

## Chapter 9

# Probit Models for Ranking Data

In 1980, the American Psychological Association (APA) conducted an election in which five candidates ( $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$ ) were running for president and voters were asked to rank order all of the candidates. Candidates  $A$  and  $B$  are research psychologists,  $C$  is a community psychologist, and  $D$  and  $E$  are clinical psychologists. Among those voters, 5738 gave complete rankings. These complete rankings are considered here (Diaconis (1988)). Note that lower rank implies more favorable. Then the average ranks received by candidates  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$  are 2.84, 3.16, 2.92, 3.09, and 2.99, respectively. This means that voters generally prefer candidate  $A$  the most, candidate  $C$  the second, etc. However, in order to make inferences on the preferences of the candidates, modeling of the ranking data is needed. In Sect. 9.1 we consider a model for this data which takes into account covariates.

In Sect. 9.2 we consider the following example for which factor analysis would be appropriate. In 1997, a mainland marketing research firm conducted a survey on people's attitude toward career and living style in three major cities in Mainland China – Beijing, Shanghai, and Guangzhou. Five hundred responses from each city were obtained. A question regarding the behavior, conditions, and criteria for job selection of the 500 respondents in Guangzhou will be discussed here. In the survey, respondents were asked to rank the three most important criteria on choosing a job among 13 criteria: (1) favorable company reputation, (2) large company scale, (3) more promotion opportunities, (4) more training opportunities, (5) comfortable working environment, (6) high income, (7) stable working hours, (8) fringe benefits, (9) well matched with employees' profession or talent, (10) short distance between working place and home, (11) challenging, (12) corporate structure of the company, and (13) low working pressure.

## 9.1 Multivariate Normal Order Statistics Models

In light of the Thurstone order statistics model mentioned in Sect. 8.1, the multivariate normal order statistics (MVNOS) model assumes that the ranking of  $t$  objects given by a judge is determined by ordering  $t$  latent utilities for the objects assigned by the judge. However, unlike the Thurstone order statistics model that assumes independent utilities, the MVNOS model assumes that the utilities are possibly correlated and the ranking  $\pi_j$  given by judge  $j$  has the following probability:

$$P(\pi_j) = P(y_{[1]_j,j} > y_{[2]_j,j} > \cdots > y_{[t]_j,j}), \quad i = 1, \dots, t, \quad (9.1)$$

where  $\langle [1]_j, \dots, [t]_j \rangle$  is the ordering of the  $t$  objects corresponding to the ranking  $\pi_j$  and the latent utility vector  $\mathbf{y}_j = (y_{1j}, \dots, y_{tj})'$  of judge  $j$  is assumed to follow multivariate normal distribution with mean utility vector  $\boldsymbol{\mu}_j = (\mu_{1j}, \dots, \mu_{tj})'$  and a general covariance matrix  $\mathbf{V}$ , i.e.,

$$\mathbf{y}_j = \boldsymbol{\mu}_j + \mathbf{e}_j \quad (9.2)$$

$$\mathbf{e}_j \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V}). \quad (9.3)$$

The MVNOS model is sometimes termed the *multinomial probit model* for ranking data.

### 9.1.1 The MVNOS Model with Covariates

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

$$\boldsymbol{\mu}_j = \mathbf{Z}_j \boldsymbol{\beta}, \quad (9.4)$$

where  $\mathbf{Z}_j$  is a  $t \times p$  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 ranking given by the respondents, some socioeconomic variables ( $s_j$ ) 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:

$$\mu_{ij} = \mathbf{a}'_i \boldsymbol{\gamma} + s'_j \boldsymbol{\delta}_i, \quad i = 1, \dots, t. \quad (9.5)$$

The parameter vector  $\boldsymbol{\gamma}$  represents the attribute effect common to all the respondents while the vector  $\boldsymbol{\delta}_i$  represents the respondents' socioeconomic background which may affect their preference of product  $i$ . It is easily seen that equation (9.5) is a particular case of the model in (9.4) when

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

In what follows, we shall consider the MVNOS model with the mean given in equation (9.4).

### 9.1.2 Parameter Identifiability 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 (9.2) while leaving the ranking probability unchanged. The location-shift problem is commonly dealt with by subtracting the first  $t - 1$  rows by the last row leading to the model

$$\mathbf{w}_j = \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j \tag{9.6}$$

$$\boldsymbol{\varepsilon}_j \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{9.7}$$

where  $w_{ij} = y_{ij} - y_{tj}$ ,  $\mathbf{X}_j = [\mathbf{I}_{t-1}, -\mathbf{1}_{t-1}] \mathbf{Z}_j$ ,  $\boldsymbol{\varepsilon}_{ij} = e_{ij} - e_{tj}$ , and

$$\boldsymbol{\Sigma} = [\mathbf{I}_{t-1}, -\mathbf{1}_{t-1}] \mathbf{V} [\mathbf{I}_{t-1}, -\mathbf{1}_{t-1}]'.$$

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  $\langle [1]_j, \dots, [t]_j \rangle$  corresponds to the event

$$E_j = \{ \mathbf{w}_j : w_{[1]_j, j} > \cdots > w_{[r-1]_j, j} > 0 > w_{[r+1]_j, j} > \cdots > w_{[t]_j, j} \} \\ \text{whenever } [r]_j = t. \tag{9.8}$$

For the sake of simplicity, we use the convention  $w_{[0]_j, j} = +\infty$  and  $w_{[t+1]_j, j} = -\infty$ . Notice that the scale-shift problem still exists in the model given by (9.6) and it can be easily resolved by adding a constraint on  $\boldsymbol{\Sigma}$  such as  $\sigma_{11} = 1$ .

Since rankings of objects only depend on utility differences,  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}$  (with  $\sigma_{11}$  fixed) are estimable, but the original parameters  $\mu_j$  and  $\mathbf{V}$  still cannot be fully identified. For example, suppose  $t = 3$  and  $\mu_j = \mu$ . Then the following three sets of parameters under the MVNOS model lead to the same ranking probabilities:

$$\begin{array}{lll}
 \text{Set A:} & \text{Set B:} & \text{Set C:} \\
 \mu_A = (1, 0, -1)' & \mu_B = (-1, -2, -3)' & \mu_C = (1, 0, -1)' \\
 V_A = \begin{pmatrix} 1 & 0 & 0.2 \\ 0 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix} & V_B = \begin{pmatrix} 0.4 & 0 & 0.4 \\ 0 & 1.6 & 1.6 \\ 0.4 & 1.6 & 2 \end{pmatrix} & V_C = \begin{pmatrix} 0.756 & -0.444 & -0.311 \\ -0.444 & 0.356 & 0.089 \\ -0.311 & 0.089 & 0.222 \end{pmatrix}
 \end{array}$$

This is because they all have the same utility differences  $y_1 - y_3$  and  $y_2 - y_3$  whose joint distribution is

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

Generally speaking, the parameter  $\beta$  can be identified in the presence of covariates  $X_j$ . However, when there are no covariates, i.e.,  $\mu_j = \mu$ , the values of the  $\mu_i$ 's will be determined only within a location shift. This indeterminacy can be eliminated by imposing one constraint on the  $\mu_i$ 's, say,  $\mu_t = 0$ .

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

$$V \longrightarrow V + c\mathbf{1}'_t + \mathbf{1}_t c', \tag{9.9}$$

for any constant vector  $c$  (Arbuckle and Nugent 1973),  $V$  can never be identified unless it is structured. In the previous example, it can be seen that  $V_A$  can be transformed to  $V_B$  and  $V_C$  by setting  $c = (-0.3, 0.3, 0.5)'$  and  $c = (-0.122, -0.322, -0.389)'$ , respectively. This identification problem is well known in the context of Thurstone order statistics models and multinomial probit models (Arbuckle and Nugent 1973; Dansie 1985; Bunch 1991; Yai et al. 1997; Train 2003).

Various solutions which impose constraints on the covariance matrix  $V$  have been proposed in the literature. Among them, the methods proposed by Chintagunta (1992) and Yu (2000) provide the most flexible form for  $V$  which does not require fixing any cell. Chintagunta's method restricts each column sum of  $V$  to zero (and  $\sigma_{11} = 1$ ), resulting in  $V = B^{-1}\Sigma(B^{-1})'$ , with  $B = [I_{t-1}, -\mathbf{1}_{t-1}]$  while Yu's method restricts each column sum of  $V$  to 1 (and  $\sigma_{11} = 1$ ), leading to

$$V = A^{-1} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & t \end{bmatrix} (A')^{-1} \text{ with } A = \begin{bmatrix} I_{t-1} & -\mathbf{1}_{t-1} \\ \mathbf{1}'_{t-1} & 1 \end{bmatrix}.$$

Note that the  $V$  identified by Chintagunta's method is singular and the associated utilities must be correlated whereas Yu's method always produces a non-singular matrix  $V$  and includes the identity matrix (or its scale shift) as a special case. In addition, it is easy to show that this non-singular matrix is an invariant transformation of the matrix used by Chintagunta (1992) under the transformation (9.9) with  $c = \frac{1}{2t}\mathbf{1}_t$ .

### 9.1.3 Bayesian Analysis of the MVNOS Model

Given a sample of  $n$  judges, the likelihood function of  $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$  is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \prod_{j=1}^n P(E_j), \tag{9.10}$$

where the event  $E_j$  is given in (9.8). Note that the evaluation of the above likelihood function requires the numerical approximation of the  $(t - 1)$ -dimensional integral (e.g., Genz 1992) which can be done relatively accurately provided that the number of objects ( $t$ ) is small, say less than 15. To avoid a high-dimensional numerical integration, limited information methods using the induced paired/triple-wise comparisons from the ranking data (e.g., structural equation models by Maydeu-Olivares and Bockenholt (2005) fitted using Mplus) have been proposed. Another approach is to use a Monte Carlo Expectation-Maximization (MCEM) algorithm (e.g., Yu et al. 2005; see also Sect. 9.2) which can avoid the direct maximization of the above likelihood function.

In this section we will consider a simulation-based Bayesian approach which can also avoid the evaluation and maximization of the above likelihood function. Recently, a number of R packages have become available for the Bayesian estimation of the MVNOS models for ranking data, including MNP (Imai and van Dyk 2005), rJAGS (Johnson and Kuhn 2013) as well as our own package StatMethRank.

#### 9.1.3.1 Bayesian Estimation and Prior Distribution

In a Bayesian approach, the first step is to specify the prior distribution of the identified parameters. As mentioned previously one constraint on  $\boldsymbol{\Sigma}$  could be added in order to fix the scale and hence to identify all the parameters. Under this condition, the usual Wishart prior distribution for the constrained  $\boldsymbol{\Sigma}$  could not be used. In the context of multinomial probit model studied by McCulloch and Rossi (1994), instead of imposing the scale constraint on  $\boldsymbol{\Sigma}$ , we may compute the full posterior distribution of  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}$  and obtain the marginal posterior distribution of the identified parameters such as  $\boldsymbol{\beta}/\sqrt{\sigma_{11}}$ ,  $\sigma_{ii}/\sigma_{11}$  and  $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ .

Let  $f(\boldsymbol{\beta}, \boldsymbol{\Sigma})$  denote the joint prior distribution of  $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ . Then the posterior density of  $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$  is

$$f(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \boldsymbol{\Pi}) \propto L(\boldsymbol{\beta}, \boldsymbol{\Sigma})f(\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{9.11}$$

where  $\boldsymbol{\Pi} = \{\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n\}$  is the data set of all  $n$  observed rankings. It is convenient to use a normal prior on  $\boldsymbol{\beta}$ ,

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

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

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

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

Although (9.11) is intractable for Bayesian calculations, we may use the method of Gibbs sampling with data augmentation. We augment the parameter  $(\boldsymbol{\beta}, \Sigma)$  by the latent variable  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)$ . Now, the joint posterior density of  $(\boldsymbol{\beta}, \Sigma, \mathbf{W})$  is

$$f(\boldsymbol{\beta}, \Sigma, \mathbf{W} | \boldsymbol{\pi}) \propto f(\boldsymbol{\Pi} | \mathbf{W}) f(\mathbf{W} | \boldsymbol{\beta}, \Sigma) f(\boldsymbol{\beta}, \Sigma), \quad (9.12)$$

which allows us to sample from the full conditional posterior distributions. The details are provided in the next section.

### 9.1.3.2 Gibbs Sampling Algorithm for the MVNOS Model

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

1. Draw  $\mathbf{w}_j$  from  $f(\mathbf{w}_j | \boldsymbol{\beta}, \Sigma, \boldsymbol{\Pi})$ , for  $j = 1, \dots, n$ .
2. Draw  $\boldsymbol{\beta}$  from  $f(\boldsymbol{\beta} | \Sigma, \mathbf{W}, \boldsymbol{\Pi}) \propto f(\boldsymbol{\beta} | \Sigma, \mathbf{W})$ .
3. Draw  $\Sigma$  from  $f(\Sigma | \boldsymbol{\beta}, \mathbf{W}, \boldsymbol{\Pi}) \propto f(\Sigma | \boldsymbol{\beta}, \mathbf{W})$ .

In step (1), it can be shown that given  $\boldsymbol{\beta}$ ,  $\Sigma$ , and  $\boldsymbol{\Pi}$ , the  $\mathbf{w}_j$ 's are independent and  $\mathbf{w}_j$  follows a truncated multivariate normal distribution,  $N(\mathbf{X}_j \boldsymbol{\beta}, \Sigma) I(\mathbf{w}_j \in E_j)$ . One may simulate  $\mathbf{w}_j$  by using the acceptance-rejection technique, but this may lead to a high rejection rate when the number of objects is fairly large. Instead of drawing the whole vector  $\mathbf{w}_j$  at one time, we successively simulate each entry of  $\mathbf{w}_j$  by conditioning on the other  $t - 2$  entries. More specifically, we replace step (1) by

1. draw  $w_{ij}$  from  $f(w_{ij} | w_{-i,j}, \boldsymbol{\beta}, \Sigma, \boldsymbol{\Pi})$ , for  $i = 1, \dots, t - 1, j = 1, \dots, n$ , where  $w_{-i,j}$  is  $\mathbf{w}_j$  with  $w_{ij}$  deleted.

Let  $\mathbf{x}'_{ij}$  be the  $i$ th row of  $\mathbf{X}_j$ ,  $\mathbf{X}_{-i,j}$  be  $\mathbf{X}_j$  with the  $i$ th row deleted, and  $\mathbf{g}_{-i,i}$  be the  $i$ th column of  $\mathbf{G}$  with  $g_{ii}$  deleted. Suppose  $\langle [1]_j, \dots, [t]_j \rangle$  is the ordering of objects corresponding to their ranks  $\boldsymbol{\pi}_j = (\pi_{1j}, \dots, \pi_{tj})$ . Then  $\pi_{ij} = r$  if and only if  $[r]_j = i$ . Now we have

$$\begin{aligned} w_{ij} | w_{-i,j}, \boldsymbol{\beta}, \Sigma, \boldsymbol{\Pi} &\sim N(m_{ij}, \tau_{ij}^2) \\ \text{subject to } w_{[r+1]_j} < w_{ij} < w_{[r-1]_j} &\text{ whenever } \pi_{ij} = r, \end{aligned} \quad (9.13)$$

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  $\tau_{ij}^2 = g_{ii}^{-1}$ .

Although it still involves simulation from a truncated univariate normal distribution, we can adopt the inverse method to sample from this distribution without using the acceptance-rejection technique which may not be efficient (Devroye 1986).

Returning to steps (2) and (3), since we are conditioning on  $\mathbf{W}$ , the MVNOS model is simply a standard Bayesian linear model setup. Therefore, the full conditional distribution of  $\boldsymbol{\beta}$  is

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}, \mathbf{W} \sim N_p(\boldsymbol{\beta}_1, \mathbf{A}_1^{-1}), \quad (9.14)$$

where

$$\mathbf{A}_1 = \mathbf{A}_0 + \sum_{j=1}^n \mathbf{X}'_j \boldsymbol{\Sigma}^{-1} \mathbf{X}_j \quad \text{and} \quad \boldsymbol{\beta}_1 = \mathbf{A}_1^{-1} (\mathbf{A}_0 \boldsymbol{\beta}_0 + \sum_{j=1}^n \mathbf{X}'_j \boldsymbol{\Sigma}^{-1} \mathbf{w}_j).$$

Finally, the full conditional distribution of  $\boldsymbol{\Sigma}$  is such that  $\boldsymbol{\Sigma} = \mathbf{G}^{-1}$  with

$$\mathbf{G} | \boldsymbol{\beta}, \mathbf{W} \sim W_{t-1} \left( \alpha + n, \mathbf{P} + \sum_{j=1}^n (\mathbf{w}_j - \mathbf{X}'_j \boldsymbol{\beta})(\mathbf{w}_j - \mathbf{X}'_j \boldsymbol{\beta})' \right). \quad (9.15)$$

With a starting value for  $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{W})$ , we draw in turn from each of the full conditional distributions given by (9.13), (9.14), and (9.15). When this process is repeated many times, the draws obtained will converge to a single draw from the full joint posterior distribution of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Sigma}$ , and  $\mathbf{W}$ . In practice, we iterate the process  $M + N$  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 independent draws from the joint posterior distribution. The value  $s$  here can be determined based on the graph of the sample autocorrelation of the Gibbs iterates.

A natural choice for a starting value for  $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$  is to use  $(0, \mathbf{I})$ . However, it is nonstandard to find a starting value for  $\mathbf{W}$ . We adopt an approach motivated by the fact that the ranking of  $\{w_{1j}, \dots, w_{t-1,j}, 0\}$  must be consistent with the observed ranking  $\{\pi_{1j}, \dots, \pi_{tj}\}$ . Using this fact, a simple choice for the starting value of the  $w$ 's is to use  $w_{ij} = (\pi_{ij} - \pi_{tj}) / \sqrt{(t^2 - 1)/12}$ , a type of standardized rank score.

It should be remarked that since Thurstone's normal order statistics model is a MVNOS model with  $V = \mathbf{I}_t$ , its parameters can be estimated by fixing  $V$  to  $\mathbf{I}_t$ , or equivalently, fixing  $\boldsymbol{\Sigma}$  to  $\mathbf{I}_{t-1} + \mathbf{1}_{t-1} \mathbf{1}'_{t-1}$  and skipping the step of generating  $\boldsymbol{\Sigma}$  in the above Gibbs sampling algorithm.

*Remark.* Although the MVNOS model discussed here considered the case of the complete ranking of  $t$  objects, it is not difficult to extend it to incorporate incomplete or partial ranking by modifying the event  $E_j$  in (9.8) and the corresponding truncation rule in (9.13) used to sample  $w_{ij}$  in the Gibbs sampling. For instance a partial ordering of 4 objects  $A, B, C$ , and  $D$  given by judge  $j$  is  $B > C > A, D$ .

The event  $E_j$  will then be modified to  $\{w_j : \max\{w_{Aj}, 0\} < w_{Cj} < w_{Bj}\}$ , and hence  $w_{Aj}$ ,  $w_{Bj}$ , and  $w_{Cj}$  will be separately simulated from truncated normal over intervals  $(-\infty, w_{Cj})$ ,  $(w_{Cj}, +\infty)$ , and  $(\max\{w_{Aj}, 0\}, w_{Bj})$ , respectively. So far, we assume that the data does not contain tied ranks or, equivalently, the observed ordering of the objects that are tied is unknown. For example we will treat the tied ranking  $B \succ C \succ A = D$  as if the partial ranking  $B \succ C \succ A, D$ . See Sect. 9.2.1 for similar treatments of incomplete rankings in the context of factor analysis.

### 9.1.4 Adequacy of the Model Fit

To test for the adequacy of the model, we may group the  $t!$  rankings into a small number of meaningful subgroups and examine the fit for each subgroup. In particular, let  $n_i$  be the observed frequency that object  $i$  is ranked as the top object. Also let

$$\hat{p}_i = P(Y_i > Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_t | \hat{\beta}, \hat{\Sigma})$$

be the estimated partial probability of ranking object  $i$  as first under the fitted MVNOS model with posterior mean estimates  $\hat{\beta}$  and  $\hat{\Sigma}$ . The fit can be examined by comparing the observed frequency  $n_i$  with the expected frequency  $n\hat{p}_i$ ,  $i = 1, 2, \dots, t$ , or by calculating the standardized residuals:

$$r_i = \frac{n_i - n\hat{p}_i}{\sqrt{n\hat{p}_i(1 - \hat{p}_i)}}, \quad j = 1, 2, \dots, t.$$

If the expected frequencies match the observed frequencies well or the absolute values of the residuals are small enough, say,  $< 2$ , the MVNOS model adequately fits the data. The same argument can be applied to other ranking models.

In performing these calculations, it is necessary to evaluate numerically the estimated probability  $\hat{p}_i$  which may be expressed as

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

where  $\mathbf{v} = (Y_1 - Y_i, \dots, Y_{i-1} - Y_i, Y_{i+1} - Y_i, \dots, Y_t - Y_i)' \sim N(\beta^*, \Sigma^*)$  and  $\beta^*$  and  $\Sigma^*$  can be obtained from  $\hat{\beta}$  and  $\hat{\Sigma}$ , respectively. We employ the Geweke-Hajivassiliou-Keane (GHK) method (see Geweke 1991; Hajivassiliou 1993; Keane 1994). Let  $\mathbf{L} = (\ell_{ij})$  be the unique lower triangular matrix obtained from the Cholesky decomposition of  $\Sigma^*$  (i.e.,  $\Sigma^* = \mathbf{L}\mathbf{L}'$ ). The GHK simulator for the estimated partial probability  $\hat{p}_i$  is constructed via the following steps:



1. Compute

$$P(v_1 < 0 | \boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*) = \Phi\left(-\frac{\beta_1^*}{\ell_{11}}\right),$$

and draw a  $\eta_1 \sim N(0, 1)$  with  $\eta_1 < -\frac{\beta_1^*}{\ell_{11}}$ .

2. For  $s = 2, \dots, t - 1$ , compute  $P(v_s < 0 | \eta_1, \eta_2, \dots, \eta_{s-1}, \boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*) = \Phi\left(-\frac{\beta_s^* + \sum_{j=1}^{s-1} \ell_{sj} \eta_j}{\ell_{ss}}\right)$ , and draw a  $\eta_s \sim N(0, 1)$  with  $\eta_s < -\frac{\beta_s^* + \sum_{j=1}^{s-1} \ell_{sj} \eta_j}{\ell_{ss}}$ .
3. Estimate  $\hat{p}_i$  by  $P(v_1 < 0 | \boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*) \prod_{s=2}^{t-1} P(v_s < 0 | \eta_1, \eta_2, \dots, \eta_{s-1}, \boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*)$ .
4. Repeat steps 1–3 a large number of times to obtain independent estimates of  $\hat{p}_i$ , and finally by taking the average of these estimates, the GHK simulator for  $\hat{p}_i$  is obtained. In a later application, we will use 10,000 replications.

### 9.1.5 Analysis of the APA Election Data

We now consider the APA election data. Let  $Y_{ij}$  be the  $j$ th voter’s utility of selecting candidate  $i$ ,  $i = A, B, C, D, E$ . We apply the MVNOS model in which (i) the  $j$ th 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:

$$Y_{ij} = \mu_i + e_{ij}, \quad i = A, B, C, D, E, \quad j = 1, \dots, 5738, \\ (e_{Aj}, e_{Bj}, e_{Cj}, e_{Dj}, e_{Ej})' \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V}),$$

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

$$w_{ij} = \beta_i + \varepsilon_{ij}, \quad i = A, B, C, D, \quad j = 1, \dots, 5738, \\ (\varepsilon_{Aj}, \varepsilon_{Bj}, \varepsilon_{Cj}, \varepsilon_{Dj})' \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

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

Using the proper priors,  $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0 = \mathbf{0}, A_0^{-1} = 100)$  and  $\boldsymbol{\Sigma}^{-1} \sim W_{t-1}(\boldsymbol{\alpha} = t + 1, \mathbf{P} = (t + 1)\mathbf{I})$ , 11,000 Gibbs iterations are generated. The first 1000 burn-in iterations were discarded. As evidenced from the sample autocorrelation of the Gibbs samples (not shown here), keeping every 20th draw in the last 10,000 Gibbs iterations gives approximately independent draws from the joint posterior distribution of the parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}$  of the MVNOS model. By imposing the constraint  $\mu_E = 0$  and our constraint for  $\mathbf{V}$  to the Gibbs sequences, we obtain estimates for  $\mu_i$  ( $i = A, B, C, D, E$ ) and  $v_{ij}$ , ( $i \leq j$ ).

**Table 9.1** 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)

**9.1.5.1 Adequacy of Model Fit and Model Comparison**

To examine the goodness of fit of the MVNOS model, Table 9.1 shows the observed proportions and estimated partial probabilities under the MVNOS model. The two statistics for Thurstone's normal order statistics model and Stern's mixture of Luce (called BTL in his/her paper) models are also listed in Table 9.1 as alternatives to the MVNOS model. Thurstone's model is fitted by repeating the Gibbs sampling with  $V$  fixed at  $I_t$ , while Stern's mixture models were fitted by Stern (1993). Stern found that the data seem to be a mixture of 2 or 3 groups of voters. This feature is also supported by Diaconis's (1989) spectral analysis and McCullagh's (1993b) model of inversions.

As seen from Table 9.1, the estimated partial probabilities for the MVNOS model match the observed proportions very well. Also the magnitudes of the standardized residuals  $r_i$  for the MVNOS model only are all very small ( $< 2$ ), indicating that among the four models considered in Table 9.1, the MVNOS model gives the best fit to the APA election data.

**9.1.5.2 Interpretation of the Fitted MVNOS Model**

Table 9.2 shows 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  $\mu_i$ 's is the same as that of the average ranks. Apart from the posterior means, the Gibbs samples can also provide estimates of the 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  $\Phi(\frac{\mu_A - \mu_C}{\sqrt{v_{AA} + v_{CC} - 2v_{AC}}})$  in the Gibbs samples, which is found to be 0.509 (posterior standard deviation = 0.006).

According to the boxplots of  $\mu_i$ ,  $v_{ii}$ , and  $r_{ij} = v_{ij} / \sqrt{v_{ii}v_{jj}}$  ( $i \neq j$ ) shown in Fig. 9.1, distributions of some parameters are fairly symmetric. In addition, a large estimate of  $v_{CC}$  indicates that voters have fairly large variation of the preference

**Table 9.2** Parameter estimates of the MVNOS model for the APA election data

Parameter	Posterior moments		
	Mean	SD	90% interval
$\mu_A$	0.086	0.015	(0.062, 0.111)
$\mu_B$	-0.071	0.014	(-0.097, -0.048)
$\mu_C$	0.067	0.018	(0.037, 0.095)
$\mu_D$	-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)

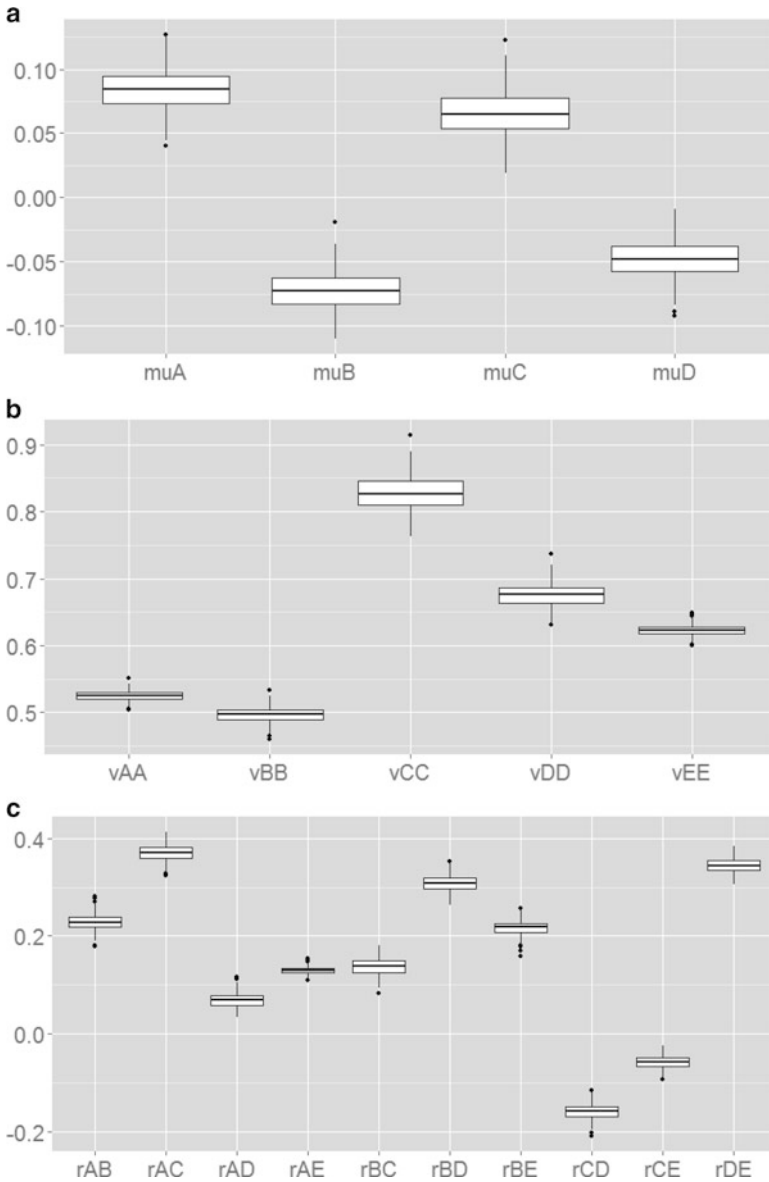
on candidate C. To further investigate the structure of the covariance matrix  $V$ , a principal components analysis of the posterior mean estimate for  $V$  is performed and the result is presented in Table 9.3.

A principal components analysis of the posterior mean estimate for  $V$  produces the utilities of the five candidates  $\{A, B, C, D, E\}$  as

$$\begin{aligned}
 \begin{bmatrix} y_A \\ y_B \\ y_C \\ y_D \\ y_E \end{bmatrix} &= \begin{bmatrix} 0.086 \\ -0.071 \\ 0.067 \\ -0.048 \\ 0 \end{bmatrix} + \sqrt{1.015}a_1z_1 + \sqrt{0.21}a_5z_2 + \sqrt{0.440}a_3z_3 \\
 &\quad + \sqrt{0.357}a_4z_4 + \sqrt{0.346}a_5z_5
 \end{aligned} \tag{9.16}$$

where the  $z$ 's are independently and identically distributed as  $N(0, 1)$  and the principal components  $a$ 's are given in Table 9.3. Since rankings of objects only depend on utility differences, the term  $\sqrt{0.21}a_5z_2$  does not affect the rankings and hence, interpretation is based on the remaining four components.

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



**Fig. 9.1** Boxplots of  $\mu_i$ ,  $v_{ii}$ , and  $r_{ij} = v_{ij} / \sqrt{v_{ii}v_{jj}}$  ( $i \neq j$ ) for the APA election data

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.

**Table 9.3** 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

## 9.2 Factor Analysis

It was mentioned in Sect. 9.1.2 that the parameters of a MVNOS model cannot be fully identified unless the covariance matrix  $V$  is structured. One possibility to resolve this problem is to impose a factor covariance structure used in factor analysis onto  $V$ .

Factor analysis is widely used in social sciences and marketing research to identify the common characteristics among a set of variables. The classical  $d$ -factor model for a set of continuous variables  $y_1, y_2, \dots, y_t$  is defined as

$$y_{ij} = \mathbf{z}'_j \mathbf{a}_i + \varepsilon_{ij}, \quad i = 1, \dots, t; \sim j = 1, \dots, n \tag{9.17}$$

where  $\mathbf{y}_j = (y_{1j}, \dots, y_{tj})'$  is a  $t$ -dimensional vector of response variables from individual  $j$ ,  $\mathbf{z}_j = (z_{1j}, \dots, z_{dj})'$  is a vector of unobserved common factors associated with individual  $j$ ,  $\mathbf{a}_i = (a_{i1}, \dots, a_{id})'$  is a vector of factor loadings associated with object  $i$  on the  $d$  factors, and  $\varepsilon_{ij}$  represents the error of the factor model. By adopting the MVNOS framework with the latent utilities satisfying the above factor model, we can generalize the classical factor model to analyze ranking data. In what follows, we shall assume that the reader has a basic familiarity with the statistical concepts of factor scores, factor loadings, and varimax rotation as can be found in most textbooks on multivariate analysis.

### 9.2.1 The Factor Model

Suppose we have a random sample of  $n$  individuals from the population and each individual is asked to rank  $t$  objects under study according to their own preferences. Within the framework of the MVNOS model, the ranking of the  $t$  objects given by individual  $j$  in the factor model is determined by the ordering of  $t$  latent utilities  $y_{1j}, \dots, y_{tj}$  which satisfies a more general  $d$ -factor model:

$$y_{ij} = \mathbf{z}'_j \mathbf{a}_i + b_i + \varepsilon_{ij} \quad j = 1, \dots, n; i = 1, \dots, t (> d) \tag{9.18}$$

where  $\mathbf{b} = (b_1, \dots, b_t)'$  is the mean utility vector reflecting the relative importance of the  $t$  objects and  $\mathbf{a}_i = (a_{i1}, \dots, a_{id})'$  represents the factor loadings. It is assumed that the latent common factors  $\mathbf{z}_1, \dots, \mathbf{z}_n$  are independent and identically distributed according to the standard  $d$ -variate normal distribution,  $N_d(\mathbf{0}, \mathbf{I})$ . The error term,  $\varepsilon_{ij}$ , is the unique factor which is assumed to follow a  $N(0, \sigma_i^2)$  distribution, independent of the  $\mathbf{z}_i$ 's.

Denote a complete ranking by  $\boldsymbol{\pi}_j = (\pi_{1j}, \dots, \pi_{tj})'$  where  $\pi_{ij}$  is the rank of object  $i$  from individual  $j$ . Smaller ranks refer to the more preferred objects and hence higher utilities. For example, if  $\boldsymbol{\pi}_j = (2, 3, 1)'$  is recorded, it corresponds to the unobservable utilities  $\mathbf{y}_j = (y_{1j}, y_{2j}, y_{3j})'$  with  $y_{2j} < y_{1j} < y_{3j}$ . Note that the only observable quantities are the  $\pi_{ij}$ 's but not the  $y_{ij}$ 's.

*Remark.* Extension of the above factor model to incorporate incomplete ranking data is quite straightforward. In the case of the top  $q$  partial rankings with the top  $q$  objects being  $[1]_j, \dots, [q]_j$  for individual  $j$ , it is natural to assign objects  $[1]_j, \dots, [q]_j$  with ranks  $1, \dots, q$ , respectively, and the rest of objects with midrank, i.e.,  $[(q + 1) + \dots + t]/(t - q)$ . The factor model can be extended to restrict the utilities  $y_{1j}, \dots, y_{tj}$  to satisfy  $y_{[1]_j j} > y_{[2]_j j} > \dots > y_{[q]_j j} > y_{[q+1]_j j}, \dots, y_{[t]_j j}$ . For subset rankings, individuals are asked to rank a subset of the  $t$  objects only. Ranking of the set of remaining objects is unknown and we can simply restrict the ordering of the utilities of objects in the subset consistent to the ranking of these objects. Generally speaking, a ranking  $\boldsymbol{\pi}$ , complete or incomplete, corresponds to an event  $\{\mathbf{y} : C\mathbf{y} < \mathbf{0}\}$ , for some contrast matrix  $C$ . For instance in the case of ranking  $t = 4$  objects, the complete ranking  $\boldsymbol{\pi}_1 = (2, 3, 1, 4)'$ , top 2 partial ranking  $\boldsymbol{\pi}_2 = (2, 3.5, 1, 3.5)'$ , and the subset ranking  $\boldsymbol{\pi}_3 = (2, \_, 1, \_)'$  refer to the events with their respective matrices  $C$  being

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad (1 \ 0 \ -1 \ 0).$$

Notationally, let

$$\mathbf{A}_{d \times t} = [\mathbf{a}_1 \cdots \mathbf{a}_t],$$

$\boldsymbol{\Psi}_{t \times t}$  be the diagonal matrix with  $diag(\boldsymbol{\Psi}) = (\sigma_1^2, \dots, \sigma_t^2)$ , and all other entries equal to zero, and

$$\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{b}, \boldsymbol{\Psi}\}$$

the set of parameters of interest. We shall discuss the maximum likelihood estimation of  $\boldsymbol{\theta}$  based on various types of ranking data via the Monte Carlo Expectation-Maximization (MCEM) algorithm in the next section.

### 9.2.2 Monte Carlo Expectation-Maximization Algorithm

In order to deal with missing data, the EM algorithm is a broadly applicable approach for the computation of maximum likelihood estimates having the advantages of simplicity and stability. It requires one to compute the conditional expectation of the complete-data log-likelihood function given the observed data (E-step) and then to maximize the likelihood function with respect to the parameters of interest (M-step).

Let  $\mathbf{Y}_{n \times t}$ ,  $\mathbf{Z}_{n \times d}$  be the matrices of the unobservable response utilities and latent common factors, respectively, with their  $j$ th rows corresponding to individual  $j$ . Denote by  $\mathbf{\Pi}_{n \times t} = [\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n]'$  the matrix of the observed ranked data. Under an EM setting, we denote by  $\{Y, Z\}$  the missing data and by  $\mathbf{\Pi}$  the observed data.

#### 9.2.2.1 Implementing the E-step via the Gibbs Sampler

Since the complete-data log-likelihood function, apart from a constant, is given by

$$\ell(\boldsymbol{\theta} | Y, Z) = -\frac{n}{2} \sum_{i=1}^t \log \sigma_i^2 - \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^n \frac{(y_{ij} - z'_j \mathbf{a}_i - b_i)^2}{\sigma_i^2}, \tag{9.19}$$

the E-step here only involves computation of the conditional expectations of the complete-data sufficient statistics  $\{Y'Y, Z'Z, Z'Y, Y'1, Z'1\}$  given  $\mathbf{\Pi}$  and  $\boldsymbol{\theta}$ . This can be done by using the Gibbs sampling algorithm which consists of drawing samples consecutively from the full conditional posterior distributions, as shown below:

1. Draw  $\mathbf{z}_j$  from  $f(\mathbf{z}_j | y_j, \boldsymbol{\pi}_j, \boldsymbol{\theta})$ .
2. Draw  $y_j$  from  $f(y_j | \mathbf{z}_j, \boldsymbol{\pi}_j, \boldsymbol{\theta})$  for  $j = 1, \dots, n$ .

For step 1, making draws from  $f(\mathbf{z}_j | y_j, \boldsymbol{\pi}_j, \boldsymbol{\theta})$  is simple because

$$f(\mathbf{z}_j | y_j, \boldsymbol{\pi}_j, \boldsymbol{\theta}) = f(\mathbf{z}_j | y_j, \boldsymbol{\theta}),$$

which is independent of  $\boldsymbol{\pi}_j$ . Draws of  $\mathbf{Z}$  can be made from the conditional distribution

$$\mathbf{z}_j | y_j, \boldsymbol{\theta} \sim N_d(\mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}(y_j - \mathbf{b}), \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}\mathbf{A}'). \tag{9.20}$$

For step 2,  $y_j$  requires to have consistent orderings with the observed ranking  $\boldsymbol{\pi}_j$ . Suppose that  $\langle [1]_j, \dots, [t]_j \rangle$  represents the ordering of the  $t$  objects with respect to the complete ranking  $\boldsymbol{\pi}_j$  such that  $[1]_j$  is the most preferred object,  $[2]_j$  is the second most preferred object, and so on. Define  $y_{[0]_j} = +\infty$  and  $y_{[t+1]_j} = -\infty$ .

### Complete Rankings

For the cases with complete rankings, we can draw  $y_{ij}$  sequentially for  $i = 1, \dots, t$  from

$$y_{ij} | y_{1j}, \dots, y_{i-1,j}, y_{i+1,j}, \dots, y_{tj}, \boldsymbol{\pi}_j, \mathbf{z}_j, \boldsymbol{\theta} \sim N(\mathbf{z}'_j \mathbf{a}_i, \sigma_i^2) \quad (9.21)$$

with the constraint  $y_{[r-1]j} > y_{ij} > y_{[r+1]j}$  for  $\pi_{ij} = r$  (or  $[r]_j = i$ ).

### Top $q$ Partial Rankings

For *top  $q$  partial* rankings, we draw the top  $q$  objects (i.e.,  $\{x_{[1]j}, \dots, x_{[q]j}\}$ ) as in the complete case and simulate the other objects by

$$y_{ij} \sim N(\mathbf{z}'_j \mathbf{a}_i, \sigma_i^2) \quad (9.22)$$

with the constraint  $-\infty < y_{ij} < y_{[q]j}$  for  $\pi_{ij} = r$  (or  $[r]_j = i$ ).

### Subset Rankings

For *subset* rankings, individuals are asked to rank a subset of the  $t$  objects only. Rankings of the set of remaining objects,  $\{y_{i'j}\}$ , are unknown and we can simulate  $\{y_{i'j}\}$  from

$$\{y_{i'j} | i' \notin \{\text{ranked objects}\}\} \sim N(\mathbf{z}'_j \mathbf{a}_i, \sigma_i^2). \quad (9.23)$$

The conditional expectation of  $\mathbf{Y}'\mathbf{1}$  and  $\mathbf{Y}'\mathbf{Y}$  can be approximated by taking the average of the random draws of  $\sum_j \mathbf{y}_j$  and the average of their product sum  $\sum_j \mathbf{y}_j \mathbf{y}'_j$ , respectively. Finally, conditional expectations of  $\mathbf{Z}'\mathbf{1}$ ,  $\mathbf{Z}'\mathbf{Z}$ , and  $\mathbf{Z}'\mathbf{Y}$  can be obtained by

$$\begin{aligned} E[\mathbf{Z}'\mathbf{1} | \boldsymbol{\Pi}, \boldsymbol{\theta}] &= \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}(E[\mathbf{Y}'\mathbf{1} | \boldsymbol{\Pi}, \boldsymbol{\theta}] - n\mathbf{b}), \\ E[\mathbf{Z}'\mathbf{Z} | \boldsymbol{\Pi}, \boldsymbol{\theta}] &= n[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}\mathbf{A}'] \\ &\quad + \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}E[(\mathbf{Y} - \mathbf{1b}')'(\mathbf{Y} - \mathbf{1b}') | \boldsymbol{\Pi}, \boldsymbol{\theta}](\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}\mathbf{A}' \\ E[\mathbf{Z}'\mathbf{Y} | \boldsymbol{\Pi}, \boldsymbol{\theta}] &= \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}E[(\mathbf{Y}' - \mathbf{b1}')\mathbf{Y} | \boldsymbol{\Pi}, \boldsymbol{\theta}]. \end{aligned}$$

#### 9.2.2.2 M-Step

By replacing the complete-data sufficient statistics  $\{\mathbf{Y}'\mathbf{Y}, \mathbf{Z}'\mathbf{Z}, \mathbf{Z}'\mathbf{Y}, \mathbf{Y}'\mathbf{1}, \mathbf{Z}'\mathbf{1}\}$  with their corresponding conditional expectations obtained in E-step, we can compute the maximum likelihood estimate of  $\boldsymbol{\theta}$  by



$$\begin{pmatrix} \hat{A} \\ \hat{b}' \end{pmatrix} = [(Z' \mathbf{1})(Z' \mathbf{1})]^{-1} (Z' \mathbf{1})' Y = \begin{bmatrix} Z'Z & Z' \mathbf{1} \\ \mathbf{1}'Z & \mathbf{1}' \mathbf{1} \end{bmatrix}^{-1} \begin{pmatrix} Z'Y \\ \mathbf{1}'Y \end{pmatrix}$$

and

$$\begin{aligned} \hat{\Psi} &= \frac{1}{n} \text{diag} \left( (Y - Z \hat{A} - \mathbf{1} \hat{b}')'(Y - Z \hat{A} - \mathbf{1} \hat{b}') \right) \\ &= \frac{1}{n} \text{diag} \left( Y'Y - 2\hat{A}'Z'Y - 2\hat{b}'\mathbf{1}'Y + \hat{A}'Z'Z\hat{A} + 2\hat{b}'\mathbf{1}'Z\hat{A} + n\hat{b}\hat{b}' \right). \end{aligned}$$

The new set of  $\theta$  is then used for calculation of the conditional expectation of the sufficient statistics in the E-step and the algorithm is iterated until convergence is attained.

### 9.2.2.3 Determining Convergence of MCEM via Bridge Sampling

To determine convergence of the EM algorithm we propose to use the bridge sampling criterion discussed by Meng and Wong (1996). The bridge sampling estimate for the likelihood ratio associated with the individual  $j$  is given by

$$\frac{L(\theta^{(s+1)}|y_j, z_j)}{L(\theta^{(s)}|y_j, z_j)} = \frac{\sum_{m=1}^M \left[ \frac{L(\theta^{(s+1)}|y_j^{(s,m)}, z_j^{(s,m)})}{L(\theta^{(s)}|y_j^{(s,m)}, z_j^{(s,m)})} \right]^{1/2}}{\sum_{m=1}^M \left[ \frac{L(\theta^{(s)}|y_j^{(s+1,m)}, z_j^{(s+1,m)})}{L(\theta^{(s+1)}|y_j^{(s+1,m)}, z_j^{(s+1,m)})} \right]^{1/2}},$$

where  $\{y_j^{(s,m)}, z_j^{(s,m)}, m = 1, \dots, M\}$  denote the  $M$  Gibbs samples from  $f(y_j|z_j, \pi_j, \theta^{(s)})$  and  $f(z_j|y_j, \pi_j, \theta^{(s)})$  with  $\theta^{(s)}$  being the  $s$ th iterate of  $\theta$ . The estimate for the log-likelihood ratio of two consecutive iterates is then given by

$$\hat{h}(\theta^{(s+1)}, \theta^{(s)}) = \sum_{j=1}^n \log \frac{L(\theta^{(s+1)}|y_j, z_j)}{L(\theta^{(s)}|y_j, z_j)}.$$

We plot  $\hat{h}(\theta^{(s+1)}, \theta^{(s)})$  against  $s$  to determine the convergence of the MCEM algorithm. A curve converging to zero indicates a convergence because the EM algorithm should increase the likelihood at each step.

### 9.2.3 Simulation

We adopt the parameter values listed in Table 9.4 used by Brady (1989) to study the MCEM algorithm for complete and incomplete rankings. Using the factor model and these parameter values, thirty sets of data with  $n = 1,000$  and utility vectors

**Table 9.4** The Parameter values of a 2-factor model for seven objects

Object	$a_1$	$a_2$	$b$	$\sigma_j$
1	6.0	7.0	30.0	8.49
2	-3.0	5.0	32.0	6.00
3	9.0	-8.0	34.0	4.24
4	-7.2	-4.0	36.0	4.24
5	12.0	10.0	38.0	8.49
6	-2.0	-9.0	40.0	5.00
7	-8.0	6.0	42.0	8.00

**Table 9.5** Incomplete block design for subset rankings

1	2	4
2	3	5
3	4	6
4	5	7
1	5	6
2	6	7
1	3	7

of  $t = 7$  objects were simulated. Three types of ranked data were *observed* from each data set. The first type corresponds to the complete rankings for seven objects by ranking the utilities of the 7 objects. The second type corresponds to the top 3 partial rankings constructed from the rankings of the three largest utilities while the third type corresponds to the subset rankings of 3 out of the 7 objects chosen according to the incomplete block design as shown in Table 9.5.

In our simulation studies, the Gibbs sampler and the MCEM algorithm both converge fairly fast. Computation time required for each MC E-step in the case of subset rankings is shorter than that in complete rankings because the number of truncated normal variates to be drawn is smaller. For each E step, we discarded the first 100 burn-in cycles and selected one  $x_i$  systematically from every fifth cycle afterward until a total of 40 draws was reached. The MCEM algorithm converged within 10 iterations for all simulation data sets. The means of the 30 sets of estimates for the complete rankings, *top 3* partial rankings and *3 out of 7* subset rankings, together with their biases and standard errors are shown in Table 9.6. Small values of biases and standard errors show that the estimation method for incomplete rankings is extremely efficient and reliable with high accuracy.

Intuitively, since more information is provided when complete rankings are observed, the estimation of the factor model should perform better than with partial or subset rankings. This is indeed the case as can be seen from Table 9.6. Larger biases and standard errors are obtained for the case of 3 out of 7 subset rankings.

**Table 9.6** Simulation results of MCEM algorithm

Object	$a_1$			$a_2$			$b$			$\sigma_j$		
	Mean	Bias	SE*	Mean	Bias	SE*	Mean	Bias	SE*	Mean	Bias	SE*
<i>Complete rankings</i>												
1	6.03	0.03	0.08	6.81	-0.19	0.06	30.09	0.09	0.07	8.38	-0.11	0.07
2	-3.12	-0.12	0.06	4.99	-0.01	0.06	32.16	0.16	0.05	6.00	0.00	0.06
3	9.14	0.14	0.04	-7.95	0.05	0.05	33.87	-0.13	0.06	4.18	-0.06	0.06
4	-7.21	-0.01	0.05	-3.89	0.11	0.04	35.88	-0.12	0.06	4.17	-0.07	0.04
5	12.02	0.02	0.10	9.90	-0.10	0.08	38.08	0.08	0.14	8.53	0.04	0.09
6	-2.01	-0.01	0.05	-8.96	0.04	0.05	39.76	-0.24	0.05	4.94	-0.06	0.05
7	-8.07	-0.07	0.09	6.11	0.11	0.08	42.16	0.16	0.10	8.02	0.02	0.07
<i>Top 3 partial rankings</i>												
1	5.96	-0.04	0.12	6.87	-0.13	0.12	29.98	-0.02	0.13	8.62	0.13	0.12
2	-2.92	0.08	0.10	4.96	-0.04	0.09	32.06	0.06	0.08	5.97	-0.03	0.11
3	8.97	-0.03	0.07	-8.01	-0.01	0.08	33.94	-0.06	0.09	4.29	0.05	0.04
4	-7.19	0.01	0.06	-4.03	-0.03	0.05	36.08	0.08	0.06	4.20	-0.04	0.07
5	11.81	-0.19	0.16	10.05	0.05	0.15	37.88	-0.12	0.11	8.53	0.04	0.15
6	-1.97	0.03	0.07	-8.89	0.11	0.08	40.03	0.03	0.07	5.07	0.07	0.09
7	-7.88	0.12	0.10	6.07	0.07	0.10	42.05	0.05	0.10	7.84	-0.16	0.09
<i>3 out of 7 subset rankings</i>												
1	6.16	0.16	0.17	7.09	0.09	0.20	30.06	0.06	0.16	8.52	0.03	0.16
2	-3.21	-0.21	0.11	4.93	-0.07	0.14	32.03	0.03	0.10	6.09	0.09	0.10
3	8.98	-0.02	0.08	-7.86	0.14	0.06	34.03	0.03	0.09	4.23	-0.01	0.01
4	-7.09	0.11	0.07	-3.98	0.02	0.06	35.96	-0.04	0.09	4.21	-0.03	0.02
5	12.26	0.26	0.18	9.81	-0.19	0.19	37.84	-0.16	0.15	8.46	-0.03	0.08
6	-2.18	-0.18	0.08	-8.84	0.16	0.11	40.12	0.12	0.11	4.94	-0.06	0.05
7	-8.15	-0.15	0.14	5.83	-0.17	0.12	41.94	0.06	0.13	7.98	-0.02	0.08

\* The standard errors are obtained empirically based on the 30 estimates

### 9.2.4 Factor Score Estimation

So far we have been interested mainly in problems concerning the parameters in factor models and their estimation. Indeed, this frequently represents the main objective of factor analysis since the loading coefficients, to a large extent, determine the reduction of observed variables into a small number of common factors in terms of meaningful phenomena. While these problems constitute the primary interest of factor analysis, it is sometimes desirable to go one step further and to estimate the scores of an individual on the common factors in terms of the realizations of the variates for that individual. Factor scores provide information concerning the relative position of each individual corresponding to each factor whereas the loadings generally remain constant for all individuals. We therefore turn our attention to the problem of factor score estimation.

With the normality assumption, estimates of the factor score can be obtained via the regression approach and the generalized least squares approach that, respectively, minimize the variation of the estimator and the sum of squared standardized residuals (see Lawley and Maxwell 1971). However, these two approaches can only be used when the utility  $Y$  can be observed. Recently, Shi and Lee (1997a) developed a Bayesian approach for estimating the factor scores in factor models with polytomous data. By constructing appropriate posterior distribution, they proposed using the posterior mean as a factor score estimate. Their method involves computation of some multiple integrals which is handled by some Monte Carlo methods. To avoid tedious computation, Shi and Lee (1997b) applied the EM algorithm to obtain a Bayesian estimate of the factor score with polytomous variables. In this section, we will estimate the factor scores with ranked data via the MCEM algorithm discussed in Sect. 9.2.2.

#### 9.2.4.1 Factor Score Estimation Using the MCEM Algorithm

The factor score  $z_j$  can be estimated by the posterior mode of the posterior distribution  $z_j | \boldsymbol{\pi}_j, \boldsymbol{\theta}$ . Hence, the MCEM algorithm can be used to find the estimate by viewing the  $z_j$ 's as parameters in the complete-data log-likelihood function  $\ell$  in (9.19) and the resulting maximum likelihood estimate of  $z_j$  will then be the posterior mode estimate. The MCEM iteration can be simplified as follows: given an initial value  $z_j^{(0)}$  and the estimate  $\boldsymbol{\theta}$ , at the  $(s + 1)^{th}$  MCEM iteration,

**E-step:** Find  $E(y_j | \boldsymbol{\pi}_j, z_j^{(s)}, \boldsymbol{\theta})$  via Gibbs sampler.

**M-step:** Update  $z_j^{(s)}$  to  $z_j^{(s+1)}$  by

$$z_j^{(s+1)} = \mathbf{A}(\mathbf{A}'\mathbf{A} + \boldsymbol{\Psi})^{-1}[E(y_j | \boldsymbol{\pi}_j, z_j^{(s)}, \boldsymbol{\theta}) - \mathbf{b}]. \quad (9.24)$$

The Monte Carlo E-step is exactly the same as finding the conditional expectation of  $y_j$  while the M-step improves the estimate of  $z_j$  in a single step only. This iterative procedure will converge to the appropriate posterior mode which will be taken as an estimate of  $z_j$ . We propose to stop the MCEM iteration when the likelihood function of  $z_j^{(s)}$  and  $z_j^{(s+1)}$  is very close to each other. A simple stopping criterion is to consider the following expression:

$$l(\mathbf{z}^{(s)}, \mathbf{z}^{(s+1)}) = \log \frac{\exp \frac{\sum_i z_j^{(s)'} z_j^{(s)}}{2}}{\exp \frac{\sum_i z_j^{(s+1)'} z_j^{(s+1)}}{2}} = \frac{1}{2} \sum_i \left( z_j^{(s)'} z_j^{(s)} - z_j^{(s+1)'} z_j^{(s+1)} \right). \quad (9.25)$$

Convergence of the MCEM iteration is attained when  $l(\mathbf{z}^{(s)}, \mathbf{z}^{(s+1)})$  becomes stationary at zero level.

Note that it is possible to estimate the factor scores using the posterior mean based on the samples generated from the Gibbs sampler. We note that the posterior mode and the posterior mean are usually very close and, moreover, the covariance matrix of the posterior mode can be obtained as a by-product of the MCEM factor score estimation.

**9.2.4.2 The Covariance Matrix of the Factor Score Estimates**

To provide more insight about the estimates and the impact of lost information from continuous to ranking measurements, it is desirable to derive the covariance matrix of the posterior distribution  $f(z_j|\pi_j, \theta)$ , which is given by the negative inverse of the Hessian matrix of  $\log[f(z_j|\pi_j, \theta)]$ . A convenient way to evaluate the Hessian matrix is via the following expression:

$$-\frac{\partial^2 \log[f(z_j|\pi_j, \theta)]}{\partial z_j \partial z_j'} = -\int \frac{\partial^2 \log[f(z_j|y_j, \theta)]}{\partial z_j \partial z_j'} f(y_j|z_j, \pi_j, \theta) dy_j - \text{Var} \left\{ -\frac{\partial \log[f(z_j|y_j, \theta)]}{\partial z_j} \right\} \tag{9.26}$$

where the variance is with respect to  $f(y_j|z_j, \pi_j, \theta)$  (Tanner (1997)).

It can be shown that the covariance matrix of the factor score estimate  $\hat{z}_j$  is equal to

$$[(I - A(A'A + \Psi)^{-1}A')^{-1} - W \text{Var}(y_j|z_j, \pi_j, \theta)W']^{-1} |_{z_j=\hat{z}_j}, \tag{9.27}$$

where  $W = (I - A(A'A + \Psi)^{-1}A')^{-1}A(A'A + \Psi)^{-1}$  and  $\text{Var}(y_j|z_j, \pi_j, \theta)$  can be approximated by the Gibbs sample variance, a by-product of the MCEM factor score estimation.

**9.2.5 Application to the Job Selection Ranking Data**

We now consider the marketing survey on people’s attitude toward career and living style in three main cities in Mainland China – Beijing, Shanghai, and Guangzhou. Five hundred responses from each city were obtained. A question regarding the behavior, conditions, and criteria for job selection of the 500 respondents in Guangzhou will be discussed here. Respondents were asked to rank the three most important criteria on choosing a job among the following 13 criteria: 1. favorable company reputation; 2. large company scale; 3. more promotion opportunities; 4. more training opportunities; 5. comfortable working environment; 6. high income; 7. stable working hours; 8. fringe benefits; 9. well matched with employees’

**Table 9.7** Summary statistics of job selection ranking data

Criteria	Sample mean	Sample variance
1. Favorable company reputation	8.16	2.18
2. Large company scale	7.01	6.27
3. More promotion opportunities	7.43	5.69
4. More training opportunities	8.12	2.43
5. Comfortable working environment	6.89	7.68
6. High income	6.03	11.14
7. Stable working hours	7.64	4.72
8. Fringe benefits	6.68	8.59
9. Well matched with employees' profession or talent	8.14	2.18
10. Short distance between working place and home	8.30	1.13
11. Challenging	8.19	1.92
12. Corporate structure of the company	8.40	0.63
13. Low working pressure	8.22	1.43

profession or talent; 10. short distance between working place and home; 11. challenging; 12. corporate structure of the company; and 13. low working pressure.

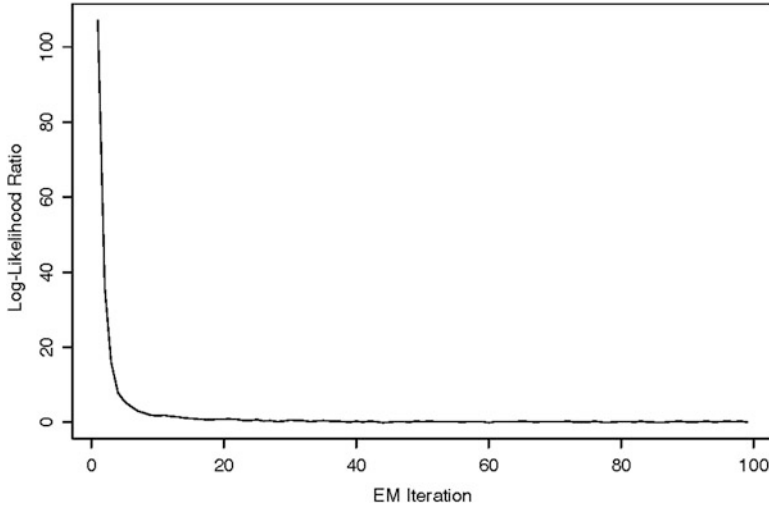
This is a typical *top 3 out of 13 objects* partial ranking problem. The values “1”, “2,” and “3” were assigned to the most, second, and third important criteria for job selection, respectively. Regarding the other less important items, it is common to define the midrank, i.e.,  $\frac{1}{t-q}[(q+1) + \dots + t]$ , as their rank. In this case the midrank is  $\frac{1}{10}[4 + \dots + 13] = 8.5$ . Table 9.7 provides some preliminary statistics, including sample mean and sample variance for each of the 13 criteria based on these 500 incomplete rankings with the midrank imputations.

The factor model is assumed and the analysis is made possible by the MCEM algorithm. Initial values of  $\theta$  were obtained by principal factor analysis and a standardized rank score,  $\frac{\pi_{ij}}{\sqrt{(t^2-1)/12}} \times \frac{1}{\sqrt{\sum_i 1/\sigma_i^2}}$ , was used as starting value of  $y_{ij}$  in the Gibbs sampler. The choice of standardized rank score was motivated by the fact that the rankings of  $y_{ij}$ 's must be consistent with the observed ranking  $\pi_j$ .

### 9.2.5.1 Model Estimation

Factor models with the number of factors ranging from zero to five were estimated. The Gibbs sampler (in the MC E-step) converged quite rapidly. We discarded the first 100 burn-in cycles and selected one  $y_j$  systematically from every fifth cycle afterward until a total of 40 draws was reached.

We used the bridge sampling criterion discussed in Sect. 9.2.2.3 to detect the convergence of the MCEM algorithm. Figure 9.2 shows the plot of the log-likelihood ratio against the number of iterations of the 3-factor model. The MCEM algorithm converged after 20 iterations.



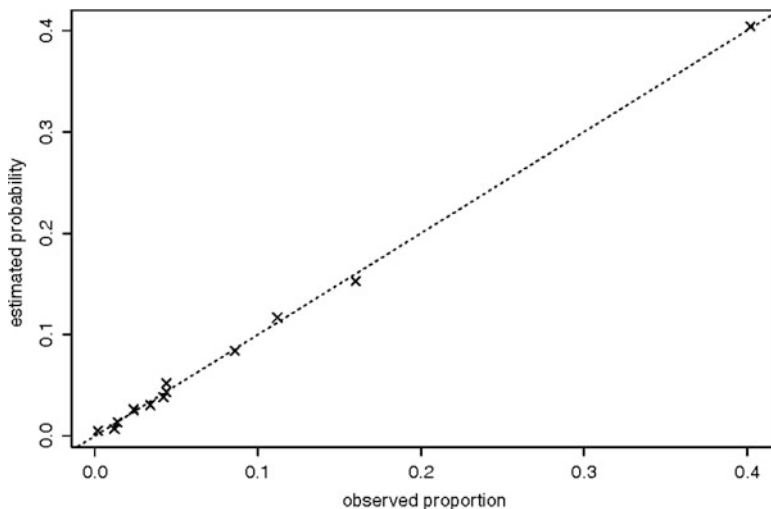
**Fig. 9.2** Bridge sampling criteria

**Table 9.8** AIC values and proportions of variance explained by the  $d$ -factor model with  $d = 0, 1, 2, 3, 4, 5$

Number of factors ( $d$ )	Number of free parameters	AIC value	Proportion of variance explained
0	24	6086.2	–
1	36	5909.0	0.1933
2	47	5894.6	0.2745
3	57	5892.8	0.4125
4	66	5896.7	0.4861
5	74	5903.8	0.5385

The Akaike information criterion (AIC) was used to determine the appropriate number of factors. The observed likelihood function which can be written as a product of multivariate normal probabilities over the rectangular region was approximated by the Geweke-Hajivassiliou-Keane (GHK) method shown to be unbiased and most reliable. Table 9.8 exhibits the values of AIC approximated by GHK methods and the proportions of variation explained by the  $d$ -factor models with  $d = 0, 1, 2, 3, 4, 5$ . It can be seen that the “best” model according to AIC is the 3-factor model and the proportion of variation explained by the 3-factor model is 41 %.

To examine the goodness of fit of the 3-factor model, we compare the top-choice probability for each of the 13 objects based on the fitted model with its corresponding observed proportions. Here, the top-choice probabilities is estimated using the GHK method. Figure 9.3 provides a plot of the estimated top-choice probabilities vs the respective observed proportions. The points appear to lie on the straight line, indicating the 3-factor model fits the data reasonably well.



**Fig. 9.3** Estimated top-choice probabilities vs observed proportions

**Table 9.9** Parameter estimates of the 3-factor model

Criteria	Factor 1	Factor 2	Factor 3	$\mathbf{b}$	$\sigma^2$
1	-0.58	-0.41	0.25	-0.79	0.73
2	-0.02	-0.21	0.79	-0.46	0.29
3	0.29	-0.65	-0.04	0.30	0.40
4	0.44	-0.38	0.24	-0.74	1.10
5	-0.10	0.09	-0.14	0.75	0.47
6	-0.05	0.07	-0.73	1.31	0.35
7	-0.09	0.69	-0.17	0.42	0.27
8	-0.22	0.36	-0.34	0.94	0.31
9	0.58	0.03	0.21	-0.26	0.53
10	0.14	0.70	-0.15	-0.05	0.19
11	0.39	-0.07	0.09	-0.46	0.88
12	-0.50	0.21	0.08	-1.10	0.82
13	-0.07	0.52	-0.23	-0.17	0.47
Cumul. prop. exp.	0.15	0.30	0.41		

Estimated values of the factor loadings were obtained by varimax rotation. The values of factor loadings expressed as the correlation between factors and utilities together with the estimated values of  $\mathbf{b}$  and  $\sigma^2$  are summarized in Table 9.9. The first factor can be regarded as a measure of career prospect. Utilization of one’s talent and job aspiration are major concerns in this factor. The second dimension represents the undemanding job nature. Short distance between working place and home, stable working hours, and low working pressure all score high loadings in



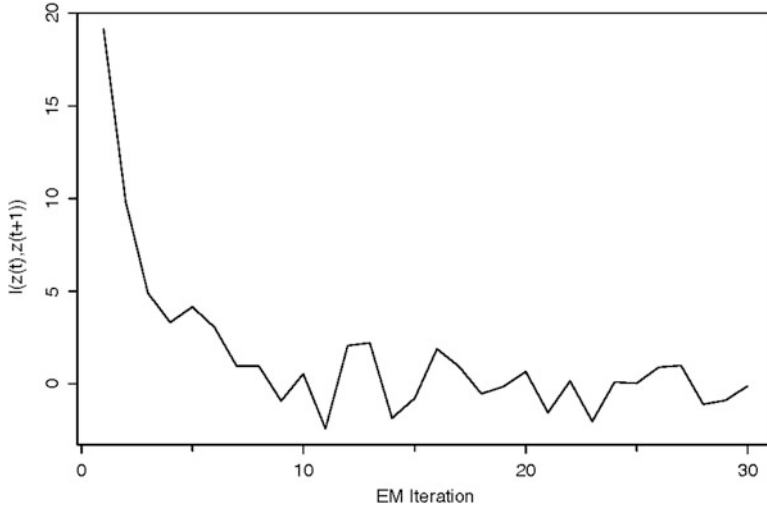


Fig. 9.4 Stopping criterion

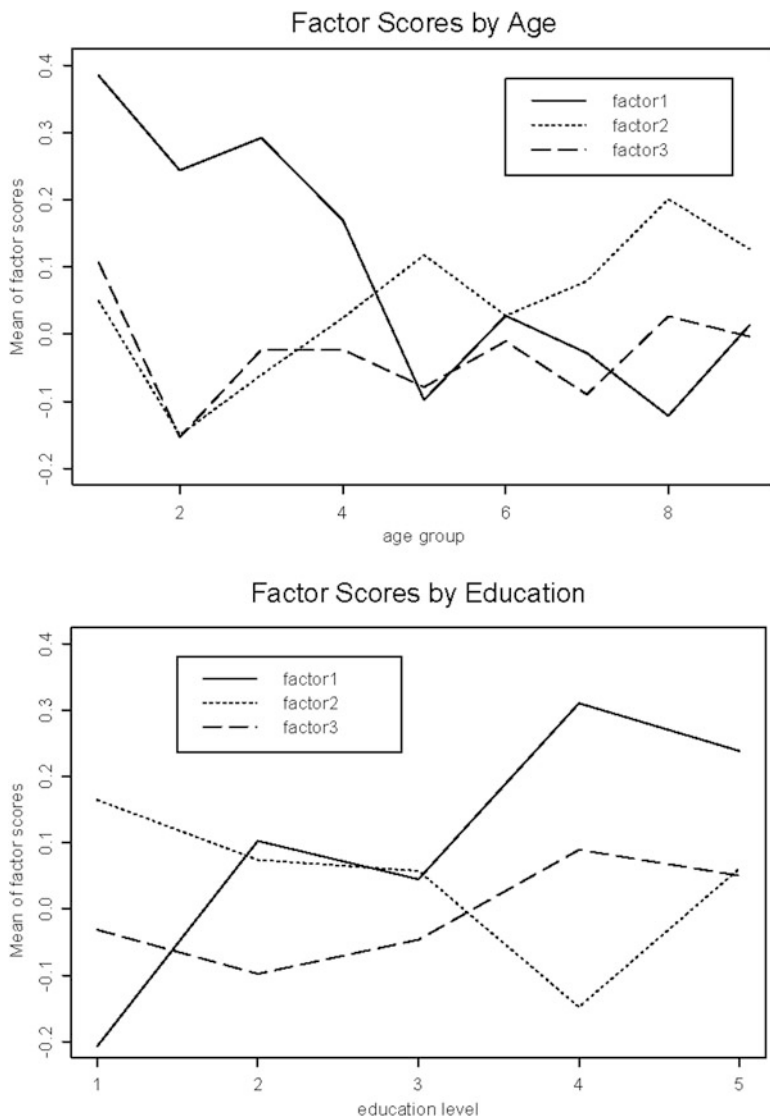
this factor. The third factor represents a contrast between the scale of the company and the salary. A large company offering lower income can be more attractive than a small company offering higher income.

Also, the mean vector  $\mathbf{b}$  reflects the overall importance of the 13 criteria. Note that the ordering of  $\hat{b}_1, \dots, \hat{b}_l$  is consistent with the average of the 500 rankings. Criterion 6 has the largest mean value which implies salary is their major concern on choosing a job while factors regarding the company itself are least important because  $\hat{b}_1, \hat{b}_4$ , and  $\hat{b}_{12}$  get large negative values.

### 9.2.5.2 Factor Score Estimation

To estimate the factor scores of the fitted 3-factor model, we applied the MCEM method. It is found that the Gibbs sampler in the E-step converged quite rapidly. We discarded the first 100 burn-in cycles and selected one  $y_j$  systematically from every fifth cycle afterward until 40 draws were reached. We simulated a total of 300 cycles for each E-step. Also, we applied the stopping criterion to detect the convergence of the MCEM algorithm. Figure 9.4 gives the plot of  $l(\mathbf{z}^{(s)}, \mathbf{z}^{(s+1)})$  against the number of iterations. According to the plot, the MCEM algorithm converged after 20 iterations.

It is often of interest to study the relationship between the factor scores and the covariates of each individual. In this survey, age group was collected in nine 5-year bands covering the ages from 15 to 59 ((1) 15–19, (2) 20–24, ..., (9) 55–59), while education level was recorded in five categories: primary (1), junior secondary (2), senior secondary (3), postsecondary (4), and university degree or above (5).



**Fig. 9.5** Factor scores vs age and education level

Figure 9.5 provides plots of the means of the factor score estimates of individuals of different age groups and education levels. From the plot of factor scores by age, a decreasing trend for factor 1 scores and an increasing trend for factor 2 scores are observed whereas from the plot of factor scores by education, an increasing trend for factor 1 scores and a decreasing trend for factor 2 scores are observed. For factor 3 scores, only a slightly increasing trend in education is observed. These observations

**Table 9.10** Standard error of factor scores

	$S_{11}$	$S_{22}$	$S_{33}$	$S_{12}$	$S_{13}$	$S_{23}$
Mean	0.1727	0.3279	0.3275	0.0119	0.0483	0.0174
SD	0.0052	0.0064	0.0140	0.0040	0.0064	0.0077
Min.	0.1606	0.3135	0.2993	0.0014	0.0354	0.0000
Max.	0.1903	0.3509	0.3825	0.0243	0.0699	0.0428

imply a young, well-educated person acquires more on career prospect while an old, less educated person may seek for a job with undemanding job nature. Finally, a better educated person is more willing to work in a large company offering lower salary.

To demonstrate the performance of our estimation on factor scores, Table 9.10 provides descriptive statistics on the covariance matrix  $S$  of  $\hat{z}_j, i = 1, \dots, 500$ . Small values of the standard error show that the estimation method is good and reliable. Also, it seems that the impact of unobservable information for this case is not serious.

## Chapter Notes

To address the robustness of the MVNOS model, Yu (2000) considered two approaches. The first one is to study the sensitivity of the parameter estimates if an outlying ranking is added to the data while the second one is to consider a more general distribution and look at the differences between the 2 sets of estimates.

In Sect. 9.1, we discussed that the parameters of a MVNOS model cannot be fully identified unless the variance-covariance matrix  $V$  is structured. Of course, factor analysis mentioned in Sect. 9.2 provides a solution for the simplified but yet flexible dependency structure for  $V$ . Other choices of dependency structure include wandering vector models (Yu and Chan 2001) and wandering ideal point models (Leung 2003).

In the factor analysis for ranking data, Yu et al. (2005) commented that apart from studying the relationship between the factors and individual's covariate via factor score estimation, we can incorporate the effect of covariates directly into the factor model:  $y_{ij} = z_j' \mathbf{a}_i + b_i + \mathbf{w}_j' \mathbf{c}_i + \varepsilon_{ij}$ , where  $\mathbf{w}_j$  is a vector of covariates of individual  $j$  such as sex and age and  $\mathbf{c}_i$  is a vector of regression parameters. The procedures of the MCEM algorithm can be implemented easily for this model but the details are omitted here. However, the number of parameters to be estimated would increase accordingly. Recently, Yu et al. (2013) further extended factor analysis to a data set of paired rankings such as rankings given by couples and identified the common factors between the individuals in each pair.

So far, we treated the ranking with ties as if the ordering of tied objects is unknown. Poon and Xu (2009) extended the MVNOS model to allow for tied objects by assuming that any two objects  $a$  and  $b$  have different ranks if and only if their utilities differ by more than a small value, i.e.,  $|y_a - y_b| \geq \delta$ . For example, the ranking  $B \succ A = C$  has utilities satisfying  $y_B - y_A \geq \delta$ ,  $y_B - y_C \geq \delta$ , and  $|y_A - y_C| < \delta$ . However, the parameter  $\delta$  in Poon and Xu (2009) must be fixed at a prespecified value because of the parameter identifiability problem.