BAYESIAN ANALYSIS OF
ORDER-STATISTICS MODELS FOR RANKING DATA

PHILIP L.H. YU

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF HONG KONG

In this paper, a class of probability models for ranking data, the order-statistics models, is investigated. We extend the usual normal order-statistics model into one where the underlying random variables follow a multivariate normal distribution. Bayesian approach and the Gibbs sampling technique are used for parameter estimation. In addition, methods to assess the adequacy of model fit are introduced. Robustness of the model is studied by considering a multivariate-t distribution. The proposed method is applied to analyze the presidential election data of the American Psychological Association (APA).

Key words: data augmentation, Gibbs sampling, order-statistics model, ranking data.

1. Introduction

The rankings of items are very common in everyday life. Gamblers want to know the rankings of the horses in horse-racing competitions; companies want to know consumers' order preference on products in the market; social and political leaders want to know which candidate will get more support in the election. Data consisting of rankings appear in many fields such as psychology, sociology, marketing, sports and econometrics.

In the literature, modeling of ranking data has received a lot of attention and many models have been proposed over the years. They could be categorized into four classes: (i) order-statistics models, (ii) distance-based models, (iii) paired-comparison models, and (iv) multistage models. A more comprehensive discussion on these probability ranking models can be found in the review paper by Critchlow, Fligner and Verducci (1991). Recently, a monograph on modeling and analyzing ranking data was published by Marden (1995). Applications of these models to real life situations can be found in the literature; for example, Koop and Poirier (1994) use rank-ordered logit models to analyze voter preferences, and Lo, Bacon-Shone and Busche (1995) apply ranking probability models to racetrack betting.

Among the four classes of probability models for ranking data, the class of order-statistics models has the longest history in the statistical and psychological literature. Dated back to 1927, Thurstone published the famous paper "A law of comparative judgement" in *Psychological Review* in which ranking of two items was considered. The basic idea behind this approach is that the judge may have tastes that fluctuate from one instant to another according to the judge's perception of each item which is not perfectly predictable and hence is a random variable. The ordering of these random variables then determines the judge's ranking of items.

Mathematically, a *ranking* of k items is defined as a permutation function $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ from $1, 2, \ldots, k$ onto $1, 2, \ldots, k$, where π_i is the rank given to item i, $i = 1, 2, \ldots, k$, and larger ranks refer to the more preferred items.

Let y_i, a random variable, be the utility of item i given by a judge ($i = 1, \ldots, k$). Then the *order statistics model* assigns the ranking π with the following probability:

$$P(\pi) = P(y_{i_1} < y_{i_2} < \cdots < y_{i_k}), \quad \pi \in \Omega, \tag{1}$$

where $i_r = t$ if and only if $\pi_t = r$, and Ω is the set of all $k!$ possible rankings. Critchlow et al. (1991) observed that if the y_i's are allowed to have arbitrary dependencies, any probability distribution on rankings can be expressed as in (1). To simplify the model, some probabilistic structures on y's are assumed. The most common one is to assume that the y_i's are independent with c.d.f. $F_i(y) = F(y - \mu_i)$. Such models are referred to as *Thurstone order-statistics models* (see Yellott 1977; and Critchlow et. al. 1991). Two famous Thurstone models that have been studied extensively in the literature are the Thurstone(1927)–Daniels(1950)–Mosteller(1951) model where F is normal, and the Bradley–Terry(1952)–Luce(1959) (BTL) model where F is just extreme value distribution, that is, $F(y) = 1 - \exp(-\exp(y))$. Because the ranking probability in (1) under the BTL model has a close form, most applications and extensions are based on the BTL model (see, for example, Beggs, Cardell, & Hausman, 1981; and Chapman & Staelin, 1982 in which the term "rank ordered logit model" is used). Henery (1983) and Stern (1990a, 1990b) generalized the BTL model to the order-statistics model with latent utility $y_i = \log(u_i)$, where $u_i \sim$ Gamma$(r, \lambda_i = \exp(\mu_i))$. Properties of the Thurstone order-statistics model are discussed in Henery (1981) and Critchlow et.al.

Despite the fact that the ranking probability (1) under BTL model has a close form, sometimes the BTL model is not suitable for use in practice. There are two main reasons. Firstly, it satisfies the independence of irrelevant alternatives (IIA) property (Tversky, 1972) which says that the conditional distribution of the utility of each item is independent of the ranking of other items. Clearly, this property is unrealistic when some items are very similar or even substitutes for the others. Secondly, when it is applied to some real data, it was shown that it does not fit the data very well (see, for example, Böckenholt, 1993; Brook & Upton, 1974; and Tallis & Dansie, 1983). Therefore, to overcome these problems, dependency structures other than those in the Thurstone order-statistics models are required. One simple dependency structure on $\mathbf{y} = (y_1, \ldots, y_k)'$ is to assume that $\mathbf{y}$ follows a multivariate normal distribution. We refer such model as a *multivariate normal order-statistics (MVNOS) model.*

Although the MVNOS model is a reasonable approach in modeling ranking data, it is not popularly used. The main reason is that estimation of parameters using the classical maximum likelihood (ML) approach is not practically feasible. The exact calculation of the likelihood function needs high-dimensional numerical integration which is too computationally involved for practical applications, especially for k being large. Hence, most applications of the MVNOS model are related to the ranking of a few items. For instance, Thurstone (1927) suggested using the MVNOS model with some special covariance structures to analyze the paired comparison data (i.e. $k = 2$). Dansie (1986) used this model to investigate the preference for $k = 4$ cars. To the best of my knowledge, no application on analyzing ranking data for $k \geq 5$ items have been found in the literature.

One major breakthrough in overcoming the computational complexity for the analysis of ranking data is to adopt a Bayesian approach and use the Monte Carlo Markov Chain (MCMC) methods (See Besag, Green, Higdon & Mengersen, 1995; Tierney, 1994; and the references therein). These powerful techniques have been demonstrated to be extremely useful in fitting the BTL model (Koop & Poirier, 1994). The aim of this paper is to apply such technique to estimate the parameters of the MVNOS model. We will see later that the proposed estimation method is generally applicable for the ranking of a batch of items of any size.

This paper is organized as follows. Section 2 defines formally the MVNOS model. Covariates can be incorporated into the model, and the identifiability of the parameters will also be considered. Section 3 shows how to apply the MCMC technique to draw sample from the joint posterior distribution of the unknown parameters of the new model. Section 4 introduces

some methods to assess the adequacy of model fit. Section 5 constitutes a simulation study to demonstrate the usefulness of the MCMC technique. The methodology is applied to analyze the data for the 1980 American Psychological Association presidential election in section 6. Section 7 discusses the robustness of the proposed model. Some concluding remarks are given in section 8.

2. The Multivariate Normal Order-Statistics (MVNOS) Models

Suppose a random sample of n judges is asked to rank k items. For $j = 1, \ldots, n$, let $\boldsymbol{\pi}_j = (\pi_{1j}, \ldots, \pi_{kj})$ be the ranking of the k items assigned by judge j and $\langle i_1, \ldots, i_k \rangle$ be the ordering of the k items corresponding to the ranking $\boldsymbol{\pi}_j$. The multivariate normal order-statistics (MVNOS) model assigns the ranking $\boldsymbol{\pi}_j$ with the following probability:

$$P(\boldsymbol{\pi}_j) = P(y_{i_1, j} < y_{i_2, j} < \cdots < y_{i_k, j}), \quad i = 1, \ldots, k, \ j = 1, \ldots, n, \tag{2}$$

where the utility vector $\mathbf{y}_j = (y_{1j}, \ldots, y_{kj})'$ of judge j satisfies

$$\mathbf{y}_j = \boldsymbol{\mu}_j + \mathbf{e}_j \tag{3}$$

with $\boldsymbol{\mu}_j = (\mu_{1j}, \ldots, \mu_{kj})'$ representing the mean utility vector for judge j and

$$\mathbf{e}_j \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V}). \tag{4}$$

Note that in (4), the diagonal elements of $\mathbf{V}$, v_{rr}, measure the dispersion of the latent utilities y_{rj}, for all j while its off-diagonals, $v_{rs} (r \neq s)$, measure the association between the latent utilities y_{rj} and y_{sj}, for all j. For example if $v_{12} > 0$, it is more likely that items 1 and 2 are close in their rankings; otherwise, it is more likely that their rankings are far apart.

2.1. The MVNOS Model with Covariates

When some covariates are associated with the judges and items, it is natural to impose the following linear model for $\boldsymbol{\mu}_j$:

$$\boldsymbol{\mu}_j = \mathbf{Z}_j \boldsymbol{\beta}, \quad j = 1, \ldots, n, \tag{5}$$

where $\mathbf{Z}_j$ is a $k \times p$ matrix of covariates associated with judge j and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters. For example in a marketing survey, respondents are asked to rank products according to their preference. Usually, apart from the rankings given by the respondents, some socio-economic variables $(\mathbf{s}_j)$ about the respondents and the attributes $(\mathbf{a}_i)$ of the products are also available. Then one may study the heterogeneity of the preference due to these variables by assuming the following model

$$\mu_{ij} = \mathbf{a}_i' \boldsymbol{\gamma} + \mathbf{s}_j' \boldsymbol{\delta}_i, \quad i = 1, \ldots, k, \ j = 1, \ldots, n. \tag{6}$$

The parameter vector $\boldsymbol{\gamma}$ represents the attribute effect common to all the respondents while the vector $\boldsymbol{\delta}_i$ represents the respondents' socio-economic background which may affect their preference for item i. It is easily seen that (6) is a particular case of model (5) when

$$\mathbf{Z}_j = \begin{pmatrix} \mathbf{a}_1' & \mathbf{s}_j' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{a}_2' & \mathbf{0} & \mathbf{s}_j' & & \\ \vdots & \vdots & & \ddots & \\ \mathbf{a}_k' & \mathbf{0} & & & \mathbf{s}_j' \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_k \end{pmatrix}.$$

Critchlow and Fligner (1993) have considered the normal order-statistics model with item covariates. In this paper, we generalize their model to include the judges' covariates.

2.2. The Likelihood Function of the MVNOS Model

Note that one can add an arbitrary constant (location shift) or multiply a positive constant (scale shift) to both sides of (3) while leaving the ranking probability in (2) unchanged. The location-shift problem is commonly dealt with by subtracting the last element of (3) from the other $k - 1$ elements. Therefore, after substituting (5) into (3), (3) and (4) become

$$\mathbf{w}_j = \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j \quad j = 1, \ldots, n, \tag{7}$$

$$\boldsymbol{\varepsilon}_j \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{8}$$

where $w_{ij} = y_{ij} - y_{kj}$, $\mathbf{X}_{rij} = \mathbf{Z}_{rij} - \mathbf{Z}_{rkj}$, $\varepsilon_{ij} = e_{ij} - e_{kj}$ and $\boldsymbol{\Sigma} = [\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]\mathbf{V}[\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]'$. Here, $\mathbf{I}$ denotes an identity matrix and $\mathbf{1}$ denotes a vector of 1's. Then the ranking $\boldsymbol{\pi}_j$ with respective ordering $\langle i_1, \ldots, i_k \rangle$ corresponds to the event,

$$E_j = \{\mathbf{w}_j : w_{i_1,j} < \cdots < w_{i_{r-1},j} < 0 < w_{i_{r+1},j} < \cdots < w_{i_k,j}\} \quad \text{whenever } i_r = k. \tag{9}$$

For the sake of simplicity, we use the convention $w_{i_0,j} = -\infty$ and $w_{i_{k+1},j} = +\infty$. Notice that the scale-shift problem still exists in the model given by (7), (8) and (9). The usual way to tackle this problem is to put a constraint on $\boldsymbol{\Sigma}$ such as $\sigma_{11} = 1$.

Since the judges are drawn randomly, the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is given by the expression

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \prod_{j=1}^{n} P(E_j). \tag{10}$$

Unless the number of items (k) is small, a direct maximization of the above likelihood function is computational demanding and numerically unstable because the evaluation of the likelihood function by quadrature of a $k - 1$-dimensional integral is not accurate, and there is no close form solution. Hence, most applications of the MVNOS model are related to the ranking of a few number of items. For instance, Dansie (1986) used this model to investigate the preference for $k = 4$ cars.

2.3. Identifiability of Model Parameters

Because rankings of items only depend on utility differences, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ (with σ_{11} fixed) are estimable, but the original parameters $\boldsymbol{\mu}_j$ and $\mathbf{V}$ cannot be fully identified. For example, as given by one of the referees, suppose $k = 3$ and $\boldsymbol{\mu}_j = \boldsymbol{\mu}$, $j = 1, 2, \ldots, n$. Then the following three sets of parameters under the MVNOS model lead to the same ranking probabilities:

Set A:	Set B:	Set C:

$$\boldsymbol{\mu}_A = (1, 0, -1)' \qquad \boldsymbol{\mu}_B = (-1, -2, -3)' \qquad \boldsymbol{\mu}_C = (1, 0, -1)'$$

$$\mathbf{V}_A = \begin{pmatrix} 1 & 0 & 0.2 \\ 0 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix} \quad \mathbf{V}_B = \begin{pmatrix} 0.4 & 0 & 0.4 \\ 0 & 1.6 & 1.6 \\ 0.4 & 1.6 & 2 \end{pmatrix} \quad \mathbf{V}_C = \begin{pmatrix} 0.756 & -0.444 & -0.311 \\ -0.444 & 0.356 & 0.089 \\ -0.311 & 0.089 & 0.222 \end{pmatrix}$$

This is because they all end up with the same joint distribution of the utility differences $y_1 - y_3$ and $y_2 - y_3$:

$$N \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1.6 & 0 \\ 0 & 0.4 \end{bmatrix} \right).$$

Generally speaking, the parameter $\boldsymbol{\beta}$ can be identified in the presence of covariates $\mathbf{X}_j$. However, when there are no covariates, i.e., $\boldsymbol{\mu}_j = \boldsymbol{\mu}$, the values of the μ_i's will be determined only within a location shift. This indeterminacy can be eliminated by imposing one constraint on the μ_i's, say, $\mu_k = 0$.

The major identification problem is due to indeterminacy of the variance-covariance matrix $\mathbf{V}$ of the utilities. Owing to the fact that the utilities $y_{ij}, i = 1, \ldots, k$, are invariant under any scale shift of $\mathbf{V}$ and any transformation of $\mathbf{V}$ of the form

$$\mathbf{V} \longrightarrow \mathbf{V} + \mathbf{c}\mathbf{1}_k' + \mathbf{1}_k\mathbf{c}', \qquad (11)$$

for any constant vector $\mathbf{c}$ (Arbuckle & Nugent, 1973), $\mathbf{V}$ can never be identified unless it is structured. In the previous example, it can be seen that $\mathbf{V}_A$ can be transformed to $\mathbf{V}_B$ and $\mathbf{V}_C$ by applying (11) with $\mathbf{c} = (-0.3, 0.3, 0.5)'$ and $\mathbf{c} = (-0.122, -0.322, -0.389)'$, respectively. This identifiability problem is well-known in the context of Thurstone's order statistics models and multinomial probit models (Arbuckle & Nugent, 1973; Bunch, 1991; Dansie, 1985; and Yai et al., 1997). Various solutions by imposing constraints to the covariance matrix $\mathbf{V}$ has been proposed in the literature and are summarized in Table 1.

Among the existing methods given in Table 1, Chintagunta's method provides the most flexible form for $\mathbf{V}$. However, it assumes that the utilities must be correlated and this is not always true. In this paper, we suggest a new approach as shown in the last row of Table 1. Its advantages over the existing methods are that it always produces a nonsingular matrix which includes the identity matrix (after a scale-shift) as a special case. In addition, it is easy to show that this nonsingular matrix is an invariant transformation of the matrix used by Chintagunta (1992) under the transformation (11) with $\mathbf{c} = \frac{1}{2k}\mathbf{1}_k$. See the application described in section 6 for comparison of the two methods.

In the next section, we shall demonstrate how to apply the Bayesian method and the MCMC techniques in estimating the parameters of the MVNOS model.

TABLE 1.
Various Solutions for Identifying $\mathbf{V}$

Example of reference	Constraints on $\mathbf{V}$	Is $\mathbf{V}$ full rank?	Conditions under which the method can be used
Albright, Lerman and Manski (1977)	Set the utility $y_{kj} = 0$, for all j, i.e. $v_{ik} = v_{ki} = 0$, $(i = 1, \cdots, k)$, and $v_{11} = 1$	No	item k is the universal benchmark and thus its utility is uncorrelated with other utilities
Dansie (1986)	Set $v_{ii} = 1 (i = 1, \cdots, k)$, and $v_{12} = 0$	Yes	similar variation of utilities, and utilities of items 1 and 2 are uncorrelated
Chintagunta (1992)	Set $\sigma_{11} = 1$ and each column of $\mathbf{V}$ sum to zero. Thus $\mathbf{V} = \mathbf{B}^-\Sigma(\mathbf{B}')^-$, where $\mathbf{B} = [\mathbf{I}_{k-1}, -\mathbf{1}_{k-1}]$	No	flexible form but the utilities must be correlated
Our approach	Set $\sigma_{11} = 1$ and each column of $\mathbf{V}$ sum to 1. Thus $\mathbf{V} = \mathbf{A}^{-1}\begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & k \end{bmatrix}(\mathbf{A}')^{-1}$ where $\mathbf{A} = \begin{bmatrix} \mathbf{I}_{k-1} & -\mathbf{1}_{k-1} \\ \mathbf{1}'_{k-1} & 1 \end{bmatrix}$	Yes	flexible form

3. Bayesian Analysis of the MVNOS Model

3.1. Prior Distribution

In a Bayesian approach, the first step is to specify the prior distributions of the identified parameters. As mentioned in section 2, one constraint on $\mathbf{\Sigma}$ could be imposed in order to fix the scale and hence to identify all the parameters. However when this constraint is added, the usual Wishart prior distribution for the constrained $\mathbf{\Sigma}$ could not be used and thus the problem becomes nonstandard. In this paper, we follow similar lines as in McCulloch and Rossi (1994). Let $f(\boldsymbol{\beta}, \mathbf{\Sigma})$ denote the joint prior distribution of $(\boldsymbol{\beta}, \mathbf{\Sigma})$. Then the posterior density of $(\boldsymbol{\beta}, \mathbf{\Sigma})$ is

$$f(\boldsymbol{\beta}, \mathbf{\Sigma}|\mathbf{\Pi}) \propto L(\boldsymbol{\beta}, \mathbf{\Sigma}) f(\boldsymbol{\beta}, \mathbf{\Sigma}), \tag{12}$$

where $\mathbf{\Pi} = \{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n\}$. We use a normal prior on $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{A}_0^{-1}),$$

and an independent Wishart prior on $\mathbf{G} \equiv \mathbf{\Sigma}^{-1}$,

$$\mathbf{G} \equiv \mathbf{\Sigma}^{-1} \sim W_{k-1}(\alpha, \mathbf{P}).$$

Note that our parameterization of the Wishart distribution is such that $E(\mathbf{\Sigma}^{-1}) = \alpha \mathbf{P}^{-1}$.

It is clear that (12) is very intractable for Bayesian calculations as it involves the complicated likelihood defined by (10). This problem can easily be solved by using Gibbs sampling with data augmentation (Tanner & Wong, 1987). We augment the parameter $(\boldsymbol{\beta}, \mathbf{\Sigma})$ by the latent variable $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$. Now, the joint posterior density of $(\boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{W})$ is

$$f(\boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{W}|\mathbf{\Pi}) \propto f(\mathbf{\Pi}|\mathbf{W}) f(\mathbf{W}|\boldsymbol{\beta}, \mathbf{\Sigma}) f(\boldsymbol{\beta}, \mathbf{\Sigma}), \tag{13}$$

which allows us to sample from the full conditional posterior distributions. The details will be given in the next section.

3.2. Gibbs Sampling Algorithm for the MVNOS Model

The Gibbs sampling algorithm for the MVNOS model, consisting of drawing samples consecutively from the full conditional posterior distributions, is as follows:

1. draw w_{ij} from $f(w_{ij}|w_{-i,j}, \boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{\Pi})$, for $i = 1, \ldots, k-1, j = 1, \ldots, n$,
2. draw $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\mathbf{\Sigma}, \mathbf{W}, \mathbf{\Pi}) \propto f(\boldsymbol{\beta}|\mathbf{\Sigma}, \mathbf{W})$ and
3. draw $\mathbf{\Sigma}$ from $f(\mathbf{\Sigma}|\boldsymbol{\beta}, \mathbf{W}, \mathbf{\Pi}) \propto f(\mathbf{\Sigma}|\boldsymbol{\beta}, \mathbf{W})$,

where $w_{-i,j}$ is $\mathbf{w}_j$ with w_{ij} deleted. To write down the above densities explicitly, we follow closely the notation used in McCulloch and Rossi (1994). Let

$$\mathbf{x}'_{ij} \quad \text{be the } i\text{-th row of } \mathbf{X}_j,$$
$$\mathbf{X}_{-i,j} \quad \text{be } \mathbf{X}_j \text{ with the } i\text{-th row deleted},$$
$$\mathbf{g}_{-i,i} \quad \text{be the } i\text{-th column of } \mathbf{G} \text{ with } g_{ii} \text{ deleted}.$$

Suppose $\langle i_1, \ldots, i_k \rangle$ is the ordering of items corresponding to their ranks $\boldsymbol{\pi}_j = (\pi_{1j}, \ldots, \pi_{kj})$. Then $\pi_{tj} = r$ if and only if $i_r = t$.

In Step 1, we have,

$$w_{ij}|w_{-i,j}, \boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{\Pi} \sim N(m_{ij}, \tau_{ij}^2)$$

subjecting to $w_{i_{r-1}j} < w_{ij} < w_{i_{r+1}j}$ whenever $\pi_{ij} = r$,

$$\tag{14}$$

where $m_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} - g_{ii}^{-1}\mathbf{g}'_{-i,i}(w_{-i,j} - \mathbf{X}_{-i,j}\boldsymbol{\beta})$ and $\tau_{ij}^2 = g_{ii}^{-1}$. To simulate from this truncated univariate normal distribution, we can adopt the inverse CDF method (see Devroye, 1986 for the details of data generation).

Now we go to Steps 2 and 3. Since we are conditioning on $\mathbf{W}$, the MVNOS model is simply a standard Bayesian linear model set-up (Box & Tiao, 1973). Therefore, the full conditional distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}|\boldsymbol{\Sigma}, \mathbf{W} \sim N_p(\boldsymbol{\beta}_1, \mathbf{A}_1^{-1}), \tag{15}$$

where $\mathbf{A}_1 = \mathbf{A}_0 + \sum_{j=1}^n \mathbf{X}'_j\boldsymbol{\Sigma}^{-1}\mathbf{X}_j$ and $\boldsymbol{\beta}_1 = \mathbf{A}_1^{-1}(\mathbf{A}_0\boldsymbol{\beta}_0 + \sum_{j=1}^n \mathbf{X}'_j\boldsymbol{\Sigma}^{-1}\mathbf{w}_j)$. Finally, the full conditional distribution of $\boldsymbol{\Sigma}$ is such that $\boldsymbol{\Sigma} = \mathbf{G}^{-1}$ with

$$\mathbf{G}|\boldsymbol{\beta}, \mathbf{W} \sim W_{k-1}\left(\alpha + n, \mathbf{P} + \sum_{j=1}^n (\mathbf{w}_j - \mathbf{X}'_j\boldsymbol{\beta})(\mathbf{w}_j - \mathbf{X}'_j\boldsymbol{\beta})'\right). \tag{16}$$

To simulate a sample from this Wishart distribution, we employ a simple transformation method suggested by Johnson (1987, p. 204). It is found that this method is efficient especially when the sample size is large (see also Anderson, 1984, p. 245–247).

In practice, with a starting value for $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{W})$, the Gibbs sampling of drawing successively from the full conditional distributions is iterated for $M + N$ times. The first M burn-in iterations are discarded. Because the iterates in the Gibbs sample are autocorrelated, we keep every s-th draw in the last N iterates so that the resulting sample contains approximately N/s independent draws from the joint posterior distribution. The value s here can be determined based on the graph of sample autocorrelation of the Gibbs iterates.

It is rather simple to choose a starting value for $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. A natural choice is to use $(0, \mathbf{I})$. However, it is nonstandard to find a starting value for $\mathbf{W}$. In this paper, we adopt the approach which is motivated by the fact that the ranking of $\{w_{1j}, \ldots, w_{k-1,j}, 0\}$ must be set to be consistent with the observed ranking $\{\pi_{1j}, \ldots, \pi_{kj}\}$. Using this fact, a simple choice for the starting value of the w's is to use $w_{ij} = (\pi_{ij} - \pi_{kj})/((k^2 - 1)/12)^{1/2}$, a type of standardized rank score.

Finally, we go back to the scale-shift problem mentioned in section 3.1. If a diffuse prior of the parameters is used in the Gibbs sampling algorithm, it can be seen that the full posterior distributions stated in (14), (15) and (16) will be improper. As a result, the Gibbs sequence of draws will follow a random walk and will not converge. Hence, a proper prior should be used but we could still set the proper prior rather diffuse.

It should be remarked that Thurstone's normal order statistics model belongs to MVNOS model with $\mathbf{V} = \mathbf{I}_k$, its parameters can be estimated by fixing $\mathbf{V}$ to $\mathbf{I}_k$ or equivalently, fixing $\boldsymbol{\Sigma}$ to $\mathbf{I}_{k-1} + \mathbf{1}_{k-1}\mathbf{1}'_{k-1}$ and skipping the step of generating $\boldsymbol{\Sigma}$ in the above Gibbs sampling algorithm.

4. Adequacy of the Model Fit

We now consider the problem of assessing the goodness-of-fit of the model. One may consider all possible rankings ($k!$ cases) as the events in a multinomial distribution (see Cohen & Mallows, 1983) and then compute the likelihood-ratio G^2 or Pearson χ^2 statistics using the observed and expected frequencies of the $k!$ rankings to assess the goodness of fit. However, it is quite often that we may encounter many empty multinomial cells, especially when k is large and the sample size is comparatively small. This problem is commonly dealt with by grouping the $k!$ rankings into a small number of meaningful subgroups and examining the fit for each subgroup. In particular, let n_i be the observed frequency that item i is ranked as the top item. Also let $\hat{p}_i = P(y_i > y_1, \ldots, y_{i-1}, y_{i+1}, \cdots, y_k | \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ be the estimated probability of ranking item i as first under the fitted MVNOS model with posterior mean estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$. The fit can be examined by comparing the observed frequency n_i with the expected frequency $n\hat{p}_i$, $i = 1, 2, \ldots, k$,

or by calculating the standardized residuals:

$$r_i = \frac{n_i - n\hat{p}_i}{(n\hat{p}_i(1 - \hat{p}_i))^{1/2}}, \quad j = 1, 2, \ldots, k.$$

In practice, the above assessment of model adequacy is more applicable when there are no covariates. Otherwise, most of the observed frequencies would be either 1 or 0 and hence large G^2, χ^2 and standardized residuals would usually be obtained even under a valid model.

One complication of facilitating the use of the above method to assess goodness-of-fit of a MVNOS model is the difficulty of evaluating numerically the estimated probability $\hat{p}_i$ which may be expressed as

$$\int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \phi(\mathbf{v}|\boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*) \, d\mathbf{v},$$

where $\mathbf{v} = (y_1 - y_i, \ldots, y_{i-1} - y_i, y_{i+1} - y_i, \ldots, y_k - y_i) \sim N(\boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*)$, and $\boldsymbol{\beta}^*$ and $\boldsymbol{\Sigma}^*$ can be obtained from $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ respectively. Instead of computing this integral by high-dimensional numerical quadrature, we employ a reliable Monte Carlo simulation approach called Geweke(1992)-Hajivassiliou(1993)-Keane(1994) (GHK) method (See also Hajivassiliou, McFadden, & Ruud, 1996). In later application, 10,000 replications are used to obtain the GHK simulator for $\hat{p}_i$.

5. Simulation Study

We now apply the method described in the previous sections to a simulated data set containing $n = 2000$ rankings of $k = 6$ items with a covariate Z, which are generated from the following model: for $j = 1, \ldots, 2000$,

the j-th ranking $\equiv$ the ranking of $\{w_{1j}, w_{2j}, w_{3j}, w_{4j}, w_{5j}, 0\}$ with

$$w_{ij} = -2X_{ij} + \varepsilon_{ij}, \quad i = 1, 2, 3, 4, 5$$

where $X_{ij} = Z_{ij} - Z_{6j}$ with Z_{ij} simulated from the uniform distribution over $(-1, 1)$ and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \varepsilon_{2j}, \varepsilon_{3j}, \varepsilon_{4j}, \varepsilon_{5j})'$ is simulated from a multivariate normal distribution with zero mean vector and covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{4} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{5} \end{pmatrix} \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{4} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{5} \end{pmatrix}.$$

Clearly, estimation based on maximum likelihood method is time-consuming and practically impossible. Here, we consider the Bayesian method. Although we must use the proper priors for β and $\boldsymbol{\Sigma}$ as explained in section 3.2, we choose a rather diffuse but proper prior: $\beta \sim N(\beta_0 = 0, A_0^{-1} = 100)$ and $\boldsymbol{\Sigma}^{-1} \sim W_{k-1}(\alpha = k + 1), \mathbf{P} = (k + 1)\mathbf{I}$. The first 1000 burn-in iterations were discarded. A further 10,000 iterations are taken to be a sample from the joint posterior distribution of the unknown parameters as well as the latent utilities w_{ij}'s. Figure 1 shows the sample autocorrelations (SACFs) of $\beta/\sqrt{\sigma_{11}}, \sigma_{22}/\sigma_{11}, \sigma_{55}/\sigma_{11}, \rho_{14}, \rho_{35}$ based on the full sample of 10,000 iterates and a reduced sample by keeping every 20th draw from the full sample. The substantial reduction of SACFs for the reduced sample over the full sample indicates that the reduced sample gives approximately independent draws from the posterior distribution. Table 2 shows the posterior means, standard deviations, and 90% posterior intervals of the parameters based on the reduced sample. The posterior intervals are constructed from the 0.05 and 0.95 percentiles of the Gibbs samples.
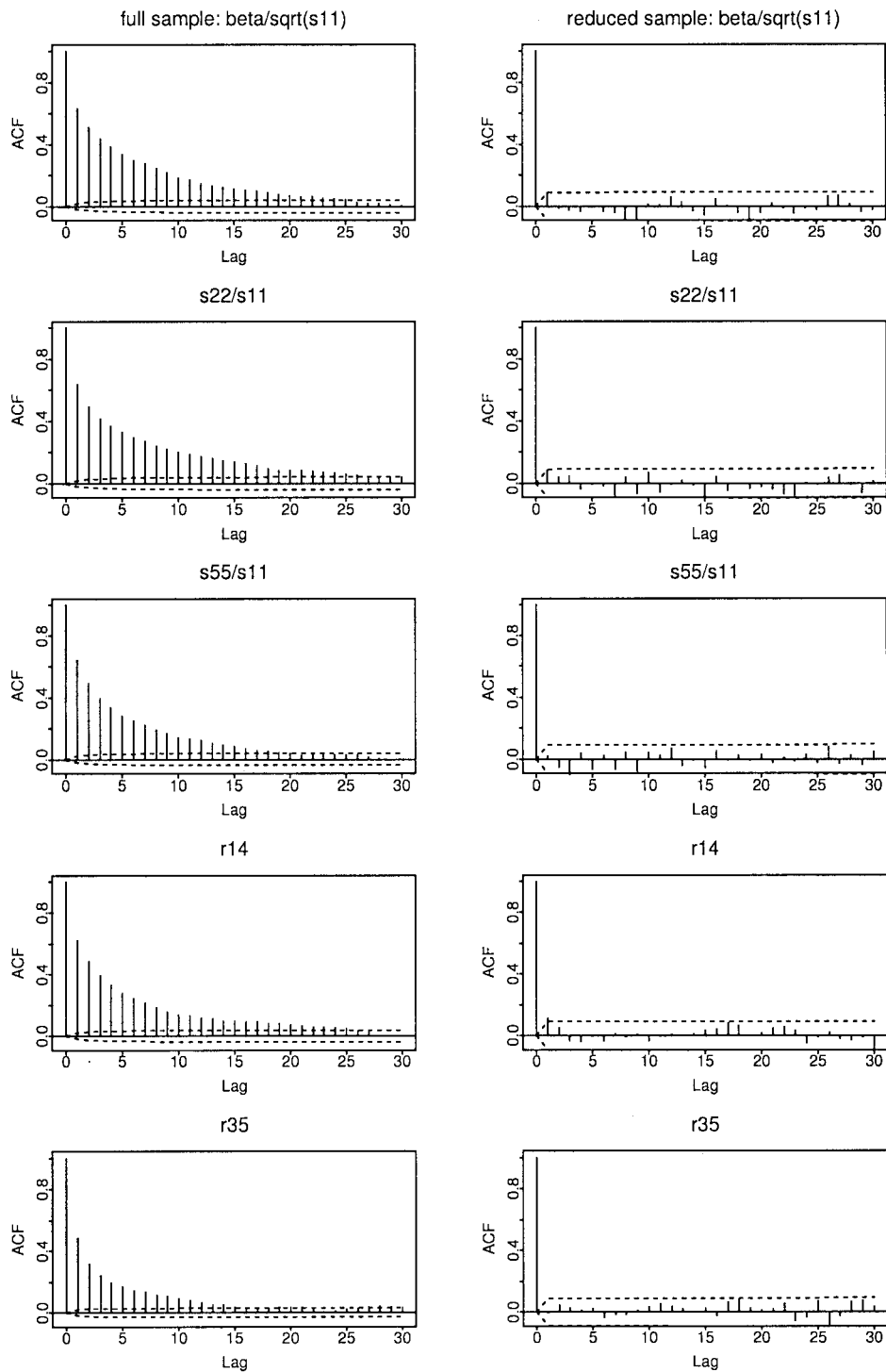
FIGURE 1.

SACFs of $\beta/\sqrt{\sigma_{11}}$, σ_{22}/σ_{11}, σ_{55}/σ_{11}, ρ_{14} and ρ_{35} using the full and reduced Gibbs samples for the simulated dataset.

TABLE 2.
Results from the Simulated Data Set

parameter	true value	posterior moments		90% interval
		mean	std. dev.	
$\beta/\sqrt{\sigma_{11}}$	-2	-2.015	0.056	$(-2.107, -1.922)$
σ_{22}/σ_{11}	2	2.009	0.126	$(1.801, 2.217)$
σ_{33}/σ_{11}	3	3.050	0.183	$(2.749, 3.350)$
σ_{44}/σ_{11}	4	4.035	0.255	$(3.616, 4.454)$
σ_{55}/σ_{11}	5	5.025	0.308	$(4.519, 5.532)$
ρ_{12}	0.5	0.514	0.028	$(0.468, 0.559)$
ρ_{13}	0.5	0.501	0.028	$(0.456, 0.546)$
ρ_{14}	0.5	0.491	0.028	$(0.444, 0.537)$
ρ_{15}	0.5	0.526	0.026	$(0.483, 0.569)$
ρ_{23}	0.5	0.492	0.024	$(0.453, 0.532)$
ρ_{24}	0.5	0.477	0.027	$(0.433, 0.521)$
ρ_{25}	0.5	0.499	0.025	$(0.458, 0.540)$
ρ_{34}	0.5	0.515	0.023	$(0.478, 0.553)$
ρ_{35}	0.5	0.517	0.023	$(0.478, 0.556)$
ρ_{45}	0.5	0.502	0.023	$(0.464, 0.540)$

It can be seen from Table 2 that all posterior means are very close to their correspond-ing true parameter values and all the posterior 90% intervals cover the true parameter val-ues. This indicates that the posterior means are good estimators for the parameters in the MVNOS model. Figure 2 displays the traces of the Gibbs sequences and the Gaussian ker-nel density estimates (see Venables & Ripley, 1994, p. 136) of the posterior densities for $\beta/\sqrt{\sigma_{11}}, \sigma_{22}/\sigma_{11}, \sigma_{55}/\sigma_{11}, \rho_{14}, \rho_{35}$. The traces of the Gibbs sequences are structureless, im-plying that the Gibbs iteration is converged. Moreover, some posterior densities are slightly asymmetric, indicating that asymptotic normality may not be satisfied for the maximum likeli-hood estimation in this situation.

6. Analysis of the APA Election Data

In 1980, the American Psychological Association conducted an election at which five can-didates $(A, B, C, D$ and $E)$ were running for president and voters were asked to rank order all of the candidates. Among those voters, 5738 gave complete rankings. These complete rankings are considered here. The data are given in Diaconis (1988). Suppose that higher rank implies more favorable. Then the average ranks received by candidates A, B, C, D and E are 3.16, 2.84, 3.08, 2.91 and 3.01 respectively. This means that voters prefer candidate A the most, candidate C the second, and then candidates E, D and B. However, in order to make inferences on the preference of the voters, modeling of the ranking data is needed. Let y_{ij} be the jth voter's utility of selecting candidate $i, i = A, B, C, D, E$. Therefore we consider the MVNOS model in which (i) the jth voter's ranking is assumed to be formed by the relative ordering of $y_{Aj}, y_{Bj}, y_{Cj}, y_{Dj}, y_{Ej}$; and (ii) the y's satisfy the following model:

$$y_{ij} = \mu_i + e_{ij}, \quad i = A, B, C, D, E, \quad j = 1, \ldots, 5738,$$

$$(e_{Aj}, e_{Bj}, e_{Cj}, e_{Dj}, e_{Ej})' \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V}),$$

or equivalently, the model could be formed by the relative ordering of $w_{Aj}, w_{Bj}, w_{Cj}, w_{Dj}, 0$, and the w's satisfy
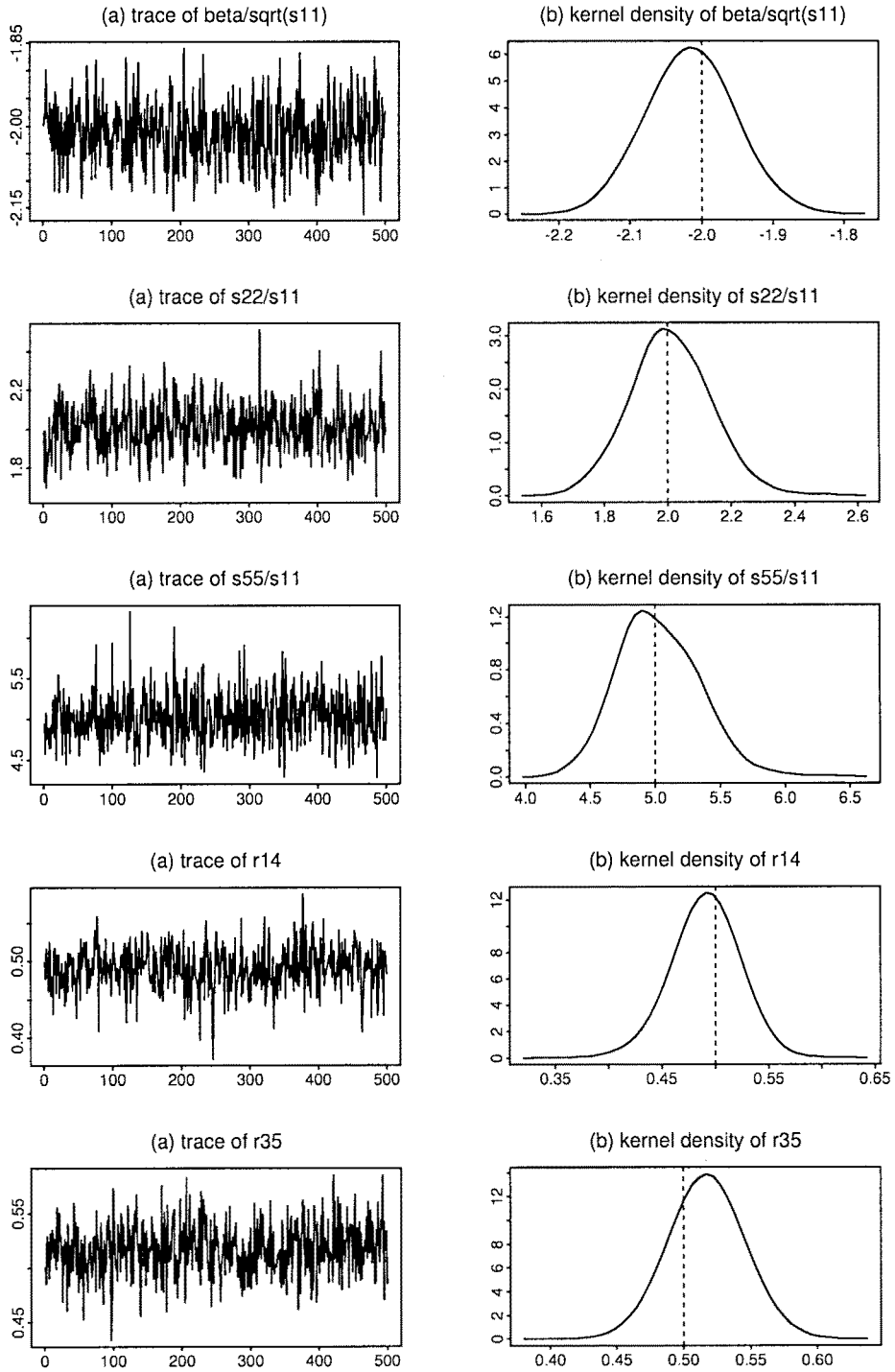
FIGURE 2.

(a) Traces of the Gibbs sequences, and (b) Posterior densities for $\beta/\sqrt{\sigma_{11}}$, σ_{22}/σ_{11}, σ_{55}/σ_{11}, ρ_{14} and ρ_{35} for the simulated dataset. The dotted line indicates the position of the true value.

$$w_{ij} = \beta_i + \varepsilon_{ij}, \quad i = A, B, C, D, \ j = 1, \ldots, 5738,$$

$$(\varepsilon_{Aj}, \varepsilon_{Bj}, \varepsilon_{Cj}, \varepsilon_{Dj})' \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{\Sigma}),$$

where $\beta_i = \mu_i - \mu_E$ and $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ij} = v_{ij} + v_{EE} - v_{iE} - v_{jE}$.

Similar to the situation in the simulation study, classical maximum likelihood estimation method would not work well to these data. Thus, we apply the proposed method to estimate the parameters $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$ of the MVNOS model. Using the priors specified in the simulation study and every 20th draw in the last 10,000 of 11,000 Gibbs iterations, we obtain the Gibbs sequences for $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$. By imposing the constraint $\mu_E = 0$ and our constraint for $\mathbf{V}$ proposed in section 2.3 to the Gibbs sequences, estimates for $\mu_i (i = A, B, C, D, E)$ and $v_{ij}, (i \le j)$ can be obtained.

6.1. Adequacy of Model Fit and Model Comparison

To examine the goodness-of-fit of the MVNOS model, Table 3 shows the observed proportions and estimated partial probabilities under the MVNOS model. The results for Thurstone's normal order statistics model and Stern's mixture of BTL models are also listed in Table 3 as alternatives to the MVNOS model. Thurstone's model is fitted by repeating the Gibbs sampling with $\mathbf{V}$ fixed at $\mathbf{I}_k$ while the parameter estimates of Stern's mixture models can be found in Stern (1993). Stern found that the data seem to be a mixture of 2 to 3 groups of voters. This feature is also supported by Diaconis (1989)'s spectral analysis and McCullagh (1993)'s model of inversions.

As seen from Table 3, the estimated first-choice probabilities for the MVNOS model match the observed proportions quite well. The absolute standardized residuals for the MVNOS model are all fairly small (< 2) but not for the other models. This indicates that the MVNOS model provides a good fit to the first-choice probabilities. On the other hand, since no sparse cells are encountered in the data, we can compute the likelihood ratio G^2 and Pearson χ^2 statistics and their values for the four models are shown in Table 3. We found that all four models do not fit the data very well. As mentioned by Stern (1993), it is difficult to draw conclusions about appropriate models for the APA election data because the large sample size could produce a great change in the likelihood even though there is a minor improvement in the model.

TABLE 3.
Observed Proportions and Estimated Probabilities that a Candidate Is Ranked as First under Various Models for the APA Election Data (*the value in bracket is the residual,* r_i)

Candidate	Observed proportion, n_i/n	Estimated probabilities, $\hat{p}_i$, under the following models			
		MVNOS model	Thurstone's model	Stern's mixture model	
				2 components	3 components
A	0.184	0.193 (−1.87)	0.170 (2.78)	0.199 (−2.89)	0.189 (−1.16)
B	0.135	0.130 (1.17)	0.231 (−17.24)	0.153 (−3.84)	0.155 (−4.27)
C	0.280	0.276 (0.70)	0.179 (20.12)	0.276 (0.80)	0.272 (1.19)
D	0.204	0.198 (1.25)	0.220 (−2.93)	0.186 (3.47)	0.192 (2.28)
E	0.197	0.200 (−0.58)	0.200 (−0.65)	0.186 (2.12)	0.189 (1.41)
Likelihood ratio G^2		334.13	1589.47	453.70	325.63
Pearson χ^2		348.13	1941.87	485.09	339.90
No. of identified parameters		13	4	9	14

TABLE 4.
Parameter Estimates of the MVNOS Model for the APA Election Data

parameter	posterior moments		90% interval
	mean	std. dev.	
μ_A	0.086	0.015	(0.018, 0.111)
μ_B	−0.071	0.014	(−0.097, −0.048)
μ_C	0.067	0.018	(0.037, 0.095)
μ_D	−0.048	0.014	(−0.071, −0.026)
v_{AA}	0.524	0.008	(0.511, 0.537)
v_{AB}	0.116	0.006	(0.106, 0.126)
v_{AC}	0.246	0.008	(0.233, 0.257)
v_{AD}	0.041	0.008	(0.027, 0.053)
v_{AE}	0.074	0.004	(0.067, 0.081)
v_{BB}	0.498	0.011	(0.479, 0.516)
v_{BC}	0.087	0.009	(0.072, 0.100)
v_{BD}	0.178	0.007	(0.166, 0.191)
v_{BE}	0.121	0.007	(0.109, 0.132)
v_{CC}	0.833	0.024	(0.795, 0.870)
v_{CD}	−0.123	0.014	(−0.146, −0.101)
v_{CE}	−0.043	0.010	(−0.060, −0.026)
v_{DD}	0.679	0.018	(0.651, 0.708)
v_{DE}	0.224	0.008	(0.212, 0.239)
v_{EE}	0.624	0.008	(0.610, 0.638)

6.2. Interpretation of the Fitted MVNOS Model

Table 4 tabulates the posterior means, standard deviations and 90% posterior intervals for the parameters of the MVNOS model. It is not surprising to see that the ordering of the posterior means of the μ_i's are the same as that of the average ranks. Apart from the posterior means, the Gibbs samples can also provide other information such as the estimated probability that candidate i is more favorable than candidate j. For instance, the probability that candidate A is more favorable than candidate C is estimated by the sample mean of $\Phi\left(\frac{\mu_A - \mu_C}{\sqrt{v_{AA} + v_{CC} - 2v_{AC}}}\right)$ in the Gibbs samples, which is found to be 0.509 (posterior standard deviation = 0.006).

According to the boxplots of μ_i, v_{ii} and $r_{ij} = v_{ij}/\sqrt{v_{ii}v_{jj}}$ ($i \neq j$) shown in Figure 3, distributions of some parameters are slightly skewed. In addition, a large estimate of v_{CC} indicates that voters have fairly large variation on the preference of candidate C. To further investigate the structure of the covariance matrix $\mathbf{V}$, a principal components analysis of the posterior mean estimate for $\mathbf{V}$ is performed and the result is presented in Table 5.

Using properties of principal components and the fact that $\mathbf{a}_2 = \sqrt{0.2}\mathbf{1}_5$, the utilities of the five candidates $\{A, B, C, D, E\}$ can be characterized by

$$\begin{bmatrix} y_A \\ y_B \\ y_C \\ y_D \\ y_E \end{bmatrix} = \begin{bmatrix} 0.086 \\ -0.071 \\ 0.067 \\ -0.048 \\ 0 \end{bmatrix} + \sqrt{1.015}\mathbf{a}_1 z_1 + \sqrt{0.2}\mathbf{1}_5 z_2 + \sqrt{0.440}\mathbf{a}_3 z_3 + \sqrt{0.357}\mathbf{a}_4 z_4 + \sqrt{0.346}\mathbf{a}_5 z_5,$$

$$(17)$$

where the z's are independently and identically distributed as $N(0, 1)$ and the principal components $\mathbf{a}$'s are given in Table 5. Because rankings of items only depend on utility differences, the term $\sqrt{0.2}\mathbf{1}_5 z_2$ in (17) does not affect the rankings and hence, interpretation would be based on the remaining four components.
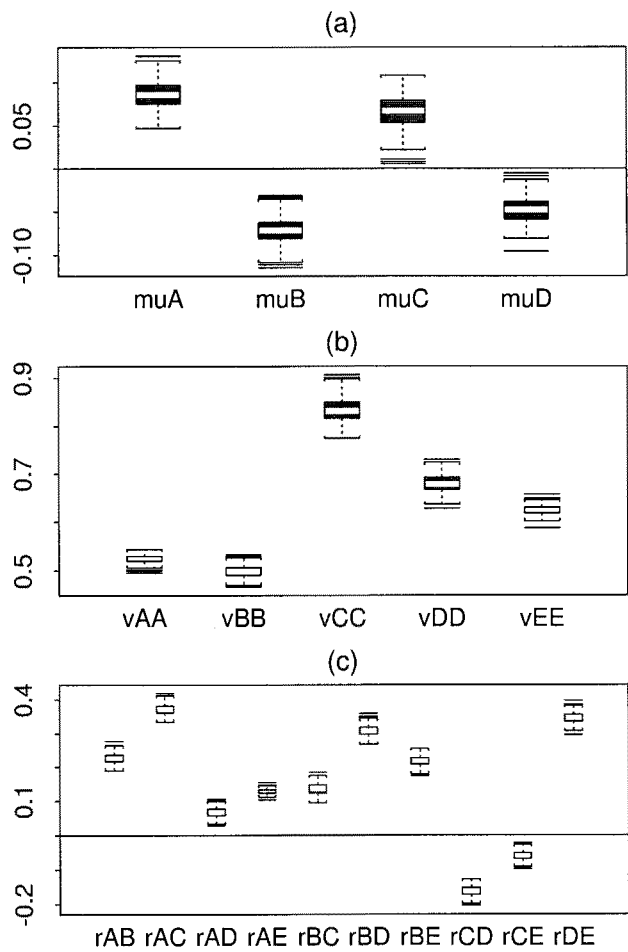
FIGURE 3.
Boxplots of μ_i, v_{ii} and $r_{ij} = v_{ij}/\sqrt{v_{ii}v_{jj}}(i \neq j)$ for the APA election data.

TABLE 5.
Principal Components Analysis of the Posterior Mean Estimate for **V**

Candidate	Principal component, $\mathbf{a}_i$				
	1	2	3	4	5
A	0.245	0.447	0.046	0.789	−0.340
B	−0.087	0.447	−0.412	0.133	0.778
C	0.726	0.447	0.012	−0.515	−0.081
D	−0.524	0.447	−0.442	−0.281	−0.502
E	−0.361	0.447	0.796	−0.125	0.145
Variance	1.015	1.000	0.440	0.357	0.346

Component 1 separates two groups of candidates: {A, C} and {D, E}, implying that there are two groups of voters: voters who prefer candidates A and C more, and those who prefer candidates D and E more. Component 3 contrasts candidate E with candidates B and D, indicating that voters either prefer B and D to E, or prefer E to B and D. For instance, if voters like B, they prefer D to E. Finally, components 4 and 5 indicate a contrast between A and C and a contrast between B and D, respectively. Based on the variances of the components, we can see that component 1 dominates and hence it plays a major role on ranking the five candidates.

So far, the analysis is based on our proposed constraints on $\mathbf{V}$ stated in section 2.3. In fact, it can be shown that the same result can be obtained by using Chintagunta's method, except that the term $\sqrt{0.2}\mathbf{1}_5 z_2$ in (17), which does not affect the rankings, will not appear. However, it should be noticed that different parameterizations of $\mathbf{V}$ are essentially equivalent in the sense that they produce the same ranking proabilities but it is uncertain that they could lead to the same interpretation of the parameters and the principal components. Moreover, since the form of $\mathbf{V}$ is not identifiable, it is impossible to specify the parameterization of $\mathbf{V}$ on the basis of ranking data. Therefore, an appropriate parameterization is important for interpretation of the components and is usually selected subjectively.

7. Robustness of the MVNOS Model

One way to address this issue is to study the sensitivity of the parameter estimates if an outlying ranking is added to the data. First of all, 5 X_i's are simulated in the same way as described in section 5. Then the outlying ranking constructed by reversing the order of $\{-2X_1, -2X_2, \ldots, -2X_5\}$ is added to the simulated dataset created in section 5. We apply the same Gibbs sampling procedure to this dataset with an outlier. The posterior moments of the parameters given in Table 2 are then reproduced. We found that the changes in the posterior means are little, less than 0.1% of the true values of parameters in magnitude whereas the changes of the posterior standard deviations are small and mainly positive.

Note that the MVNOS model mainly relies on the assumption of multivariate normal distribution of the latent variables $\mathbf{y}_j$, $j = 1, \ldots, n$. Therefore, an alternative way to address the robustness of the model is to consider a more general distribution and look at the differences between the 2 sets of estimates. A well-known class of distribution which includes multivariate normal distribution as a special case is the multivariate-t distribution.

7.1. Multivariate-t Latent Distribution and the MVNOS(v) Model

If the random vector $\mathbf{y}_j$ of judge j is multivariate-t distributed with location vector $\boldsymbol{\mu}_j$, scale matrix $\mathbf{V}$ and degrees of freedom v, its probability density function is given by

$$f(\mathbf{y}_j) = \frac{\Gamma(v+k)/2}{\Gamma(v/2)(v\pi)^{k/2}} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{1}{v}(\mathbf{y}_j - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_j) \right)^{-(v+k)/2}.$$

Note that in the case of $k = 2$, this model becomes the one for binary data proposed by Albert and Chib (1993) in which a student-t link function is used. When the d.f. v tends to infinity, this model becomes the MVNOS model. Therefore, this model allows us to study the sensitivity of the parameter estimates when there are extreme utilities. Note that it is sufficient to assume a known d.f. v.

With the multivariate-t distribution replacing the multivariate normal distribution, it is not difficult to modify the Gibbs sampling algorithm because the multivariate-t distribution can be treated as a scale-mixture of multivariate normal distribution (see Andrews & Mallows, 1974; and Carlin & Polson, 1991). More specifically, the vector $\mathbf{y}_j$ have the following representation:

$$\mathbf{y}_j = \boldsymbol{\mu}_j + \mathbf{e}_j, \tag{18}$$

$$\text{where} \quad \mathbf{e}_j | \lambda_j \overset{\text{iid}}{\sim} N(\mathbf{0}, \lambda_j \mathbf{V}), \tag{19}$$

$$\text{and} \quad \nu/\lambda_j \overset{\text{iid}}{\sim} \chi^2(\nu). \tag{20}$$

After conditioning on λ_j, the distribution of the utility vector $\mathbf{y}_j$ follows a multivariate normal distribution. Hence, we shall denote this model as the MVNOS(ν) model.

7.2. A Latent-Structure Interpretation of the MVNOS(ν) Model

With reference to the rationale of generating rankings from the MVNOS model as stated in (3) and (4), the utilities given by the judges are assumed to be measured on the same scale. However, this is not true in general. The MVNOS(ν) model provides a solution to this problem. By viewing λ_j in (19) as a level of scale employed by judge j, the MVNOS(ν) model extends the MVNOS model to provide a latent structure for describing the scale-heterogeneity of the judges. In other words, the ranking $\mathbf{R}_j$ given by judge j under the MVNOS(ν) model can be generated by the following steps:

1. Draw randomly from the distribution (20) a level of scale, denoted by λ_j;
2. Draw independently a utility vector $\mathbf{y}_j$ from $N(\boldsymbol{\mu}_i, \lambda_j \mathbf{V})$; and
3. Order the utilities $y_{1j}, \ldots, y_{kj}$ to produce the ranking $\mathbf{R}_j$.

Notice that the scales λ's are unobserved and assumed to be randomly sampled from the inverse chi-square distribution with ν degrees of freedom. The larger the ν, the smaller the variation of the scales given by the judges is.

7.3. Gibbs Sampling Algorithm for the MVNOS(ν) Model

Following the Bayesian approach to the MVNOS model, we apply the $\mathbf{w}$-transformation to the $\mathbf{y}_j$'s. The Gibbs sampling algorithm for the MVNOS(ν) model becomes very simple when we augment the vector $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{W})$ by the latent vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$. That means the Gibbs sampling algorithm consists of four main steps. Using the same priors defined in section 3.1 and the linear model for $\boldsymbol{\mu}_j$ defined in (5), these four steps are summarized as follows:

1. Draw $w_{ij} | w_{-i,j}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\Pi} \sim N(m_{ij}, \tau_{ij}^2)$ with $w_{i_{r-1}j} < w_{ij} < w_{i_{r+1}j}$ whenever $\pi_{ij} = r$, where $m_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} - g_{ii}^{-1} \mathbf{g}_{-i,i}' (w_{-i,j} - \mathbf{X}_{-i,j} \boldsymbol{\beta})$ and $\tau_{ij}^2 = \lambda_j g_{ii}^{-1}$.
2. Draw $\boldsymbol{\beta} | \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \mathbf{W} \sim N_p(\boldsymbol{\beta}_1, \mathbf{A}_1^{-1})$, where $\mathbf{A}_1 = \mathbf{A}_0 + \sum_{j=1}^n \mathbf{X}_j' \boldsymbol{\Sigma}^{-1} \mathbf{X}_j / \lambda_j$ and $\boldsymbol{\beta}_1 = \mathbf{A}_1^{-1}(\mathbf{A}_0 \boldsymbol{\beta}_0 + \sum_{j=1}^n \mathbf{X}_j' \boldsymbol{\Sigma}^{-1} \mathbf{w}_j / \lambda_j)$.
3. Draw $\boldsymbol{\Sigma} = \mathbf{G}^{-1}$ such that
   $\mathbf{G} | \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W} \sim W_{k-1}(\alpha + n, \mathbf{P} + \sum_{j=1}^n (\mathbf{w}_j - \mathbf{X}_j' \boldsymbol{\beta})(\mathbf{w}_j - \mathbf{X}_j' \boldsymbol{\beta})' / \lambda_j)$.
4. Draw $\lambda_j | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{W} \sim IG\left((\nu + k - 1)/2, \nu/2 + (\mathbf{w}_j - \mathbf{X}_j' \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{w}_j - \mathbf{X}_j' \boldsymbol{\beta})/2 \right)$,
   where $IG(a, b)$ represents the inverse-gamma distribution with shape parameter a and scale parameter b.

7.4. APA Election Data—Revisited

Consider the APA election data again. To evaluate the robustness of the MVNOS model, we examine six cases of degrees of freedom: $\nu = 1, 2, 3, 5, 10, 20$. Based on the same initial setting that we used in section 6, the Gibbs sequences of the unknown parameters are obtained from six

TABLE 6.
Parameter Estimates of the MVNOS(v) Model for the APA Election Data

parameter	Posterior mean (Posterior std. dev.)						
	$v = 1$	$v = 2$	$v = 3$	$v = 5$	$v = 10$	$v = 20$	$v = \infty$
μ_A	0.107	0.099	0.094	0.088	0.088	0.086	0.086
	(0.018)	(0.017)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
μ_B	−0.092	−0.080	−0.077	−0.076	−0.073	−0.072	−0.071
	(0.017)	(0.016)	(0.016)	(0.015)	(0.014)	(0.014)	(0.014)
μ_C	0.084	0.077	0.074	0.070	0.069	0.067	0.067
	(0.021)	(0.020)	(0.020)	(0.019)	(0.019)	(0.018)	(0.018)
μ_D	−0.063	−0.054	−0.052	−0.051	−0.050	−0.049	−0.048
	(0.016)	(0.015)	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)
v_{AA}	0.522	0.523	0.523	0.524	0.524	0.524	0.524
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
v_{AB}	0.119	0.117	0.117	0.116	0.116	0.116	0.116
	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)
v_{AC}	0.242	0.243	0.243	0.244	0.244	0.245	0.246
	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
v_{AD}	0.043	0.042	0.041	0.042	0.041	0.041	0.041
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)
v_{AE}	0.074	0.074	0.075	0.074	0.074	0.074	0.074
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
v_{BB}	0.495	0.497	0.497	0.498	0.498	0.499	0.498
	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)
v_{BC}	0.089	0.088	0.088	0.087	0.088	0.087	0.087
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
v_{BD}	0.176	0.177	0.178	0.178	0.178	0.178	0.178
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
v_{BE}	0.120	0.120	0.120	0.121	0.121	0.120	0.121
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
v_{CC}	0.833	0.833	0.834	0.833	0.834	0.832	0.833
	(0.023)	(0.023)	(0.024)	(0.025)	(0.024)	(0.023)	(0.024)
v_{CD}	−0.121	−0.121	−0.122	−0.122	−0.122	−0.121	−0.123
	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
v_{CE}	−0.044	−0.044	−0.044	−0.043	−0.044	−0.043	−0.043
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
v_{DD}	0.679	0.679	0.679	0.678	0.679	0.679	0.679
	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)
v_{DE}	0.223	0.223	0.223	0.224	0.224	0.224	0.224
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
v_{EE}	0.626	0.626	0.626	0.624	0.625	0.625	0.624
	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)

The values in the last column are extracted from Table 4.

sampling chains driven by the same stream of random numbers. Table 6 shows that the posterior means and standard deviation of the parameters for various values of v.

Generally speaking, the posterior means and standard deviations are fairly robust with respect to the degrees of freedom v, except a slow decay of the magnitudes of the posterior moments of the μ's. Therefore, it is reasonable to say that the MVNOS model is fairly insensitive to an extreme utility or as a result an unusual ranking.

8. Concluding Remarks

This paper studied the order-statistics model for ranking data in which the latent utilities follow a multivariate normal distribution. The Bayesian approach via the Gibbs sampling technique is proposed for parameters estimation. The simulation study and the example given in sections 5 and 6 show that the proposed method is computationally efficient and more flexible than the maximum likelihood method in estimating the MVNOS model, especially in the situation where the number of items is moderately large, say ≥ 5.

In this paper, adequacy of model fit is assessed based on the comparisons between observed and expected frequencies of all possible rankings or a small number of non-overlapped subgroups of rankings such as subgroup of rankings in which item i is ranked as first ($i = 1, \ldots, k$). Of course, we might consider other choice of non-overlapped subgroups such as subgroup of rankings where items i and j are the top two, or subgroup of rankings where item i is the first and item j is the second. As we adopt a Bayesian approach for the estimation of the MVNOS model, it would be better to consider some Bayesian assessment methods of model fitness (See Albert & Chib, 1995; Gelman, Meng, & Stern, 1996; and Yao & Böckenholt, 1999 for some recent developments in this aspect). Further investigation is needed.

It has been mentioned in section 2.3 that the parameters of a MVNOS model cannot be fully identified unless the variance-covariance matrix $\mathbf{V}$ is structured. In the context of multinomial probit model, some structured $\mathbf{V}$ are proposed in the literature, see for examples, Hausman and Wise (1978), Elrod and Keane (1993), and Yai, Iwakura, and Morichi (1997) and references therein. Further research on extending the MVNOS model to incorporate a structured $\mathbf{V}$ could be considered. All these interesting problems will be deferred to later papers.

References

Albert, J.H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association, 88*, 669–679.
Albert, J.H. & Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika, 82*, 747–759.
Anderson, T.W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: John Wiley.
Andrews, D.F., & Mallows, C.L. (1974). Scale mixtures of normality. *Journal of Royal Statistical Society, Series B, 36*, 99–102.
Arbuckle, J., & Nugent, J.H. (1973). A general procedure for parameter estimation for the law of comparative judgement. *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, 17*, 1–19.
Besag, J. Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science, 10*, 3–66.
Böckenholt, U. (1993). Applications of Thurstonian models to ranking data. In M.A. Fligner & J.S. Verducci (Eds.), *Probability models and statistical analyses for ranking data*. New York: Springer-Verlag.
Box, G.E.P., & Tiao, G.C. (1973). *Bayesian inference in statistical analysis*. New York: John Wiley.
Bradley, R.A., & Terry, M.E. (1952). Rank analysis of incomplete block designs, I: The method of paired comparisons. *Biometrika, 39*, 324–345.
Brook, D., & Upton, G.J.G. (1974). Biases in local government elections due to position on the ballot paper. *Applied Statistics, 23*, 414–419.
Bunch, D.S. (1991). Estimability in the multinomial probit model. *Transportation Research, Series B, 25*(1), 1–12.
Carlin, B.P., & Polson, N.G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler. *Canadian Journal of Statistics, 19*, 399–405.
Chapman, R.G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research, 19*, 288–301.
Chintagunta, P.K. (1992). Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science, 11*(4), 386–407.
Cohen, A., & Mallows, C.L. (1983). Assessing goodness of fit of ranking models to data. *The Statistician, 32*, 361–373.
Critchlow, D.E., & Fligner, M.A. (1993). Ranking models with item covariates. In M.A. Fligner & J.S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 1–19). New York: Springer-Verlag.
Critchlow, D.E., Fligner, M.A., & Verducci, J.S. (1991). Probability models on rankings. *Journal of Mathematical Psychology, 35*, 294–318.
Daniels, H.E. (1950). Rank correlation and population models. *Biometrika, 33*, 129–135.
Dansie, B.R. (1985). Parameter estimability in the multinomial probit model. *Transportation Research, Series B, 19*(6), 526–528.

Dansie, B.R. (1986). Normal order statistics as permutation probability models. *Applied Statistics, 35*, 269–275.

Devroye, L. (1986). *Non-uniform random variate generation.* New York: Springer-Verlag.

Diaconis, P. (1988). *Group Representations in Probability and Statistics* (IMS Lecture Notes, Volume 11). Hayward, CA: Institute of Mathematical Statistics.

Diaconis, P. (1989). A generalization of spectral analysis with application to ranked data. *Annals of Statistics, 17*, 949–979.

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

Gelman, A., Meng, X.L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6*(4), 733–760.

Geweke, J. (1992). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface* (pp. 571–578). Alexandria, VA: American Statistical Association.

Hajivassiliou, V. (1993). Simulation estimation methods for limited dependent variable models. In G.S. Maddala, C.R. Rao, & H.D. Vinod (Eds.), *Handbook of Statistics (Econometrics)* (Volume 11, pp. 519–543). Amersterdam: North-Holland.

Hajivassiliou, V., McFadden, D., & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics, 72*, 85–134.

Hausman, J.A., & Wise, D.A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica, 46*(2), 403–426.

Henery, R.J. (1981). Permutation probabilities as models for horse races. *Journal of Royal Statistical Society, Series B, 43*, 86–91.

Henery, R.J. (1983). Permutation probabilities for gamma random variables. *Joural of Applied Probability, 20*, 822–834.

Johnson, M.E. (1987). *Multivariate statistical simulation.* New York: John Wiley.

Keane, M.P. (1994). A computationally practical simulation estimator for panel data. *Econometrica, 62*, 95–116.

Koop, G., & Poirier, D.J. (1994). Rank-ordered logit models: An empirical analysis of Ontario voter preferences. *Journal of Applied Econometrics, 9*, 369–388.

Lo, V.S.Y., Bacon-Shone, J., & Busche, K. (1995). The application of ranking probability models to racetrack betting. *Management Science, 41*, 1048–1059.

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

Marden, J.L. (1995). *Analyzing and modeling rank data.* New York: Chapman and Hall.

McCullagh, P. (1993). Permutations and regression models. In M.A. Fligner & J.S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 196–215). New York: Springer-Verlag.

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

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika, 16*, 3–9.

Stern, H. (1990a). A continum of paired comparisons models. *Biometrika, 77*, 265–273.

Stern, H. (1990b). Models for distributions on permutations. *Journal of the American Statistical Association, 85*, 558–564.

Stern, H. (1993). Probability models on rankings and the electoral process. In M.A Fligner & J.S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 173–195). New York: Springer-Verlag.

Tallis, G.M., & Dansie, B.R. (1983). An alternative approach to the analysis of permutations. *Applied Statistics, 32*, 110–114.

Tanner, T., & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*, 528–549.

Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Review, 34*, 273–287.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion and rejoinder). *Annals of Statistics, 22*, 1701–1762.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review, 79*, 281–299.

Venables, W.N., & Ripley, B.D. (1994). *Modern applied statistics with S-Plus.* New York: Springer-Verlag.

Yai, T., Iwakura, S., & Morichi, S. (1997). Multinomial probit with structured covariance for route choice behavior. *Transportation Research, Series B, 31*(3), 195–207.

Yao, K.G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology, 52*(1), 79–92.

Yellot, J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology, 15*, 109–144.