

A new class of multivariate skew distributions with applications to Bayesian regression models

Sujit K. SAHU, Dipak K. DEY and Márcia D. BRANCO

Key words and phrases: Bayesian inference; elliptical distributions; Gibbs sampler; heavy tailed error distribution, Markov chain Monte Carlo; multivariate skewness.

MSC 2000: Primary 62F15, 62J05; secondary 62E15, 62H12.

Abstract: The authors develop a new class of distributions by introducing skewness in multivariate elliptically symmetric distributions. The class, which is obtained by using transformation and conditioning, contains many standard families including the multivariate skew-normal and t distributions. The authors obtain analytical forms of the densities and study distributional properties. They give practical applications in Bayesian regression models and results on the existence of the posterior distributions and moments under improper priors for the regression coefficients. They illustrate their methods using practical examples.

Une nouvelle classe de lois multivariées asymétriques et ses applications dans le cadre de modèles de régression bayésiens

Résumé : Les auteurs engendrent une nouvelle classe de lois en introduisant un facteur d'asymétrie dans la famille des distributions multivariées elliptiquement symétriques. La classe, qui est obtenue par transformation et conditionnement, inclut plusieurs familles de lois connues, dont la Student et la normale multivariées asymétriques. Les auteurs donnent la forme explicite de la densité de ces lois et en examinent les propriétés. Ils en présentent des applications pratiques dans le cadre des modèles de régression bayésiens, où ils démontrent l'existence de lois a posteriori et de leurs moments lorsque les lois a priori des paramètres de la régression sont impropres. Ils illustrent en outre leurs méthodes dans des cas concrets.

1. INTRODUCTION

Advances in Bayesian computation and Markov chain Monte Carlo have extended and broadened the scope of statistical models to fit actual data. Surprisingly, the methodologies and techniques of data augmentation and computation can also be used for developing new sets of flexible models for data. The main motivation of this article comes from this observation. A simple but powerful method of generating a class of multivariate skew elliptical distributions is obtained with a view to finding easily implementable fitting methods.

The class of elliptical distributions, introduced by Kelker (1970), includes a vast set of known symmetric distributions, for example, normal, Student t and Pearson type II distributions. These ideas are quite well developed; see for example Fang, Kotz & Ng (1990). A major focus of the current paper is the introduction of skewed versions of these distributions that are suitable for practical implementations. A general transformation technique together with a conditioning argument is used to obtain skewed versions of the multivariate distributions. In univariate cases, similar ideas have been studied by many authors; see for example, Aigner, Lovell & Schmidt (1977) and Chen, Dey & Shao (1999).

The conditioning arguments on some unobserved variables used to develop the models are commonly used in regression models. The resulting models are often called the hidden truncation models; see, e.g., Arnold & Beaver (2000, 2002). Consider the following motivating example. In order to gain admission to a medical school, applicants are often screened by both academic and nonacademic criteria. Only the candidates meeting several academic criteria (e.g., overall grades and grades in science) are evaluated by nonacademic criteria such as commitment and caring, sense of responsibility, etc. A response variable called the *nonacademic total* is the sum

of scores from seven such nonacademic headings which is used to screen applicants for the next stage of the admission process. Thus meeting the academic criteria acts as a conditioning variable for the response to the nonacademic total. Moreover, some variables (components) for meeting the academic criteria are yet unobserved, since the admission process is often initiated before the applicants take their final qualifying examinations. This example is discussed in more detail in Section 6.

The methodology developed here is also useful in modelling stock market returns. The expected rate of returns on risky financial assets, e.g., stocks, bonds, options and other securities, are often assumed to be normally distributed but are subject to shocks in either positive or negative directions; positive shocks lead to positively skewed models and negative shocks lead to negatively skewed models; see, for example, Adcock (2002). In the related area of capital asset pricing models, the assumption of multivariate normality is often hard to justify in real-life examples (Huang & Litzenberger 1988) and the proposed skew models can be used instead.

In many practical regression problems, a suitable transformation for symmetry is often considered for skewed data. The proposed models eliminate the need for such ad hoc transformations. Instead of transforming the data, our methods transform the error distributions to accommodate skewness.

In the case of normal distributions, our setup provides a new family of multivariate skew-normal distributions. The distributions are different from the ones obtained by Azzalini and his colleagues. See, e.g., Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999); see also Arnold & Beaver (2000) for a generalization. They obtain the multivariate distribution by conditioning on one suitable random variable being greater than zero, while we condition on as many random variables as the dimension of the multivariate distribution. Thus in the univariate case, the new distributions are the same as the ones obtained by Azzalini & Dalla Valle (1996). In the multivariate setup, however, the two sets of distributions are quite different. Also our method extends to other distributions, e.g., the t and the Pearson type II distributions.

There are some other variants of skewed distributions available in the literature. For example, Jones (2001) (and see the references to his earlier work therein) provides an alternative skew- t distribution which in the limiting case is a scaled inverse χ distribution. Fernandez & Steel (1998) consider an alternative form where two t distributions (with different scale parameters) in the positive and negative domains are combined to form a skew- t distribution. The distributions developed in this article, however, are much easier to work with and implement than others.

Bayesian analysis of regression problems under heavy-tailed error distributions has received considerable attention in recent statistical literature. A pioneering work in this area is Zellner (1976), in which a study based on the multivariate t distribution is considered. Extensions of those results for elliptical distributions are considered in Chib, Tiwari & Jammalamadaka (1988), Osiewalski & Steel (1993) and Branco, Bolfarine, Iglesias & Arellano-Valle (2000). More about Bayesian regression under heavy tailed error distributions can be found in Geweke (1993), Fernandez & Steel (1998) and references therein. However, these methodologies do not generally extend to multivariate skew distributions.

The plan of the remainder of this paper is as follows. Section 2 develops the multivariate skew elliptical distributions. Sections 3 and 4 consider the particular cases of normal and t distributions. In Section 5, we develop regression models for the skewed distributions obtained in the preceding sections. Results on the propriety of the associated posterior distributions in the univariate case are also obtained here. In Section 6.2, we illustrate our methods when the response variable is univariate. A multivariate example is discussed in Section 6.3. We give a few summary remarks in Section 7. Technical proofs of our results are in the Appendix.

2. MULTIVARIATE DISTRIBUTIONS

2.1. Elliptical distribution.

Let Ω be a positive definite matrix of order k and $\boldsymbol{\theta} \in \mathbb{R}^k$. Consider a k -dimensional random vector \mathbf{X} having probability density function of the form

$$f(\mathbf{x} \mid \boldsymbol{\theta}, \Omega; g^{(k)}) = |\Omega|^{-1/2} g^{(k)} \{(\mathbf{x} - \boldsymbol{\theta})' \Omega^{-1} (\mathbf{x} - \boldsymbol{\theta})\}, \quad \mathbf{x} \in \mathbb{R}^k \quad (1)$$

where $g^{(k)}(u)$ is a function from \mathbb{R}^+ to \mathbb{R}^+ defined by

$$g^{(k)}(u) = \frac{\Gamma(k/2)}{\pi^{k/2}} \frac{g(u; k)}{\int_0^\infty r^{k/2-1} g(r; k) dr}, \quad (2)$$

where $g(u; k)$ is a function from \mathbb{R}^+ to \mathbb{R}^+ such that the integral $\int_0^\infty r^{k/2-1} g(r; k) dr$ exists. In this paper we shall always assume the existence of the probability density function (1). The function $g^{(k)}$ is often called the *density generator* of the random vector \mathbf{X} . Note that the function $g(u; k)$ provides the kernel of \mathbf{X} and other terms in $g^{(k)}$ constitute the normalizing constant for the density f . In addition the function g , hence $g^{(k)}$, may depend on other parameters which would be clear from the context. For example, for t distributions the additional parameter will be the degrees of freedom. The density f defined above represents a broad class of distributions called the *elliptically symmetric distributions* for which we will use the notation

$$\mathbf{X} \sim \text{El}(\boldsymbol{\theta}, \Omega; g^{(k)}),$$

in this article. Let $F(\mathbf{x} \mid \boldsymbol{\theta}, \Omega; g^{(k)})$ denote the cumulative density function of \mathbf{X} where $\mathbf{X} \sim \text{El}(\boldsymbol{\theta}, \Omega; g^{(k)})$.

We consider two examples, namely the multivariate normal and t distributions, which will be used throughout this paper.

Example 1 (Multivariate normal). Let $g(u; k) = \exp(-u/2)$. Then straightforward calculation yields

$$g^{(k)}(u) = e^{-u/2} / (2\pi)^{k/2}.$$

Accordingly,

$$f(\mathbf{x} \mid \boldsymbol{\theta}, \Omega; g^{(k)}) = \frac{1}{(2\pi)^{k/2}} |\Omega|^{-1/2} \exp\{-1/2(\mathbf{x} - \boldsymbol{\theta})' \Omega^{-1} (\mathbf{x} - \boldsymbol{\theta})\}, \quad \mathbf{x} \in \mathbb{R}^k$$

which is the probability density function of the k -variate normal distribution with mean vector $\boldsymbol{\theta}$ and covariance matrix Ω . We denote this distribution by $N_k(\boldsymbol{\theta}, \Omega)$ and the probability density function by $N_k(\mathbf{x} \mid \boldsymbol{\theta}, \Omega)$ henceforth.

Example 2 (Multivariate t). Let

$$g(u; k, \nu) = \left(1 + \frac{u}{\nu}\right)^{-(\nu+k)/2}, \quad \nu > 0.$$

Here g depends on the additional parameter ν , the degrees of freedom. Straightforward calculation yields

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

Hence

$$f(\mathbf{x} \mid \boldsymbol{\theta}, \Omega, g^{(k)}) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{k/2}} |\Omega|^{-1/2} \left\{ 1 + \frac{(\mathbf{x} - \boldsymbol{\theta})' \Omega^{-1} (\mathbf{x} - \boldsymbol{\theta})}{\nu} \right\}^{-(\nu+k)/2}, \quad \mathbf{x} \in \mathbb{R}^k$$

which is the density of the k -variate t distribution with parameters $\boldsymbol{\theta}$, Ω and degrees of freedom ν . We denote this distribution by $t_{k,\nu}(\boldsymbol{\theta}, \Omega)$ and the density by $t_{k,\nu}(\mathbf{x} \mid \boldsymbol{\theta}, \Omega)$ henceforth. The subscript k will be omitted when it is equal to 1.

2.2. Skew elliptical distribution.

Let $\boldsymbol{\varepsilon}$ and \mathbf{Z} denote m -dimensional random vectors. Let $\boldsymbol{\mu}$ be an m -dimensional vector and Σ be an $m \times m$ positive definite matrix. Assume that

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{Z} \end{pmatrix} \sim \text{El}\left(\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \Omega = \begin{pmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}; g^{(2m)}\right),$$

where $\mathbf{0}$ is the null matrix and I is the identity matrix. We consider a skew elliptical class of distributions by using the transformation

$$\mathbf{Y} = D\mathbf{Z} + \boldsymbol{\varepsilon}, \quad (3)$$

where D is a diagonal matrix with elements $\delta_1, \dots, \delta_m$, though we can work with any nonsingular square matrix. Let $\boldsymbol{\delta}' = (\delta_1, \dots, \delta_m)$. We develop the class by considering the random variable $\mathbf{Y} \mid \mathbf{Z} > \mathbf{0}$, where $\mathbf{Z} > \mathbf{0}$ means $Z_i > 0$ for $i = 1, \dots, m$. Note that if $\boldsymbol{\delta} = \mathbf{0}$, then we retrieve the original elliptical distribution. The construction (3) with the conditioning introduces skewness. For positive values of components of $\boldsymbol{\delta}$, we obtain positively (right) skewed distributions and for negative values, we obtain negatively (left) skewed distributions. The conditional density of Y is obtained in the following theorem.

THEOREM 1. Let $\mathbf{y}_* = \mathbf{y} - \boldsymbol{\mu}$. Then the probability density function of $\mathbf{Y} \mid \mathbf{Z} > \mathbf{0}$ is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, D; g^{(m)}) = 2^m f_{\mathbf{Y}}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma + D^2; g^{(m)}) \text{P}(\mathbf{V} > \mathbf{0}), \quad (4)$$

where

$$\mathbf{V} \sim \text{El}\left(D(\Sigma + D^2)^{-1} \mathbf{y}_*, I - D(\Sigma + D^2)^{-1} D; g_{q(\mathbf{y}_*)}^{(m)}\right)$$

and

$$g_a^{(m)}(u) = \frac{\Gamma(m/2)}{\pi^{m/2}} \frac{g(a+u; 2m)}{\int_0^\infty r^{m/2-1} g(a+r; 2m) dr}, \quad a > 0 \quad (5)$$

and

$$q(\mathbf{y}_*) = \mathbf{y}_*' (\Sigma + D^2)^{-1} \mathbf{y}_*.$$

This density matches with the one obtained by Branco & Dey (2001) only in the univariate case. We denote the random variable \mathbf{Y} by using the notation $Y \sim SE(\boldsymbol{\mu}, \Sigma, D; g^{(m)})$. In Sections 3 and 4, we provide two examples of the density (4). In general, the cumulative density function in (4) is hard to evaluate. However, for practical MCMC model fitting the cumulative density function need not be calculated; see Section 5.

In the univariate case, i.e., when $m = 1$, we take $\Sigma = \sigma^2$ and $D = \delta$. The density (4) then simplifies to

$$f(y \mid \mu, \sigma^2, \delta; g^{(1)}) = \frac{2}{\sqrt{\sigma^2 + \delta^2}} g^{(1)}\left(\frac{y - \mu}{\sigma^2 + \delta^2}\right) F\left(\frac{\delta}{\sigma} \frac{y - \mu}{\sqrt{\sigma^2 + \delta^2}} \mid 0, 1; g_a^{(1)}\right), \quad (6)$$

where $g^{(1)}(u)$ is given in (2), $a = (y - \mu)^2/(\sigma^2 + \delta^2)$ and $g_a^{(1)}(u)$ are given in (5).

Using the arguments in the proof of Theorem 1, we can obtain the marginal distribution of subsets of components of \mathbf{Y} . We derive the marginal distributions using the construct $Y_i | \mathbf{Z} > 0$, and not using the construct $Y_i | Z_i > 0$. Suppose that it is desired to obtain the marginal density of first m_1 components of \mathbf{Y} . The marginal density will be

$$f(\mathbf{y}^{(1)} | \boldsymbol{\mu}^{(1)}, \Sigma_{11}, D_{11}; g^{(m_1)}) = 2^{m_1} f_{\mathbf{Y}}^{(1)}(\mathbf{y}^{(1)} | \boldsymbol{\mu}^{(1)}, \Sigma_{11} + D_{11}^2; g^{(m_1)}) P(\mathbf{V} > \mathbf{0}),$$

where the symbols have their usual meanings and

$$\mathbf{V} \sim \text{El} \left(D_{11}(\Sigma_{11} + D_{11}^2)^{-1} \mathbf{y}^{(1)}, I - D_{11}(\Sigma_{11} + D_{11}^2)^{-1} D_{11}; g_{q(\mathbf{y}_*^{(1)})}^{(m_1)} \right).$$

It is straightforward to observe that the above marginal density is of the same form as (4). Hence coherence with respect to marginalization is preserved under (4). The conditional density of any subset of variables can be obtained from the joint and marginal densities.

3. THE SKEW-NORMAL DISTRIBUTION

3.1. Density.

Let $g(u; m) = \exp(-u/2)$. Then it is easy to see that $g^{(m)}(u) = (2\pi)^{-m/2} \exp(-u/2)$ and $g_{q(\mathbf{y}_*)}^{(m)}$ is free of $q(\mathbf{y}_*)$; see equation (5). Now the probability density function of the skew-normal distribution is given by

$$f(\mathbf{y} | \boldsymbol{\mu}, \Sigma, D) = 2^m |\Sigma + D^2|^{-1/2} \phi_m \left\{ (\Sigma + D^2)^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \right\} P(\mathbf{V} > \mathbf{0}) \tag{7}$$

where

$$\mathbf{V} \sim N_m \{ D(\Sigma + D^2)^{-1} (\mathbf{y} - \boldsymbol{\mu}), I - D(\Sigma + D^2)^{-1} D \}$$

and ϕ_m is the density of the m -dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix identity. (We drop the subscript m in ϕ_m when $m = 1$.) We denote the above distribution by $\text{SN}(\boldsymbol{\mu}, \Sigma, D)$. An appealing feature of (7) is that it gives independent marginals when $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$. The density (7) then reduces to

$$f(\mathbf{y} | \boldsymbol{\mu}, \sigma^2, D) = \prod_{i=1}^m \left\{ 2(\sigma_i^2 + \delta_i^2)^{-1/2} \phi \left(\frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right) \Phi \left(\frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right) \right\},$$

where Φ is the cumulative density function of the standard normal distribution.

3.2. Moments and skewness.

For the multivariate distribution $\text{SN}(\boldsymbol{\mu}, \Sigma, D)$, we provide the first two moments. These are obtained with the moment generating function

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^m \Phi_m(D\mathbf{t}) \exp\{ \mathbf{t}' \boldsymbol{\mu} + \mathbf{t}' (\Sigma + D^2) \mathbf{t} / 2 \}, \tag{8}$$

where Φ_m is the cumulative density function of the m dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix identity. The Appendix contains the derivation of this. The mean and variance of $\text{SN}(\boldsymbol{\mu}, \Sigma, D)$ are given by

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{2}{\pi} \right)^{1/2} \boldsymbol{\delta} \quad \text{and} \quad \text{cov}(\mathbf{Y}) = \Sigma + \left(1 - \frac{2}{\pi} \right) D^2.$$

Since the matrix D is assumed to be diagonal, the introduction of skewness does not affect the correlation structure. It changes the values of correlations but the structure remains the same.

Thus the mutual independence of the components, when Σ is diagonal, is preserved under (3) for the normal distribution. However, this is not true for the skew-normal distribution of Azzalini & Capitanio (1999). The introduction of skewness in their setup changes the correlation structure.

As mentioned previously, $SN(\boldsymbol{\mu}, \Sigma, D)$ coincides with the skew-normal distribution obtained by Azzalini & Dalla Valle (1996), and Azzalini & Capitanio (1999) in the *univariate* case only. Hence, the skewness properties of the univariate distributions are not investigated here. Instead, we consider the bivariate distributions for comparison. We use the skewness measure $\beta_{1,2}$ introduced by Mardia (1970) for such comparisons. We choose the bivariate densities where each component has zero mean and unit variance. We achieve this by linear transformation, which does not change the skewness measure $\beta_{1,2}$, i.e., the $\beta_{1,2}$ for the original y variables is the same as the $\beta_{1,2}$ for the linearly transformed variables, z in the following discussion.

We consider the following simpler form of (7),

$$f(\mathbf{y} \mid \delta) = \frac{4}{1 + \delta^2} \phi\left(\frac{y_1}{\sqrt{1 + \delta^2}}\right) \phi\left(\frac{y_2}{\sqrt{1 + \delta^2}}\right) \Phi\left(\delta \frac{y_1}{\sqrt{1 + \delta^2}}\right) \Phi\left(\delta \frac{y_2}{\sqrt{1 + \delta^2}}\right). \quad (9)$$

The two components are independent and the marginal distributions are identical with mean

$$\mu(\delta) = \delta \sqrt{\frac{2}{\pi}} \quad \text{and variance} \quad \sigma^2(\delta) = 1 + \left(1 - \frac{2}{\pi}\right) \delta^2.$$

Mardia's skewness measure is given by

$$\beta_{1,2}(\delta) = 4(4 - \pi)^2 \left\{ \frac{\delta^2}{\pi + \delta^2(\pi - 2)} \right\}^3.$$

We use the standardization transformation

$$z_i = \frac{y_i - \mu(\delta)}{\sigma(\delta)}, \quad i = 1, 2,$$

and the density of $\mathbf{Z} = (Z_1, Z_2)$ for comparison purposes.

A version of the bivariate skew-normal distribution obtained by Azzalini & Dalla Valle (1996) is the following:

$$f(\mathbf{y} \mid \alpha) = 2\phi(y_1)\phi(y_2)\Phi\left\{\frac{\alpha}{\sqrt{1 - 2\alpha^2}}(y_1 + y_2)\right\}, \quad (10)$$

where α is the skewness parameter. This distribution also provides identical marginal distributions each with mean

$$\mu(\alpha) = \alpha \sqrt{\frac{2}{\pi}} \quad \text{and variance} \quad \sigma^2(\alpha) = 1 - \frac{2}{\pi} \alpha^2.$$

The correlation coefficient between the two components is given by

$$\rho(\alpha) = -\frac{2\alpha^2}{(\pi - 2\alpha^2)}.$$

The skewness measure $\beta_{1,2}$ is given by

$$\beta_{1,2}(\alpha) = 16(4 - \pi)^2 \left(\frac{\alpha^2}{\pi - 4\alpha^2} \right)^3.$$

When $\delta = \alpha$, we have $\mu(\delta) = \mu(\alpha)$, but the variances $\sigma^2(\delta)$ and $\sigma^2(\alpha)$ fail to coincide. A fair graphical comparison between two densities is not possible if they have unequal variances

because the variance also affects the tail of the density. We apply the above standardization transformation so that components have zero mean and unit variance, i.e., we set

$$z_i = \frac{y_i - \mu(\alpha)}{\sigma(\alpha)}, \quad i = 1, 2.$$

This transformation does not remove the correlation $\rho(\alpha)$ between the components. The Editor and an anonymous Associate Editor have pointed out that the densities of the standardized random variables corresponding to (9) and (10) are still not comparable because of the presence of correlation $\rho(\alpha)$ in the latter. That is why we also compare the linearly orthogonalized version of (10) using the following transformation:

$$z_1 = \frac{y_1 - \mu(\alpha)}{\sigma(\alpha)}, \quad z_2 = \frac{1}{\sqrt{1 - \rho^2(\alpha)}} \left\{ \frac{y_2 - \mu(\alpha)}{\sigma(\alpha)} - \rho(\alpha) \frac{y_1 - \mu(\alpha)}{\sigma(\alpha)} \right\}.$$

This linear transformation does not change the skewness measure $\beta_{1,2}$; see Mardia (1970).

In Figure 1, we first plot the density (9) linearly transformed to have zero mean and unit variance for each component for $\delta = 1, 3, 5$ and 10 . The corresponding values of the skewness measure $\beta_{1,2}(\delta)$ are $0.04, 0.89, 1.45$ and 1.82 . Clearly, the bivariate distribution gets more right skewed as δ increases.

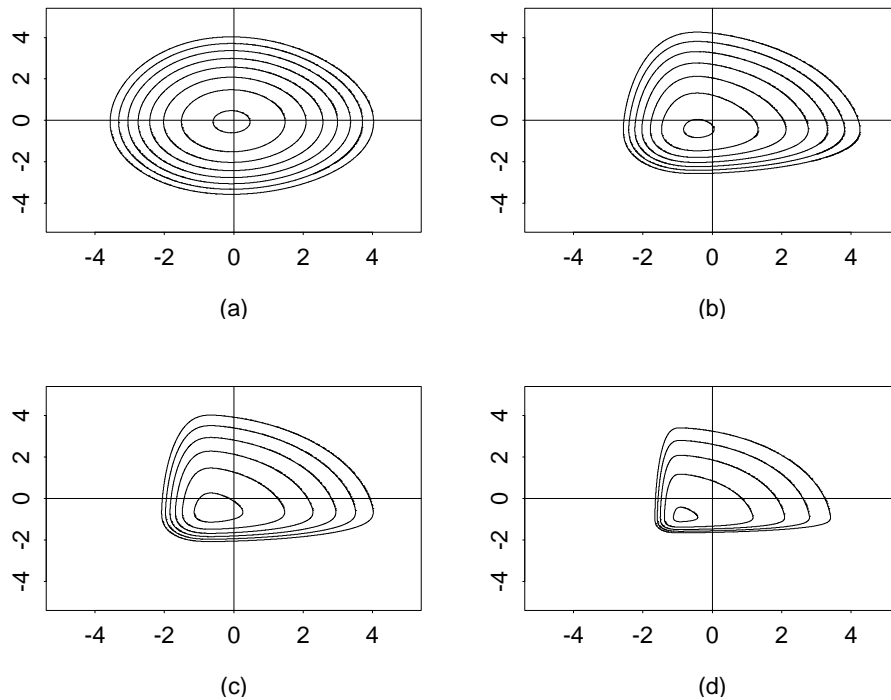


FIGURE 1: Contour plots of bivariate skew-normal distribution (9) linearly transformed so that components have zero mean and unit variance. In panel (a): $\delta = 1, \beta_{1,2} = 0.04$; in (b): $\delta = 3, \beta_{1,2} = 0.89$; in (c): $\delta = 5, \beta_{1,2} = 1.45$; in (d): $\delta = 10, \beta_{1,2} = 1.83$.

It now remains to compare the shape of (9) with that of the skew-normal density (10) of Azzalini & Dalla Valle (1996). We plot the densities of the transformed random variables in Figure 2 for $\beta_{1,2}(\delta) = \beta_{1,2}(\alpha) = 0.72$ and 0.98 . We obtained these values by first choosing $\alpha^2 = 0.48$ and 0.4995 . Note that the constraint $0 < \alpha^2 < 0.5$ is required by (10).

The two plots in the first column correspond to the skew-normal density (9) and the plots in the second and third column correspond to (10). The second column plots the standardized density without orthogonalization and the plots in the third column are for the orthogonalized version. The implied values of δ and α are labeled in the plot. The differences in the plots are explained by the fact that the density (10) results from conditioning on *one* random variable, while (9) results from conditioning on *two* random variables. The two conditioning random variables in (9) allow two lines to effectively bound the left tail of the bivariate distribution, while the only conditioning random variable limits the left tail of (10) by using only one line.

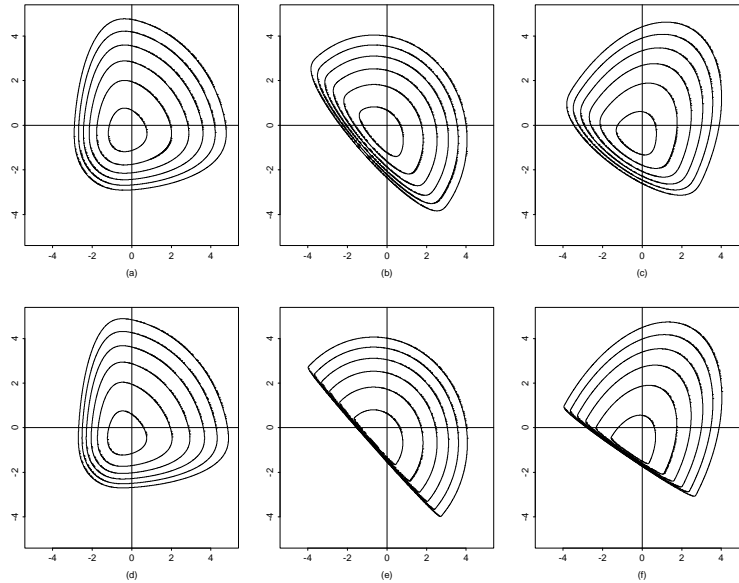


FIGURE 2: Contour plots of standardized versions bivariate skew-normal distributions. The first column corresponds to the density (9). The last two columns correspond to the density (10). The two plots in the second column are for the nonorthogonalized version while the two plots in the last column are for the orthogonalized version. In panel (a): $\delta = 2.6085, \beta_{1,2} = 0.72$; in (b) and (c): $\alpha = 0.6928, \beta_{1,2} = 0.72$; in (d): $\delta = 3.2321, \beta_{1,2} = 0.98$; in (e) and (f): $\alpha = 0.7068, \beta_{1,2} = 0.98$.

4. THE SKEW- t DISTRIBUTION

4.1. Density.

Let

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

Note that the two ingredients of (4) require a marginal and a cumulative conditional density. For an m -dimensional marginal density, we have

$$g^{(m)}(u) = \frac{\Gamma(m/2)}{\pi^{m/2}} \frac{g(u; m, \nu)}{\int_0^\infty r^{m/2-1} g(r; m, \nu) dr},$$

following Theorem 3.7 in Fang, Kotz & Ng (1990, p. 83). Therefore, in (4) the marginal density is $t_{m,\nu}(\mathbf{y} \mid \mu, \Sigma + D^2)$. For the cumulative conditional density we first obtain $g_a^{(m)}$. From (5) and ingredients in Lemma A.1, we have

$$g_a^{(m)}(u; \nu) = \Gamma(m/2)\pi^{-m/2}g(a+u; 2m, \nu) \left\{ \int_0^\infty r^{m/2-1}g(a+r; 2m, \nu) dr \right\}^{-1}$$

$$= \Gamma\left(\frac{m}{2}\right)\{\pi(\nu+m)\}^{-m/2}\left(\frac{\nu+m}{\nu+a}\right)^m\left(1+\frac{u}{\nu+m}\frac{\nu+m}{\nu+a}\right)^{-(\nu+2m)/2}$$

Hence, the conditional density is given by

$$t_{m,\nu+m}\left[z \mid D(\Sigma+D^2)^{-1}\mathbf{y}_*, \frac{\nu+q(\mathbf{y}_*)}{\nu+m}\{I-D(\Sigma+D^2)^{-1}D\}\right]$$

The generator function calculation gave two extra quantities that were not present in the multivariate skew-normal distribution, namely, (i) the degrees of freedom of the conditional density is $\nu+m$; and (ii) the factor $(\nu+q(\mathbf{y}_*)) / (\nu+m)$ in the scale parameter.

Summarizing the preceding discussion, we have the density of the multivariate skew- t distribution given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma, D, \nu) = 2^m t_{m,\nu}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma+D^2)P(\mathbf{V} > \mathbf{0}), \tag{11}$$

where \mathbf{V} follows the t -distribution $t_{m,\nu+m}$. We denote this distribution by $ST_\nu(\boldsymbol{\mu}, \Sigma, D)$. For $\Sigma = \sigma^2 I$ and $D = \delta I$, the above simplifies to

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\mu}, \sigma^2, \delta, \nu) &= 2^m (\sigma^2 + \delta^2)^{-m/2} \frac{\Gamma((\nu+m)/2)}{\Gamma(\nu/2)(\nu\pi)^{m/2}} \left\{ 1 + \frac{\mathbf{y}'_* \mathbf{y}_*}{\nu(\sigma^2 + \delta^2)} \right\}^{-(\nu+m)/2} \\ &\times T_{m,\nu+m} \left[\left\{ \frac{\nu+q(\mathbf{y}_*)}{\nu+m} \right\}^{-1/2} \frac{\delta}{\sigma \sqrt{\sigma^2 + \delta^2}} \mathbf{y}_* \right], \end{aligned}$$

where $T_{m,\nu+m}(\cdot)$ denotes the cumulative density function of $t_{m,\nu+m}(\mathbf{0}, I)$. However, unlike in the skew-normal case, the above density cannot be written as the product of univariate skew- t densities. Here the Y_i are dependent but uncorrelated.

4.2. Moments and skewness.

The moments of the skew- t distribution $ST_\nu(\boldsymbol{\mu}, \Sigma, D)$ are not straightforward to obtain with the density (11). Here we derive the first two moments by viewing $ST_\nu(\boldsymbol{\mu}, \Sigma, D)$ as a scale mixture of $SN(\boldsymbol{\mu}, \Sigma, D)$. We obtain the following results by using the expression for the moment generating function given in the Appendix.

The mean and variance of the skew- t distribution $ST_\nu(\boldsymbol{\mu}, \Sigma, D)$ are given by

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)} \boldsymbol{\delta}$$

and

$$\text{cov}(\mathbf{Y}) = (\Sigma + D^2) \frac{\nu}{\nu-2} - \frac{\nu}{\pi} \left[\frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)} \right]^2 D^2$$

when $\nu > 2$.

We have calculated the multivariate skewness measure $\beta_{1,m}$ (Mardia 1970) in analytic form for the skew- t distribution. The expression does not simplify and involves nonlinear interactions between the degrees of freedom (ν) and the skewness parameter δ where $D = \delta I$. However, $\beta_{1,m}$ approaches ± 1 as $\delta \rightarrow \pm\infty$.

5. REGRESSION MODELS WITH SKEWNESS

5.1. Models for univariate response.

We consider the regression model where the error distribution follows the skew elliptical distribution. Let X be an $n \times p$ design matrix (with full column rank) and β be a (p -variate) vector of regression parameters.

Suppose that we have n independent observed one-dimensional response variables y_i . Further $y_i \sim \text{SE}(\mu_i, \sigma^2, \delta; g^{(1)})$ independently. Let $\mu = (\mu_1, \dots, \mu_n)'$. For the regression model, we assume that $\mu_i = \mathbf{x}_i' \beta$, where \mathbf{x}_i' denotes the i th row of the matrix X . Thus the assumed regression model is $\mu = X\beta$. The likelihood function of β, σ^2 and δ and any other parameter involved in $\text{SE}(\mu_i, \sigma^2, \delta; g^{(1)})$ is given by the product of densities of the form (6). Hence we write

$$L(\beta, \sigma^2, \delta, g^{(1)}; y_1, \dots, y_n) = \prod_{i=1}^n f(y_i | \mu_i, \sigma^2, \delta; g^{(1)}),$$

where $f(y | \mu, \sigma^2, \delta; g^{(1)})$ is given in (6). The above likelihood may also depend on additional parameters. For example, for the t distributions, the additional parameter is ν .

Often the error distribution in a regression model is taken to have mean zero. The regression model developed here can be forced to satisfy this requirement when the intercept parameter is suitably adjusted. See Section 6.2 for particulars.

To completely specify the Bayesian model, we need to specify prior distributions for all the parameters. As a default prior for β , we take the constant prior $\pi_\beta \propto 1$ in \mathbb{R}^p . For $\tau = 1/\sigma^2$ we assume a gamma prior distribution $\Gamma(\kappa, \kappa)$, where the parameterization has mean 1 and κ is assumed to be a known parameter. In other words, σ^2 is given an inverse gamma distribution. The parameter δ is given a normal prior distribution. Specific parameter values of this prior distribution will be discussed in particular examples. When the t models are considered, we need a prior distribution for the degrees of freedom parameter ν . For this, we use the exponential distribution with parameter 0.1 truncated in the region $\nu > 2$, so that the underlying t distribution (skew or not) has finite mean and variance.

Now the joint posterior density is given by

$$\pi(\beta, \sigma^2, \delta, \nu | y_1, \dots, y_n) \propto L(\beta, \sigma^2, \delta, g^{(1)}; y_1, \dots, y_n) \pi_\beta \pi_{\sigma^2} \pi_\delta \pi_\nu, \quad (12)$$

where π on the right-hand side denotes the prior density of its argument. Note that the parameter ν is omitted for the normal distributions.

In many actual examples, it is possible to decide a priori the type of skewed distributions that are appropriate for data. For example, either the positively skewed or the negatively skewed distributions may be considered to be appropriate for data. Thus it is reasonable to assume proper prior distributions for the skewness parameter.

In many examples, however, we may not have precise information about β and σ^2 and we will be required to use the default prior distributions, as is often done in practice. A natural question in such a case is whether the full posterior distribution is proper. In the following theorem, we answer this question in the affirmative for the skew-normal and the skew- t error distributions.

THEOREM 2. *Suppose that π_δ and π_ν are proper distributions and $\pi_\beta \propto 1$. Then the posterior (12) is proper under the skew normal or skew- t model if $n > p$.*

In the Appendix, we provide a proof of this theorem. In fact a more general theorem is proved and the above result is obtained under the special cases of normal and t distributions. As a consequence of the proof of Theorem 2, we have the following result on the existence of the posterior moments of σ^2 .

THEOREM 3. *Suppose that π_δ and π_ν are proper distributions and $\pi_\beta \propto 1$. Then $E\{(\sigma^2)^k | \mathbf{y}\}$ exists under the skew-normal or skew- t model if $n - p > 2k$.*

A similar result is obtained by Geweke (1993) for the t model with unequal variance assumption. Theorem 3 extends his result to the skewed models which include several other distributions.

5.2. MCMC specification.

In order to specify model (3) for MCMC computation, we use the hierarchical setup of $f(\mathbf{y} | \mathbf{z})$ and $f(\mathbf{z})I(\mathbf{z} > \mathbf{0})$, where f is a generic notation denoting the density of the random variable in the argument. We obtain these two distributions from (A.1) in the Appendix,

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \sim \text{El}\left(\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \Omega = \begin{pmatrix} \Sigma + D^2 & D \\ D & I \end{pmatrix}; g^{(2m)}\right).$$

Here we have

$$\mathbf{Y} | \mathbf{Z} = \mathbf{z} \sim \text{El}(\boldsymbol{\mu} + D\mathbf{z}, \Sigma; g_{q(\mathbf{z})}^{(m)})$$

where $q(\mathbf{z}) = \mathbf{z}'\mathbf{z}$. For the skew-normal model, this is simply a multivariate normal distribution with mean $\boldsymbol{\mu} + D\mathbf{z}$ and covariance matrix Σ , since $g_{q(\mathbf{z})}^{(m)}$ is independent of $q(\mathbf{z})$. However, for the skew- t model, this is not so, and

$$\mathbf{Y} | \mathbf{Z} = \mathbf{z} \sim t_{m, \nu+m}\left(\boldsymbol{\mu} + D\mathbf{z}, \frac{\nu + \mathbf{z}'\mathbf{z}}{\nu + m}\Sigma\right).$$

The marginal specification for \mathbf{Z} for the skew-normal case is simply the $N_m(\mathbf{0}, I)$ distribution. The same for the skew- t case is $t_{m, \nu}(\mathbf{0}, I)$. Lastly, the distribution of \mathbf{Z} is truncated in the space $\mathbf{z} > \mathbf{0}$.

5.3. Multivariate response.

Regression models for multivariate response variables are constructed as follows. Let $\mathbf{Y}_i \sim \text{SE}(\boldsymbol{\mu}_i, \Sigma, D; g^{(m)})$ for $i = 1, \dots, n$. For each data point with covariate information assumed in a $p \times m$ matrix X_i , we can specify the linear model

$$\boldsymbol{\mu}_i = X_i'\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a p -vector of regression coefficients. The coefficients are given a multivariate normal $N_p(\boldsymbol{\beta}_0, \Lambda)$ prior distribution, where Λ is a known positive definite matrix and $\boldsymbol{\beta}_0$ is a vector of constants to be chosen later. The matrix Σ is assigned an independent conjugate Wishart prior distribution

$$\Sigma^{-1} = Q \sim W_m(2r, 2\kappa),$$

where $2r$ is the assumed prior degrees of freedom ($\geq m$) and κ is a positive definite matrix. We say that \mathbf{X} has the Wishart distribution $W_m(k, A)$ if its density is proportional to

$$|A|^{k/2} |y|^{1/2(k-p-1)} e^{-1/2\text{tr}(Ay)}, \tag{13}$$

if \mathbf{x} is an $m \times m$ positive definite matrix. [Here, $\text{tr}(A)$ is the trace of a matrix A .] This is the parameterization used by, for example, the BUGS software (Spiegelhalter, Thomas & Best 1996). The skewness parameters in D , vectorized as $\boldsymbol{\delta}$, are given a normal prior distribution $N_m(\mathbf{0}, \Gamma)$, where Γ is a positive definite matrix.

In the remainder of this section, we develop a computational procedure for the multivariate skew- t distribution and obtain the methods for the multivariate skew-normal as a special case. The full likelihood specification is given as follows. We introduce n i.i.d. random variables w_i

for each data point to obtain the t models. For the normal distributions, each of these will be set at 1.

$$\begin{aligned} \mathbf{Y}_i | \mathbf{z}_i, \boldsymbol{\beta}, X_i, \Sigma, D, w_i &\sim N_m \left(X_i' \boldsymbol{\beta} + D \mathbf{z}_i, \frac{\Sigma}{w_i} \right), \\ \mathbf{Z}_i &\sim N_m(\mathbf{0}, I) I(\mathbf{z} > \mathbf{0}), \quad \boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \Lambda), \\ Q = \Sigma^{-1} &\sim W_m(2r, 2\kappa), \quad \boldsymbol{\delta} \sim N_m(\mathbf{0}, \Gamma), \\ w_i &\sim \Gamma(\nu/2, \nu/2), \quad \nu \sim \Gamma(1, 0.1) I(\nu > 2). \end{aligned}$$

The last two distributional specifications are omitted in the normal distribution case. All of the full conditional distributions for Gibbs sampling are straightforward to derive and sample from except for \mathbf{z}_i and $\boldsymbol{\delta}$. Their full conditional distributions are given by

$$\mathbf{Z}_i | \dots \sim N_m(A_i^{-1} \mathbf{a}_i, A_i^{-1}) I(\mathbf{z}_i > \mathbf{0}), \quad \boldsymbol{\delta} | \dots \sim N_m(B^{-1} \mathbf{b}, B^{-1})$$

where

$$\begin{aligned} A_i &= I + w_i D Q D \quad \text{and} \quad \mathbf{a}_i = w_i D Q (\mathbf{y}_i - X_i' \boldsymbol{\beta}), \\ B &= \Gamma^{-1} + \sum_{i=1}^n \text{diag}(\mathbf{z}_i) Q \text{diag}(\mathbf{z}_i) \quad \text{and} \quad \mathbf{b} = \sum_{i=1}^n \text{diag}(\mathbf{z}_i) Q (\mathbf{y}_i - X_i' \boldsymbol{\beta}), \end{aligned}$$

where $\text{diag}(\mathbf{a})$ is a diagonal matrix with diagonal elements being the components of \mathbf{a} .

6. EXAMPLES

6.1. Interview data.

In order to gain admissions to a certain medical school, the applicants are screened for both academic qualifications and nonacademic characteristics. Each applicant meeting some observed and some predicted academic criteria receives a nonacademic total which is the sum of seven scores. These seven scores are assigned on the basis of work experience, sense of responsibility, commitment and caring, motivation, study skills, interest and referees' comments. Applicants are subsequently selected for interviews based on their nonacademic totals. The interviewed applicants are given scores, which are the sums of two individual scores given by each member of a two-member interview committee.

In our univariate skewed regression model setup, the nonacademic totals are considered to be realizations of the response variable. Here, the academic scores from the final qualifying examination (called the *A-level* examination in Great Britain) of the applicants work as the unobserved conditioning variables leading to our regression model. The true academic scores of applicants are yet unobserved because the admission process takes place before the applicants sit for the *A-level* examinations in Great Britain.

The data set to be analyzed here is obtained as part of a large cohort data set giving the details of candidates who have applied for a medical degree from a certain school in Great Britain. For the univariate analysis, we have the nonacademic totals of 584 applicants categorized by race and sex. We work with a larger data set for our bivariate analysis. The response variables are the nonacademic total and a composite score in a secondary science examination for each of 731 applicants. The 584 applicants for the univariate analysis are the applicants who were called for interviews among the 731 initial applicants.

6.2. Univariate regression.

The response variable nonacademic total is influenced by several academic, socio-economic and demographic factors, as expected. In our current study, we only consider the influence of race and gender of the applicants; these characteristics of the applicants were known to their evaluators. Although the applicants are classified as having come from six combined ethnic types (namely

white, black, Indian, Pakistani and Bangladeshi, other Asian and others), for our purposes we classify candidates as to whether they were white or nonwhite. We are then interested to compare four groups of applicants: white female, white male, nonwhite female and nonwhite male. The first group has higher average nonacademic totals than the other groups. Simple t tests on the data also show significant differences between the groups.

The data are not expected to be heavily skewed since the individual data points are sums of components as mentioned previously. However, the observations are sums of only seven components, so the central limit theorem does not ensue for such a small sample size. Initial exploratory plots (not shown) confirm that the left tail of the underlying distribution descends more slowly than the right tail. Our explicit skew-regression models will estimate and test for the skewness in the data more formally.

Let y_i denote the nonacademic total of the i th applicant for $i = 1, \dots, n = 584$. In order to compare among the four groups, white female, white male, nonwhite female and nonwhite male, we code three binary regressors taking the values 0 and 1 as described below. The first regressor takes the value 1 for white male, the second takes the value 1 for nonwhite female, while the third takes the value 1 for nonwhite male. The resulting regression coefficients allow comparison of the last three groups, with the white female as the base group. Thus we have the regression model

$$y_i = \alpha' + \sum_{j=1}^3 \beta_j x_{ij} + \delta z_i + \varepsilon_i, \quad i = 1, \dots, n.$$

We calculate the true intercept $\alpha = \alpha' - \delta E(z_i)$ which corresponds to the regression model where the error distribution has mean zero. For the normal and t models, expressions for $E(z_i)$ are given in the preceding sections.

We assume throughout independent diffuse prior distributions $N(0, 10^4)$ for the regression parameters α' and β_j . For $\tau = 1/\sigma^2$, we assume a limiting noninformative gamma prior distribution $\Gamma(0.01, 0.01)$, where the parameterization has mean 1. When δ is not assumed to be zero, it is given a normal prior distribution with mean zero and variance 100. Thus δ is assigned a proper prior distribution, which is a requirement of Theorems 2 and 3.

1. Normal linear model: We take $\delta = 0$ and $\varepsilon \sim N(0, \sigma^2)$.
2. Skew normal model: We assume that z_i , given all the other random quantities in the model, follows the standard half-normal distribution.
3. t -model: We assume that $\varepsilon \sim t_\nu$, where t_ν is the standard t -distribution with ν degrees of freedom. We further assume $\delta = 0$. Although the parameter ν is traditionally taken as an integer, it can be treated as a continuous parameter with positive values since the associated densities are well defined in this case. We assume a truncated ($\nu > 2$) exponential distribution with parameter 0.1. The truncation assures the finiteness of the mean and variance of the associated t error distribution.
4. Skew t -model: We assume that $\varepsilon \sim t_\nu$, where t_ν is the t -distribution with ν degrees of freedom. We also assume that z_i follows i.i.d. $|t_\nu|$, conditional on other random quantities in the model. The prior distributions for the remaining parameters are assumed to be the same as in the previous cases.

The Gibbs sampler has been implemented using the BUGS software; the codes are available from the authors upon request. We use the 10,000 iterates after discarding the first 5000 iterates to make inference. The regression model has an intercept α' and three regression parameters: β_1 for white male applicants, β_2 for nonwhite female applicants and β_3 for the nonwhite male applicants. The resulting parameter estimates (posterior means) are given in Table 1.

TABLE 1: Posterior mean, standard deviation (sd) and 95% probability intervals for the parameters under the four fitted models in the univariate example.

	Normal model				Skew-normal model			
	Mean	sd	2.5%	97.5%	Mean	sd	2.5%	97.5%
α	26.18	0.14	25.90	26.47	30.40	0.62	28.96	31.44
β_1	-0.65	0.24	-1.13	-0.18	-0.63	0.24	-1.10	-0.15
β_2	-0.87	0.36	-1.58	-0.17	-0.89	0.35	-1.58	-0.19
β_3	-0.98	0.39	-1.74	-0.21	-1.11	0.39	-1.86	-0.33
σ^2	6.46	0.38	5.75	7.25	3.88	0.65	2.77	5.34
δ					-2.64	0.38	-3.28	-1.77
	Student t model				Skew- t model			
	Mean	sd	2.5%	97.5%	Mean	sd	2.5%	97.5%
α	26.21	0.14	25.93	26.49	26.47	2.75	23.11	30.80
β_1	-0.57	0.24	-1.04	-0.10	-0.56	0.24	-1.04	-0.07
β_2	-0.89	0.35	-1.58	-0.19	-0.89	0.35	-1.57	-0.21
β_3	-0.97	0.37	-1.70	-0.25	-1.00	0.38	-1.72	-0.23
σ^2	5.39	0.44	4.54	6.29	4.10	0.98	2.28	5.72
ν	13.85	5.61	6.71	28.74	9.99	3.95	5.17	20.69
δ					-0.16	1.58	-2.70	1.76

The estimates of the regression parameters across the models agree broadly. All three regression parameters β_j are significant in all the models since the associated 95% probability intervals do not include the value zero. The negative estimates show that the base group of white female receives significantly higher average nonacademic totals than the remaining three groups. The difference between the white female and nonwhite male is the most significant. Thus the latter group seems to have performed most poorly nonacademically, even though they met all the academic criteria.

The estimates of the parameter σ^2 are smaller for the corresponding skewed model. This is expected since high variability, heaviness of the tails and skewness are interchangeable to a certain extent. The nonskewed symmetric error models endeavour to capture skewness by having larger tails. The important question is whether high variability can completely replace skewness. In the next paragraph, we answer this negatively.

The skewness parameter δ is estimated to be negative in both the skew-normal and skew- t model; this confirms the left skewness of the response mentioned previously. Moreover, δ is significant under the skew-normal model since the 95% probability interval is $(-3.28, -1.77)$. Thus we can conclude that significant skewness is required to model the data.

The parameter δ does not turn out to be significant under the skew- t models. This is explained as follows. Observe that the fitted symmetric t -error distribution is lighter tailed (estimated $df = 13.85$) with larger dispersion parameter σ^2 than the fitted skew- t model (estimated $df = 9.99$). With such heavy tailed error distribution, it was not possible to see significant skewness in the data. This, however, does not necessarily reduce the predictive power of the skew- t model.

To compare the four models informally, we compute the effective number of parameters ρ_D and the deviance information criterion (DIC) as presented by Spiegelhalter, Best, Carlin & van der Linde (2002). They claim that the (DIC) as implemented in the BUGS software can be used to compare complex models and large differences in a criterion can be attributed to real predictive differences in the models, although these claims have been questioned. Using the DIC values

shown in Table 2, we see that the skewed models improve the corresponding symmetric models; the symmetric normal and t models are very similar; the skew- t model is the best model for the data. For the symmetric normal and t models the effective number of parameters ρ_D roughly indicates the number of parameters in the regression model. Spiegelhalter, Best, Carlin & van der Linde (2002) mention that ρ_D can be negative for nonlog-concave densities, the present example with skewed distributions provides a case in point. Thus ρ_D is not meaningful in our example.

TABLE 2: The effective number of parameters, p_D and DIC for the four fitted models.

	p_D	DIC
Normal	4.9	2750.6
Skew-normal	217.8	2658.1
Student t	5.6	2742.1
Skew- t	-187.5	2387.4

The same conclusions (that the skewed models are better and the skew- t model is the best) are also arrived at using more formal Bayesian predictive model choice criteria, e.g., the Bayes factors (DiCiccio, Kass, Raftery & Wasserman 1997). We, however, omit the details. Instead, we compare the residuals from the symmetric and the corresponding skew models to examine if indeed the skew models were able to improve upon the symmetric models. In Figure 3, we plot kernel density estimates of the standardized residuals with the same smoothing parameter. Clearly the density plots for the skewed models have thinner tails than the corresponding symmetric models. We also provide normal Q-Q plots of the residuals to examine the four fitted distributions in Figure 4. All four plots show the existence of outliers, but the skew- t model is seen to be the best fitted model. This confirms the model choice and diagnostic results based on the DIC criterion and the kernel density plot provided in Figure 3.

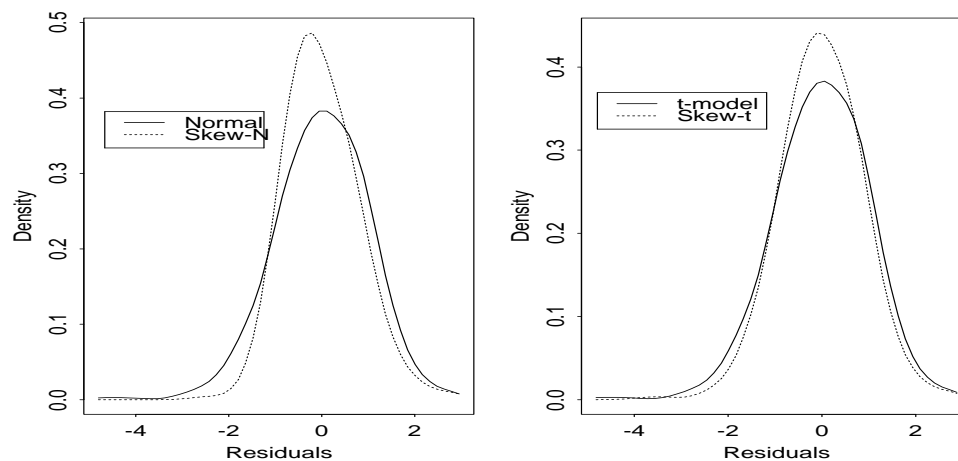


FIGURE 3: Kernel density estimates of the residuals under four regression models.

6.3. A multivariate illustration.

The nonacademic totals and the scores in science from the secondary examination of 731 candidates are plotted in Figure 5. From the plot, it is clear that symmetric distributions should not be fitted to this data set. We proceed with the multivariate models of Section 5.3. We adopt

the following values of the hyperparameters. Let ξ be the two component vector where each element is the mid-point of the corresponding component of the bivariate data. Further, let R denote the diagonal matrix where each diagonal entry is the squared range of the corresponding component in the data. Since we do not consider any covariate for this example, the regression parameter β is the mean parameter μ . For this, we assume a normal prior distribution with mean $\beta_0 = \xi$ and covariance matrix $\Lambda = 100 \times R$. The degrees of freedom parameter $2r$ in the Wishart distribution is set at 3 which corresponds to the noninformative prior distribution; see the Wishart density in (13). The matrix κ in the Wishart distribution is taken as $100/(2r)R^{-1}$. Finally, each component of δ is given an independent normal prior distribution with mean zero and variance 100.

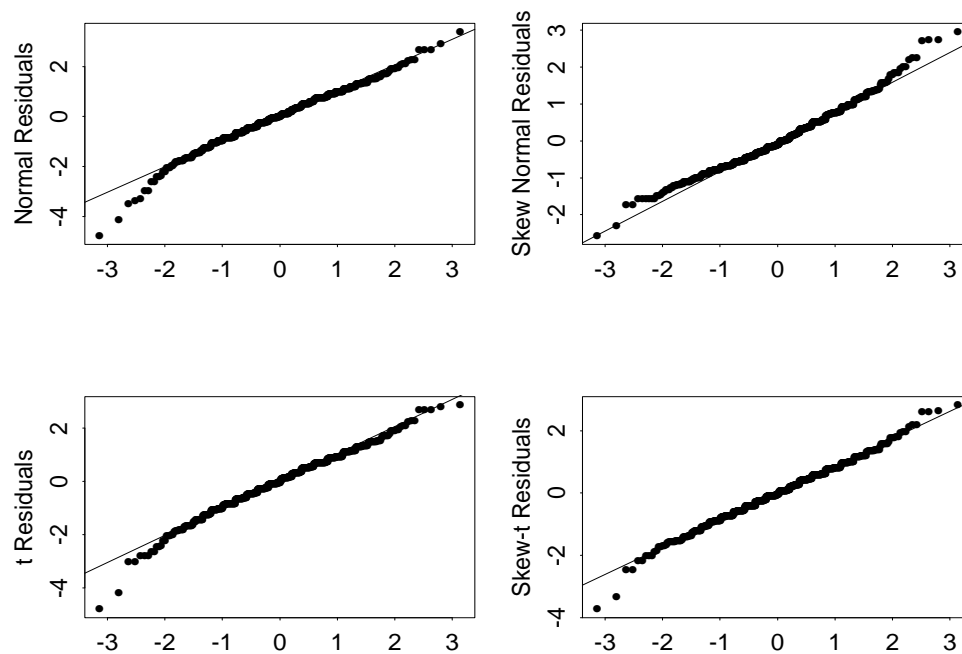


FIGURE 4: Normal Q-Q plots of residuals.

The estimates of the marginal likelihood using the approach of Gelfand & Dey (1994) are -3502.1 , -3439.7 , -3493.9 and -3443.5 for the normal, skew-normal, t and the skew- t model, respectively. The skewed bivariate models are large improvements over the corresponding symmetric models. However, the data favour the skew-normal model when compared with the skew- t model. Other model comparison criteria can also be used. We, however, use the Gelfand & Dey (1994) method for this multivariate example since it is reasonably easy to implement and provides a quick comparison between competing models.

7. CONCLUSION

The new class of skewed distributions obtained in this article is very general, quite flexible and widely applicable. The skewed distributions are shown to provide an alternative to symmetric distributions that are often assumed in regression. Although the associated density functions are quite difficult to handle, we show that the models can be easily fitted using MCMC methods. Moreover, the univariate models are fitted using the software BUGS that is available publicly.

This makes our approach quite powerful and accessible to practicing statisticians. Other variants of skewed distributions currently available are not so easy to implement.

In this article, we obtain the skewed distributions by transformation and then conditioning on the same number m of random variables, as in Theorem 1. As mentioned in the Introduction, Azzalini & Capitanio (1999) condition on one random variable being positive. It is certainly possible to impose the nonnegativity condition on any other number of random variables, although we have not pursued this.

Observe that the exact form of the densities of skewed distributions obtained in Theorem 1 need not be calculated if the sole purpose is to perform model fitting. However, model comparison using the Bayes factors can be easily performed if it was possible to calculate the density. The augmented variables used in model fitting can be ignored when calculating the marginal likelihood since the marginal density of the data is available analytically.

Although we have not discussed this option, the Bayes factors can be used to solve the associated problems of variable selection. Moreover, other existing Bayesian techniques of variable selection and model averaging can be implemented with the models developed here.

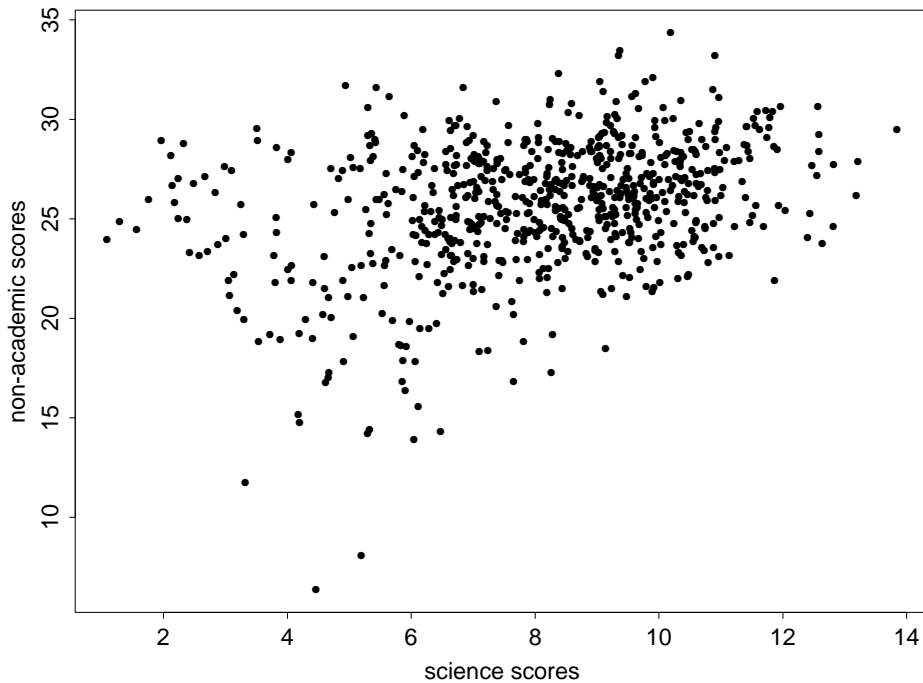


FIGURE 5: Scatter plot of the bivariate data used in model fitting.

APPENDIX: PROOFS OF THE THEOREMS

Before proving Theorem 1, we consider the following lemma, which is Theorem 2.18 in Fang, Kotz & Ng (1990, p. 45). Partition \mathbf{X} , $\boldsymbol{\theta}$, Ω into

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}^{(1)} \\ \boldsymbol{\theta}^{(2)} \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

where $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are respectively k_1 - and k_2 -dimensional random vectors, and $k_1 + k_2 = k$. The parameters $\boldsymbol{\theta}$ and Ω are partitioned accordingly.

LEMMA A.1. Let $\mathbf{X} \sim \text{El}(\boldsymbol{\theta}, \Omega; g^{(k)})$. Then

$$\mathbf{X}^{(1)} \mid \mathbf{X}^{(2)} = \mathbf{x}^{(2)} \sim \text{El}(\boldsymbol{\theta}_{1.2}, \Omega_{11.2}; g_{q(\mathbf{x}^{(2)})}^{(k_1)}),$$

where

$$\begin{aligned} \boldsymbol{\theta}_{1.2} &= \boldsymbol{\theta}^{(1)} + \Omega_{12}\Omega_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)}), \quad \Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}, \\ q(\mathbf{x}^{(2)}) &= (\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)})'\Omega_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)}), \quad g_a^{(k_1)}(u) = \frac{\Gamma(k_1/2)}{\pi^{k_1/2}} \frac{g(a+u; k)}{\int_0^\infty r^{k_1/2-1} g(a+r; k) dr}. \end{aligned}$$

Proof of Theorem 1. Consider the transformation

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} I & D \\ 0 & I \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{Z} \end{pmatrix},$$

where $\mathbf{0}$ is the null matrix. Using Theorem 2.16 of Fang, Kotz & Ng (1990), we see that

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{V} \end{pmatrix} \sim \text{El}\left(\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \Omega = \begin{pmatrix} \Sigma + D^2 & D \\ D & I \end{pmatrix}; g^{(2m)}\right). \quad (\text{A.1})$$

From this joint distribution, we aim to obtain the conditional density of $\mathbf{Y} \mid \mathbf{V} > \mathbf{0}$. Since $\mathbf{V} \sim \text{El}(\mathbf{0}, I; g^{(m)})$ marginally, we have $P(\mathbf{V} > \mathbf{0}) = 2^{-m}$. By standard arguments,

$$f(\mathbf{y} \mid \mathbf{V} > \mathbf{0}) = 2^m f_{\mathbf{Y}}(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma + D^2; g^{(m)}) P(\mathbf{V} > \mathbf{0} \mid \mathbf{y}).$$

In order to calculate $P(\mathbf{V} > \mathbf{0} \mid \mathbf{y})$, we obtain the conditional density of $\mathbf{V} \mid \mathbf{Y} = \mathbf{y}$ from the joint distribution (A.1). Using Lemma A.1, we have

$$\mathbf{V} \mid \mathbf{Y} = \mathbf{y} \sim \text{El}(\boldsymbol{\theta} = D(\Sigma + D^2)^{-1}\mathbf{y}_*, \Omega = I - D(\Sigma + D^2)^{-1}D; g_{q(\mathbf{y}_*)}^{(m)}),$$

where $q(\mathbf{y}_*) = \mathbf{y}_*'(\Sigma + D^2)^{-1}\mathbf{y}_*$. Hence the proof is complete. \square

LEMMA A.2. If $\mathbf{Y} \sim \text{SN}(\boldsymbol{\mu}, \Sigma, D)$, its moment generating function is of the form (8).

Proof of Lemma A.2. Note that $M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{t}) \exp(\mathbf{t}'\boldsymbol{\mu})$, where $\mathbf{X} \sim \text{SN}(\mathbf{0}, \Sigma, D)$. Let $Q = (\Sigma + D^2)^{-1}$ and $B = I - DQD$ for notational convenience. Also for conciseness we assume that B is diagonal in the following calculations. The general case follows similarly. Now

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= 2^m \int_{\mathbf{R}^m} |Q|^{1/2} (2\pi)^{-m/2} e^{-1/2\mathbf{x}'Q\mathbf{x} + \mathbf{t}'\mathbf{x}} \Phi_m(B^{-1/2}DQ\mathbf{x}) d\mathbf{x} \\ &= e^{\mathbf{t}'Q^{-1}\mathbf{t}/2} 2^m \int_{\mathbf{R}^m} |Q|^{1/2} (2\pi)^{-m/2} e^{-1/2(\mathbf{x}-Q^{-1}\mathbf{t})'Q(\mathbf{x}-Q^{-1}\mathbf{t})} \Phi_m(B^{-1/2}DQ\mathbf{x}) d\mathbf{x} \\ &= e^{\mathbf{t}'Q^{-1}\mathbf{t}/2} 2^m \int_{\mathbf{R}^m} |Q|^{1/2} (2\pi)^{-m/2} e^{-1/2\mathbf{z}'Q\mathbf{z}} \Phi_m\{B^{-1/2}DQ(\mathbf{z} + Q^{-1}\mathbf{t})\} d\mathbf{z} \\ &= e^{\mathbf{t}'Q^{-1}\mathbf{t}/2} 2^m \Phi_m(D\mathbf{t}) \end{aligned}$$

based on the following result.

PROPOSITION A.1. If $\mathbf{Z} \sim N_m(\mathbf{0}, \Sigma)$, then $E_{\mathbf{Z}}\{\Phi_m(\mathbf{a} + G\mathbf{Z})\} = \Phi_m\{(I + G\Sigma G')^{-1/2}\mathbf{a}\}$.

The following lemma obtains the moment generating function of scale mixture of skew-normal distribution.

LEMMA A.3. Let $\mathbf{X} \sim \text{SN}(\mathbf{0}, \Sigma, D)$ and $\mathbf{Y} = w^{-1/2}\mathbf{X}$, given $W = w$, where $W \sim \Gamma(\nu/2, \nu/2)$ and the parameterization has mean 1. Then the moment generating function of the marginal distribution of \mathbf{Y} is given by

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^m \int_0^\infty e^{\mathbf{t}'Q^{-1}\mathbf{t}/(2w)} \Phi_m(Dw^{-1/2}\mathbf{t}) dG(w),$$

where $G(w)$ denotes the cumulative distribution function of $\Gamma(\nu/2, \nu/2)$.

THEOREM A.1. Assume that

$$h(w, \mathbf{y}, X) = \int_{\mathbb{R}^p} \prod_{i=1}^n g^{(1)}\{(y_i - \mathbf{x}'_i\beta)^2/w\} d\pi_\beta \leq M(\mathbf{y}, X)w^{p/2},$$

where $M(\mathbf{y}, X)$ is a constant free of w, δ and ν . Also assume that π_δ and π_ν are proper distributions. Then the posterior (12) is proper under the skew normal or skew- t model if $n > p$.

LEMMA A.4. Under the skew-normal model, $h(w, \mathbf{y}, X) \leq M(\mathbf{y}, X)w^{p/2}$.

Proof of Lemma A.4. Here $g^{(1)}(u) = e^{-u/2}/\sqrt{2\pi}$. Hence $\prod_{i=1}^n g^{(1)}(u_i) = (2\pi)^{-n/2} e^{-\sum_{i=1}^n u_i/2}$. However, note that

$$\begin{aligned} \sum_{i=1}^n (y_i - \mathbf{x}'_i\beta)^2 &= (\mathbf{y} - X\beta)'(\mathbf{y} - X\beta) \\ &= (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) + \mathbf{y}'\{I - X(X'X)^{-1}X'\}\mathbf{y}, \end{aligned}$$

where $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$. Let $H = I - X(X'X)^{-1}X'$ and $Q = X'X$. Since we assume that X has full column rank, we have that Q is a positive definite matrix. Also H is idempotent (and hence a nonnegative definite) matrix. Let $\pi_\beta = 1$. Now we have

$$\begin{aligned} h(w, \mathbf{y}, X) &= \int_{\mathbb{R}^p} \prod_{i=1}^n g^{(1)}\{(y_i - \mathbf{x}'_i\beta)^2/w\} d\pi_\beta \\ &= \int_{\mathbb{R}^p} (2\pi)^{-n/2} \exp\{-(\beta - \hat{\beta})'Q(\beta - \hat{\beta})/2w + \mathbf{y}'H\mathbf{y}/2w\} d\beta \\ &= (2\pi)^{(p-n)/2} w^{p/2} |Q|^{-1/2} \exp(-\mathbf{y}'H\mathbf{y}/2w) \\ &= w^{p/2} \times M(\mathbf{y}, X), \end{aligned}$$

since $\mathbf{y}'H\mathbf{y}$ is a nonnegative definite quadratic form. Hence the proof is complete. □

LEMMA A.5. Under the skew- t model, $h(w, \mathbf{y}, X) \leq M(\mathbf{y}, X)w^{p/2}$.

Proof of Lemma A.5. Let

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

Here, $g^{(1)}(u; \nu) = J(1 + u/\nu)^{-(\nu+1)/2}$. Now

$$\begin{aligned} \prod_{i=1}^n g^{(1)}(u_i) &= J^n \left\{ \prod_{i=1}^n (1 + u_i/\nu) \right\}^{-(\nu+1)/2} \leq J^n \left(1 + \sum_{i=1}^n u_i/\nu \right)^{-(\nu+1)/2} \\ &= J^n \left\{ 1 + \sum_{i=1}^n (y_i - \mathbf{x}'_i\beta)^2/(\nu w) \right\}^{-(\nu+1)/2} \\ &= J^n \left\{ 1 + (\beta - \hat{\beta})'Q(\beta - \hat{\beta}) + \mathbf{y}'H\mathbf{y}/(\nu w) \right\}^{-(\nu+1)/2} \\ &\leq J^n \left\{ 1 + (\beta - \hat{\beta})'Q(\beta - \hat{\beta})/(\nu w) \right\}^{-(\nu+1)/2}, \end{aligned}$$

where the first inequality follows from

$$\prod_{i=1}^n (1 + u_i) \geq 1 + \sum_{i=1}^n u_i, \quad u_1 > 0, \dots, u_n > 0.$$

Now we have

$$\begin{aligned} h(w, \mathbf{y}, X) &= \int_{\mathbb{R}^p} \prod_{i=1}^n g^{(1)} \{ (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 / w \} d\pi_{\boldsymbol{\beta}} \\ &\leq \int_{\mathbb{R}^p} J^n \{ 1 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' Q (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) / \nu w \}^{-(\nu+1)/2} d\boldsymbol{\beta} \\ &= J^n \Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{p/2} w^{p/2} |Q|^{-1/2} / \Gamma\left(\frac{\nu+p}{2}\right) \\ &= w^{p/2} a(\nu) |Q|^{-1/2}, \end{aligned}$$

where

$$a(\nu) = \left\{ \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{1/2}} \right\}^n \frac{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{p/2}}{\Gamma\left(\frac{\nu+p}{2}\right)} \leq C$$

with C a constant free of ν . The last inequality follows by using the following bound for the gamma function (see, for example, Whittaker & Watson 1927, chap. 12):

$$\Gamma(z) = (2\pi)^{1/2} z^{z-1/2} e^{-z+b(z)}, \quad z > 0,$$

with $0 < b(z) < K/z$ for some positive constant K . Hence the proof is complete. □

Proof of Theorem A.1. We first consider the integral of the likelihood times the prior. Let

$$A = \int \dots \int L(\boldsymbol{\beta}, \sigma^2, \delta, \nu; \mathbf{y}) d\pi_{\boldsymbol{\beta}} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu}.$$

In the following derivation, the value of the constant C may not be the same from line to line:

$$\begin{aligned} A &= C \int \dots \int \frac{2^n}{(\sigma^2 + \delta^2)^{-n/2}} \\ &\quad \times \prod_{i=1}^n \left[g^{(1)} \left\{ \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\sigma^2 + \delta^2} \right\} F \left(\frac{\delta}{\sigma} \frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sqrt{\sigma^2 + \delta^2}} \mid 0, 1; g_a^{(1)} \right) \right] d\pi_{\boldsymbol{\beta}} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \int \dots \int (\sigma^2 + \delta^2)^{-n/2} \prod_{i=1}^n \left[g^{(1)} \left\{ \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\sigma^2 + \delta^2} \right\} \right] d\pi_{\boldsymbol{\beta}} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \iiint (\sigma^2 + \delta^2)^{-n/2} \left[\int_{\mathbb{R}^p} \prod_{i=1}^n \left[g^{(1)} \left\{ \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\sigma^2 + \delta^2} \right\} \right] d\pi_{\boldsymbol{\beta}} \right] d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \iiint (\sigma^2 + \delta^2)^{-n/2} (\sigma^2 + \delta^2)^{p/2} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \iiint (\sigma^2 + \delta^2)^{-(n-p)/2} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \iiint (\sigma^2)^{-(n-p)/2} d\pi_{\sigma^2} d\pi_{\delta} d\pi_{\nu} \\ &\leq C \iiint \left\{ \int_0^\infty (\sigma^2)^{-(n-p)/2 + \kappa + 1} e^{-\kappa/\sigma^2} d\sigma^2 \right\} d\pi_{\delta} d\pi_{\nu}. \end{aligned}$$

Now as $\kappa \rightarrow 0$, the innermost integral is finite if $n > p$. Also it is assumed that π_{δ} and π_{ν} are proper. Hence A is finite and the result follows.

Proof of Theorem 2. The proof follows by using Lemma A.4., Lemma A.5. and Theorem A.1.

Proof of Theorem 3. Follows from the last displayed inequality in the proof of Theorem A.1.

ACKNOWLEDGEMENTS

We would like to thank Adelchi Azzalini and Chris Jones for many insightful discussions. We also thank J. T. A. S. Ferreira and Mark Steel for pointing out a mistake in an earlier version. Also we thank the Editor, an Associate Editor and two referees for many helpful comments and suggestions. We are specially indebted to the Editor, Professor Richard A. Lockhart, for his help with the S-PLUS code that produced Figures 1 and 2.

REFERENCES

- C. J. Adcock (2002). *Asset Pricing and Portfolio Selection Based on the Multivariate Skew-Student Distribution*. Technical Report, University of Sheffield, England.
- D. J. Aigner, C. A. K. Lovell & P. Schmidt (1977). Formulation and estimation of stochastic frontier production function model. *Journal of Econometrics*, 12, 21–37.
- B. C. Arnold & R. J. Beaver (2000). Hidden truncation models. *Sankhyā, Series A*, 62, 23–35.
- B. C. Arnold & R. J. Beaver (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11, 7–54.
- A. Azzalini & A. Capitanio (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B*, 61, 579–602.
- A. Azzalini & A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika*, 8, 715–726.
- J. M. Bernardo & A. F. M. Smith (1994). *Bayesian Theory*. Wiley, New York.
- M. D. Branco, H. Bolfarine, P. Iglesias & R. B. Arellano-Valle (2000). Bayesian analysis of calibration problem under elliptical distributions. *Journal of Statistical Planning and Inference*, 90, 69–85.
- M. D. Branco & D. K. Dey (2001). A general class of multivariate skew elliptical distributions. *Journal of Multivariate Analysis*, 79, 99–113.
- M.-H. Chen, D. K. Dey & Q.-M. Shao (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94, 1172–1186.
- S. Chib, R. C. Tiwari & S. R. Jammalamadaka (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics*, 38, 349–360.
- T. J. DiCiccio, R. E. Kass, A. E. Raftery & L. A. Wasserman (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92, 903–915.
- K.-T. Fang, S. Kotz & K.-W. Ng (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.
- C. Fernandez & M. F. J. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93, 359–371.
- A. E. Gelfand & D. K. Dey (1994). Bayesian model choice—Asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*, 56, 501–514.
- J. Geweke (1993). Bayesian treatment of the independent Student-*t* linear model. *Journal of Applied Econometrics*, 8, 519–540.
- O. Huang & R. H. Litzenberger (1988). *Foundations for Financial Economics*. North Holland, New York.
- M. C. Jones (2001). A skew-*t* distribution. In *Probability and Statistical Models with Applications* (Ch. A. Charalambides, M. V. Koutras & N. Balakrishnan, eds.), Chapman & Hall/CRC, Boca Raton, pp. 269–278.
- D. Kelker (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā*, 32, 419–430.
- K. V. Mardia (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530.

- J. Osiewalski & M. F. J. Steel (1993). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, 57, 345–363.
- D. J. Spiegelhalter, N. G. Best, B. P. Carlin & A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*, 64, 583–639.
- D. J. Spiegelhalter, A. Thomas & N. G. Best (1996). Computation on Bayesian graphical models. In *Bayesian Statistics 5: Proceedings of the 5th Valencia International Meeting Held in Alicante, June 5–9, 1994* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds.), Oxford University Press, pp. 407–426.
- E. T. Whittaker & G. N. Watson (1927). *A Course of Modern Analysis: An Introduction to the General Theory of Infinite Processes and of Analytic Functions*. Cambridge University Press.
- A. Zellner (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error term. *Journal of the American Statistical Association*, 71, 400–405.

Received 3 December 2001

Accepted 2 May 2003

Sujit K. SAHU: S.K.Sahu@maths.soton.ac.uk

Faculty of Mathematical Studies, University of Southampton
Highfield SO17 1BJ, England, UK

Dipak K. DEY: dey@stat.uconn.edu

Department of Statistics, University of Connecticut
Storrs, CT 06269-3120, USA

Márcia D. BRANCO: mbranco@ime.usp.br

Departamento de Estatística, Instituto de Matemática e Estatística
Universidade de São Paulo, São Paulo, Brasil