Regression Model under Skew Elliptical Error Distribution

Márcia D. Branco
University of São Paulo, Brasil

Dipak K. Dey
University of Connecticut, USA

Summary. *We consider a Bayesian regression model under a skewed heavy tailed error distribution. A general class of skewed elliptical distribution is developed which includes heavy tailed distribution and a skewness parameter. Dependent and independent skew elliptical error distributions are considered for the regression problem. The results provide a unified parametric approach to non-normal regression problem with flexible error distributions. The proposed methodology is exemplified by a regression problem with real data using a skewed Student-t error distribution.*

keywords: Elliptical distributions, folded distribution, regression, multivariate t distribution, skewness.

1 Introduction

Regression analysis problems under elliptical error distributions have received considerable attention in recent statistical literature. A pioneering work in this area is due to Zellner(1976), in which a study based on multivariate Student-t distribution is considered. Extensions of those results for elliptical distributions are considered in Chib, Tiwari and Jammalamadaka (1988) and Osiewalski and Steel(1993). The class of the elliptical distributions was introduced by Kelker (1970) and include a vast set of known symmetric distributions, e.g., scale mixtures of normal and Pearson type II distributions. More about Bayesian regression under heavy tailed distribution can be seen in Geweke(1993) and Branco, Bolfarine, Iglesias and Arellano-Valle(2000). Recently, Fernandez and Steel(1998)considered a skew-t regression model in a Bayesian setup. In this paper first we develop a very simple but powerful approach to develop a multivariate skew-elliptical distribution which include several distributions including the multivariate skew-normal distribution as developed in Azzalini and Dalla-Valle(1996). We then consider the regression problem under a skew-elliptical error distribution and develop a methodology for inference of the regression parameters.

The major goal of this paper is to consider a wider class of distributions for the error in a regression problem which encompass several existing methods in the literature. The rest of the paper is organized as follows. In section 2 we derive our proposed skew-elliptical distribution. Several examples are provided for the univariate and multivariate distributions. In particular skew-t distribution is studied in detail including derivation of its moments. Section 3 provides the development of regression models with skew-t error distribution. In section 4, proposed regression methodology is applied to a data set on regression of skinfold thickness on body fat. The paper is concluded in section 5 with brief discussions.

2 Skew Elliptical Distribution

A $p \times 1$ random vector Y is said to have an elliptical distribution with parameters μ (the location vector) and Σ (the dispersion matrix) of dimension $p \times 1$ and $p \times p$, respectively, with Σ being positive definite, if Y has density function

$$f(y \mid \mu, \Sigma; g) = |\Sigma|^{-\frac{1}{2}} g\left[(y-\mu)'\Sigma^{-1}(y-\mu)\right], \quad y \in \mathcal{R}^p$$

where $g(u), u \geq 0$ is called a generator fuction, which satisfies the equation

$$\int_0^\infty u^{p/2-1} g(u) du = \frac{\Gamma(p/2)}{\pi^{p/2}}$$

a standardization constant. We use the notation $Y \sim El_p(\mu, \Sigma, g)$.

2.1 Univariate Distribution

First we develop a class of skewed distribution using the approach given in Chen, Dey and Shao(1999), where the skewed random variable evolves from a sum of a symmetric and a positive random variables, and is given as

$$Y = \delta Z + \varepsilon. \tag{1}$$

The important point here is to have ε 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, when $\delta > 0 (\delta < 0)$ means right(left) skewness. To see this, we assume that Z and ε have upto third moments. Let σ_Z^2 and σ_ε^2 be the variances of Z and ε, respectively, and

μ_Z^3 be the standardized third moment of Z, i.e., $\mu_Z^3 = E[(Z - E(Z))/\sigma_Z]^3$. Then, the standardized third moment of Y is given by

$$\mu_Y^3 = E\left[\frac{Y - E(Y)}{\sigma_Y}\right]^3 = \frac{\delta^3 \sigma_Z^3 \mu_Z^3}{\sigma_Y^3}, \tag{2}$$

where $\sigma_Y^2 = Var(Y) = \delta^2\sigma_Z^2 + \sigma_\varepsilon^2$. We are assuming that Z is skewed to the right, therefore $\mu_Z^3 > 0$. Then, (2) implies that the marginal distribution of Y is skewed to the right (left) when $\delta > 0 (\delta < 0)$.

As defined before, the notation $El_p(\mu, \Sigma, g)$ represents a p-dimensional elliptical distribution, with location and scale parameters μ and Σ, respectively, and generator function g (see Fang and Zhang 1990). We omit p in the notation when p=1. We only consider elliptical distribution with density, i.e., absolutely continous with respect to a Lebesgue measure.

Next we extend results to the elliptical class, where $\varepsilon \sim El(0, 1, g_2)$ and $z \sim El^+(0, 1, g_1)$, a positive elliptical distribution. A particular case is the skew-normal distribution, which was considered in Azzalini and Dalla-Valle(1996) and Azzalini and Capitanio(1999).

An interesting situation is when there exist a relationship between the generator functions g_1 and g_2, such that, we can obtain a nice closed form expression for the density function of the random variable Y. This is considered in the next theorem. First we need to consider the following lemma, through which the posterior inference under elliptical model can be considered under a conjugate prior density. The posterior updating follows the well known formal mechanics of conjugate normal-normal models.

Lemma 2.1.1: Consider $y \mid \theta \sim El_n(\theta, \Sigma; g)$ and $\theta \sim El_n(\theta_0, B_0; h_0)$. If there is a generator function f such that

$$h_0(u) = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty r^{(n-1)/2} f(r+u)dr$$

and

$$g(u) = \frac{f(u + q(\theta))}{h_0(q(\theta))}, \text{with} \;\; q(\theta) = (\theta - \theta_0)' B_0^{-1}(\theta - \theta_0),$$

then

i) the joint distribution of $(\theta, y) \sim El_{2n}(\theta_0, B; f)$,

ii) the conditional distribution of $\theta \mid y \sim El_n(b_1, B_1; h_1)$ and

iii) the marginal distribution of $y \sim El_n(\theta_0, \Sigma + B_0; g^*)$, where

$$b_1 = \theta_0 + B_0(B_0 + \Sigma)^{-1}(y - \theta_0),$$

$$B_1 = B_0 - B_0(B_0 + \Sigma)^{-1}B_0$$

and

$$B = \begin{bmatrix} B_0 + \Sigma & B_0 \\ B_0 & B_0 \end{bmatrix}.$$

The proof essentially follows, with some modifications, from Branco et al (2000).

In the next theorem we develop an ellegant expression for the skew elliptical density. We consider a relationship between the generator functions g_1 and g_2 in order to obtain a density form as a product of an elliptical density and an elliptical distribution function.

Theorem 2.1.1: Suppose there is a generator function g such that

$$g_2(u) = \frac{\pi^{1/2}}{\Gamma(1/2)} \int_0^\infty g(r+u)dr, \tag{3}$$

and

$$g_1(u) = \frac{g(u + q(\varepsilon))}{g_2(q(\varepsilon))}, \text{with } \ q(\varepsilon) = \varepsilon^2, \tag{4}$$

where g_2 and g_1 are respectively the generator function of ε and Z. Then, the p.d.f. of Y as defined in (1) is

$$f_Y(y) = 2(\delta^2+1)^{-1/2} f_{g_2}\left[\frac{y}{(\delta^2+1)^{1/2}}\right] F_{g_{q(y)}}\left[\frac{y\delta}{(\delta^2+1)^{1/2}}\right], \tag{5}$$

where f_g and F_g are respectively the p.d.f. and the c.d.f. of the $El(0,1,g)$ and

$$g_{q(y)}(u) = \frac{g(u+q(u))}{g_2(q(y))} \text{ with } \ q(y) = y^2. \tag{6}$$

Proof: Considering $\varepsilon \sim El(0,1,g_2)$ and $z \sim El^+(0,1,g_1)$, the p.d.f. are $f(\varepsilon) = g_2(\varepsilon^2)$ and $f(z) = 2g_1(z^2)I_{(0,\infty)}(z)$, respectively. Using variable transformation and the fact that $y = \delta z + \varepsilon$, we have

$$f_Y(y) = 2\int_{-\infty}^{y} \frac{1}{\delta} g_1 \left[\frac{(y-\varepsilon)^2}{\delta^2}\right] g_2(\varepsilon^2) d\varepsilon. \tag{7}$$

Now there exist a random variable, say y^* such that $y^* \mid \varepsilon, \delta \sim El(\varepsilon, \delta^2,\ g_1)$ and the p.d.f. of y^* is $\delta^{-1} g_1 \left[\frac{(y-\varepsilon)^2}{\delta^2}\right]$. Thus the expression inside the integral in (7) can be thought of a joint distribution for (y^*, ε), and using the above lemma, this has an elliptical distribution. In this case $\varepsilon \mid y^*$ and y^* have elliptical distribution with

$$\varepsilon \mid y \sim El\left(\frac{y}{(\delta^2+1)}, \frac{\delta^2}{(\delta^2+1)}; g_{q(y)}\right) \tag{8}$$

and

$$y^* \sim El(0, \delta^2+1; g^*). \tag{9}$$

Therefore,

$$f_Y(y) = \frac{2}{(\delta^2+1)^{1/2}} g^* \left[\frac{y^2}{(\delta^2+1)}\right] \int_{-\infty}^{y} [\frac{(\delta^2+1)}{\delta^2}]^{1/2} g_{q(y)} \left[\frac{(\varepsilon-y)}{\delta^2(\delta^2+1)}\right] d\varepsilon \tag{10}$$

Considering the following variable transformation

$$\varepsilon^* = \frac{(\varepsilon - z)}{(\delta^2(\delta^2+1))^{1/2}} \tag{11}$$

we get the expression (5) which completes the proof.

Now, we consider some examples which demonstrate that (5) gives a very general class of distributions which include many standard distributions.

Example 2.1.1: The skew-normal distribution.

Considering ε having a normal distribution, then the generator function $g_2(u) = (2\pi)^{1/2} exp(-u/2)$. If we consider $g(u) = 2\pi exp(-u/2)$, the generator function of a bivariate normal, we obtain

$$g_1(u) = \frac{g(u+q(\varepsilon))}{g_2(q(\varepsilon))} = (2\pi)^{1/2} exp(-u/2) \tag{12}$$

and similary

$$g_{q(y)}(u) = (2\pi)^{1/2} exp(-u/2). \tag{13}$$

In this case z is a positive normal, $f_{g_2} = \phi$ and $F_{g_{q(y)}} = \Phi$, the p.d.f and the c.d.f. of the standard normal distribution, respectively. Thus, we obtain the following density function which is given in Chen et al (1999),

$$f_Y(y) = 2(\delta^2+1)^{-1/2}\phi\left[\frac{y}{(\delta^2+1)^{1/2}}\right]\Phi\left[\frac{y\delta}{(\delta^2+1)^{1/2}}\right]. \tag{14}$$

Example 2.1.2: The skew-t distribution.

From Arellano-Valle and Bolfarine(1995) the generator function of the univariate Student-t is

$$g_2(u) = \frac{\Gamma((\nu+1)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\pi^{1/2}}[\nu+u]^{-(\nu+1)/2}. \tag{15}$$

If we consider

$$g(u) = \frac{\Gamma((\nu+2)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\pi}[\nu+u]^{-(\nu+2)/2}, \tag{16}$$

the generator function of a bivariate t distribution, we obtain

$$g_1(u) = \frac{\Gamma((\nu+2)/2)}{\Gamma((\nu+1)/2)\pi^{1/2}}[\nu+q(\varepsilon)]^{(\nu+1)/2}[\nu+q(\varepsilon)+u]^{-(\nu+2)/2}, \tag{17}$$

and

$$g_{q(y)}(u) = \frac{\Gamma((\nu+2)/2)}{\Gamma((\nu+1)/2)\pi^{1/2}}[\nu+q(y)]^{(\nu+1)/2}[\nu+q(y)+u]^{-(\nu+2)/2}. \tag{18}$$

Both expressions above are generator function of a generalized t distribuition, see also Arellano-Valle et al (1995) for further references. Consequentely, if Z is a positive generalized t with parameters $\nu_1 = \nu+1$ and $\lambda = \nu+q(\varepsilon)$; then Y in (1) has the density of the form

$$f_Y(y) = 2(\delta^2+1)^{-1/2}t_v\left[\frac{y}{(\delta^2+1)^{1/2}}\right]GT_{\nu_1,\lambda_1}\left[\frac{y\delta}{(\delta^2+1)^{1/2}}\right], \tag{19}$$

where t_v is the p.d.f. of the Student-t and GT_{ν_1,λ_1} is the c.d.f. of the generalized t distribution, with parameters $\lambda_1 = \nu+q(y)$ and ν_1 as above.

This distribution depends on two parameters that has a nice interpretation. The parameter ν (degree of freedom) controls the heavyness of the tail and the parameter

δ controls the skewness. Note that when $\nu = 1$, (19) gives the density of the skew-Cauchy distribution and as $\nu \to \infty$, (19) reduces to (10), the p.d.f. of the skew-normal distribution.

Several density plots with different values of skewness parameter δ and degrees of freedom ν are given in figure 1.

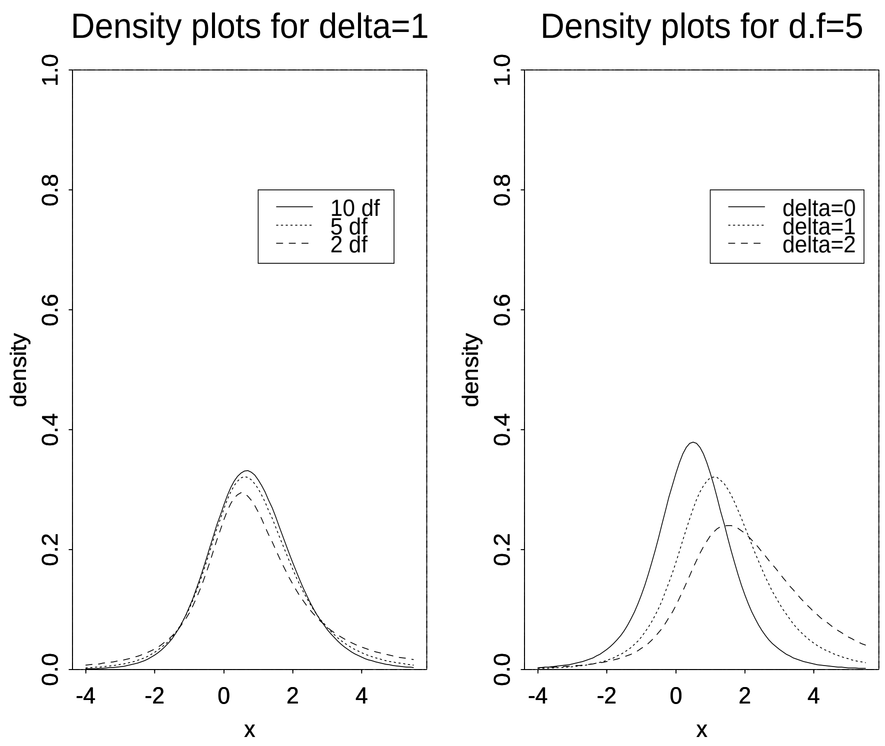


Fig. 1. Skew-t density

Example 2.1.3: The skew-Pearson type II distribution.

The generator function of a Pearson type II distribution (see Fang et al, 1990) is

$$g_2(u) = \frac{\Gamma(m+1+1/2)}{\Gamma(m+1)\pi^{1/2}}(1-u)^m, \quad 0 \le u \le 1 \ \text{ and } \ m > -1. \tag{20}$$

If we consider

$$g(u) = \frac{\Gamma(m_1+1+1)}{\Gamma(m_1+1)\pi}(1-u)^{m_1}, \tag{21}$$

then it is the generator function of the bivariate Pearson type II. Now we define $m_1 = m - 1/2$ where $m > -1/2$. In this case

$$g_1(u) = \frac{\Gamma(m_1 + 1/2 + 1)}{\Gamma(m_1 + 1)\pi^(1/2)} r^{-(m_1+1/2)}(r - u)^{m_1}, \tag{22}$$

with $r = 1 - q(\varepsilon)$. Considering $0 \leq u \leq r$ and using the generator function we obtain the following p.d.f.

$$f_Y(y) = \frac{\Gamma(m_1 + 1/2 + 1)}{\Gamma(m_1 + 1)\pi^(1/2)} r^{-(m_1+1/2)}(r - y^2)^{m_1}, \quad -r \leq y \leq r.$$

We will call this distribution as generalized Pearson type II, with parameters m_1 and $r > 0$ .Thus Z has a positive generalized Pearson type II distribution.

Similary,

$$g_{q(y)}(u) = \frac{\Gamma(m_1 + 1/2 + 1)}{\Gamma(m_1 + 1)\pi^(1/2)} r_1^{-(m_1+1/2)}(r_1 - u)^{m_1}, \tag{23}$$

with $r_1 = 1 - q(y)$. Consequentely,

$$f_Y(y) = 2(\delta^2 + 1)^{-1/2} pII_m \left[\frac{y}{(\delta^2 + 1)^{1/2}}\right] PII_{m_1,r_1} \left[\frac{y\delta}{(\delta^2 + 1)^{1/2}}\right], \tag{24}$$

where pII_m is the p.d.f. of the Pearson type II, with $m > 1/2$ and PII_{m_1,r_1} is the c.d.f. of the generalized Pearson type II distribution.

Remark 2.1.1: The distribution of Z and ε are usually dependent. It is only independent for the skew-normal distribution.

2.2 Multivariate Distributions

In this Section we will extend the result of theorem 2.1.1 for a vector of random variables. Consider

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{Z} + \boldsymbol{\varepsilon}, \tag{25}$$

where $\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\varepsilon}$ are $n \times 1$ vectors, $\boldsymbol{D} = diag(\delta)$ is a diagonal matrix of order $n \times n$ and $\delta = (\delta_1, \delta_2, ..., \delta_n)$. We define $\boldsymbol{D}^2 = diag(\delta_1^2, \delta_2^2, \ldots, \delta_n^2)$.

Theorem 2.2.1: Let $\boldsymbol{\varepsilon} \sim El_n(\boldsymbol{0}, \boldsymbol{\Omega}; g_2)$ and $\boldsymbol{Z} \sim El_n^+(\boldsymbol{0}, \boldsymbol{I}; g_1)$. Suppose there is a generator function g such that

$$g_2(u) = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty g(r+u)dr \tag{26}$$

and

$$g_1(u) = \frac{g(u+q(\boldsymbol{\varepsilon}))}{g_2(q(\boldsymbol{\varepsilon}))} \quad \text{with} \quad q(\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^t \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}, \tag{27}$$

then $f_Y(\boldsymbol{y})$ is equal to

$$2|\boldsymbol{\Omega}+\boldsymbol{D}^2|^{-1/2} f_{g_2}\left[\boldsymbol{y}^t(\boldsymbol{\Omega}+\boldsymbol{D}^2)^{-1/2}\right] F_{g_{q(y)}}\left[\boldsymbol{y}^t(\boldsymbol{I}_n - \boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{D}^2)^{-1}(\boldsymbol{\Omega}-\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{D}^2)^{-1}\boldsymbol{\Omega})^{-1/2}\right] \tag{28}$$

where f_g and F_g are respectively the p.d.f. and the c.d.f. of the n-dimensional elliptical distribution with generator function g , and

$$g_{q(y)}(u) = \frac{g(u+q(u))}{g_2(q(u))}, \quad \text{with} \quad q(y) = \boldsymbol{y}^t(\boldsymbol{\Omega}+\boldsymbol{D}^2)^{-1}\boldsymbol{y}. \tag{29}$$

Proof: The idea of the proof is similar to that in the univariate case. First considering the variable transformation $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\varepsilon}$, it follows that $\boldsymbol{z} = \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{\varepsilon})$. Note that $\boldsymbol{D}^{-1}$ exist since $\boldsymbol{D}$ is diagonal. Then, $f_Y(\boldsymbol{y})$ is

$$2\int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} |\boldsymbol{\Omega}|^{-1/2} g_2(\boldsymbol{\varepsilon}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}) g_1[(\boldsymbol{y}-\boldsymbol{\varepsilon})^t(\boldsymbol{D}^2)^{-1}(\boldsymbol{y}-\boldsymbol{\varepsilon})]|\boldsymbol{D}^2|^{-1/2} d\varepsilon_n \ldots d\varepsilon_n. \tag{30}$$

Again, we can see that the inside term of the integral is like a joint distribuiton. Under the conditions (26) and (27), it is elliptical with $2n$-dimension and can be written as a product of a marginal distribution of $\boldsymbol{y}^*$ and the conditional distribution of the $\boldsymbol{\varepsilon} \mid \boldsymbol{y}^*$. Thus

$$f_Y(\boldsymbol{y}) = 2|\boldsymbol{\Omega}+\boldsymbol{D}^2|^{-1/2} g_2(\boldsymbol{\varepsilon}^t(\boldsymbol{\Omega}+\boldsymbol{D}^2)^{-1}\boldsymbol{\varepsilon}) \tag{31}$$

$$\int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} g_{q(y)}[(\boldsymbol{\varepsilon}-\mu_1)^t\Sigma_1^{-1}(\boldsymbol{\varepsilon}-\mu_1)]|\Sigma_1|^{-1/2} d\varepsilon_n \ldots d\varepsilon_n.$$

Using the following transformation of variable

$$\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}-\mu_1)^t\Sigma_1^{-1/2} \tag{32}$$

we get the final result.

Remark 2.2.1: If $\boldsymbol{\Omega} = \boldsymbol{I}_n$ then $f_Y(\boldsymbol{y})$ is

$$\prod_{i=2}^{n} \frac{2}{(1+\delta_i^2)^{1/2}} f_{g_2}\left[\frac{y_1}{(1+\delta_1^2)^{1/2}}, \dots, \frac{y_n}{(1+\delta_n^2)^{1/2}}\right] F_{g_{q(y)}}\left[\frac{y_1\delta_1}{(1+\delta_1)^{1/2}}, \dots, \frac{y_1\delta_1}{(1+\delta_1)^{1/2}}\right], \quad (33)$$

where $y_i's$ are the components of the vector $\boldsymbol{y}$, which is a natural extension of the univariate case given in (5).

Example 2.2.1: The multivariate skew-t distribution.

Suppose $\boldsymbol{\varepsilon} \sim t_n(\boldsymbol{0}, \boldsymbol{\Omega}; \nu)$, where $t_n(\mu, \Sigma; \nu)$ denotes an n-dimensional t distribution with location μ, scale Σ and degrees of freedom ν. In this case the generator function is

$$g_2(u) = \frac{\Gamma[(\nu+n)/2]\nu^{\nu/2}[\nu+u]^{-(\nu+n)/2}}{\Gamma[\nu/2]\pi^{n/2}}. \quad (34)$$

If we consider

$$g(u) = \frac{\Gamma[(\nu+2n)/2]\nu^{\nu/2}[\nu+u]^{-(\nu+2n)/2}}{\Gamma[\nu/2]\pi^{2n/2}}, \quad (35)$$

the generator function of a 2n-dimensional t, we obtain

$$g_1(u) = \frac{\Gamma[(\nu_1+n)/2]\lambda^{\nu_1/2}[\lambda+u]^{-(\nu_1+n)/2}}{\Gamma[\nu_1/2]\pi^{n/2}}, \quad (36)$$

with $\nu_1 = \nu + n$ and $\lambda = \nu + q(\varepsilon) = \nu + \boldsymbol{\varepsilon}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$ and

$$g_{q(y)}(u) = \frac{\Gamma[(\nu_1+n)/2]\lambda_1^{\nu_1/2}[\lambda_1+u]^{-(\nu_1+n)/2}}{\Gamma[\nu_1/2]\pi^{n/2}}, \quad (37)$$

with $\lambda_1 = \nu + q(\boldsymbol{y}) = \nu + \boldsymbol{y}^t(\boldsymbol{\Omega} + \boldsymbol{D}^2)^{-1}\boldsymbol{y}$.

Again, as in the univariate case, both expressions (36) and (37) are generator functions of a generalized t distribution. Suppose $\boldsymbol{\Omega} = \sigma^2\boldsymbol{I}$ and $\delta_i = \delta, i = 1, 2, \dots, n$. This means that the components of the vector $\boldsymbol{\varepsilon}$ are uncorrelated. In this case

$$f_Y(\boldsymbol{y}) = \frac{2}{(\sigma^2+\delta^2)^{n/2}} t_n\left[\frac{\boldsymbol{y}}{(\sigma^2+\delta^2)^{1/2}}\right] GT_n\left[\frac{\boldsymbol{y}\delta}{(\sigma^2+\delta^2)^{1/2}}\right]. \quad (38)$$

Remark 2.2.2: We can easily add a location parameter μ, such that, $\boldsymbol{Y} = \mu + \boldsymbol{DZ} + \boldsymbol{\varepsilon}$, which is same as considering $\boldsymbol{Y} = \boldsymbol{DZ} + \boldsymbol{\varepsilon}^*$ with $\varepsilon^* \sim El_n(\mu, \boldsymbol{\Omega}; \nu)$.

2.3 Moments of Skew-t distribution

From (1), since $E[\varepsilon] = 0$, we know that $V[Y] = \delta^2 V[Z] + V[\varepsilon] + 2Cov[Z, \varepsilon]$ and $E[Y] = \delta E[Z]$. It is easy to calculate the moments of the skew-normal distribution, since Z and ε are independent and Z has a positive normal distribution. Therefore for the skew-normal distribution

$$E[Y] = \delta(2/\pi)^{1/2} \quad \text{and} \quad V[Y] = \delta^2[1 - 2/\pi] + 1. \tag{39}$$

However for the skew-t distribution, as defined in example 2.1.2, the mean and variance cannot be obtained directly. Observe that Z and ε are not independent anymore and the distribution of Z conditional on ε is a positive generalized t distribution. First we obtain the mean and variance for $Z \mid \varepsilon$, and then use conditional mean and variance formula to obtain the unconditional mean and variance of Z.

To obtain the moments of the generalized positive t distribution, we use the fact that

$$T^+ = W^{1/2} N^+ \tag{40}$$

where $T^+ \sim gt^+(0, 1, v, \lambda)$, $W \sim IG(\frac{v}{2}, \frac{\lambda}{2})$, with density $f(w) \propto w^{-\nu/2} exp\{-\lambda/2w\}$, and $N^+ \sim N^+(0, 1)$, respectivly, the generalized positive t, inverse gamma and positive normal distribution. After some algebraic manipulations, we obtain

$$E[T^+] = \frac{\Gamma[(v-1)/2]}{\Gamma[v/2]} \left(\frac{\lambda}{\pi}\right)^{1/2} \tag{41}$$

and

$$Var[T^+] = \frac{\nu}{\lambda - 2} - \frac{\lambda}{\pi} \left[\frac{\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]}\right]^2, \tag{42}$$

respectively, the mean and the variance of the positive generalized t distribution.

In our case (see ex.2.1.2.), $Z \mid \varepsilon \sim gt^+(0, 1, \nu + 1, \lambda(\varepsilon))$ with $\lambda(\varepsilon) = \nu + \varepsilon^2$. Then,

$$E[Z] = \frac{\Gamma[\nu/2]}{\pi^{1/2}\Gamma[(\nu+1)/2]} E[(\nu + \varepsilon^2)^{1/2}]. \tag{43}$$

Since $\varepsilon \sim t_\nu$, we can get the last expectation without any problem. Finally,

$$E[Y] = \frac{\delta\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]} \left(\frac{\nu}{\pi}\right)^{1/2}. \tag{44}$$

To obtain the expression for the variance, first we observe that $Cov[Z,\varepsilon]=0, Var(\varepsilon)=\nu/(\nu-2)$ and $Var[Z]=Var[E(Z\mid\varepsilon)]+E[Var(Z\mid\varepsilon)]$. Using (32) and (33) and after some algebraic manipulations, we obtain

$$Var[Y]=\frac{\nu}{\nu-2}+\delta^2\left\{(\nu+1)E[(\nu+\varepsilon^2-2)^{-1}]-\frac{\nu}{\pi}\left[\frac{\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]}\right]^2\right\}, \tag{45}$$

where the expectation is calculated under the distribution of ε.

For application it is easier to consider Z and ε to be independent. If we take Z independent of ε we will not get the expression (13) anymore, but the distribution will be still skewed with heavy tail. In this case, the mean and the variance can be calculated easly and will be done next.

We consider two cases and call the corresponding distributions as pseudo skew-t distribution of type I and of type II. In the first case, we consider a positive normal distribution for z for which the mean and the variances are given, respectively, as

$$E[Y]=\delta\left(\frac{2}{\pi}\right)^{1/2} \quad \text{and} \quad V[Y]=\delta^2\left[1-\frac{2}{\pi}\right]+\frac{\nu}{\nu-2}. \tag{46}$$

Note that the mean is same as that of the skew normal distribution. However the variance is bigger than that of the skew-normal distribution. Observe that when ν goes to infinity, the variance goes to the skew-normal variance.

The type II pseudo skew-t distribution is obtained by considering an independent positive t with the same degrees of freedom ν. In this case the mean and the variance are given respectively as

$$E[Y]=\frac{\delta\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]}\left(\frac{\nu}{\pi}\right)^{1/2} \tag{47}$$

and

$$V[Y]=\delta^2\left[\frac{\nu}{\nu-2}-\frac{\nu}{\pi}\left(\frac{\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]}\right)^2\right]+\frac{\nu}{\nu-2}. \tag{48}$$

This mean is equal to the mean of the original skew-t distribution, given in (44).

Table 2.1 inclues a summary from the moments results obtained in this section.

Table 2.1: Mean and variance for some skewed distributions.

Distribution	Mean	Variance
Skew-normal	$\delta(\frac{2}{\pi})^{1/2}$	$\delta^2(1-\frac{2}{\pi})+1$
Skew-t typeI	$\delta(\frac{2}{\pi})^{1/2}$	$\delta^2(1-\frac{2}{\pi})+\frac{\nu}{\nu-2}$
Skew-t typeII	$\frac{\delta\Gamma[(\nu-1)/2](\nu/2)^{1/2}}{\Gamma[\nu/2]}$	$\delta^2\left[\frac{\nu}{\nu-2}-\frac{\nu}{\pi}\left(\frac{\Gamma[(\nu-1)/2]}{\Gamma[\nu/2]}\right)^2\right]+\frac{\nu}{\nu-2}$

3 Regression Analysis for the Skew Elliptical Error

3.1 Introdution

In this Section we consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^* \tag{49}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of the response variables, $\boldsymbol{X}$ is a $n \times p$ known design matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters and $\boldsymbol{\varepsilon}^*$ is a $n \times 1$ vector of the errors. The idea is to assume for the error vector a class of distribution that included skewed and heavy tailed distribution. When we work with the elliptical distribution for the errors, two models can be considered. The first one is called the dependent model, in this case we consider a joint distribution for the vector $\boldsymbol{\varepsilon}^*$, such that, the components have no correlation. The second, is called the independent model and in this case each component ε_i^* are independent and identically distributed, $i = 1, 2, \ldots, n$. Note that, the two such models coincide only in the normal case.

In this section we consider the regression model with skew-t distribution for the errors.

3.2 The dependent model

Consider

$$\boldsymbol{y} = \boldsymbol{X\beta} + \delta\boldsymbol{z} + \boldsymbol{\varepsilon} \tag{50}$$

with

$$\boldsymbol{\varepsilon} \sim t_n(\mathbf{0}, \sigma^2 \boldsymbol{I}; \nu), \tag{51}$$

and $\boldsymbol{z}$ is positive random variable. When $\boldsymbol{z} \mid \boldsymbol{\varepsilon} \sim gt_n^+(\mathbf{0}, \boldsymbol{I}, \nu_1, \lambda)$ with $\nu_1 = \nu + n$ and $\lambda = \nu + (\boldsymbol{\varepsilon}^t \boldsymbol{\varepsilon})/\sigma^2$, it follows from (38) that the density for $\boldsymbol{y}$ is

$$f_Y(\boldsymbol{y}) = \frac{2}{(\sigma^2 + \delta^2)^{n/2}} t_n \left[\frac{(\boldsymbol{y} - \boldsymbol{X\beta})}{(\sigma^2 + \delta^2)^{1/2}} \right] GT_n \left[\frac{(\boldsymbol{y} - \boldsymbol{X\beta})\delta}{(\sigma^2 + \delta^2)^{1/2}} \right]. \tag{52}$$

However, for the proposed application, we consider $\boldsymbol{z}$ to be positive standard normal and is independent of $\boldsymbol{\varepsilon}$ (Y has a pseudo skew-t type I). Then $\boldsymbol{z}$ can be seen like a latent variable. In this case

$$\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2, \delta, \nu, \boldsymbol{z} \sim t_n(\boldsymbol{X\beta} + \delta \boldsymbol{z}, \sigma^2 \boldsymbol{I}; \nu). \tag{53}$$

It is well known that the t distribution can be written as a scale mixture of normals where the mixing measure has an inverse gamma distribution. That is consider one more step and add another latent variable w such that

$$w \sim IG(\frac{\nu}{2}, \frac{\nu}{2}) \tag{54}$$

and hence

$$\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2, \delta, \nu, \boldsymbol{z}, w \sim N_n(\boldsymbol{X\beta} + \delta \boldsymbol{z}, w\sigma^2 \boldsymbol{I}). \tag{55}$$

Therefore the likelihood is proportional to

$$(\sigma^2)^{-n/2} w^{-(n+\nu)/2-1} exp\{-\frac{(\boldsymbol{y} - \boldsymbol{X\beta} - \delta\boldsymbol{z})^t(\boldsymbol{y} - \boldsymbol{X\beta} - \delta\boldsymbol{z})}{2\sigma^2 w} - \frac{\boldsymbol{z}^t\boldsymbol{z}}{2} - \frac{\nu}{2w}\} \tag{56}$$

with $z_i \geq 0, \quad i = 1, 2, \ldots, n$.

3.3 The independent model

Consider

$$y_i = \boldsymbol{\beta}^t \boldsymbol{x}_i + \delta z_i + \varepsilon_i \tag{57}$$

with

$$\varepsilon_i \sim t(0, \sigma^2; \nu) \qquad i = 1, 2, \ldots, n, \tag{58}$$

are independent and z_i is a positive valued random variable. In the case when $z_i \mid \varepsilon_i \sim gt^+(0,1,\nu^*,\lambda_i)$ with $\nu^* = \nu + 1$ and $\lambda_i = \nu + \varepsilon_i^2$, the likelihood function is

$$2(\sigma^2+\delta^2)^{-n/2}\prod_{i=1}^{n} t_\nu\left[\frac{y_i-\boldsymbol{\beta}^t x_i}{(\sigma^2+\delta^2)^{1/2}}\right] T_{\nu^*,\lambda_i^*}\left[\frac{\delta(y_i-\boldsymbol{\beta}^t x_i)^t}{\sigma(\sigma^2+\delta^2)^{1/2}}\right] \tag{59}$$

with $\lambda_i^* = \nu + (y_i - \boldsymbol{\beta}^t x_i)^2/(\sigma^2+\delta^2)$.

In general the likelihood is very dificult to work with. For application purpose we consider z_i a positive standard normal random variable and is independent of ε_i, for $i = 1, 2, \ldots, n$. Thus we can treat $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)^t$ as a vector of latent variables.

The next step is to consider each t distribution as a scale mixture of normal distributions. For this we need to add n more latent variables, $w_1, w_2, \ldots, w_n$ such that

$$w_i \sim IG(\frac{\nu}{2}, \frac{\nu}{2}), \qquad i = 1, 2, \ldots, n, \tag{60}$$

and are independent. Therefore the likelihood is proportional to

$$\prod_{i=1}^{n} w_i^{-(\nu+1)/2-1}(\sigma^2)^{-1/2}\frac{\nu^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2)} exp\{\frac{-(y_i-\boldsymbol{\beta}^t x_i\delta z_i)^2}{2w_i\sigma^2} - \frac{z_i^2}{2} - \frac{\nu}{2w_i}\} \tag{61}$$

with $z_i \geq 0,\ i = 1, 2, \ldots, n$.

Each w_i has an inverse gamma distribution, i.e.,

$$w_i \mid \boldsymbol{\beta}, \sigma^2, \delta, \nu, \boldsymbol{z}, \boldsymbol{y} \sim IG(\frac{\nu+1}{2}, \frac{(y_i - \boldsymbol{\beta}\boldsymbol{x}_i\delta z_i)^2}{2\sigma^2} + \frac{\nu}{2}). \tag{62}$$

4 Application

To illustrate our proposed model as described in section 3, we consider a data set obtained at the Pediatric Department of the Pontificia Universidad Catolica de Chile. Two measures of body fatness are considered for evaluating the percentage of body fatness: *DEXA*(dual-energy X-ray absortiometry) and the *Slaughter* index, relating skinfold thickness and density. Table 4.1. presents the measurements for a group of n=24 women aged 16 - 19 years. We consider the regression model

$$Slaughter = \alpha + \beta DEXA + \varepsilon,$$

with independent skew-t error distribution as developed in the section 3.3. The maximum likelihood estimates are obtained using an EM algorithm. The Bayesian analysis is also straightforward to implement for this model using Gibbs sampling algorithm.

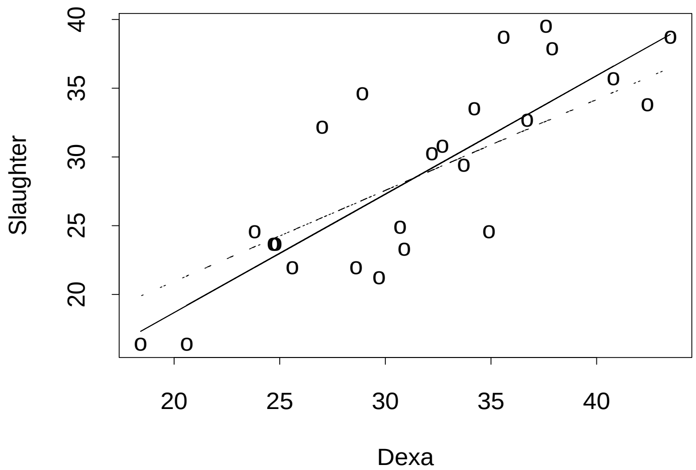
Table 4.1: The Measurements of Body Fatness for a Group of 24 Women.

Woman	DEXA(x)	Slaughter(y)	Woman	DEXA(x)	Slaughter(y)
1	28.6	21.93	13	24.7	23.62
2	24.8	23.62	14	32.7	30.75
3	30.9	23.29	15	34.2	33.49
4	32.2	30.21	16	28.9	34.58
5	35.6	38.68	17	23.8	24.55
6	18.4	16.35	18	25.6	21.93
7	37.9	37.86	19	34.9	24.55
8	27.0	32.12	20	30.7	24.85
9	37.6	39.50	21	36.7	32.67
10	40.8	35.67	22	42.4	33.76
11	33.7	29.38	23	43.5	38.68
12	20.6	16.35	24	29.7	21.21

The maximum likelihood estimates of the parameters in the model are $\hat{\alpha} = 1.069$, $\hat{\beta} = 0.661$, $\hat{\delta} = 8.359$, $\hat{\sigma}^2 = 0.101$ and $\hat{\nu} = 13.030$. We cleary see the strong evidence of positive skewness. The estimate of ν indicates that heavy tailed error model is appropriate.

Figure 2 indicates a plot of the data along with the m.l.e. fitted line and least squares line. Note that, in the skew-t model the error has mean different than zero, as we can see in (46). Then $E[Slaughter] = \alpha + \beta DEXA + \delta(2/\pi)^{1/2}$. The m.l.e fitted line considers the m.l.e. for each one of the parameters on the last expression. As we can see (Figure 2) the lines are not coincident.

Fig. 2. The two fitted lines.



Solid line is the l.s. line and dot line is the m.l. fitted curve

5 Discussion

This article has presented a new skewed elliptical regression model. The distribution theory developed in this paper includes many well known univariate and multivariate distributions. The skewness parameter introduces more flexibility in modelling the data. There are, however, few aspects that warrant further discussion.

The most important issue for the regression model is the choice of the distribution of $Z \mid \varepsilon$ and that of ε. In practice, choice of the distribution of ε is usually restricted to standard distribution, e.g., normal or Student-t. However there are many plausible choices of g. Choice of $Z \mid \varepsilon$ being positive normal or positive t are analytically tractable and easily implementable from a computational point of view. Nonparametric mixture models such as Dirichlet process mixture models can be implemented in modeling ε and $Z \mid \varepsilon$. This research will be pursued in future.

In this article, since our primary concern is to propose the new skewed elliptical model, besides model fitting, we have not pursued any model diagnostics and comparison.

Acknowledgements

The authors acknowledge Professor Pilar Iglesias, Department of Statistics of Catholic University of Chile, Doctor Gloria Icaza of the INTA, University of Chile, and Doctor Andreina Cattani of the Pediatric Department of Catholic University of Chile for providing the data set used in Section 4. The first author acknowledges to *Fundação de Amparo a Pesquisa do Estado de São Paulo* (FAPESP) for the financial support in her visit to the University of Connecticut.

References

Arellano-Valle, R.B. and Bolfarine, H.(1995). On some charecterizations of t-distribution. *Statistics and Probability Letters,* **25**, 79-85.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society,* B, **61**, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* **83**, 715-726.

Branco, M.D., Bolfarine, H., Iglesias, P. and Arellano-Valle, R.(2000). Bayesian Analysis of the Calibration Problem under Elliptical Distributions. *Journal of Statistical Planning and Inference,***90,**69-85.

Chen, M-H., Dey, D. and Shao, Q-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association,* **?,** 1172-1186.

Chib, S., Tiwari, R.C. and Jammalamadaka, S.R. (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics,* **38**, 349-360.

Fang, K.T. and Zhang, Y.T. (1990).*Generalized Multivariate Analysis.* Springer-Verlag.

Fernandez, C. and Steel, M.F.J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association,***93**, 359-371 .

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics,* **8**, 519-540.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā*,**32**, 419-430.

Osiewalski J. and Steel, M.F.J.(1993). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**, 345-363.

Zellner, A. (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.