Journal of the Korean Statistical Society (2003), **32**: *4, pp 425–448*

BAYESIAN HIERARCHICAL MODEL WITH SKEWED ELLIPTICAL DISTRIBUTION†

YOUNSHIK CHUNG[1], DIPAK K. DEY[2], TAEYOUNG YANG[3] AND JUNGHOON JANG[1]

ABSTRACT

Meta-analysis refers to quantitative methods for combining results from independent studies in order to draw overall conclusions. We consider hierarchical models including selection models under a skewed heavy tailed error distribution proposed originally by Chen *et al.* (1999) and Branco and Dey (2001). These rich classes of models combine the information of independent studies, allowing investigation of variability both between and within studies, and incorporate weight function. Here, the testing for the skewness parameter is discussed. The score test statistic for such a test can be shown to be expressed as the posterior expectations. Also, we consider the detail computational scheme under skewed normal and skewed Student-t distribution using MCMC method. Finally, we introduce one example from Johnson (1993)'s real data and apply our proposed methodology. We investigate sensitivity of our results under different skewed errors and under different prior distributions.

AMS 2000 subject classifications. Primary 62F15; Secondary 62F03.
Keywords. Bayesian meta-analysis, density generator, elliptical distribution, Gibbs sampler, hierarchical selection model, Metropolis-Hastings algorithm, score test, skewed elliptical distribution, weight function.

Received September 2002; accepted August 2003.
†This work is supported by the grant No. 2000-042-D00010 from the Korea Research Foundation.

[1]Department of Statistics, Pusan National University, Pusan 609-735, Korea (e-mail : yschung@pusan.ac.kr)
[2]Department of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A. (e-mail : dey@stat.uconn.edu)
[3]Department of Mathematics, Myongji University, Yongin 120-728, Korea (e-mail : tyang@mju.ac.kr)

1. Introduction

Meta-analysis is a quantitative method for combining results from independent studies and combining information which may be used to evaluate cumulative effectiveness, plan new studies and so on, with wide application in the field of medicine. There are two main problems in meta-analysis. One is that the study effects are heterogeneous and usually account for the random effect or hierarchical models (Morris and Normand, 1992). The other is that meta-analysis may have the publication bias for example, only studies with significant results are observed. When there exists the publication bias, the weight function can be used to account for such bias (Larose and Dey, 1996). To solve such problem, Silliman (1997) introduced hierarchical selection models (HSM) which incorporate weight function into the general hierarchical model and Chung *et al.* (2002) presented semiparametric hierarchical selection model with Dirichlet process prior which is the extended version of Silliman's HSM.

In this paper, we consider hierarchical model including selection model with skewed elliptical error for Bayesian meta-analysis. Such non-normal disturbance in statistical model has been investigated by several authors for theoretical and applied interest. Especially as the pioneer of this area, Zellner (1976) considered a Bayes and classical analysis of linear multivariate Student-t regression models. Azzalini and Dalla-Valle (1996) present a general theory for the multivariate version of skew-normal distribution which extends the class of normal distributions by the addition of a shape parameter. Recently, Branco and Dey (2001) proposed a general class of multivariate skew-elliptical distributions which contain the multivariate normal, Student-t, exponential power and Pearson type II, but with an extra parameter to regulate skewness. Sahu *et al.* (2001) considered the regression problem under a skew-elliptical error distribution and developed a Bayesian methodology for the inference of regression parameters.

The rest of this article is organized as follows. Section 2 reviews the multivariate skew-elliptical distribution. The particular cases of normal and Student-t distributions are explained as examples. Also, we develop Bayesian hierarchical model with skew-elliptical errors. The posterior propriety is studied. In particular, Section 3 includes the Bayesian hierarchical selection models with skew-elliptical errors. Then the testing for the skewness parameter is discussed. The score test can be expressed as the posterior expectations. Also, we consider the detail computational scheme under skew-normal and Student-t distributions using MCMC method. In Section 4, we introduce one example from Johnson

(1993)'s real data and apply our proposed methodology and investigate sensitivity of our results under different skewed errors. Finally in Section 5, we discuss our results and propose directions for future works.

2. BAYESIAN HIERARCHICAL MODEL WITH SKEWED ELLIPTICAL ERROR

Morris and Normand (1992) consider a hierarchical model as follows: for $i = 1, \ldots, n$,

$$\begin{gathered} Y_i = \alpha_i + \epsilon_i, \\ \epsilon_i \sim N(0, \sigma_i^2), \quad \alpha_i | \mu, \sigma_\alpha^2 \sim N(\mu, \sigma_\alpha^2), \quad (\mu, \sigma_\alpha^2) \sim \pi(\mu, \sigma_\alpha^2). \end{gathered} \tag{2.1}$$

For meta-analysis, here we can interpret that Y_i is the observed study effect, α_i is the true study effect, σ_i^2 is the within-study variance, μ is the average study effect and σ_α^2 is between-study variance. The joint prior $\pi(\mu, \sigma_\alpha^2)$ can be specified both proper and improper. Assuming the improper prior distributions still guarantee the proper posterior distribution and which is proved later. Note that we may fix σ_i because it is usually the standard error of an estimate.

In the model (2.1), the error terms are assumed to be symmetric. As mentioned in Section 1, we will assign the skew-elliptical distribution to the error terms ϵ_i in (2.1) based on Chen *et al.* (1999), Sahu *et al.* (2001), and Chung and Jang (2003). Since the error terms are univariate, we deal with the univariate random variables.

2.1. Univariate skew-elliptical distribution

In this subsection, we present a class of skewed elliptical distribution using the approach given in Chen *et al.* (1999), where the skewed random variable evolves from a sum of a symmetric and positive random variables, and is given as

$$\epsilon = U + \delta Z. \tag{2.2}$$

The important point here is to have U having a symmetric, unimodal and Z having a positive skewed distribution. When $\delta = 0$, we get the original symmetric distribution. The parameter δ has an easy interpretation and is called the skewness parameter. It can be shown that when $\delta > 0$ ($\delta < 0$), then the distribution is right (left) skewed.

Following Branco and Dey (2001), we modify their results to the simple elliptical class, where $U \sim El(\mu, \sigma; g_2)$ and $Z \sim El(0, 1; g_1)$ with $El(\mu, \sigma; g)$ denotes the elliptical distribution with location parameter μ, dispersion parameter σ and density generator g. That is, if $Y \sim El(\mu, \sigma; g)$ then the density of Y, f_{el}, is given by

$$f_{el}(y|\mu, \sigma; g^{(1)}) = \sigma^{-1/2} g^{(1)} \left(\frac{(y-\mu)^2}{\sigma} \right), \tag{2.3}$$

for a non-increasing function $g^{(1)}$, $u \geq 0$, such that

$$g^{(1)}(u) = \frac{g(u)}{\int_0^\infty u^{-1/2} g(u) du}, \tag{2.4}$$

where $g(u)$ is a non-increasing function such that the integral $\int_0^\infty u^{-1/2} g(u) du$ exists. Such g is called the density generator function. From now onwards, we consider the following transformation (2.2). Following Azzalini and Dalla-Valle (1996), the skew-elliptical class is developed by considering the random variable

$$\epsilon| (Z > 0), \tag{2.5}$$

denoted by $SE(\mu, \sigma, \delta; g)$.

THEOREM 2.1 (Sahu *et al.*, 2001). *Under the above assumption, the probability density function of* $\epsilon| (Z > 0)$ *defined in (2.5) is given by*

$$\begin{aligned} f(\epsilon|\mu, \sigma, \delta; g^{(1)}) = \; & 2 f_{el}(\epsilon|\mu, \sigma + \delta^2; g^{(1)}) \\ & \times F_{el} \left(\frac{\delta y_*}{(\sigma + \delta^2)\sqrt{1 - \delta^2 (\sigma + \delta^2)^{-1}}} \middle| 0, 1; g^{(1)}_{q(y_*)} \right) \end{aligned} \tag{2.6}$$

where $f_{el}(\epsilon|\mu, \sigma; g)$ *is the same as defined in (2.3) and* F_{el} *denotes the elliptical distribution corresponding to the elliptical density* f_{el},

$$g_a^{(1)}(u) = \frac{\Gamma(1/2)}{\pi^{1/2}} \cdot \frac{g(a+u)}{\int_0^\infty r^{1/2-1} g(a+r) dr}, \quad a > 0, \tag{2.7}$$

and

$$q(y_*) = y_*^t (1 + \delta^2)^{-1} y_*, \; y_* = y - \mu. \tag{2.8}$$

EXAMPLE 2.1 (*Skew-normal distribution*). Let $g(u) = e^{-u/2}$. Then it is easy to show that $g^{(1)}(u) = e^{-u/2}/\sqrt{2\pi}$ and $g^{(1)}_{q(y_*)}$ is free of $q(y_*)$. Now, the probability density function of the skew-normal distribution is given by

$$f(\epsilon|\mu,\sigma,\delta) = 2\phi\left(\frac{\epsilon-\mu}{\sqrt{\sigma+\delta^2}}\right) \times \Phi\left(\frac{\delta(\epsilon-\mu)}{\sqrt{\sigma(\sigma+\delta^2)}}\right), \tag{2.9}$$

where ϕ and Φ denote the density and *cdf* of standard normal distribution, respectively. This density in (2.9) is exactly the same as the density of $\epsilon = U + \delta Z$ with $U \sim N(\mu,\sigma)$ and $Z \sim N^+(0,1)$ where $N^+(0,1)$ denotes the folded normal distribution to the right at zero. Therefore, we can express the skew-normal distribution of ϵ as $\epsilon = U + \delta Z$ with $U \sim N(\mu,\sigma)$ and $Z \sim N^+(0,1)$. This result corresponds to that of Azzalini and Dalla-Valle (1996).

EXAMPLE 2.2 (*Skew-t distribution*). Let $g(u;\nu) = (1+u/\nu)^{-(\nu+2)/2}$. Then

$$g_a^{(1)}(u;\nu) = \frac{\Gamma((\nu+2)/2)}{\Gamma((\nu+1)/2)}\{\pi(\nu+1)\}^{-1/2}\left(\frac{\nu+1}{\nu+a}\right)^{1/2}\left(1+\frac{u}{\nu+1}\frac{\nu+1}{\nu+a}\right)^{-(\nu+2)/2}. \tag{2.10}$$

Therefore, the density of the skew-t distribution is given by

$$f(\epsilon|\mu,\sigma,\nu) = 2t_\nu(y|\mu,1+\delta^2) \times T_{\nu+1}\left(\left\{\frac{\nu+q(y_*)}{\nu+p}\right\}^{-1/2} \cdot \frac{\delta(\epsilon-\mu)}{\sqrt{\sigma(\sigma+\delta^2)}}\right), \tag{2.11}$$

where t_ν and T_ν denote the density and its *cdf* of t-distribution with *df* ν , respectively. This density is the same as the density of $\epsilon = U + \delta Z$ with $U \sim t_\nu(\mu,\sigma)$ and $Z \sim t^+_{\nu_z}(0,1)$ where $t_\nu(\mu,\sigma)$ denotes the t-distribution with mean μ and variance σ and *df* ν and $t^+_{\nu_z}(0,1)$ denotes the truncated t-distribution to the right at zero with *df* ν_z.

2.2. Bayesian hierarchical models with skewed elliptical error

Now, our Bayesian hierarchical model with skew-elliptical distribution is proposed by replacing the normal error term in (2.1) with the skewed elliptical error term in (2.5) as follows: for $i = 1,\ldots,n$,

$$\begin{gathered} Y_i = \alpha_i + \epsilon_i, \\ \epsilon_i \sim SE(0,\sigma_i,\delta;g), \quad \alpha_i \sim N(\mu,\sigma^2_\alpha), \quad (\mu,\sigma^2_\alpha,\sigma_i,\delta) \sim \pi(\mu,\sigma^2_\alpha,\sigma_i,\delta), \end{gathered} \tag{2.12}$$

where SE denotes the skew-elliptical distribution in (2.6). If $\delta = 0$, the error term has the symmetric distribution.

To completely specify the Bayesian model, we need to specify prior distributions for all the parameters. Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^t$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_n^2)^t$. When the skew-$t$ models are considered, we need prior distribution for the degrees of freedom parameter ν. Now the posterior density is given by

$$\begin{aligned}&\pi(\mu, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2, \sigma_\alpha^2, \delta, \nu | \mathbf{y}) \\ &\propto \prod_{i=1}^{n} \left\{ SE(y_i | \alpha_i, \sigma_i^2, \delta; g^{(1)}) N(\alpha_i | \mu, \sigma_\alpha^2) \right\} \times \pi(\mu, \boldsymbol{\sigma}, \sigma_\alpha^2, \delta, \nu), \qquad (2.13)\end{aligned}$$

where $\mathbf{y} = (y_1, \dots, y_n)^t$ and $\pi(\mu, \boldsymbol{\sigma}, \sigma_\alpha^2, \delta, \nu)$ is the joint prior density of $\mu, \boldsymbol{\sigma}, \sigma_\alpha^2, \delta$ and ν. Note that for the skew-normal models, the distribution of ν is omitted.

In our example in Section 4, Johnson (1993) mentioned that σ_i^2 can be estimated by maximum likelihood estimate (MLE). Therefore the information for $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_n^2)^t$ is assumed to be available and we assign the prior distribution of $\boldsymbol{\sigma}^2$ to be informative. In practice, we may experiment with improper prior distributions for μ and σ_α^2. A natural question in such a case is whether the full posterior distribution is proper. The following theorem gives a sufficient condition for the propriety of posterior.

THEOREM 2.2. *Suppose that $\boldsymbol{\sigma}^2$, δ and ν are independent and their priors are proper. Then the posterior density in (2.12) is proper under the skew-normal or skew-t model if $n > 1$.*

The proof is given in Appendix.

In particular, from Examples 2.1 and 2.2, ϵ_i can be expressed as

$$\epsilon_i = u_i + \delta z_i, \qquad (2.14)$$

where z_i is the truncated normal (or truncated t) random variable, u_i has the normal distribution (or t-distribution) and $E(z_i)$ denotes the mean of z_i. In these cases, our model in (2.1) can be expressed as

$$Y_i = \alpha_i + \delta\{z_i - E(z_i)\} + u_i. \qquad (2.15)$$

The reason for $E(z_i)$ to be in (2.15) is that $E(Y_i)$ is equal to the predictor α_i as the structure of the linear model. Therefore, $E(z_i)$ is called the *correction factor* in this set-up.

3. HIERARCHICAL SELECTION MODELS WITH SKEWED ELLIPTICAL ERROR

In this section, we introduce Silliman (1997)'s hierarchical selection models and then extend it to the skew-elliptical model. Silliman (1997)'s hierarchical selection models (HSM) are as follows: for $i = 1, \ldots, n$,

$$Y_i|\alpha_i, \sigma_i \sim f^w(y_i|\alpha_i, \sigma_i), \quad \alpha_i|\mu, \sigma_\alpha^2 \sim N(\mu, \sigma_\alpha^2), \quad (\mu, \sigma_\alpha^2) \sim \pi(\mu, \sigma_\alpha^2), \tag{3.1}$$

where $f^w(y_i|\alpha_i, \sigma_i)$ denotes the weighted density defined by

$$f^w(y_i|\alpha_i, \sigma_i) = \frac{w(y_i)\ f(y_i|\alpha_i, \sigma_i)}{C(\alpha_i, \sigma_i)}, \tag{3.2}$$

with the normalizing constant $C(\alpha_i, \sigma_i) = \int w(x)\ f(x|\alpha_i, \sigma_i)dx$ where $f(x|\alpha_i, \sigma_i)$ is unweighted density of x given (α_i, σ_i). In particular, if $C(\alpha_i, \sigma_i) = 1$, then the model in (3.1) is exactly the same as the hierarchical model in (2.1). For meta-analysis, the interpretation of variables is the same as in (2.1). More detail results and explanations are given in Silliman (1997) and Chung and Jang (2001).

If we choose the normal distribution as $f(y_i|\alpha_i, \sigma_i)$ in (3.2), then the model in (3.1) is called the *hierarchical selection model with normal error* (HSMN). Similarly, the model in (3.1) is called the *hierarchical selection model with t error* (HSMT) if choosing the Student-t distribution as $f(y_i|\alpha_i, \sigma_i)$ in (3.2). From (2.12) and (2.15), our Bayesian hierarchical model with skew-elliptical distribution is considered as follows: for $i = 1, \ldots, n$,

$$\begin{gathered} Y_i = \alpha_i + \delta\{z_i - E(z_i)\} + u_i, \ \alpha_i \sim N(\mu, \sigma_\alpha^2), \\ (\mu, \sigma_\alpha^2, \boldsymbol{\sigma}, \delta) \sim \pi(\mu, \sigma_\alpha^2, \boldsymbol{\sigma}, \delta), \end{gathered} \tag{3.3}$$

where u_i has a symmetric, unimodal distribution and z_i has the positive skewed distribution. If $\delta = 0$, the error term has the symmetric distribution.

From now on, the weight function $w(y)$ is considered for the selection model as in (3.1). The model (3.3) is called the *hierarchical selection model with skew-normal* (HSMSN) if the normal and truncated normal distributions are used for the distribution of the error term u_i and z_i, respectively and also, the model (3.3) is said to be the *hierarchical selection model with skew-t* (HSMST) if the Student-t and truncated Student-t distributions are chosen as the distribution of the error term u_i and z_i, respectively.

3.1. Hierarchical selection model with skew-normal error (HSMSN)

Assume that z_i and u_i in (3.3) are distributed as the folded normal and the standardized normal distribution, respectively. That is, from (3.3), our Bayesian hierarchical model with skewed normal error (HSMSN) can be written as follows: for $i = 1, \dots, n$,

$$\begin{gathered} Y_i = \alpha_i + \delta\{z_i - E(z_i)\} + u_i,\ z_i \sim N^+(0,1),\ u_i \sim N(0,\sigma_i^2), \\ \alpha_i|\mu,\sigma_\alpha \sim N(\mu,\sigma_\alpha^2),\ (\mu,\sigma_\alpha,\boldsymbol{\sigma},\delta) \sim \pi(\mu,\sigma_\alpha,\boldsymbol{\sigma},\delta), \end{gathered} \tag{3.4}$$

where $E(z_i)$ is defined as (2.15) under skew-normal which is evaluated as $2/\pi$. Under the weight function $w(y)$, for $i = 1, \dots, n$,

$$y_i|\alpha_i,\sigma_i,\delta,z_i \sim \{C(\alpha_i,\sigma_i,\delta,z_i)\}^{-1} w(y_i)\ N\big(\alpha_i + \delta\{z_i - E(z_i)\},\sigma_i^2\big), \tag{3.5}$$

where $C(\alpha_i,\sigma_i,\delta,z_i)$ is the normalizing constant defined by (3.2). From now, brackets denote the densities for the notational convenience. For example, $[X,Y]$, $[X|Y]$ and $[X]$ mean the joint, conditional and marginal density, respectively. Therefore, the complete-likelihood function is expressed as follows:

$$\begin{aligned} &[y_i,\alpha_i,z_i|\mu,\sigma_\alpha,\sigma_i,\delta] \\ &= [y_i|\alpha_i,\sigma_i,\delta,z_i]\ [\alpha_i|\mu,\sigma_\alpha]\ [z_i] \\ &\propto \{C(\alpha_i,\sigma_i,\delta,z_i)\}^{-1} (\sigma_i^2)^{-1/2} \exp\left(-\frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{2\sigma_i^2}\right) \\ &\quad \times (\sigma_\alpha^2)^{-1/2} \exp\left\{-\frac{(\alpha_i - \mu)^2}{2\sigma_\alpha^2}\right\} \exp\left(-\frac{{z_i}^2}{2}\right) I(z_i > 0). \end{aligned} \tag{3.6}$$

In particular, if we take $\delta = 0$, the model in (3.4) is called hierarchical selection model with normal error (HSMN) as in Chung *et al.* (2002). Here, the following priors are assumed: for $i = 1, \dots, n$,

$$\mu \sim N(a,b), \quad \delta \sim N(m,\tau), \quad \sigma_\alpha^2 \sim IG(c_1,d_1), \quad \sigma_i^2 \sim IG(c_2,d_2), \tag{3.7}$$

where $a, b, m, \tau, c_1, c_2, d_1$ and d_2 are assumed to be known. Then the joint posterior density, which is proportional to the product of the complete likelihood function and the prior density functions, is given as follows:

$$\prod_{i=1}^{n} \left[\{C(\alpha_i,\sigma_i,\delta,z_i)\}^{-1} (\sigma_i^2)^{-1/2} \exp\left(-\frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{2\sigma_i^2}\right) \right]$$

$$\times \prod_{i=1}^{n} \left[(\sigma_\alpha^2)^{-1/2} \exp\left\{ -\frac{(\alpha_i - \mu)^2}{2\sigma_\alpha^2} \right\} \exp\left(-\frac{z_i^2}{2} \right) (\sigma_i^2)^{-(c_2+1)} \exp\left(-\frac{d_2}{\sigma_i^2} \right) \right]$$
$$\times \exp\left\{ -\frac{(\mu - a)^2}{2b} \right\} \exp\left\{ -\frac{(\delta - m)^2}{2\tau} \right\} (\sigma_\alpha^2)^{-(c_1+1)} \exp\left(-\frac{d_1}{\sigma_\alpha^2} \right). \tag{3.8}$$

The sampling scheme for the MCMC method can be applied, for which the following full conditional distributions based on the joint posterior in (3.8) are needed:

$$[\alpha_i | \alpha_j (j \neq i), \mathbf{y}, \mathbf{z}, \mu, \delta, \sigma_\alpha, \boldsymbol{\sigma}] \propto \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} N(\beta_1(\alpha_i), \gamma_1(\alpha_i)),$$
$$[\mu | \boldsymbol{\alpha}, \mathbf{y}, \mathbf{z}, \delta, \sigma_\alpha, \boldsymbol{\sigma}] = N\left(\frac{b \sum_{i=1}^{n} \alpha_i + \sigma_\alpha^2 a}{bn + \sigma_\alpha^2}, \frac{b\sigma_\alpha^2}{bn + \sigma_\alpha^2} \right),$$
$$[\delta | \boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \sigma_\alpha, \boldsymbol{\sigma}] \propto \prod_{i=1}^{n} \left[\{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} \right] N(\beta_1(\delta), \gamma_1(\delta)),$$
$$[z_i | \boldsymbol{\alpha}, \mathbf{y}, \mu, z_j (j \neq i), \delta, \sigma_\alpha, \boldsymbol{\sigma}] \propto \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} N^+(\beta_1(z_i), \gamma_1(z_i)),$$
$$[\sigma_\alpha^2 | \boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \boldsymbol{\sigma}] = IG\left(c_1 + \frac{n}{2}, \frac{1}{2} \sum_{i=1}^{n} (\alpha_i - \mu)^2 + d_1 \right),$$
$$[\sigma_i^2 | \boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \sigma_\alpha, \sigma_j (j \neq i)] \propto \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} IG(\xi_1(\sigma_i), \rho_1(\sigma_i)),$$

where

$$\beta_1(\alpha_i) = \frac{\sigma_\alpha^2 [y_i - \delta\{z_i - E(z_i)\}] + \sigma_i^2 \mu}{\sigma_\alpha^2 + \sigma_i^2}, \quad \gamma_1(\alpha_i) = \frac{\sigma_\alpha^2 \sigma_i^2}{\sigma_\alpha^2 + \sigma_i^2},$$
$$\beta_1(\delta) = \frac{\sum_{i=1}^{n} \{z_i - E(z_i)\} (y_i - \alpha_i) \sigma_i^{-2} + m\tau^{-1}}{\sum_{i=1}^{n} \{z_i - E(z_i)\}^2 \sigma_i^{-2} + \tau^{-1}},$$
$$\gamma_1(\delta) = \left[\sum_{i=1}^{n} \frac{\{z_i - E(z_i)\}^2}{\sigma_i^2} + \frac{1}{\tau} \right]^{-1}, \quad \beta_1(z_i) = \frac{\delta\{y_i - \alpha_i + \delta E(z_i)\}}{\sigma_i^2 + \delta^2},$$
$$\gamma_1(z_i) = \left(\frac{\sigma_i^2}{\sigma_i^2 + \delta^2} \right)^{-1}, \quad \xi_1(\sigma_i) = \frac{1}{2} + c_2 \quad \text{and}$$
$$\rho_1(\sigma_i) = \frac{1}{2} [y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2 + d_2.$$

Then, the Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970) is needed for sampling of $(\alpha_i, \delta, z_i, \sigma_i)$. Therefore, this computation of the posterior distribution is all that is required for making the desired inferences, such as the computation of quantities, means, standard deviations, credible sets and predictions.

3.2. Hierarchical selection model with skew-t error (HSMST)

In this subsection, we consider the model with the skew-t error. Assume z_i and u_i in (3.4) have standard folded t distribution and standard t-distribution, respectively. Then, from (3.3), our Bayesian hierarchical model with skewed t error (HSMST) is given by

$$\begin{gathered} Y_i = \alpha_i + \delta\{z_i - E(z_i)\} + u_i, \\ z_i \sim t^+_{\nu_2}(0,1), \quad u_i \sim t_{\nu_1}(0,\sigma_i^2), \quad \alpha_i \sim N(\mu,\sigma_\alpha^2), \\ (\mu,\sigma_\alpha,\boldsymbol{\sigma},\delta) \sim \pi(\mu,\sigma_\alpha,\boldsymbol{\sigma},\delta), \end{gathered} \tag{3.9}$$

where $E(z_i)$ is also expectation of latent variable z_i under skew-t distribution and its value is as follows:

$$E(z_i) = \frac{\Gamma((\nu_2+1)/2)}{\Gamma(\nu_2/2)} \cdot \frac{2}{\nu_2 - 1}\sqrt{\frac{\nu_2}{\pi}}.$$

Then, for $i = 1, \dots, n$,

$$y_i|\alpha_i,\sigma_i,\delta,z_i \sim \{C_t(\alpha_i,\sigma_i,z_i,\delta)\}^{-1} w(y_i)\ t_{\nu_1}(\alpha_i + \delta\{z_i - E(z_i)\}, \sigma_i^2) \tag{3.10}$$

where $C_t(\alpha_i,\sigma_i,z_i,\delta)$ is defined by

$$C_t(\alpha_i,\sigma_i,z_i,\delta) = \int w(x)\ t_{\nu_1}(x;\alpha_i+\delta\{z_i - E(z_i)\},\sigma_i^2)dx. \tag{3.11}$$

In particular, if $\delta = 0$, the model in (3.9) is called hierarchical selection model with standard t error (HSMT) in Chung *et al.* (2002). Also using the scale mixtures of normal distribution, we can get for $i = 1, \dots, n$,

$$\begin{gathered} y_i|\alpha_i,\sigma_i,\delta,z_i,w_i \sim \{C_t(\alpha_i,\sigma_i,z_i,\delta)\}^{-1} w(y_i)\, N\left(\alpha_i+\delta\{z_i - E(z_i)\}, w_i\sigma_i^2\right), \\ w_i \sim IG\left(\frac{\nu_1}{2},\frac{\nu_1}{2}\right), \quad z_i|\lambda_i \sim N^+(0,\lambda_i), \quad \lambda_i \sim IG\left(\frac{\nu_2}{2},\frac{\nu_2}{2}\right). \end{gathered} \tag{3.12}$$

The prior distributions for $(\mu,\sigma_\alpha^2,\delta,\sigma)$ are assumed the same as in (3.7). Then, the full conditional distributions for MCMC can be expressed as follows:

$$\begin{aligned} [\alpha_i|\alpha_j(j\neq i),\mathbf{y},\mathbf{z},\mu,\delta,\sigma_\alpha,\boldsymbol{\sigma},\mathbf{w},\boldsymbol{\lambda}] &\propto \{C_t(\alpha_i,\sigma_i,\delta,z_i)\}^{-1} N(\beta_2(\alpha_i),\gamma_2(\alpha_i)) \\ [\mu|\boldsymbol{\alpha},\mathbf{y},\mathbf{z},\delta,\sigma_\alpha,\boldsymbol{\sigma},\mathbf{w},\boldsymbol{\lambda}] &= N\left(\frac{b\sum_{i=1}^n \alpha_i + \sigma_\alpha^2 a}{bn+\sigma_\alpha^2}, \frac{b\sigma_\alpha^2}{bn+\sigma_\alpha^2}\right) \\ [\delta|\boldsymbol{\alpha},\mathbf{y},\mu,\mathbf{z},\sigma_\alpha,\boldsymbol{\sigma},\mathbf{w},\boldsymbol{\lambda}] &\propto \prod_{i=1}^n \{C_t(\alpha_i,\sigma_i,\delta,z_i)\}^{-1} N(\beta_2(\delta),\gamma_2(\delta)), \end{aligned}$$

$$
\begin{aligned}
[z_i|\boldsymbol{\alpha}, \mathbf{y}, \mu, z_j(j \neq i), \delta, \sigma_\alpha, \boldsymbol{\sigma}, \mathbf{w}, \boldsymbol{\lambda}] &\propto \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} N^+(\beta_2(z_i), \gamma_2(z_i)), \\
[\sigma_\alpha^2|\boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \boldsymbol{\sigma}, \mathbf{w}, \boldsymbol{\lambda}] &= IG\left(n + c_1, \frac{1}{2}\sum_{i=1}^n (\alpha_i - \mu)^2 + d_1\right), \\
[\sigma_i^2|\boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \sigma_\alpha, \sigma_j(j \neq i), \mathbf{w}, \boldsymbol{\lambda}] &\propto \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} IG(\xi_2(\sigma_i), \rho_2(\sigma_i)), \\
[w_i|\boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \sigma_\alpha, w_j(j \neq i), \boldsymbol{\sigma}, \boldsymbol{\lambda}] &= IG(\xi_2(w_i), \rho_2(w_i))
\end{aligned}
$$

and

$$
[\lambda_i|\boldsymbol{\alpha}, \mathbf{y}, \mu, \mathbf{z}, \delta, \sigma_\alpha, \lambda_j(j \neq i), \boldsymbol{\sigma}, \mathbf{w}] = IG\left(\frac{\nu_2 + 1}{2}, \frac{1}{2}(z_i^2 + \nu_2)\right)
$$

where

$$
\begin{aligned}
\beta_2(\alpha_i) &= \frac{\sigma_\alpha^2\,[y_i - \delta\{z_i - E(z_i)\}] + \sigma_i^2 \mu w_i}{\sigma_\alpha^2 + \sigma_i^2 w_i}, \quad \gamma_2(\alpha_i) = \frac{\sigma_\alpha^2 \sigma_i^2 w_i}{\sigma_\alpha^2 + \sigma_i^2 w_i}, \\
\beta_2(\delta) &= \frac{\sum_{i=1}^n \{z_i - E(z_i)\}(y_i - \alpha_i)(\sigma_i^2 w_i)^{-1} + m\tau^{-1}}{\sum_{i=1}^n \{z_i - E(z_i)\}^2 (\sigma_i^2 w_i)^{-1} + \tau^{-1}}, \\
\gamma_2(\delta) &= \left[\sum_{i=1}^n \frac{\{z_i - E(z_i)\}^2}{\sigma_i^2 w_i} + \frac{1}{\tau}\right]^{-1}, \\
\beta_2(z_i) &= \frac{\lambda_i \delta\{y_i - \alpha_i + \delta E(z_i)\}}{\sigma_i^2 w_i + \delta^2 \lambda_i}, \quad \gamma_2(z_i) = \frac{w_i \lambda_i \sigma_i^2}{\sigma_i^2 w_i + \delta^2 \lambda_i}, \\
\xi_2(\sigma_i) &= \frac{1}{2} + c_2, \quad \rho_2(\sigma_i) = \frac{1}{2w_i}\,[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2 + d_2, \\
\xi_2(w_i) &= \frac{\nu_1 + 1}{2} \text{ and } \rho_2(w_i) = \frac{1}{2}\left(\frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{\sigma_i^2} + \nu_1\right).
\end{aligned}
$$

3.3. Testing the skewness parameter $\delta = 0$

Now, our interest is to derive a test statistic for the null hypothesis

$$
H_0 : \delta = 0
$$

using the likelihood in (3.6) along with the priors in (3.7). In the various statistical testing, we select the score test which is based on the viewpoint of Ibrahim *et al.* (1998). This test statistic is based on the marginal likelihood of δ, denoted by $L(\delta)$ which is obtained as follows:

$$
L(\delta) = \int \prod_{i=1}^n \{[y_i, \alpha_i, z_i|\mu, \delta, \sigma_\alpha, \sigma_i][\sigma_i]\}[\mu, \sigma_\alpha] d\boldsymbol{\alpha} d\boldsymbol{\sigma} d\mathbf{z} d\mu d\sigma_\alpha^2. \tag{3.13}
$$

The score test depends only on posterior expectations of the likelihood

$$[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\delta,\mathbf{z}] = \prod_{i=1}^{n} \{[y_i|\alpha_i,\sigma_i,\delta,z_i]\}$$

evaluated at $\delta = 0$. In other words, we need the somewhat different joint posterior distribution in (3.8). This joint posterior density is given by

$$[\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z}|\mathbf{y},\delta=0] \propto \int [\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta=0][\boldsymbol{\alpha}|\mu,\sigma_\alpha^2][\mathbf{z}][\mu,\sigma_\alpha^2][\boldsymbol{\sigma}]d\mu d\sigma_\alpha^2. \tag{3.14}$$

Following Ibrahim *et al.* (1998), the score statistic for testing $H_0 : \delta = 0$ takes the form

$$U = \frac{S^2}{V}$$

where

$$S = \frac{d}{d\delta}\log L(\delta)\Big|_{\delta=0} \quad \text{and} \quad V = -\frac{d^2}{d\delta^2}\log L(\delta)\Big|_{\delta=0}.$$

Under the null hypothesis, it is well-known fact that U has an asymptotic chi-squared distribution with degree of freedom 1.

THEOREM 3.1. *Under the null hypothesis $H_0 : \delta = 0$, the score test statistic is expressed as*

$$U = \frac{S^2}{V},$$

where

$$V = S^2 - E^\theta\left[\frac{d^2}{d\delta^2}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]\Big|_{\delta=0}\right] - E^\theta\left[\left(\frac{d}{d\delta}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]\Big|_{\delta=0}\right)^2\right],$$

$$S = E^\theta\left[\frac{d}{d\delta}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]\Big|_{\delta=0}\right],$$

$\theta = (\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z})$ and E^θ denotes the expectation with respect to the joint posterior density in (3.8).

Appendix gives the proof of this theorem.

To compute the estimates of S and V, we need the first and second derivatives of the log likelihood function, $\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]$ of the skew-normal case and skew-t case.

3.3.1. Skew-normal error. From (3.5), for $i = 1, \dots, n$,

$$[y_i|\alpha_i, \sigma_i, \delta, z_i] = \{C(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} \frac{w(y_i)}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{2\sigma_i^2} \right), \quad (3.15)$$

where $C(\alpha_i, \sigma_i, \delta, z_i)$ is the normalizing constant. For example,

$$C(\alpha_i, \sigma_i, \delta, z_i) = \begin{cases} \eta \left\{ 1 - 2\Phi\left(-\frac{\eta}{\sigma_i}\right)\right\} + \sqrt{\frac{2}{\pi}}\sigma_i \exp\left(-\frac{\eta^2}{2\sigma_i^2}\right), & \text{if } w(y) = |y| , \\ \eta^2 + \sigma_i^2, & \text{if } w(y) = |y|^2, \end{cases} \quad (3.16)$$

where $\eta = \alpha_i + \delta\{z_i - E(z_i)\}$ and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Therefore, the log-likelihood of the full data, $\mathbf{y} = (y_1, \dots, y_n)^t$, is given by

$$\begin{aligned} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] &= -\sum_{i=1}^{n} \log(C(\alpha_i, \sigma_i, \delta, z_i)) + \sum_{i=1}^{n} \log w(y_i) - \frac{n}{2} \log 2\pi \\ &\quad - \sum_{i=1}^{n} \log \sigma_i - \sum_{i=1}^{n} \frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{2\sigma_i^2}. \end{aligned} \quad (3.17)$$

Now we can get the first and second derivatives of (3.17) evaluated at $\delta = 0$ which are given by

$$\begin{aligned} \frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta]\Big|_{\delta=0} &= \sum_{i=1}^{n} \frac{(y_i - \alpha_i)\{z_i - E(z_i)\}}{\sigma_i^2} \\ &\quad - \sum_{i=1}^{n} \left\{ \frac{d}{d\delta} C(\alpha_i, \sigma_i, z_i, \delta)\Big|_{\delta=0} \right\} \left\{ C(\alpha_i, \sigma_i, z_i, \delta)\Big|_{\delta=0} \right\}^{-1} \end{aligned} \quad (3.18)$$

and

$$\frac{d^2}{d\delta^2} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta]\Big|_{\delta=0} = \sum_{i=1}^{n} \left\{ \frac{d^2}{d\delta^2} C(\alpha_i, \sigma_i, z_i, \delta)\Big|_{\delta=0} \right\} \left\{ C(\alpha_i, \sigma_i, z_i, \delta)\Big|_{\delta=0} \right\}^{-1}$$

$$
- \sum_{i=1}^{n} \left[\left\{ \frac{d}{d\delta} C(\alpha_i, \sigma_i, z_i, \delta) \Big|_{\delta=0} \right\} \left\{ C(\alpha_i, \sigma_i, z_i, \delta) \Big|_{\delta=0} \right\}^{-1} \right]^2
- \sum_{i=1}^{n} \frac{\{z_i - E(z_i)\}^2}{\sigma_i^2} . \tag{3.19}
$$

Therefore, it follows from (A.1), (A.4) in Appendix, (3.18) and (3.19) that the estimates of S and V are respectively

$$
\widehat{S} = \frac{1}{G} \sum_{g=1}^{G} \left[\sum_{i=1}^{n} \frac{(y_i - \alpha_i^{(g)})\{z_i^{(g)} - E(z_i)\}}{(\sigma_i^{(g)})^2} - \sum_{i=1}^{n} \left\{ \frac{d}{d\delta} C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\} \left\{ C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\}^{-1} \right]
$$

and

$$
\begin{aligned}
\widehat{V} = \widehat{S}^2 + \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} & \left[\left\{ \frac{d^2}{d\delta^2} C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\} \left\{ C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\}^{-1} + \frac{\{z_i^{(g)} - E(z_i)\}^2}{(\sigma_i^2)^{(g)}} \right] \\
& - \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} \left[\left\{ \frac{d}{d\delta} C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\} \left\{ C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\}^{-1} \right]^2 \\
& - \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} \left[\frac{(y_i - \alpha_i^{(g)})\{z_i^{(g)} - E(z_i)\}}{(\sigma_i^2)^{(g)}} - \left\{ \frac{d}{d\delta} C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\} \right. \\
& \left. \times \left\{ C(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)}, \delta) \Big|_{\delta=0} \right\}^{-1} \right]^2 ,
\end{aligned}
$$

where $\left\{(\alpha_i^{(g)}, \sigma_i^{(g)}, z_i^{(g)})\right\}_{g=1}^{G}$ is the MCMC output in Section 3.1.

3.3.2. Skew-t error. Since the skew-t distribution in (3.9) is assumed as the error term, then the function in (3.10) is expressed as

$$
\begin{aligned}
[y_i|\alpha_i, \sigma_i, \delta, z_i] = \{C_t(\alpha_i, \sigma_i, \delta, z_i)\}^{-1} w(y_i) \frac{\Gamma((\nu_1 + 1)/2)}{\Gamma(\nu_1/2)\sqrt{\nu_1 \pi \sigma_i^2}} \\
\times \left(1 + \frac{[y_i - \alpha_i - \delta\{z_i - E(z_i)\}]^2}{\nu_1 \sigma_i^2}\right)^{-(\nu_1+1)/2}
\end{aligned} \tag{3.20}
$$

Here, the normalizing constant is quite different from (3.17) and is given by

$$C_t(\alpha_i,\sigma_i,\delta,z_i) = \begin{cases} \eta\left\{1-2T_{\nu_1}\left(-\frac{\eta}{\sigma_i}\right)\right\}+2J(\nu_1)\sigma_i\left(\nu_1+\frac{\eta^2}{\sigma_i^2}\right)^{-(\nu_1-1)/2}, & \text{if } w(y)=|y|, \\ \eta^2+\frac{\nu_1}{\nu_1-2}\sigma_i^2, & \text{if } w(y)=|y|^2. \end{cases} \tag{3.21}$$

Also, η is the same as defined in (3.16), T_{ν_1} denotes the *cdf* of standard t-distribution with degrees of freedom of ν_1 and $J(\nu_1)$ is given by

$$J(\nu_1)=\frac{\Gamma((\nu_1+1)/2)}{\Gamma(\nu_1/2)\sqrt{\pi}}\cdot\frac{\nu_1{}^{\nu_1/2}}{\nu_1-1}.$$

Like the skew-normal case, the log-likelihood of the full data is given by

$$\begin{aligned}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta] &= -\sum_{i=1}^n \log(C_t(\alpha_i,\sigma_i,\delta,z_i)) + n\log\left(\frac{\Gamma((\nu_1+1)/2)}{\Gamma(\nu_1/2)\sqrt{\nu_1\pi\sigma_i^2}}\right) \\ &\quad -\frac{\nu_1+1}{2}\sum_{i=1}^n \log\left(1+\frac{[y_i-\alpha_i-\delta\{z_i-E(z_i)\}]^2}{\nu_1\sigma_i^2}\right) \\ &\quad +\sum_{i=1}^n \log\left(\frac{w(y_i)}{\sigma_i}\right). \end{aligned} \tag{3.22}$$

The first and the second derivatives of (3.22) evaluated at $\delta=0$ have the form:

$$\begin{aligned}\frac{d}{d\delta}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]\Big|_{\delta=0} &= -\sum_{i=1}^n\left\{\frac{d}{d\delta}C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}\left\{C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}^{-1} \\ &\quad +(\nu_1+1)\sum_{i=1}^n\frac{(y_i-\alpha_i)\{z_i-E(z_i)\}}{\nu_1\sigma_i^2+(y_i-\alpha_i)^2}, \end{aligned} \tag{3.23}$$

$$\begin{aligned}\frac{d^2}{d\delta^2}\log[\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\sigma},\mathbf{z},\delta]\Big|_{\delta=0} &= -\sum_{i=1}^n\left\{\frac{d^2}{d\delta^2}C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}\left\{C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}^{-1}\end{aligned}$$

$$-\sum_{i=1}^{n}\left[\left\{\frac{d}{d\delta}C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}\left\{C_t(\alpha_i,\sigma_i,\delta,z_i)\Big|_{\delta=0}\right\}^{-1}\right]^2$$
$$+(\nu_1+1)\sum_{i=1}^{n}\left[-\frac{\{z_i-E(z_i)\}^2}{\nu_1\sigma_i^2+(y_i-\alpha_i)^2}\right]$$
$$+(\nu_1+1)\sum_{i=1}^{n}\left[\frac{2\{z_i-E(z_i)\}^2(y_i-\alpha_i)^2}{\{\nu_1\sigma_i^2+(y_i-\alpha_i)^2\}^2}\right]. \tag{3.24}$$

Therefore, the estimates of S and V are as follows:

$$\widehat{S}=\frac{\nu_1+1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\left[\frac{(y_i-\alpha_i^{(g)})\{z_i^{(g)}-E(z_i)\}}{\nu_1(\sigma_i^{(g)})^2+(y_i-\alpha_i^{(g)})^2}\right.$$
$$\left.-\left\{\frac{d}{d\delta}C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}\left\{C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}^{-1}\right],$$

$$\widehat{V}=\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\left[\left\{\frac{d^2}{d\delta^2}C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}\left\{C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}^{-1}\right.$$
$$\left.-\left\{\frac{d}{d\delta}C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}\left\{C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}^{-1}\right]^2$$
$$+(\nu_1+1)\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\left[\frac{\{z_i^{(g)}-E(z_i)\}^2}{\nu_1(\sigma_i^{(g)})^2+(y_i-\alpha_i^{(g)})^2}-\frac{2\{z_i^{(g)}-E(z_i)\}^2(y_i-\alpha_i^{(g)})^2}{\{\nu_1(\sigma_i^{(g)})^2+(y_i-\alpha_i^{(g)})^2\}^2}\right]$$
$$-\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\left[\frac{(y_i-\alpha_i^{(g)})\{z_i^{(g)}-E(z_i)\}}{(\sigma_i^{(g)})^2}-\left\{\frac{d}{d\delta}C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}\right.$$
$$\left.\times\left\{C(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)},\delta)\Big|_{\delta=0}\right\}^{-1}\right]^2+\widehat{S}^2,$$

where $\{(\alpha_i^{(g)},\sigma_i^{(g)},z_i^{(g)})\}_{g=1}^{G}$ is the MCMC output in Section 3.2.

4. An Illustrative Example

Johnson (1993) reviewed 12 studies comparing the effectiveness of two different types of fluoride, NaF which means a sodium fluoride and SMFP which means a sodium monofluorophosphate, in preventing tooth cavities. For each study, the observed average difference in effect y_i (given by the average increment in decayed, missing, and/or filled surfaces for teeth for patients using SMFP minus the average increment for patients using NaF), the corresponding standard error

TABLE 4.1 *The original Johnson data*

	Studies											
	1	2	3	4	5	6	7	8	9	10	11	12
y_i	0.86	0.33	0.47	0.50	−0.28	0.04	0.80	0.19	0.49	0.49	0.01	0.67
$\hat{\sigma}_i$	0.57	0.56	0.35	0.25	0.54	0.28	0.78	0.13	0.28	0.24	0.08	0.17
N	247	326	277	363	343	1490	418	2273	1352	2762	2222	2126

$\hat{\sigma}_i$, and its sample size N are given by Table 4.1. If the average increment is positive, then NaF is more effective than SMFP. In this study we want to know the overall effect, $\hat{\mu}$, estimated by weighted average. Johnson found that the estimate of μ is 0.32 and 95% confidence interval is (0.13, 0.52), supporting the hypothesis that NaF is better.

Although Johnson's result represented the superiority of NaF, her meta-analysis may have some problem. In particular, she considered 12 studies which were selected through a world-wide literature search. Actually she found 13 studies but used only 12, because the thirteenth failed to provide an estimate of the variance of the study effect and all the study effects except the 5^{th} are positive. For this reason, it is reasonable that her meta-analysis may have a publication bias and thus Silliman (1997) applied her HSM in (3.1) into Johnson's data.

Now, we apply our four Bayesian hierarchical selection models such as HSMN, HSMT, HSMSN and HSMST in Section 3 to Johnson's data in Table 4.1. To analyze Johnson's data in Table 4.1, we must specify both weight functions and prior distributions. First, two weight functions will be used, $w(y) = |y|$ and $w(y) = |y|^2$. Clearly, $w(y) = |y|$ indicates that the larger study effects are more likely to be observed, while $w(y) = |y|^2$ is used for more extreme publication bias. Under above weight functions, the exact forms of the normalizing constant are available which are given in (3.16) and (3.21). Now we specify the prior distributions of parameters of interest. Since the skewness parameter, δ, has a real line support, we assume the Normal distribution with zero mean and variance 10^2 based on Branco and Dey (2001). Both informative and improper priors of the overall mean, μ, and the between variance, σ_α^2, are considered. As informative priors, we assume that $\mu \sim N(0, 0.04)$ and σ_α^2 has the inverse gamma distribution with mean 0.04 and variance 1, This priors are based on the noniterative estimates of DerSimonian and Laird (1986). This is referred to as the clinical informative prior in Johnson (1993). As improper priors, Jeffreys' priors are used in Silliman

TABLE 4.2 *Geweke statistics for convergence diagnostics*

prior	*parameter*	*Geweke statistics*	*prior*	*parameter*	*Geweke statistics*
Inform	α	-1.280	Jeffreys	α	-1.520
	μ	-1.190		μ	-1.690
	σ_α^2	-1.320		σ_α^2	0.344
	δ	-1.620		δ	1.870
	z	0.298		z	-0.772
	σ^2	1.390		σ^2	-0.803

(1997) which is given as

$$\pi(\mu, \sigma_\alpha^2) \propto \left[\frac{1}{2}\sum_{i=1}^{n} \frac{1}{(\sigma_i^2 + \sigma_\alpha^2)}\right]. \tag{4.1}$$

We monitor the values of the parameters to four Bayesian hierarchical selection models under two weight functions and two prior distributions for (μ, σ_α^2). Gibbs samplers were run for 5,000 iterations with Metropolis algorithm of 10,000 iterations, where the first 3,000 being discarded as a burn-in period. Convergence of the Gibbs sampler was assessed *via* Geweke (1992) method, using the CODA (Best *et al.*, 1995), with suitable diagnostics in S-PLUS. Then all of the parameters had Geweke statistics within $\pm$ 1.96, indicating convergence is achieved. The Geweke statistics of each parameter is given by Table 4.2 and the Table 4.3 reports the estimates of μ, σ_α^2 and δ.

Also the results of the score test for $H_0 : \delta = 0$ are given in Table 4.4 depending on the choices of priors. In the Table 4.3, the values in parentheses of the column containing $\hat{\mu}$ denote the estimated posterior probability, $Pr(\mu > 0|\text{data})$, the values in parentheses of the column containing $\hat{\delta}$ denote the 95% credible interval, the column of $\pi(\mu, \sigma_\alpha^2)$ denotes the form of the prior distributions for (μ, σ_α^2), *i.e.*, *inform* means the informative prior distributions are assumed for (μ, σ_α^2) and *Jeffreys* does the Jeffreys' priors and E denotes the 10^{-4}, for example, $57E$ denotes 0.0057, respectively. N/A means that the value in the cell is not appeared. Tables 4.3 says that NaF is more effective since the posterior probabilities that μ is positive are almost around 0.9. It is reasonable to decide whether NaF is more effective or not by computing the posterior probability that the average treatment difference μ is positive. For informative prior of (μ, σ_α^2), in Table 4.3, the estimates of μ under skew-normal error (HSMSN) and skew-t error (HSMST) are smaller than those under normal error (HSMN) and t error (HSMT), respectively regardless of the weight functions. The all estimates of

TABLE 4.3 *Parameter estimates of* μ, σ_α^2 *and* δ

$w(y)$	$\pi(\mu,\sigma_\alpha^2)$	*Models*	$\hat{\mu}$	$\hat{\delta}$	$\hat{\sigma}_\alpha^2$
$\|y\|$	Inform	HSMN	0.214(0.999)	N/A	57E
		HSMT	0.344(0.999)	N/A	58E
		HSMSN	0.073(0.915)	0.001(−0.195, 0.212)	55E
		HSMST	0.266(0.999)	−0.088(−0.294, 0.184)	58E
	Jeffreys	HSMN	0.304(0.885)	N/A	0.35
		HSMT	0.347(0.970)	N/A	0.37
		HSMSN	0.313(0.915)	0.011(−0.205, 0.163)	0.44
		HSMST	0.327(0.940)	−0.019(−0.221, 0.190)	0.34

TABLE 4.4 *Results of the* $H_0 : \delta = 0$

$w(y)$	$\pi(\mu,\sigma_\alpha^2)$	*Models*	U	*p-value*
$\|y\|$	Inform	HSMSN	1.365	0.2430
		HSMST	0.548	0.4594
	Jeffreys	HSMSN	2.244	0.1336
		HSMST	0.472	0.4926

the between-study variance σ_α^2 are almost 0.005. The estimated posterior probabilities that $\mu > 0$ are almost 0.99 which means that NaF is more effective in combatting cavities than SMFP. Similar results also happen except the estimated of σ_α^2 when the Jeffreys' prior is assumed. Therefore, our limited experience says that the evidence in favor of NaF is robust to choice of priors.

Next, we consider the skewness of error, *i.e.*, whether $\delta = 0$ or not. Table 4.4 contains the observed values of a score statistic U and these significant probabilities (p-value). For example, in Table 4.4, the observed value of the score test statistic under HSMSN with $w(y) = |y|$ and the informative priors of $\pi(\mu, \sigma_\alpha^2)$ is 1.365 and its p-value is 0.2430. In Table 4.3, the estimates of δ are almost close to zero in any case and its 95% credible intervals contains zero. Therefore, our limited experience shows that this data set supports no skewness in our model. Furthermore, Tables 4.4 tells us that our data in Table 4.1 supports there is no skewness, *i.e.*, $\delta = 0$ in every skewed models since p-values are larger than the general significant level ($\alpha = 0.05$).

5. Concluding Remarks

In this paper, we introduce the hierarchical selection models including skewed errors developed by Silliman (1997)'s which can be applied meta-analysis. A general class of skew-elliptical distribution is reviewed, the posterior propriety under assuming skewed error and improper prior for some parameters is proved. Using the method of Chen *et al.* (1999) we construct sampling structure so that MCMC can be conducted. Also, the score test for skewness parameter is available. Our proposed methodology is applied Johnson (1993)'s data.

The hierarchical selection models with skewed errors are big class including symmetric error models so it can be applied various statistical field such as regression problems and reliability analysis and so on.

Further work may include following things: first, the model selection procedure is required. The general Bayes factor cannot have a closed form so we must find reasonable measure which can be adapted our situation to compare model selection. Second, it is solved the problems of estimating degrees of freedom, ν_1 and ν_2 under skew-t models. Though Branco and Dey (2001) and Sahu *et al.* (2001) assume random degrees of freedom but their sampling methods such as adaptive-rejection are still inefficient which is needed more efficient sampling algorithm. Finally, for more effective analysis it is considered creating the simulating dataset under selection models.

Appendix : Proofs of Theorems

Proof of Theorem 2.2. Assume $\pi(\mu) \propto 1$ and $\sigma_\alpha^2 \sim IG(c,d)$ where $c, d \to 0$. Let

$$A = \int \cdots \int \prod_{i=1}^{n} \left[SE(y_i|\alpha_i, \sigma_i^2, \delta : g^{(1)}) N(\alpha_i|\mu, \sigma_\alpha^2) \right] \times \pi(\mu)\pi(\sigma_\alpha^2)\pi(\boldsymbol{\sigma}^2, \delta, \nu) d\boldsymbol{\alpha} d\mu d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta d\nu.$$

It is sufficient to show that A is finite. Now,

$$A = C_1 \int \frac{2^n}{\prod_{i=1}^n (\sigma_i^2 + \delta^2)^{1/2}} \prod_{i=1}^{n} \left\{ g^{(1)}\left(\frac{(y_i - \alpha_i)^2}{\sigma_i^2 + \delta^2} \right) F_{el}\left(\frac{\delta}{\sigma} \frac{y_i - \alpha_i}{\sqrt{\sigma_i^2 + \delta^2}} \middle| 0, 1, g_a^{(1)} \right) \right\}$$
$$\times \frac{1}{(\sigma_\alpha^2)^{n/2}} \prod_{i=1}^{n} \left[\exp\left\{ -\frac{(\alpha_i - \mu)^2}{2\sigma_\alpha^2} \right\} \right] \pi(\mu)\pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta)\pi(\nu) d\boldsymbol{\alpha} d\mu d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta d\nu$$

$$\leq C_2 \int \frac{2^n}{\prod_{i=1}^n (\sigma_i^2 + \delta^2)^{1/2}} \cdot \frac{1}{(\sigma_\alpha^2)^{n/2}} \prod_{i=1}^n \left\{ g^{(1)} \left(\frac{(y_i - \alpha_i)^2}{\sigma_i^2 + \delta^2} \right) \right\}$$
$$\times \left[\int_R \exp\left\{ -\frac{\sum_{i=1}^n (\alpha_i - \mu)^2}{2\sigma_\alpha^2} \right\} d\mu \right] \pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta)\pi(\nu) d\boldsymbol{\alpha} d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta d\nu,$$

where C_1 and C_2 are some constants. Note that

$$\int_R \exp\left\{ -\frac{\sum_{i=1}^n (\alpha_i - \mu)^2}{2\sigma_\alpha^2} \right\} d\mu \leq \sqrt{2\pi \frac{\sigma_\alpha^2}{n}}.$$

First, for the skew-normal model, we do not need the parameter ν. It follows from Example 2.1 that

$$A \leq C_3 \int \frac{2^n}{\prod_{i=1}^n (\sigma_i^2 + \delta^2)^{1/2}} \cdot \frac{1}{(\sigma_\alpha^2)^{n/2}} \sqrt{\frac{\sigma_\alpha^2}{n}}$$
$$\times \prod_{i=1}^n \int_R \exp\left\{ -\frac{(y_i - \alpha_i)^2}{2(\sigma_i^2 + \delta^2)} \right\} d\alpha_i \, \pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta) d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta$$
$$\leq C_4 \int \frac{1}{(\sigma_\alpha^2)^{(n-1)/2}} \pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta) d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta$$
$$\leq C_4 \iint \left\{ \int \frac{1}{(\sigma_\alpha^2)^{(n-1)/2}} \cdot \frac{1}{(\sigma_\alpha^2)^{c+1}} \exp\left(-\frac{d}{\sigma_\alpha^2} \right) d\sigma_\alpha^2 \right\} \pi(\boldsymbol{\sigma}^2)\pi(\delta) d\boldsymbol{\sigma}^2 d\delta$$

where C_3 and C_4 are appropriate constants, too. Now as $c, d \to 0$ the inner integral is finite if $n > 1$. Hence A is finite.

Finally, consider the skew-t model. It follows from Example 2.2 that

$$g^{(1)}(u; \nu) = \frac{\Gamma\big((\nu+1)/2\big)}{\Gamma\big(\nu/2\big)(\nu\pi)^{-(1/2)}} \left(1 + \frac{u}{\nu}\right)^{-(\nu+1)/2}.$$

Therefore,

$$A \leq C_5 \int \frac{1}{\prod_{i=1}^n (\sigma_i^2 + \delta^2)^{1/2}} J^n \prod_{i=1}^n \left[\int_R \left(1 + \frac{(y_i - \alpha_i)^2}{\nu(\sigma_i^2 + \delta^2)} \right)^{-(\nu+1)/2} d\alpha_i \right]$$
$$\times (\sigma_\alpha^2)^{-n/2} \sqrt{2\pi \frac{\sigma_\alpha^2}{n}} \pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta)\pi(\nu) d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta d\nu,$$

where $J = \Gamma\big((\nu+1)/2\big) \left\{ \Gamma\big(\nu/2\big) \ (\nu\pi)^{-(1/2)} \right\}^{-1}$. After some straightforward algebra, we can get the product of inner integral, *i.e.*,

$$\prod_{i=1}^n \left[\int_R \left\{ 1 + \frac{(y_i - \alpha_i)^2}{\nu(\sigma_i^2 + \delta^2)} \right\}^{-(\nu+1)/2} d\alpha_i \right] = J^{-n} \prod_{i=1}^n (\sigma_i^2 + \delta^2)^{1/2}.$$

Therefore,

$$A \le C_6 \int \frac{1}{(\sigma_\alpha^2)^{(n-1)/2}} \; \pi(\boldsymbol{\sigma}^2)\pi(\sigma_\alpha^2)\pi(\delta)\pi(\nu)d\boldsymbol{\sigma}^2 d\sigma_\alpha^2 d\delta d\nu$$
$$\le C_6 \iiint \left\{ \int \frac{1}{(\sigma_\alpha^2)^{(n-1)/2}} \cdot \frac{1}{(\sigma_\alpha^2)^{c+1}} \exp\left(-\frac{d}{\sigma_\alpha^2}\right) d\sigma_\alpha^2 \right\} \pi(\boldsymbol{\sigma}^2)\pi(\delta)\pi(\nu)d\boldsymbol{\sigma}^2 d\delta d\nu$$

where C_5 and C_6 are appropriate constants. Also $\pi(\nu)$ is proper, A is finite as $c, d \to 0$ if $n > 1$. The proof is complete. □

PROOF OF THEOREM 3.1. Let $S = d\{\log L(\delta)\}/d\delta|_{\delta=0}$. Then, it can be shown by direct computations that

$$S = \int \frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0]\pi(\theta|\mathbf{y}, \delta = 0)d\boldsymbol{\alpha}d\mathbf{z}d\boldsymbol{\sigma}^2,$$

where

$$\pi(\theta|\mathbf{y}, \delta = 0) = \frac{[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]}{\int [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]d\boldsymbol{\alpha}dzd\boldsymbol{\sigma}^2}$$

denotes the joint posterior distribution of $\theta = (\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z})$ given $(\mathbf{y}, \delta = 0)$. Thus,

$$S = \frac{d}{d\delta} \log L(\delta)\Big|_{\delta=0} = E^\theta \left[\frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta]\Big|_{\delta=0} \right]. \tag{A.1}$$

By using the similar computations as above, we can derive the form of V as follows:

$$\begin{aligned} V &= -\frac{d^2}{d\delta^2} \log L(\delta)\Big|_{\delta=0} \\ &= -\frac{d}{d\delta}\left[\left\{ \int \left(\frac{d}{d\delta}[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta]\Big|_{\delta=0} \right) [\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]d\boldsymbol{\alpha}dzd\boldsymbol{\sigma}^2 \right\} \right. \\ &\qquad \left. \times \left\{ \int [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]d\boldsymbol{\alpha}dzd\boldsymbol{\sigma}^2 \right\}^{-1} \right] \\ &= S^2 - W, \end{aligned} \tag{A.2}$$

where

$$W = \int \left(\frac{d^2}{d\delta^2}[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta]\Big|_{\delta=0} \right) \times \frac{[\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]}{\int [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]d\boldsymbol{\alpha}d\mathbf{z}d\boldsymbol{\sigma}^2} d\boldsymbol{\alpha}d\mathbf{z}d\boldsymbol{\sigma}^2.$$

Then, W can be expressed as follows:

$$\begin{aligned} W &= \int \left[\left\{ \frac{d^2}{d\delta^2} [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0]^{-1} \right\} \right. \\ &\quad \left. \times \frac{[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}]}{\int [\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta = 0][\mathbf{z}][\boldsymbol{\sigma}^2][\boldsymbol{\alpha}] d\boldsymbol{\alpha} d\mathbf{z} d\boldsymbol{\sigma}^2} \right] d\boldsymbol{\alpha} d\mathbf{z} d\boldsymbol{\sigma}^2 \\ &= \int \left(\frac{d^2}{d\delta^2} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right) \pi(\theta|\mathbf{y}, \delta = 0) d\boldsymbol{\alpha} d\mathbf{z} d\boldsymbol{\sigma}^2 \\ &\quad + \int \left(\frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right)^2 \pi(\theta|\mathbf{y}, \delta = 0) d\boldsymbol{\alpha} d\mathbf{z} d\boldsymbol{\sigma}^2 \\ &= E^{\theta} \left[\frac{d^2}{d\delta^2} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right] + E^{\theta} \left[\left(\frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right)^2 \right] \end{aligned} \tag{A.3}$$

where $\pi(\theta|\mathbf{y}, \delta = 0)$ is the same as defined above. Therefore, it follows from (A.2) and (A.3) that

$$\begin{aligned} V &= S^2 - W \\ &= S^2 - E^{\theta} \left[\frac{d^2}{d\delta^2} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right] - E^{\theta} \left[\left(\frac{d}{d\delta} \log[\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}, \mathbf{z}, \delta] \Big|_{\delta=0} \right)^2 \right]. \end{aligned} \tag{A.4}$$

This completes the proof. □

REFERENCES

AZZALINI, A. AND DALLA-VALLE, A. (1996). "The multivariate skew-normal distribution", *Biometrika*, **83**, 715–726.

BEST, N. G., COWLES, M. K. AND VINES, S. K. (1995). "CODA convergence diagnosis and output analysis software for Gibbs sampler output : Version 0.3", Unpublished technical report, MRC Biostatistics Unit, Cambridge.

BRANCO M. D. AND DEY, D. K. (2001). "A general class of multivariate skew-elliptical distributions", *Journal of Multivariate Analysis*, **79**, 99–113.

CHEN, M. H., DEY, D. K. AND SHAO, Q. M. (1999). "A new skewed link model for dichotomous quantal response data", *Journal of the American Statistical Association*, **94**, 1172–1186.

CHUNG, Y., DEY, D. K. AND JANG, J. (2002). "Semiparametric hierarchical selection models for Bayesian meta-analysis", *Journal of Statistical Computation and Simulation*, **72**, 825–839.

CHUNG, Y. AND JANG, J. (2001). "A Bayesian method to semiparametric hierarchical selection models", *The Korean Journal of Applied Statistics*, **14**, 161–175.

CHUNG, Y. AND JANG, J. (2003). "Semiparametric Bayesian hierarchical selection models with skewed elliptical distributions", *The Korean Journal of Applied Statistics*, **16**, 101–115.

DERSIMONIAN, R. AND LAIRD, N. (1986). "Meta-analysis in clinical trials", *Controlled Clinical Trials*, **7**, 177–188.

GEWEKE, J. (1992). "Evaluating the accuracy of sampling-based approaches to calculating posterior moments", In *Bayesian Statistics IV* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 169–193, Oxford University Press, Oxford.

HASTINGS, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, **57**, 97–109.

IBRAHIM, J. G., RYAN, L. M. AND CHEN, M. H. (1998). "Using historical controls to adjust for covariates in trend tests for binary data", *Journal of the American Statistical Association*, **93**, 1282–1293.

JOHNSON, M. F. (1993). "Comparative efficacy of NaF and SMFP dentifrices in caries prevention : A meta-analytic overview", *Journal of European Organization for Caries Research*, **27**, 328–336.

LAROSE, D. AND DEY, D. K. (1996). "Weighted distributions viewed in the context of model selection : A Bayesian perspective", *Test*, **5**, 227–246.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. AND TELLER, E. (1953). "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, **21**, 1087–1092.

MORRIS, C. N. AND NORMAND, S. L. (1992). "Hierarchical models for combining information and for meta-analysis", In *Bayesian Statistics IV* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 321–344, Oxford University Press, Oxford.

SAHU, S. K., DEY, D. K. AND BRANCO, M. D. (2001). "A new class of multivariate skew distributions with applications to Bayesian regression models", Unpublished technical report, University of Southampton.

SILLIMAN, N. P. (1997). "Hierarchical selection models with applications in meta-analysis", *Journal of the American Statistical Association*, **92**, 926–936.

ZELLNER, A. (1976). "Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms", *Journal of the American Statistical Association*, **71**, 400–405.