Statistica Sinica **11**(2001), 445-461

BAYESIAN ANALYSIS OF WANDERING VECTOR MODELS FOR DISPLAYING RANKING DATA

Philip L. H. Yu and Lucia K. Y. Chan

The University of Hong Kong

Abstract: In a process of examining k objects, each judge provides a ranking of them. The aim of this paper is to investigate a probabilistic model for ranking data—the wandering vector model. The model represents objects by points in a d-dimensional space, and the judges are represented by latent vectors emanating from the origin in the same space. Each judge samples a vector from a multivariate normal distribution; given this vector, the judge's utility assigned to an object is taken to be the length of the orthogonal projection of the object point onto the judge vector, plus a normally distributed random error. The ordering of the k utilities given by the judge determines the judge's ranking. A Bayesian approach and the Gibbs sampling technique are used for parameter estimation. The method of computing the marginal likelihood proposed by Chib (1995) is used to select the dimensionality of the model. Simulations are done to demonstrate the proposed estimation and model selection method. We then analyze the Goldberg data, in which 10 occupations are ranked according to the degree of social prestige.

Key words and phrases: Bayesian approach, Gibbs sampling, marginal likelihood, ranking data, wandering vector model.

1. Introduction

Suppose a judge examines a set of k objects (such as people, products and occupations) and ranks the objects according to a certain preference criterion. This judge provides a ranking of k objects which is simply a permutation of the integers $1, \ldots, k$. More specifically, a ranking $\mathbf{R}$ of k objects is defined as a vector of ranks, $(R_1, \ldots, R_k)$, where R_i denotes the rank given to object i, and larger ranks refer to more preferred objects. We are interested in data sets composed of the rankings of a set of judges.

In general the first step in exploring high-dimensional data is to use some projection method, such as biplots in principal component analysis (Gabriel (1971)) and multidimensional scaling (Cox and Cox (1994)). In the context of visualizing ranking data, the first projection method is the multidimensional preference analysis (MDPREF) proposed by Carroll (1972, 1980) and Cohen and Mallows (1980). This method assumes that each object is represented by a point in an

d-dimensional space and each judge is represented by a vector in the same space. The vectors and the points are chosen so that the projections of the object points onto any one vector indicate the ranking given by the corresponding judge, as closely as possible. The technique of singular value decomposition is used to determine the judge vectors and object points by reducing the dimensionality of the data, while hopefully retaining as much information as possible. Recently, Alvo and Ertas (1992) used a similar technique to display rankings obtained from several populations. For a survey of these projection methods and other graphical methods for ranking data, see the monograph written by Marden (1995).

A random sample of judges may exhibit different standards or perception about the objects. This leads to a probabilistic version of the projection method in which the judge vectors are assumed to be random. The model suggested by this is called the *wandering vector model*.

This model was originally proposed by Carroll (1980), and further elaborated upon by De Soete and Carroll (1983), to visualize paired comparison data—the case where each judge is allowed to rank two objects each time. Applications in De Soete and Carroll (1983) showed that the wandering vector model can adequately represent the paired comparison data. However, the model is not commonly used for displaying ranking data. The main reason for this is that maximum likelihood estimation of parameters is not practically feasible. Exact calculation of the likelihood function requires a multidimensional numerical integration and can be inaccurate when the number of objects (k) is large. One major breakthrough in this aspect is to adopt a Bayesian approach and use the Gibbs sampling method (see Tierney (1994), Besag, Green, Higdon and Mengersen (1995), and the references therein). The aims of this paper are two-fold. First, we apply these techniques to fit the wandering vector model for ranking data. Second, we employ the method of computing the marginal likelihood proposed by Chib (1995) to select the optimal dimensionality of the model. We see later (Section 5) that the proposed methods are flexible and generally applicable for displaying the ranking of any number of objects.

The organisation of this paper is as follows. Section 2 defines the wandering vector model for ranking data. Section 3 discusses the identifiability of the model parameters. Section 4 shows how to apply the Gibbs sampling technique to draw a sample from the joint posterior distribution of the unknown parameters of the model. Section 5 introduces the marginal likelihood approach of Chib (1995) to select the dimensionality of the model. Section 6 reviews some simulation studies to demonstrate the usefulness of the estimation technique and the performance of the marginal likelihood method in choosing the dimensionality of the model. The methods are applied in Section 7 to analyze the Goldberg data, in which 10 occupations are ranked according to the degree of social prestige.

2. The Wandering Vector Model

Suppose each of n judges is asked to rank k objects. In the wandering vector model, object i $(i = 1, \ldots, k)$ is represented by a point $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{id})'$ in a d-dimensional space $(d < k - 1)$ and judge j $(j = 1, \ldots, n)$ is represented by a vector $\mathbf{x}_j$ emanating from the origin in the same space. Different judges have different standards or perception about the objects and the model takes $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as random.

In the wandering vector model, each judge independently samples a vector $\mathbf{x}_j$ from $N_d(\boldsymbol{\mu}, \mathbf{I})$. Given this vector, the judge's utility U_{ij} assigned to object i is

$$U_{ij} = \boldsymbol{\theta}_i' \mathbf{x}_j + \varepsilon_{ij}, \qquad i = 1, \ldots, k, \ j = 1, \ldots, n. \tag{1}$$

Here, the term $\boldsymbol{\theta}_i' \mathbf{x}_j$ measures the length of the orthogonal projection of the object point $\boldsymbol{\theta}_i$ onto the judge vector $\mathbf{x}_j$ (apart from a scale) while the ε_{ij}'s represent errors accounting for the utility variation not explained by dimensions absent in the model, for example the utility variation over time. It is assumed that the ε_{ij}'s are i.i.d. $N(0, \sigma^2)$ and independent of the $\mathbf{x}_j$'s. Finally, the ordering of the utilities $\{U_{ij}, i = 1, \ldots, k\}$, determines the jth judge's ranking. Note that all utilities are unobservable and what we can observe is just their ranking.

A 2-dimensional model is shown in Figure 1. In the figure, three objects are positioned at $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, and judges j and ℓ are represented by $\mathbf{x}_j$ and $\mathbf{x}_\ell$, independently sampled from $N_2(\boldsymbol{\mu}, \mathbf{I})$. We can see that the orthogonal projection of $\boldsymbol{\theta}_1$ onto $\mathbf{x}_j$ is larger than that of $\boldsymbol{\theta}_2$ onto $\mathbf{x}_j$, which in turn is larger than that of $\boldsymbol{\theta}_3$ onto $\mathbf{x}_j$. Therefore, it is more likely that judge j would on this particular occasion prefer the ranking: object 1, object 2, object 3. However, judge ℓ would prefer object 2 to object 1 to object 3.

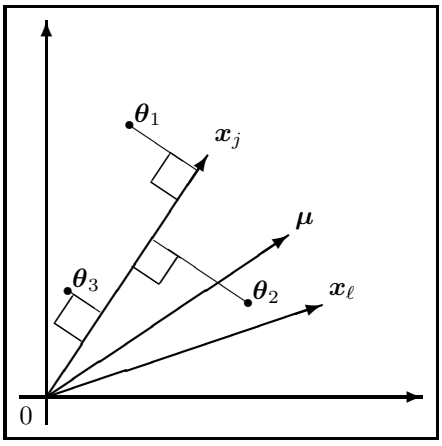


Figure 1. An illustration of a 2-dimensional wandering vector model.

As each judge independently samples a new vector $\mathbf{x}$ from $N_d(\boldsymbol{\mu}, \mathbf{I})$, the judges' vectors *wander* over the mean $\boldsymbol{\mu}$, giving rise to the name chosen for the model.

It is not necessary to generalize the covariance matrix of $\mathbf{x}$ to a general positive definite matrix, $\boldsymbol{\Sigma}$ say, since (1) remains unchanged when $\boldsymbol{\theta}_i$, $\mathbf{x}_j$ and $\boldsymbol{\mu}$ are replaced by $\mathbf{L}'\boldsymbol{\theta}_i$, $\mathbf{L}^{-1}\mathbf{x}_j$ and $\mathbf{L}^{-1}\boldsymbol{\mu}$ respectively, where $\mathbf{L}$ is the unique lower triangular matrix with positive diagonal elements obtained from the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$ (see Schott (1997), p.138-140).

2.1. The likelihood function of the proposed model

For $j = 1, \ldots, n$, let $< i_1, \ldots, i_k >$ be the ordering of the k objects corresponding to the ranking $\mathbf{R}_j = (R_{1j}, \ldots, R_{kj})$. The wandering vector model assigns the ranking $\mathbf{R}_j$ the probability

$$P(\mathbf{R}_j) = P(U_{i_1 j} < U_{i_2 j} < \cdots < U_{i_k j}), \qquad i = 1, \ldots, k,\ j = 1, \ldots, n, \tag{2}$$

where the utility vector $\mathbf{U}_j = (U_{1j}, \ldots, U_{kj})'$ of judge j is given in (1). Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k]'$. Under this model, the utility vectors $\mathbf{U}_j$'s are i.i.d. with mean $\boldsymbol{\Theta}\boldsymbol{\mu}$ and a factor-like covariance matrix: $\boldsymbol{\Theta}\boldsymbol{\Theta}' + \sigma^2\mathbf{I}$.

Now the likelihood function of $(\boldsymbol{\mu}, \boldsymbol{\Theta}, \sigma^2)$ is $\mathbf{L}(\boldsymbol{\mu}, \boldsymbol{\Theta}, \sigma^2) = \prod_{j=1}^{n} \mathbf{P}(\mathbf{R}_j)$. Unless the number of objects is small, a direct maximization of the likelihood function is computationally demanding and numerically unstable because evaluation by quadrature of a $(k-1)$-dimensional integral may not be accurate. Hence, most applications of the wandering vector model in the literature are related to paired comparison data (see for example De Soete and Carroll (1983)).

3. Identifiability Problems for Model Parameters

The wandering vector model involves three distinct sets of parameters: the mean judge vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$; the object points $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{id})'$, $i = 1, \ldots, k$; the variance parameter σ^2. However, not all parameters can be identified. More specifically, the ranking probabilities in (2) are invariant under the following transformations.

(a) *Location-shift of the object points*
Adding an arbitrary $d \times 1$ constant vector $\mathbf{c}$ to all object points does not affect the ranking probabilities. (If $\boldsymbol{\theta}_i \to \boldsymbol{\theta}_i + \mathbf{c}$ for all i, all utilities for judge j are increased by the same amount $\mathbf{c}'\mathbf{x}_j$ and their ranking remains unchanged.)

(b) *Scale-shift of the object points*
Multiplying all object points by a positive constant b does not affect the ranking probabilities. (All U_{ij}'s and σ, the standard deviation of the error term, are reduced by the same factor.)

(c) *Orthogonal rotation of the object points and judge vectors*
It is easy to see that $U_{ij} = (\mathbf{T}\boldsymbol{\theta}_i)'(\mathbf{T}\mathbf{x}_j) + \varepsilon_{ij} = \boldsymbol{\theta}_i'\mathbf{x}_j + \varepsilon_{ij}$ for any $d \times d$ orthogonal matrix $\mathbf{T}$.

The location-shift problem (a) is commonly dealt with by assuming $\sum_{i=1}^{k} \boldsymbol{\theta}_i = 0$ or, equivalently, setting

$$\boldsymbol{\theta}_1 = -(\boldsymbol{\theta}_2 + \cdots + \boldsymbol{\theta}_k). \tag{3}$$

A natural way to tackle the scale-shift problem (b) is to set $\sigma^2 = 1$. The rotation problem (c) is well known in the context of maximum likelihood factor analysis and is usually solved by imposing the computationally convenient constraint: $\boldsymbol{\Theta}'\boldsymbol{\Theta} = \mathbf{D}$, a diagonal matrix (see Johnson and Wichern (1992, p.411)). Since this constraint is not tractable for Bayesian calculations, we use the following constraints:

(i) set all $d(d-1)/2$ cells below the diagonal of the lower-left $d \times d$ submatrix of $\boldsymbol{\Theta}$ to zero,

$$\theta_{it} = 0, \quad i = k-d+2, \ldots, k, \quad t = 1, \ldots, i-(k-d+1). \tag{4}$$

(ii) restrict the mean vector $\boldsymbol{\mu}$ to be positive.

The justification for choosing these constraints is given in the Appendix.

To sum up, we only need to estimate $d + dk + 1 - d - d(d-1)/2 - 1 = d(k-(d-1)/2)$ unknown identified parameters : $\boldsymbol{\mu}(> \mathbf{0})$ and the constrained $\boldsymbol{\Theta}$.

4. Bayesian Estimation of the Wandering Vector Model

We demonstrate how to apply a Bayesian approach and the Gibbs sampling technique to estimate the unknown parameters of the wandering vector model.

4.1. Prior distribution

We begin with a truncated normal prior for $\boldsymbol{\mu}$, i.e., $\boldsymbol{\mu} \sim N_d(\boldsymbol{\mu}_0, \mathbf{A}_0^{-1})$ subjecting to $\boldsymbol{\mu} > \mathbf{0}$. As mentioned in Section 3, some constraints on $\boldsymbol{\Theta}$ are imposed in order to fix the location and rotation, and hence the prior for $\boldsymbol{\Theta}$ is nonstandard. For $i = k-d+2, \ldots, k$, let $\boldsymbol{\theta}_{(i)}$ be a column vector containing the last $k-i+1$ cells of $\boldsymbol{\theta}_i$. Then, after imposing the constraints (3) and (4) on $\boldsymbol{\Theta}$, the identified $\boldsymbol{\Theta}$ consists of $\boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{k-d+1}, \boldsymbol{\theta}_{(k-d+2)}, \ldots, \boldsymbol{\theta}_{(k)}$, and can be rearranged into a $(kd - d(d+1)/2) \times 1$ vector $\boldsymbol{\lambda}$, with

$$\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_{k-d+1} \\ \boldsymbol{\theta}_{(k-d+2)} \\ \vdots \\ \boldsymbol{\theta}_{(k)} \end{pmatrix}. \tag{5}$$

Once $\boldsymbol{\lambda}$ is known, $\boldsymbol{\Theta}$ is known and vice versa. We use a normal prior on $\boldsymbol{\lambda}$,

$$\boldsymbol{\lambda} \sim N_{kd-d(d+1)/2}(\boldsymbol{\lambda}_0, \mathbf{P}_0^{-1}), \tag{6}$$

and assume that $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are independent.

4.2. Gibbs sampling algorithm for the wandering vector model

Let $\mathbf{R}$ be the collection of rankings $\mathbf{R}_j$'s. The Gibbs sampling algorithm draws samples from the full conditional posterior distributions as follows:

(1) draw $\mathbf{U}_j$ from $f(\mathbf{U}_j|\mathbf{R}, \boldsymbol{\Theta}, \mathbf{X}, \boldsymbol{\mu}) \propto f(\mathbf{U}_j|\mathbf{R}, \boldsymbol{\Theta}, \mathbf{X})$, for $j = 1, \ldots, n$,
(2) draw $\boldsymbol{\Theta}$ from $f(\boldsymbol{\Theta}|\mathbf{R}, \mathbf{U}, \mathbf{X}, \boldsymbol{\mu}) \propto f(\boldsymbol{\Theta}|\mathbf{U}, \mathbf{X})$,
(3) draw $\mathbf{x}_j$ from $f(\mathbf{x}_j|\mathbf{R}, \mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu}) \propto f(\mathbf{x}_j|\mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu})$, for $j = 1, \ldots, n$, and
(4) draw $\boldsymbol{\mu}$ from $f(\boldsymbol{\mu}|\mathbf{R}, \mathbf{U}, \boldsymbol{\Theta}, \mathbf{X}) \propto f(\boldsymbol{\mu}|\mathbf{X})$.

In step (1), it can be easily shown that, given $\boldsymbol{\Theta}$ and $\mathbf{X}$, the $\mathbf{U}_j$'s are independent with $\mathbf{U}_j|\boldsymbol{\Theta}, \mathbf{X} \sim N_d(\boldsymbol{\Theta}\mathbf{x}_j, \mathbf{I})$ for $j = 1, \ldots, n$ and

$$U_{i_1,j} < U_{i_2,j} < \cdots < U_{i_k,j}, \tag{7}$$

where $R_{i_r,j} = r$, for $r = 1, \ldots, k$. To draw a sample from this truncated multivariate normal distribution, we can apply the acceptance-rejection technique by drawing samples from $N_d(\boldsymbol{\Theta}\mathbf{x}_j, \mathbf{I})$ until condition (7) is satisfied. For more efficiency, instead of drawing the whole vector $\mathbf{U}_j$ at one time, we successively simulate each entry of $\mathbf{U}_j$ by conditioning on the other $k-1$ entries. More specifically, we replace step (1) by

(1') draw $\mathbf{U}_{ij}$ from $f(U_{ij}|U_{-i,j}, \mathbf{R}, \boldsymbol{\Theta}, \mathbf{X}, \mu)$ for $i = 1, \ldots, k,\ \ j = 1, \ldots, n$,

where $U_{-i,j}$ is $\mathbf{U}_j$ with $U_{i,j}$ deleted. Suppose $< i_1, \ldots, i_k >$ is the ordering of objects corresponding to their ranks $(R_{1j}, \ldots, R_{kj})$. Then $R_{tj} = r$ if and only if $i_r = t$. Now we have

$$U_{i,j}|U_{-i,j}, \mathbf{R}, \boldsymbol{\Theta}, \mathbf{X}, \boldsymbol{\mu} \sim N(\boldsymbol{\theta}_i'\mathbf{x}_j, 1) \tag{8}$$

subject to $U_{i_{r-1},j} < U_{i,j} < U_{i_{r+1},j}$ whenever $R_{ij} = r$, where $U_{i_0,j} = -\infty$ and $U_{i_{k+1},j} = +\infty$. It is clear from (8) that this simulation method is much simpler as only univariate draws are needed, and truncated normal samples can be obtained by the inverse CDF method (see Devroye (1986)).

In step (2), we need to sample $\boldsymbol{\Theta}$ from the fullconditional density $f(\boldsymbol{\Theta}|\mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu})$. Since $\boldsymbol{\Theta}$ is in one-to-one correspondence to $\boldsymbol{\lambda}$, sampling a $\boldsymbol{\Theta}$ from $f(\boldsymbol{\Theta}|\mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu})$ is equivalent to drawing a $\boldsymbol{\lambda}$ from $f(\boldsymbol{\lambda}|\mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu})$ and then transforming back to $\boldsymbol{\Theta}$ using (5) and the constraints (3) and (4). Analogous to $\boldsymbol{\theta}_{(i)}$, let $\mathbf{x}_{j(i)}$ be a column vector containing the last $k-i+1$ cells of $\mathbf{x}_j$, for $i = k-d+2, \ldots, k$,

$j = 1, \ldots, n$. The utility equation in (1) can be rewritten as

$$\mathbf{U}_j = \begin{pmatrix} U_{1j} \\ U_{2j} \\ \vdots \\ U_{k-d+1,j} \\ U_{k-d+2,j} \\ \vdots \\ U_{kj} \end{pmatrix} = \begin{pmatrix} -\mathbf{x}'_j \ldots - \mathbf{x}'_j & -\mathbf{x}'_{j(k-d+2)} \ldots - \mathbf{x}'_{j(k)} \\ \mathbf{x}'_j & \\ \ddots & \mathbf{0} \\ \mathbf{x}'_j & \\ & \mathbf{x}'_{j(k-d+2)} \\ \mathbf{0} & \ddots \\ & \mathbf{x}'_{j(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_{k-d+1} \\ \boldsymbol{\theta}_{(k-d+2)} \\ \vdots \\ \boldsymbol{\theta}_{(k)} \end{pmatrix} + \boldsymbol{\varepsilon}_j$$
$$= \mathbf{X}_{(j)}\boldsymbol{\lambda} + \boldsymbol{\varepsilon}_j.$$

Since we are conditioning on $\mathbf{R}$ and $\mathbf{X}$, the wandering vector model is simply a standard Bayesian linear model set-up (Box and Tiao (1973)). Therefore the full conditional distribution of $\boldsymbol{\lambda}$ is

$$\boldsymbol{\lambda}|\mathbf{R}, \mathbf{U}, \mathbf{X}, \boldsymbol{\mu} \sim N_{kd-d(d+1)/2}(\boldsymbol{\lambda}_1, \mathbf{P}_1^{-1}), \tag{9}$$

where $\mathbf{P}_1 = \mathbf{P}_0 + \sum_{j=1}^n \mathbf{X}'_{(j)}\mathbf{X}_{(j)}$ and $\boldsymbol{\lambda}_1 = \mathbf{P}_1^{-1}(\mathbf{P}_0\boldsymbol{\lambda}_0 + \sum_{j=1}^n \mathbf{X}'_{(j)}\mathbf{U}_j)$. After a $\boldsymbol{\lambda}$ is drawn, $\boldsymbol{\Theta}$ can be determined by using (5) and the constraints (3) and (4).

Now we go to steps (3) and (4) using the samer argument. Since the $\mathbf{x}_j$ are i.i.d. $N_d(\boldsymbol{\mu}, \mathbf{I})$, the full conditional distribution of $\mathbf{x}_j$ is

$$\mathbf{x}_j|\mathbf{R}, \mathbf{U}, \boldsymbol{\Theta}, \boldsymbol{\mu} \sim N_d(\boldsymbol{\omega}_j, \mathbf{W}^{-1}), \quad j = 1, \ldots, n, \tag{10}$$

where $\boldsymbol{\omega}_j = \mathbf{W}^{-1}(\boldsymbol{\mu} + \boldsymbol{\Theta}'\mathbf{U}_j)$ and $\mathbf{W} = \boldsymbol{\Theta}'\boldsymbol{\Theta} + \mathbf{I}$. Finally, the full conditional distribution of $\boldsymbol{\mu}$ is $\boldsymbol{\mu}|\mathbf{R}, \mathbf{U}, \boldsymbol{\Theta}, \mathbf{X} \sim N_d((n\mathbf{I} + \mathbf{A}_0)^{-1}(\boldsymbol{\mu}_0 + \sum_{j=1}^n \mathbf{x}_j), (n\mathbf{I} + \mathbf{A}_0)^{-1})$ subject to $\boldsymbol{\mu} > \mathbf{0}$. To avoid drawing from this truncated multivariate normal distribution, we consider the special case $\mathbf{A}_0 = a_0\mathbf{I}$. In this case, it can be shown that the full conditional distribution of $\boldsymbol{\mu}$ is such that the μ_t's are independently distributed with

$$\mu_t \sim N(\frac{\mu_{0t} + \sum_{j=1}^n x_{jt}}{n + a_0}, \frac{1}{n + a_0}) \text{ subject to } \mu_t > 0, \quad t = 1, \ldots, d, \tag{11}$$

and can be sampled easily by using the inverse CDF method.

With a starting value for $(\mathbf{U}, \boldsymbol{\Theta}, \mathbf{X}, \boldsymbol{\mu})$, we iterate steps (1'), (2), (3) and (4) in turn by using the full conditional distributions (8), (9), (10) and (11). Under mild conditions, when this process is repeated many times, the draws obtained will converge to a single draw from the joint posterior distribution of $\mathbf{U}, \boldsymbol{\Theta}, \mathbf{X}$ and $\boldsymbol{\mu}$ (Tierney (1994)). In practice, we iterate the process $M + N$ times. The first M burn-in iterations are discarded. The last N iterates are taken to be an approximate sample from the joint posterior distribution.

It is rather simple to choose a starting value for $(\boldsymbol{\Theta}, \boldsymbol{\mu})$. A natural choice is to use $(\mathbf{0}, \mathbf{1})$. It is also convenient to draw a starting value of $\mathbf{X}$ from $N_d(\mathbf{1}, \mathbf{I})$. However, there is no standard rule to choose a starting value for $\mathbf{U}$. Knowing that the ranking of $\{U_{1j}, \ldots, U_{kj}\}$ must be set to be consistent with the observed ranking $\{R_{1j}, \ldots, R_{kj}\}$, a simple choice for the startingvalue of the U_{ij}'s is to use $U_{ij} = R_{ij}/\sqrt{(k^2-1)/12}$, a type of standardized rank score.

5. Choice of Dimensionality of the Model

Before fitting the wandering vector model, the dimensionality d of the model must be pre-specified. Obviously, different choices of dimensionality may lead to different models. In the Bayesian literature, the problem of model selection has proved extremely challenging and has received a lot of attention recently. For a review of this topic, see Kass and Raftery (1995), Raftery (1995) and the references therein.

In this paper, we perform the dimensionality selection by computing the marginal likelihoods of models with different dimensionalities and then choosing the largest one. Note that the marginal likelihood of a model is obtained by integrating the likelihood function with respect to the prior density. However, as mentioned in Section 1, the likelihood function for the wandering vector model is so complicated that marginal likelihood cannot easily be determined. Recently, Chib (1995) proposed a method of computing marginal likelihood by using Gibbs output. We discuss how to apply the Chib's method to select the dimensionality of the wandering vector model.

5.1. The marginal likelihood of the wandering vector model

Let M_d denote a d-dimensional wandering vector model. Also let $f(\mathbf{R}|M_d, \boldsymbol{\Theta}, \boldsymbol{\mu})$ be the likelihood function of the ranking data $\mathbf{R}$ under model M_d given the model-specific parameter vectors $\boldsymbol{\Theta}$ and $\boldsymbol{\mu}$, $\pi(\boldsymbol{\Theta}, \boldsymbol{\mu}|M_d)$ the prior density. Then the marginal likelihood under model M_d is $m(\mathbf{R}|M_d) = \int f(\mathbf{R}|\boldsymbol{\Theta}, \boldsymbol{\mu}, M_d)\, \pi(\boldsymbol{\Theta}, \boldsymbol{\mu}|M_d)\, d(\boldsymbol{\Theta}, \mu)$. The basic idea of Chib's method is, in our case, to consider the so-called the *basic marginal likelihood identity* form of the marginal likelihood:

$$m(\mathbf{R}|M_d) = \frac{f(\mathbf{R}|\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*, M_d)\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|M_d)}{\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R}, M_d)}, \tag{12}$$

where the numerator is a product of the augmented likelihood function and the prior density, and the denominator is the posterior density of $(\boldsymbol{\Theta}, \boldsymbol{\mu}, \mathbf{U})$, with all evaluated at the posterior mean $(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*)$. The advantage of using Chib's marginal likelihood estimation method is that it requires no new inputs except some additional Gibbs iterations.

First of all, it can be seen from (12) that the augmented likelihood is always 1 because the ranking $\mathbf{R}$ is completely known when $\mathbf{U}^*$ is given. In addition, since the priors of $\boldsymbol{\mu}$ and $\boldsymbol{\Theta}$ are independent, the prior density can be written as

$$\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*) = \pi(\mathbf{U}^*|\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)\pi(\boldsymbol{\Theta}^*)\pi(\boldsymbol{\mu}^*). \tag{13}$$

Since all three terms in (13) have a closed form, the computation of the prior density is not difficult (see Section 5.2). It remains to compute the posterior density $\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R}, M_d)$. It will be seen in Section 5.3 that this can be estimated by using the output of the Gibbs sampling and some further Gibbs iterations. As a result, the estimate of (12) can be simplified and expressed in a convenient logarithm-scale (with the model label M_d suppressed):

$$\begin{aligned}\ln \hat{m}(\mathbf{R}) &= \ln \pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*) - \ln \hat{\pi}(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R}) \\ &= \ln \pi(\mathbf{U}^*|\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*) + \ln \pi(\boldsymbol{\Theta}^*) + \ln \pi(\boldsymbol{\mu}^*) - \ln \hat{\pi}(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R}). \end{aligned} \tag{14}$$

5.2. Computation of the prior density

To calculate the prior density $\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*)$, we want to know the exact distribution of $(\mathbf{U}^*|\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$, and the prior densities of $\boldsymbol{\Theta}^*$ and $\boldsymbol{\mu}^*$. First of all, using (6) and (11), $\pi(\boldsymbol{\Theta}^*)$ and $\pi(\boldsymbol{\mu}^*)$ can be easily determined. Also, it is not difficult to see that

$$\mathbf{U}_j|\boldsymbol{\Theta}^*, \boldsymbol{\mu}^* \sim N_k \left(\begin{pmatrix} \boldsymbol{\mu}^{*'}\boldsymbol{\theta}_1^* \\ \boldsymbol{\mu}^{*'}\boldsymbol{\theta}_2^* \\ \vdots \\ \boldsymbol{\mu}^{*'}\boldsymbol{\theta}_k^* \end{pmatrix}, \begin{pmatrix} \boldsymbol{\theta}_1^{*'}\boldsymbol{\theta}_1^* + 1 & \boldsymbol{\theta}_1^{*'}\boldsymbol{\theta}_2^* & \cdots & \boldsymbol{\theta}_1^{*'}\boldsymbol{\theta}_k^* \\ \boldsymbol{\theta}_2^{*'}\boldsymbol{\theta}_1^* & \boldsymbol{\theta}_2^{*'}\boldsymbol{\theta}_2^* + 1 & \cdots & \boldsymbol{\theta}_2^{*'}\boldsymbol{\theta}_k^* \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\theta}_k^{*'}\boldsymbol{\theta}_1^* & \boldsymbol{\theta}_k^{*'}\boldsymbol{\theta}_2^* & \cdots & \boldsymbol{\theta}_k^{*'}\boldsymbol{\theta}_k^* + 1 \end{pmatrix} \right) \quad j = 1, \ldots, n.$$

Using (13), the prior density $\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*)$ can be computed. We now move on to compute the posterior density.

5.3. Computation of the posterior density

The computation of the posterior density $\pi(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R})$ involves the calculation of the conditional densities (a) $\pi(\boldsymbol{\mu}^*|\mathbf{R})$, (b) $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \boldsymbol{\mu}^*)$, and (c) $\pi(\mathbf{U}^*|\mathbf{R}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$. To estimate these three densities, sampling additional draws from all or some steps of the Gibbs sampling algorithm is required.

(a) Estimation of $\pi(\boldsymbol{\mu}^*|\mathbf{R})$

Notice that $\pi(\boldsymbol{\mu}^*|\mathbf{R})$ can be expressed as

$$\pi(\boldsymbol{\mu}^*|\mathbf{R}) = \int \pi(\boldsymbol{\mu}^*|\mathbf{R}, \boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}) f(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}|\mathbf{R}) \; d(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}), \tag{15}$$

where $\pi(\boldsymbol{\mu}^*|\mathbf{R}, \boldsymbol{\Theta}, \mathbf{U}, \mathbf{X})$ is the full conditional density (11) specified in the Gibbs sampling algorithm, and $f(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}|\mathbf{R})$ is the posterior density of $(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X})$. If we sample G additional draws, $\{\boldsymbol{\Theta}^{(g)}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)}, \boldsymbol{\mu}^{(g)}\}_{g=1}^{G}$, and run (1)-(4), the draws $\{\boldsymbol{\Theta}^{(g)}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)}\}_{g=1}^{G}$, can be viewed as an independent sample from $f(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}|\mathbf{R})$. Hence, $\pi(\boldsymbol{\mu}^*|\mathbf{R})$ can be estimated as $\hat{\pi}(\boldsymbol{\mu}^*|\mathbf{R}) = G^{-1}\sum_{g=1}^{G} \pi(\boldsymbol{\mu}^*|\mathbf{R}, \boldsymbol{\Theta}^{(g)}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)})$.

(b) Estimation of $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \boldsymbol{\mu}^*)$

As in (15), since $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \boldsymbol{\mu}^*) = \int \pi(\boldsymbol{\Theta}^*|\mathbf{R}, \mathbf{U}, \mathbf{X}, \boldsymbol{\mu}^*) f(\mathbf{U}, \mathbf{X}|\mathbf{R}, \boldsymbol{\mu}^*) d(\mathbf{U}, \mathbf{X})$, $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \boldsymbol{\mu}^*)$ can be estimated by first sampling G additional draws, $(\boldsymbol{\Theta}^{(g)}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)})_{g=1}^{G}$, from $f(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}|\mathbf{R}, \boldsymbol{\mu}^*)$ and then taking the sample average of the G densities $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)}, \boldsymbol{\mu}^*)$, $g = 1, \ldots, G$, where $\pi(\boldsymbol{\Theta}^*|\mathbf{R}, \mathbf{U}^{(g)}, \mathbf{X}^{(g)}, \boldsymbol{\mu}^*)$ is governed by (9). At the first glance, it seems that it is difficult to sample from $f(\boldsymbol{\Theta}, \mathbf{U}, \mathbf{X}|\mathbf{R}, \boldsymbol{\mu}^*)$. However, this can be easily done by slightly modifying the Gibbs sampling algorithm stated in Section 4.2. The trick is to iterate steps (1'), (2) and (3) without passing through step (4) to generate $\boldsymbol{\mu}$ in the Gibbs iterations while fixing $\boldsymbol{\mu}$ at $\boldsymbol{\mu}^*$.

(c) Estimation of $\pi(\mathbf{U}^*|\mathbf{R}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$

As in cases (a) and (b), $\pi(\mathbf{U}^*|\mathbf{R}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$ is estimated by $\hat{\pi}(\mathbf{U}^*|\mathbf{R}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)) = \prod_{j=1}^{n}\prod_{i=1}^{k} \hat{\pi}(U_{i,j}^*|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$, where $\hat{\pi}$ $(U_{i,j}^*|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*) = G^{-1}\sum_{g=1}^{G} \pi(U_{i,j}^*|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, U_{i+1,j}^{(g)}, \ldots, U_{k,j}^{(g)}, \boldsymbol{\Theta}^*, \mathbf{X}^{(g)}, \boldsymbol{\mu}^*)$, with the $U_{i+1,j}^{(g)}, \ldots, U_{k,j}^{(g)}, \mathbf{X}^{(g)}$ drawn from a modified Gibbs sampling from the densities $\pi(U_{i,j}|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, U_{i+1,j}, \ldots, U_{k,j}, \boldsymbol{\Theta}^*, \mathbf{X}, \boldsymbol{\mu}^*)$, $\pi(U_{i+1,j}|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, U_{i,j}, U_{i+2,j}, \ldots, U_{k,j}, \boldsymbol{\Theta}^*, \mathbf{X}, \boldsymbol{\mu}^*)$, $\ldots$, $\pi(U_{k,j}|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, U_{i,j}, \ldots, U_{k-1,j}, \boldsymbol{\Theta}^*, \mathbf{X}, \boldsymbol{\mu}^*)$, and $\pi(\mathbf{X}|\mathbf{R}, U_{1,j}^*, \ldots, U_{i-1,j}^*, U_{i,j}, \ldots, U_{k,j}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)$.

Finally, the estimated posterior density $\hat{\pi}(\boldsymbol{\Theta}^*, \boldsymbol{\mu}^*, \mathbf{U}^*|\mathbf{R})$ can be computed from as $\hat{\pi}(\mathbf{U}^*|\mathbf{R}, \boldsymbol{\Theta}^*, \boldsymbol{\mu}^*)\hat{\pi}(\boldsymbol{\Theta}^*|\mathbf{R}, \boldsymbol{\mu}^*)\hat{\pi}(\boldsymbol{\mu}^*|\mathbf{R})$.

6. Simulation Studies

We generate $n = 1000$ rankings of $k = 5$ objects from the following model: for $j = 1, \ldots, 1000$, the jth ranking is the ranking of $\{U_{1j}, \ldots, U_{5j}\}$ with $\mathbf{U}_j = \boldsymbol{\Theta}\mathbf{x}_j + \boldsymbol{\varepsilon}_j$, where $\mathbf{x}_j$ and $\boldsymbol{\varepsilon}_j$ are simulated from $N_2(\boldsymbol{\mu}, \mathbf{I})$ and $N_5(\mathbf{0}, \mathbf{I})$ respectively, and

$$\boldsymbol{\mu} = \begin{pmatrix} 1.5 \\ 0.8 \end{pmatrix}, \qquad \boldsymbol{\Theta} = \begin{pmatrix} -2.0 & 0.5 \\ 1.0 & -1.0 \\ 0.5 & 1.0 \\ 0.5 & -2.5 \\ 0.0 & 2.0 \end{pmatrix}.$$

A 2-dimensional wandering vector model, M_2, is fitted to this simulated data set. Using a rather diffuse but proper prior: $\boldsymbol{\lambda} \sim N_{kd-d(d+1)/2}(\boldsymbol{\lambda}_0 = \mathbf{0}, \mathbf{P}^{-1} = 1000\mathbf{I})$, and $\boldsymbol{\mu} \sim N_d(\boldsymbol{\mu}_0 = \mathbf{0}, \mathbf{A}_0^{-1} = 1000\mathbf{I})I(\boldsymbol{\mu} > \mathbf{0})$, we iterated steps (1') to (4) 30,000 times. The first 20,000 burn-in iterations were discarded. The remaining 10,000 iterations were taken to be a sample from the joint posterior distribution of the unknown parameters, the latent variables U_{ij}'s, and the $\mathbf{x}_j$'s. Note that if the standard error of the posterior moment estimates are required, one may use the robust variance estimation method suggested by Andrews (1991). Table 1 shows the posterior means, standard deviations and 90% posterior interval of the parameters of M_2 based on the reduced sample. The posterior intervals are constructed from the 0.05 and 0.95 percentiles of the Gibbs samples.

We can see from Table 1 that all posterior means are fairly close to their true values and all posterior 90% intervals cover the true values. This indicates that the posterior means are good estimators for the parameters in the wandering vector model. Figure 2 shows the traces of the Gibbs sequences and the Gaussian kernel density estimates of the posterior densities for μ_1, μ_2, θ_{21}, θ_{31}, θ_{42}, and θ_{52}. The traces of the Gibbs sequences do not show any special pattern, implying that the Gibbs iteration converged. Furthermore, all posterior densities are roughly symmetric.

Table 1. Results of fitting M_2 to the simulated data set.

		posterior moments		
parameter	true value	mean	std. dev.	90% interval
μ_1	1.5	1.432	0.079	(1.309, 1.567)
μ_2	0.8	0.764	0.045	(0.691, 0.839)
θ_{11}	−2.0	−2.005	0.136	(−2.231, −1.768)
θ_{12}	0.5	0.398	0.102	(0.229, 0.564)
θ_{21}	1.0	1.027	0.070	(0.907, 1.145)
θ_{22}	−1.0	−0.940	0.081	(−1.076, −0.808)
θ_{31}	0.5	0.484	0.045	(0.410, 0.560)
θ_{32}	1.0	1.058	0.086	(0.920, 1.204)
θ_{41}	0.5	0.494	0.066	(0.387, 0.603)
θ_{42}	−2.5	−2.527	0.126	(−2.741, −2.324)
θ_{52}	2.0	2.010	0.097	(1.851, 2.175)

Consider fitting a 2-dimensional (M_2) and a 3-dimensional (M_3) wandering vector models to the simulated data set. Applying Chib's method with $G = 5000$, the logarithm of marginal likelihoods of models M_2 and M_3 are found to be -6513.7 and -6578.5 respectively. The 2-dimensional wandering vector model gives the larger estimate of marginal likelihood.

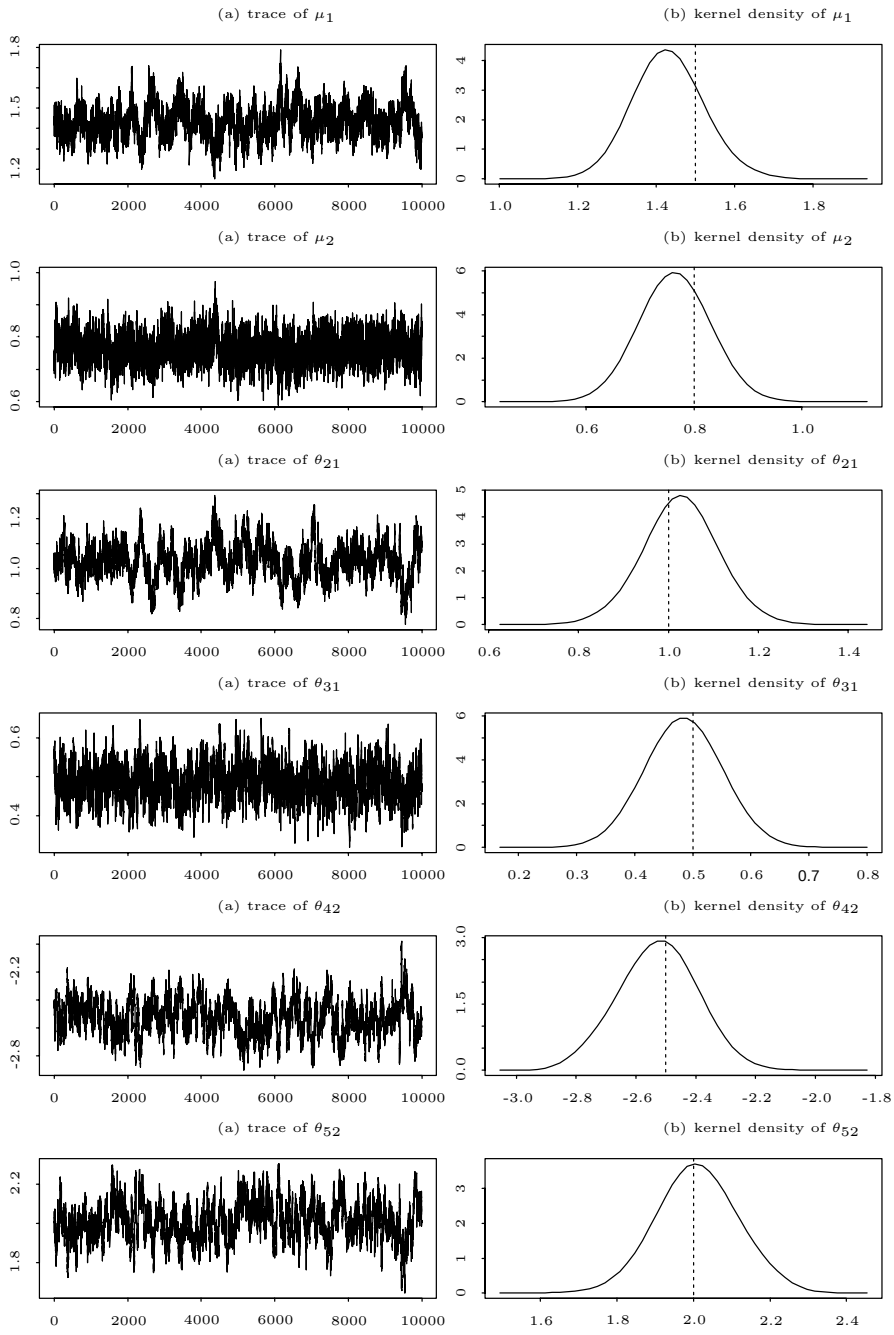


Figure 2. (a) Traces of the Gibbs sequences and (b) Posterior densities for $\mu_1, \mu_2, \theta_{21}, \theta_{31}, \theta_{42}$, and θ_{52} for the simulated data set. The dotted line indicates the position of the true value.

7. Application to the Goldberg Data

Consider the Goldberg (1976) data in which $n = 143$ graduates were asked to rank $k = 10$ occupations according to the degree of social prestige. These 10 occupations are: (i) Faculty member in an academic institution (`Fac`), (ii) Mechanical engineer (`ME`), (iii) Operation researcher (`OR`), (iv) Technician (`Tech`), (v) Section supervisor in a factory (`Sup`), (vi) Owner of a company employing more than 100 workers (`Own`), (vii) Factory foreman (`For`), (viii) Industrial engineer (`IE`), (ix) Manager of a production department employing more than 100 workers (`Mgr`) and (x) Applied scientist (`Sci`). The data are given in Cohen and Mallows (1980) and have been analyzed by many researchers. Fligner and Verducci (1988) and Marden (1992) summarized the findings of these analyses.

Feigin and Cohen (1978) analyzed the Goldberg data and found three outliers due to the fact that the corresponding graduates wrongly presented rankings in reverse order. After reversing these 3 rankings, the average ranks received by the 10 occupations are 8.57, 4.90, 6.29, 1.90, 4.34, 8.13, 1.47, 6.27, 5.29, 7.85, with the convention that higher rank means more prestige. Then the preference of graduates is in the order: `Fac` $>$ `Own` $>$ `Sci` $>$ `OR` $>$ `IE` $>$ `Mgr` $>$ `ME` $>$ `Sup` $>$ `Tech` $>$ `For`.

We consider a graphical method of displaying ranking data by fitting the wandering vector models with dimensions $d = 2, 3$ and 4. Using the same prior specified in the simulation studies, and the last 10,000 of 40,000 Gibbs iterations, we obtain three sets of Gibbs samples from which the marginal likelihoods of the three fitted models are estimated in a logarithm scale, as shown in Table 2. It can be seen that the 3-dimensional model gives the largest marginal likelihood.

Table 2. The logarithm of marginal likelihoods for the Goldberg data.

model fitted	log(marginal)
M_2	-1945.88
M_3	-1791.43
M_4	-1904.48

Based on the Gibbs samples generated from the fitted 3-dimensional model, posterior means are used as estimators for the model parameters. Figure 3 displays the 3-dimensional solution for the Goldberg data after a varimax rotation. The interpretation is as follows.

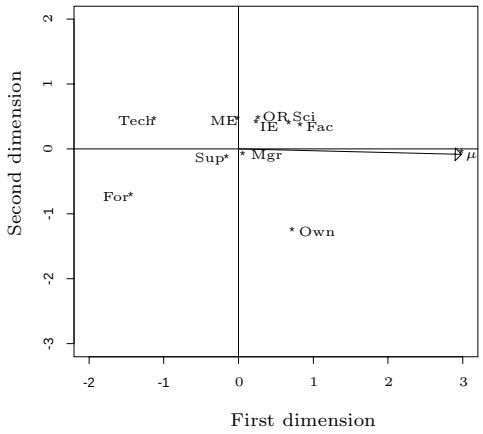
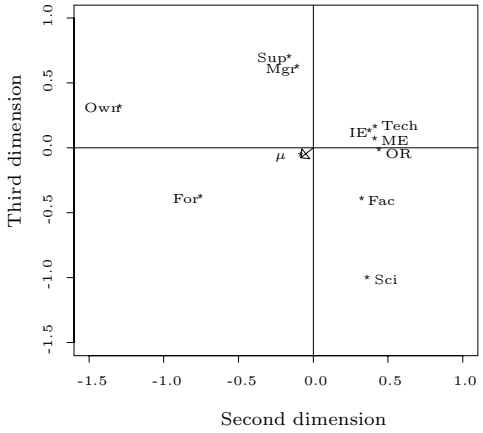


Figure 3. (a) Plot of the first and second dimensions of the fitted 3-dimensional model.



(b) Plot of the second and third dimensions of the fitted 3-dimensional model.

1. The first dimension represents an overall prestige and indicates the mean direction of the judges' preference. The larger the value in this direction, the more prestigious the occupation. Not surprisingly, when the occupation points are projected onto this dimension, the ordering of the lengths of the projections are the same as that of the average ranks.
2. The second dimension can be regarded as the Technical/Managerial dichotomy. It separates the occupations into two groups: Technical group {`Fac, Sci, OR, IE, ME, Tech`} and Managerial group {`Own, Mgr, Sup, For`}.
3. The third dimension represents a measure of social skill required for an occupation. Clearly, people working in the group {`Mgr, Sup`} require more social skill than those working as `Sci`.

Apart from the posterior means, the expected utilities $\boldsymbol{\mu}'\boldsymbol{\theta}_i$ can also provide

useful information. If $\boldsymbol{\mu}'\boldsymbol{\theta}_i$ is larger than $\boldsymbol{\mu}'\boldsymbol{\theta}_j$, it is more likely that occupation i is preferred to occupation j. According to the boxplots of the expected utilities of the occupations shown in Figure 4, the occupations can be roughly divided into three groups: high prestige group {`Fac, Own, Sci`}, middle prestige group {`OR, IE, Mgr, ME, Sup`}, and low prestige group {`Tech, For`}. When we focus on the posterior means of the expected utilities of the occupations, it is not hard to see that the ordering of these posterior means also agrees with that of the average ranks.

Finally, it should be noted that most of the above findings are similar to those found in the literature except in the optimal number of dimensions for displaying Goldberg data. Only two-dimensional graphical displays have been reported (see for example, Cohen and Mallows (1980)). We have suggested a new dimension—social skill.

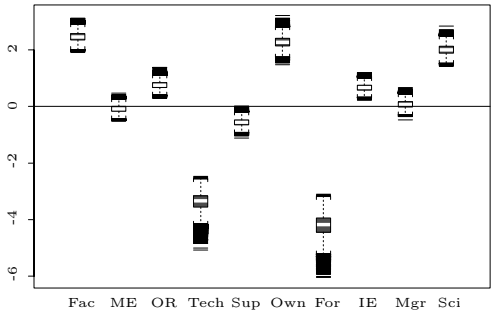


Figure 4. Boxplots of the expected utilities for the fitted 3-dimensional wandering vector model.

8. Comments

As commented by a referee, since only the partial information (i.e., ranks) is used, this method provides a good example of what Doksum and Lo (1986) have termed Bayesian robustness.

In our simulation studies and the application presented in Sections 6 and 7, the CPU times run on an IBM 9076 SP2 machine were all less than one hour. One way to shorten the running time is to develop some block Gibbs samplings instead of separate Gibbs samplings. Further research in this aspect is needed.

Appendix. Justification for using the constraints on Θ and μ for fixing the rotation problem

Following Section 3.1 of Thisted (1988), we first construct a $d \times d$ orthogonal matrix $\mathbf{Q}$ as a product of Householder matrices $\mathbf{H}_d \ldots \mathbf{H}_1$ such that

$$\mathbf{Q}\boldsymbol{\Theta}' = \begin{pmatrix} \tilde{\theta}_{11} & \cdots & \tilde{\theta}_{k-d+1,1} & 0 & \cdots & 0 \\ \tilde{\theta}_{12} & \cdots & \tilde{\theta}_{k-d+1,2} & \tilde{\theta}_{k-d+2,2} & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & 0 \\ \tilde{\theta}_{1,d} & \cdots & \tilde{\theta}_{k-d+1,d} & \tilde{\theta}_{k-d+2,d} & \cdots & \tilde{\theta}_{k,d} \end{pmatrix} \equiv \tilde{\boldsymbol{\Theta}}', \tag{A.1}$$

where each $\mathbf{H}_i$ aims to zero the first $d-i$ positions of $\boldsymbol{\theta}_{k-i+1}$, the $(k-i+1)$th column of $\boldsymbol{\Theta}'$. As mentioned in Thisted (1988), a special property of the Householder matrices is that each $\mathbf{H}_i$ does not change any of the last $i-1$ columns to which it is applied.

Now, we need to show that if we set the constrained $\boldsymbol{\Theta}$ to the matrix $\tilde{\boldsymbol{\Theta}}$ in (A.1) and assume $\mu_1, \ldots, \mu_d > 0$, the rotation problem is fixed. Mathematically, it is equivalent to show that if there exist an orthogonal matrix $\mathbf{T}$ and a vector $\hat{\boldsymbol{\mu}} > \mathbf{0}$ such that (i) $\hat{\boldsymbol{\Theta}}' \equiv \mathbf{T}\tilde{\boldsymbol{\Theta}}'$ has the same form of $\tilde{\boldsymbol{\Theta}}'$ and (ii) $\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Theta}}\boldsymbol{\mu}$, where $\boldsymbol{\mu} > \mathbf{0}$, then $\mathbf{T} = \mathbf{I}$.

Here is a proof. By comparing each entry of both sides of $\hat{\boldsymbol{\Theta}}' \equiv \mathbf{T}\tilde{\boldsymbol{\Theta}}'$, it can be shown that $\mathbf{T}$ is a diagonal matrix. Since $\mathbf{T}$ is orthogonal, $t_{ii} = \pm 1$ for $i = 1, \ldots, d$. Using (ii), we have $(\hat{\theta}_{1i}, \ldots, \hat{\theta}_{ki}) = t_{ii}(\tilde{\theta}_{1i}, \ldots, \tilde{\theta}_{ki})$ for all i, and $\hat{\mu}_i = \mu_i > 0$ if $t_{ii} = 1$ or $\hat{\mu}_i = -\mu_i < 0$ if $t_{ii} = -1$. Since $\hat{\boldsymbol{\mu}} > \mathbf{0}$, $t_{ii} = 1$ for all i. In other words, $\mathbf{T} = \mathbf{I}$. This completes the proof.

Acknowledgements

The authors would like to thank the referees for their helpful comments and suggestions, and Douglas E. Critchlow for providing the Goldberg data. Philip L. H. Yu's research was substantially supported by a block grant from The University of Hong Kong and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7169/98H).

References

Alvo, M. and Ertas, K. (1992). Graphical methods for ranking data. *Canad. J. Statist.* **20**, 469-482.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley, New York.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10**, 1-19.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. 1 (Edited by R. N. Shepard, A. K. Romney and S. B. Nerlove), 105-155. Seminar Press, New York.

Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice or other dominance data. In *Similarity and Choice* (Edited by E. D. Lantermann and H. Feger). Huber, Bern, Switzerland.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90**, 1313-1321.

Cohen, A. and Mallows, C. L. (1980). *Analysis of Ranking Data*. Bell Laboratories Technical Memorandum.

Cox, T. F. and Cox, M. A. A. (1994). *Multidimensional Scaling*. Chapman and Hall, London.

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

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Doksum, K. A. and Lo, A. Y. (1986). Comments on 'On the consistency of Bayes estimates' by Diaconis and Freedman. *Ann. Statist.* **14**, 42-45.

Feigin, P. and Cohen, A. (1978). On a model of concordance between judges. *J. Roy. Statist. Soc. Ser. B* **40**, 203-213.

Fligner, M. A. and Verducci, J. S. (1988). Multistage ranking models. *J. Amer. Statist. Assoc.* **83**, 892-901.

Gabriel, K. R. (1971). The biplot graphic display of matrices with applications to principal component analysis. *Biometrika* **58**, 453-468.

Goldberg, A. I. (1976). The relevance of cosmopolitan/local orientations to professional values and behavior. *Sociology of Work and Occupation* **3**, 331-356.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. 3rd edition. Prentice Hall, Upper Saddle River, New Jersey.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773-795.

Marden, J. I. (1992). Use of nested orthogonal contrasts in analyzing rank data. *J. Amer. Statist. Assoc.* **87**, 307-318.

Marden, J. I. (1995). *Analyzing and Modeling Rank Data*. Chapman and Hall, London.

Raftery, A. E. (1995). Hypothesis testing and model selection with posterior simulation. In *Practical Markov Chain Monte Carlo* (Edited by W. R. Gilks, D. J. Spiegelhalter and S. Richardson). Chapman and Hall, London.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. John Wiley, New York.

Thisted, R. A. (1988). *Elements of Statistical Computing*. Chapman and Hall, London.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion and rejoinder). *Ann. Statist.* **22**, 1701-1762.

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: plhyu@hku.hk

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: luciac@hongkong.com

(Received March 1998; accepted September 2000)