}_{22}(\mathbf{Z}_i \otimes \mathbf{C}_{k^*}\mathbf{G}\boldsymbol{\gamma}^*)' + \mathbf{I}_m \otimes \sigma^2(\mathbf{1}_{p-1}\mathbf{1}_{p-1}' + \mathbf{I}_{p-1}),\end{aligned}$$

[6] Throughout this discussion let $\mathbf{I}_d$ denote an identity matrix of order d and let $\mathbf{1}_d$ denote a d-dimensional unit vector.

where $\mathbf{\Phi}_{11}$ is the covariance matrix of $\boldsymbol{\xi}_i^*$, $\mathbf{\Phi}_{12}$ is the covariance matrix between $\boldsymbol{\xi}_i^*$ and $\boldsymbol{\zeta}_i$, and $\mathbf{\Phi}_{22}$ is the covariance matrix of $\boldsymbol{\zeta}_i$. $\mathbf{\Phi}_{11}$ and $\mathbf{\Phi}_{12}$ are not identified because the covariance structure is unchanged after replacing $\mathbf{\Phi}_{11}$ and $\mathbf{\Phi}_{12}$ by $\mathbf{\Phi}_{11} + \mathbf{f}\mathbf{1}_p' + \mathbf{1}_p\mathbf{f}'$ and $\mathbf{\Phi}_{12} = \mathbf{\Phi}_{12} + \mathbf{1}_p\mathbf{g}'$, respectively, where $\mathbf{f}$ and $\mathbf{g}$ are arbitrary vectors of appropriate length (see Tsai, 2003, for a related identification problem in the context of models for paired comparisons). Thus $\mathbf{\Phi}_{11}$ and $\mathbf{\Phi}_{12}$ are identified only up to the covariance matrix

$$\operatorname{Cov}\begin{pmatrix}\mathbf{C}_{k^*}\mathbf{A}\boldsymbol{\xi}_i^* \\ \boldsymbol{\zeta}_i\end{pmatrix} = \begin{pmatrix}\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{11}\mathbf{A}'\mathbf{C}_{k^*}' & \mathbf{\Phi}_{21}\mathbf{A}'\mathbf{C}_{k^*}' \\ \mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{12} & \mathbf{\Phi}_{22}\end{pmatrix}. \tag{8}$$

Symmetry Constraints for the Multinomial Probit Stereotype Model

The structures of the identified covariance matrices $\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{11}\mathbf{A}'\mathbf{C}_{k^*}'$ and $\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{12}$ defined in the previous section are constrained if symmetry constraints are imposed on $\boldsymbol{\xi}_i$. Specifically, the symmetry constraints imply that $\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{11}\mathbf{A}'\mathbf{C}_{k^*}'$ is symmetric with respect to its *minor* diagonal, and $\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{12}$ is "row-symmetric" in the sense that its kth row equals its $(p-k+1)$th row. This symmetry also implies that the identified covariance matrix (8) is positive *semi*definite because there are unit correlations corresponding to the covariances on the minor diagonal of $\mathbf{C}_{k^*}\mathbf{A}\mathbf{\Phi}_{11}\mathbf{A}'\mathbf{C}_{k^*}'$.

Inference

Let $\boldsymbol{\theta}$ and $\mathbf{y}$ denote the vectors of unknown parameters and observed responses, respectively, and let $\mathbf{R}_{ij}$ denote a matrix such that $\mathbf{R}_{ij} = -\mathbf{I}_{p-1}$ if $Y_{ij} = k^*$, a columnwise permutation of

$$\mathbf{P} = \begin{pmatrix}-1 & \mathbf{0}_{p-2}' \\ -\mathbf{1}_{p-2} & \mathbf{I}_{p-2}\end{pmatrix}$$

such that the yth column of $\mathbf{P}$ is $-\mathbf{1}_{p-1}$ if $Y_{ij} < k^*$, or a columnwise permutation of $\mathbf{P}$ such that the $(y+1)$th column of $\mathbf{P}$ is $-\mathbf{1}_{p-1}$ if $Y_{ij} > k^*$ where k^* is the specified reference category. Provided that the responses are conditionally independent given $\ddot{\boldsymbol{\zeta}}_i$, the likelihood function for the generalized stereotype model is

$$L(\boldsymbol{\theta};\mathbf{y}) = \prod_{i=1}^{n}\int\cdots\int \phi_{p+q-1}(\ddot{\boldsymbol{\zeta}}_i;\mathbf{0},\mathbf{\Phi})\prod_{j=1}^{m} P(\mathbf{R}_{ij}\mathbf{d}_{ij} < \mathbf{0}|\ddot{\boldsymbol{\zeta}}_i)\,\mathbf{d}\ddot{\boldsymbol{\zeta}}_i, \tag{9}$$

where ϕ_{p+q-1} denotes the probability density function of a $(p+q-1)$-dimensional normal random vector with given mean vector and covariance matrix. The term $P(\mathbf{R}_{ij}\mathbf{d}_{ij} < \mathbf{0}|\ddot{\boldsymbol{\zeta}}_i)$ involves a $(p-1)$-dimensional integral. For the mixed multinomial logit model this can be expressed in closed form as

$$P(\mathbf{R}_{ij}\mathbf{d}_{ij} < \mathbf{0}|\ddot{\boldsymbol{\zeta}}_i) \propto \begin{cases}\exp[E(U_{ijk} - U_{ijk'}|\boldsymbol{\xi}_i,\boldsymbol{\zeta}_i)] & \text{if } Y_{ij} = k \neq k', \\ 1 & \text{otherwise.}\end{cases}$$

But for the multinomial probit model

$$P(\mathbf{R}_{ij}\mathbf{d}_{ij} < \mathbf{0}|\ddot{\boldsymbol{\zeta}}_i) = F_{p-1}\big[z;\mathbf{R}_{ij}\mathbf{C}_{k^*}\boldsymbol{\alpha}^* + \mathbf{R}_{ij}\mathbf{C}_{k^*}\boldsymbol{\gamma}^*\otimes(\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_i),\sigma^2\mathbf{R}_{ij}\mathbf{C}_{k^*}\mathbf{C}_{k^*}'\mathbf{R}_{ij}'\big],$$

where F_{p-1} denotes the distribution function of a $(p-1)$-dimensional normal random vector with given mean vector and covariance matrix. The marginal distribution of $\mathbf{d}_i = (\mathbf{I}_m \otimes \mathbf{C}_{k^*})\mathbf{u}_i$ is also normal so that the likelihood can be written as

$$L(\boldsymbol{\theta};\mathbf{y}) = \prod_{i=1}^{n} F_{m(p-1)}[z;\boldsymbol{\mu}_i(\boldsymbol{\theta}),\mathbf{\Sigma}_i(\boldsymbol{\theta})]\,\mathbf{d}z, \tag{10}$$

where

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \mathbf{R}_i \ddot{\mathbf{X}}_i(\boldsymbol{\gamma})\ddot{\boldsymbol{\beta}} \quad \text{and} \quad \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \mathbf{R}_i\big[\ddot{\mathbf{Z}}_i(\boldsymbol{\gamma})\boldsymbol{\Phi}\ddot{\mathbf{Z}}_i'(\boldsymbol{\gamma}) + \mathbf{I}_m \otimes \sigma^2(\mathbf{1}_{p-1}\mathbf{1}_{p-1}' + \mathbf{I}_{p-1})\big]\mathbf{R}_i',$$

respectively, and where $\mathbf{R}_i$ is a block-diagonal matrix formed from $\mathbf{R}_{i1}, \mathbf{R}_{i2}, \ldots, \mathbf{R}_{im}$. Because of the multiple integrals optimization of the likelihood can be problematic computationally. In (9) the dimension of the outer integral is $p + q - 1$. Provided the dimension of the outer integral is not too large it can be approximated using quadrature, and for the probit model the integrand can be approximated numerically (e.g., Genz, 1992; Geweke, Keane, & Runkle, 1994). But directly numerical optimization of (9) or (10) is computationally feasible only if the dimension(s) of integration(s) is/are relatively low, which is not necessarily the case if there are more than a few categories and category-specific random effects are assumed.

In cases where the dimension of integration is prohibitively high for quadrature, a more viable approach is to use simulation-based methods. The multinomial probit stereotype model is particularly amenable to a variety of approaches that have been developed for multinomial probit models. These include simulated maximum likelihood (Hajivassiliou & Ruud, 1994), simulated method of moments (McFadden, 1989), the method of simulated scores (Hajivassiliou & McFadden, 1998), Monte Carlo expectation-maximization (MCEM) algorithms (Natarajan, McCulloch, & Kiefer, 2000), and Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference (McCulloch & Rossi, 1994). The inferences reported in the IRT application in this paper were done using a minor extension of the MCEM algorithm outlined by Natarajan et al. to obtain maximum likelihood estimates, a Monte Carlo approximation to the observed information matrix to obtain standard errors (e.g., Tanner, 1996), and Geweke's (1991) probability sampler to obtain an approximation of the likelihood at the estimates.

References

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

Alwin, D.F., & Krosnick, J.A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly*, *49*, 535–552.

Anderson, J.A. (1984). Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society. Series B*, *46*, 1–30.

Baumgartner, H., & Steenkamp, J.E.M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*, 143–156.

Böckenholt, U. (2001a). Thresholds and intransitivities in pairwise judgments: A multilevel analysis. *Journal of Educational and Behavioral Statistics*, *26*, 269–282.

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

Clogg, C.C., & Shihadeh, E.S. (1994). *Statistical models for ordinal variables*. Sage: Thousand Oaks.

Cronbach, L.J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*, 475–494.

Cronbach, L.J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*, 3–31.

De Boeck, P., & Wilson, M. Eds. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Fielding, A., Yang, M., & Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Statistical Modelling*, *3*, 127–153.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *1*, 141–150.

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-*t* distributions subject to linear constraints and the evaluation of constraint probabilities. In E.M. Keramidas (Ed.), *Computer science and statistics: Proceedings of the twenty-third symposium on the interface* (pp. 571–578). Fairfax: International Foundation of North America.

Geweke, J., Keane, M., & Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics*, *76*, 609–632.

Geweke, J., Keane, M.P., & Runkle, D.E. (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, *80*, 125–165.

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine*, *13*, 1665–1677.

Hajivassiliou, V., & McFadden, D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, *66*, 863–896.

Hajivassiliou, V., & Ruud, P. (1994). Classical estimation methods for LDV models using simulation. In R. Engle & D. McFadden (Eds.), *Handbook of Econometrics* (pp. 2383–2441). Amsterdam: North-Holland.

Hamilton, D.L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192–203.

Holtbrügge, W., & Schuhmacher, M. (1991). A comparison of regression models for the analysis of ordered categorical data. *Applied Statistics*, *40*, 249–259.

Johnson, T.R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, *68*, 563–583.

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

Laird, N.M., & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lenk, P., Wedel, M., & Böckenholt, U. (2006). Bayesian estimation of circumplex models subject to prior theory constraints and scale-usage bias. *Psychometrika*, *71*, 33–55.

Liu, I., & Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, *14*, 1–73.

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

McCulloch, R., & Rossi, P.E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 207–240.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics*. New York: Academic Press.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*, 995–1027.

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

Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, *19*, 91–100.

Messick, S. (1991). Psychology and methodology of response styles. In R.E. Snow & D.E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 161–200). Hillsdale: Lawrence Erlbaum.

Natarajan, R., McCulloch, C.E., & Kiefer, N.M. (2000). A Monte Carlo EM method for estimating multinomial probabilities. *Computational Statistics & Data Analysis*, *34*, 33–50.

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

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, *69*, 167–190.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–205.

Singelis, T.M., Triandis, H.C., Bhawuk, D.P.S., & Gelfand, M.J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, *29*, 240–275.

Skrondal, A., & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, *68*, 267–287.

Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, *52*, 371–392.

Tanner, M. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (3rd ed.). New York: Springer.

Triandis, H.C., & Gelfand, M.J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*, 118–128.

Tsai, R.-C. (2003). Remarks on the identifiability of Thurstonian paired comparison models under multiple judgment. *Psychometrika*, *68*, 361–372.

Tsai, R.-C., & Böckenholt, U. (2001). Maximum likelihood estimation of factor and ideal point models for paired comparisons. *Journal of Mathematical Psychology*, *45*, 795–811.

Tutz, G., & Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, *22*, 537–557.

Vansteelandt, K. (2000). *Formal models for contextualized personality psychology*. Unpublished doctoral dissertation, K.U. Leuven, Belgium

Weeks, M. (1997). The multinomial probit model revisited: A discussion of parameter estimability. *identification and specification testing, Journal of Economic Surveys*, *11*, 297–320.

Wolfe, R., & Firth, D. (2002). Modelling subjective use of an ordinal response scale in a many period crossover experiment. *Journal of the Royal Statistical Society. Series C*, *51*, 245–255.

Manuscript received 18 NOV 2005
Final version received 8 MAR 2007
Published Online Date: 28 JUL 2007