International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 7 Issue 12, December - 2023, Pages: 173-180

Bayesian Estimation on the Epsilon-Skew Generalized Error distribution

Asaad Naser Hussein Mzedawee

asaad.nasir@qu.edu.iq

University Of AL-Qadisiya-College of Administration and Economics

Abstract: *Practical applied statistics reveals that the analysis of many real data that exhibit both fat-tailness and skewness indicates significant departures from normality assumptions. In these circumstances the adoption of more flexible models that cope with near normal data may be appropriate in place of adopt the robust approach, semi parametric or nonparametric models, and Box-Cox transformation. An alternative approach is to consider using the Epsilon Skew Generalized Error (ESGE) is a special case of Skewed Generalized T (GT) distribution proposed by Theodossiou (1998) which is embeds the normal, fat-tailness, and skewness distributions as special cases. This paper focus on the Bayesian estimation for the parameter of the epsilon –skew GED , near formula have derived based on the reference prior to estimate the parameters . Simulation study have conducted to find the parameters estimates of the proposed models the results based on MSE criteria has show that the Bayesian estimation are accrued based on the reference priors that have proposed*

1-Introduction

Why Epsilon Skew Generalized Error (ESGE) distribution? It is well-known that many datasets and processes often show some skewness present and heavy-tails behavior, and the analysis of the near normal data indicates significant departure from

normality assumptions. So, on the applied side the symmetric distributions are not practical for modeling skewed and heavy-tailed datasets. Therefore, there is a strong motivation to construct suitable distributions that can accommodate behaviors such as skeweness and heavy-tails.

Flexible nested asymmetric models can be adopted to deal with the situations when the classes of symmetric distribution indicate poor properties. The attractive features of a flexible nested-skew model are the encompassing of the normal distribution as a special model and allowing in various situations a continuous variation from normality to no-normality. Therefore, such flexible models can be useful to fit the

near normal data.

Subbotin (1923) formulated the Generalized Error (GE) distribution as a model for the random errors; this distribution can approach the normal distribution as the shape parameter tends to an appropriate value. Box and Tiao (1962) considered the Generalized Error distribution as more leptokurtic and more platykurtic than normal distribution. Azzalini (1986) considered the asymmetric exponential power (AEP) that accommodate both skeweness and heavy-tails properties. West (1984) considered using the Generalized Error (exponential power) family as heavy-tailed distribution of the errors in the linear regression model to accommodate outliers. Azzalini (1985) introduced the skew normal (SN) distribution as new model to accommodate asymmetry; the SN is allowing a continuous variation from normality to non-normality.

Mudholkar and Huston (2000) introduced the Epsilon skew normal (ESN) distribution by parameterizing the skew family that proposed by Fernandez and Steel (1998). Epsilon-Skew Generalized Error (ESGE) or Skewed Generalized Error (SGE) is a special case of the Skewed Generalized T distribution proposed by Theodossiou (1998), skewed GT is a flexible and accommodating the heavy-tails (Leptokurtic) and skewness properties.

The ESGE distribution is attractive and flexible because it allows continuous variation from normality to non-normality and nested with many models especially with the normal distribution, that is mean the ESGE includes normal distribution as a special case and hence is a “robust model”, this is why ESGE distribution has certain benefits. One major and the most important benefit in the ESGE distribution, is to use it where the normal distribution is used in common statistical applications.

We propose another extension; the ESGE regression model .The ESGE can be considered a suitable model to random errors of regression models that may exhibit both heavy-tails and skewness behavior. Despite the flexibility feature of the ESGE for accommodating both heavy-tails and skewness behavior, there is no literature on the linear regression model when the error term is assumed to have ESGE distribution as well as on the noninformative Bayesian inference for regression technique with ESGE errors. The earliest use of noninformative priors, due to Laplace (1812). Jeffreys (1961) proposed the prior proportional to the positive square root of the determinant of the Fisher information matrix to overcome the problem of invariance of the uniform prior. However, despite its success of the in one parameter problem, Jeffreys' prior often comes with difficulty in multiparameter problems, where there are nuisance parameters.

Bernardo (1979) introduced the so called "reference prior" to overcome the difficulty of apply the Jeffrey prior in case of multiparameter problem. Berger and Bernardo (1989, 1992 a,b) extended and generalized the algorithm proposed by Bernardo (1979). Datta and Ghosh (1996) simplified the calculation of the reference prior proposed by Berger and Bernardo (1992b) under the same conditions. There are hundreds of literatures on the noninformative priors, e.g. Salaza et al. (2009) introduced the noninformative Bayesian regression analysis under the exponential power (generalized error) error term. Fonseca et al. (2008) developed noninformative Bayesian regression with student-t disturbance term. We propose Noninformative Bayesian Analysis for ESGE Linear Regression Model.

2-Epsilon Skew Generalized Error (ESGE) distribution

Theodossiou (1998) introduced the Skewed Generalized T (GT) distribution that can accommodate the heavy-tails (Leptokurtic) and skewness properties. The p.d.f. of the skewed GT is defined by

$$f(y) = K \begin{cases} \left[1 + \frac{h\lambda^{-h}}{n-2}\left|\frac{y}{\sigma(1+\theta)}\right|^{h}\right]^{\frac{-(n+1)}{h}} & ; y \geq 0 \\ \left[1 + \frac{h\lambda^{-h}}{n-2}\left|\frac{y}{\sigma(1-\theta)}\right|^{h}\right]^{\frac{-(n+1)}{h}} & ; y < 0 \end{cases} \qquad (1)$$

Where h n, θ and σ are scaling parameters and K and λ are normalizing constants.

Theodossiou (1998) considered that the Subbotin's generalized error distribution is a special case of Skewed GT. By reparameterize (1) in a fashion similar to Box and Tiao (1992), Berger et al. (2009), and Mudholkar and Huston (2000) we obtain the following ESGE density function which denoted by $ESGE(\mu, \sigma, p, \varepsilon)$

$$f(y) = \frac{1}{\sigma} \begin{cases} exp\left[-\left(\frac{\Gamma(1+1/p)(y-\mu)}{\sigma(1+\varepsilon)}\right)^{p}\right] & ; \quad y \geq \mu \\ exp\left[-\left(\frac{\Gamma(1+1/p)(y-\mu)}{\sigma(1-\varepsilon)}\right)^{p}\right] & ; \quad y < \mu \end{cases} \qquad (2)$$

Where $-1 < \varepsilon < 1$ is the skewness parameter, $-\infty < \mu < \infty$is the location parameter,

$\sigma_P > 0$ is the scale parameter and p > 0is the shape parameter. Moreover, the density

function (2) known also as the Epsilon skew normal distribution of order p .

Theodossiou (1998) referred that the Kurtosis parameter β_2 of the ESGE is directly linked to the shape and the epsilon parameters. In fact, the shape and the epsilon parameter determines the thickness of the ESGE tail and the shape of the ESGE curve. Some special cases of the ESGE are the epsilon skew Laplace distribution (p =1), the epsilon skew normal distribution (p = 2), the normal distribution (p = = 2, ε=0), the Laplace distribution (p = = 1, ε=0), the uniform distribution (p → ∞ = , ε=0), the generalized error distribution (ε = 0).

From (2) the probability density function p.d.f. and the cumulative distribution function c.d.f. of the standard form $ESGE(0,1,p,\varepsilon)$are respectively:

$$f_0(y) = \begin{cases} exp\left[-\left(\frac{\Gamma(1+1/p)(y)}{\sigma(1+\varepsilon)}\right)^{p}\right] & ; \quad y \geq 0 \\ exp\left[-\left(\frac{\Gamma(1+1/p)(-y)}{\sigma(1-\varepsilon)}\right)^{p}\right] & ; \quad y < 0 \end{cases} \qquad (3)$$

$$F_0(y) = \begin{cases} 1 - \frac{1+\varepsilon}{\Gamma(1/p)}\Gamma\big(1/p\,, g(y)\big) & ; \; y \geq 0 \\ 1 - \frac{1-\varepsilon}{\Gamma(1/p)}\Gamma\big(1/p\,, h(y)\big) & ; \; y < 0 \end{cases} \quad (4)$$

Here, $\Gamma(p, y)$ is incomplete gamma function.

The general form for the density function, denoted $ESGE(\mu, \sigma, p, \varepsilon)$,is$\frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}\right)$,

here $f_0$is defined by (3), and the general form of the c.d.f. of $ESGE(\mu, \sigma, p, \varepsilon)$is $F_0(\frac{y-\mu}{\sigma})$here is$F_0$ defined by (4). The Fisher information matrix of (2) based on Zhu and Zinda-Walsh (2009) is defined as follow

$$I(\mu, \sigma, p, \varepsilon) = \begin{bmatrix} \frac{\Gamma(1/p)\Gamma(2-1/p)}{\sigma^2(1-\varepsilon^2)} & 0 & 0 & \frac{-p}{\sigma(1-\varepsilon^2)} \\ 0 & \frac{p}{\sigma^2} & \frac{-1}{\sigma p} & 0 \\ 0 & \frac{-1}{\sigma p} & \frac{p+1}{p^4}\acute{\Psi}(1+\frac{1}{p}) & 0 \\ \frac{-p}{\sigma(1-\varepsilon^2)} & 0 & 0 & \frac{p+1}{1-\varepsilon^2} \end{bmatrix}$$

3- Noninformative Bayesian Analysis for ESGE Linear Regression Model

Bayesian analysis combines prior information about the parameters of certain model with information from observed data to come up with a posterior distribution. Thus, the researcher needs to determine the likelihood function, as well as the prior distribution which is representing his beliefs before observing the outcomes of the experiment. Usually, it is not easy for the statistician to come up with the distribution of the parameters based on specify prior beliefs, that is, the lack of the subjective beliefs in formulating such a prior .Therefore, to overcome this difficulty noninformative prior is used in place of subjective prior when little or no prior information is available.

Jeffreys (1946, 1961) derived a method to generate noninformative (objective) priors which takes into account the invariant structure under the transformations of the parameters of the model .For $0\epsilon\mathcal{R}^*$The Jeffreys general rule is noninformative prior defined as

$$\pi^*(\theta) \propto [\det(I(\theta)]^{1/2}$$

Where the θ is the vector of the parameters and $I(\theta)$ is the expected value of the Fisher information matrix under a probability distribution which is taken to be $I(\theta) = \{I_{i,j}\}$ where

$$I_{i,j} = -E\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} IogL(\theta)\right), where\ L(\theta) is\ the\ likelihood\ function$$

In multiparameter models with the presence of nuisance parameters, Jeffreys general rule does not work well , see Tibshirani (1989), to overcome this lack of work in Jeffreys general rule prior we consider the methodology of Bernardo (1979) that so called reference prior for multidimensional case (multiparameter) by splitting the parameter vector into nuisance parameters and interest parameters. Berger and Bernardo (1992) extended the idea of dividing the vector of parameter into two or more groups according to their order of inferential importance. The noninformative prior is proportional to the product of an arbitrary function of the nuisance parameters and the jeffreys general rule for the parameter of interest (Tibshirani R. 1989).Datta and Ghosh (1995) simplified the calculation of the reference priors under some conditions, such as the Fisher information matrix is a block diagonal matrix. In this section we shall investigate the following linear regression model,

$$y = x\beta + e$$

$y_i = (y_1, \dots, y_n)'$:Observed variable vector

$e_i = (e_1, \dots, e_n)'$: Error random variable vector such that e 's are independent, identically Distributed according to (2) with e = 0

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 7 Issue 12, December - 2023, Pages: 173-180

$x = (x_1, \dots, x_n)'$: $n \times p$Matrix of regressor variables

$\beta = (\beta_1, \dots, \beta_n)' \in \mathcal{R}^q$:Linear regression coefficient vector

The likelihood function of the parameters is defined by:

$$L(\theta) = \prod_{i=1}^{n} f(y_i / x_i' \beta, \sigma, \varepsilon, p)$$

$$= (\sigma)^{-n} \begin{cases} \exp\left[-\sum_{i=1}^{n} \left(\frac{\Gamma(1+1/p)(y_i - x_i'\beta}{\sigma(1+\varepsilon)} \right)^p \right] ; y \geq x_i'\beta \\ exp\left[-\sum_{i=1}^{n} \left(\frac{\Gamma(1+1/p)(y_i - x_i'\beta}{\sigma(1+\varepsilon)} \right)^p \right] ; y < x_i'\beta \end{cases}$$

Where $\theta = (\varepsilon, p, \beta, \sigma)$

The Jeffrey's General Rule Prior π^{GR} (Jeffreys 1961) is defined by:

$$\pi^{GR}(\theta) \propto [\det(I(\theta)]^{1/2}$$

θ is the vector of the ESEP parameters and $I(\theta)$is the expected value of the Fisher information matrix under ESEP distribution which is taken to be $I(\theta) = \{I_{i,j}\}$ where

$$I_{i,j} = -E\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} IogL(\theta) \right); i,j = 1,2,3,4$$

Proposition 1: The Fisher information matrix I(θ) of the sample $(y_1, y_2, \dots, y_n)$ with standard form of ESGE probability density function (2) , which has the parameter vector, $\theta = (p, \beta, \varepsilon)$is given by

$$I(\theta) = \begin{bmatrix} \frac{n(p+1)}{p^4} \acute{\Psi}(1+\frac{1}{p}) & 0 & 0 \\ 0 & \frac{\Gamma(\frac{1}{p})\Gamma(2-\frac{1}{p}) \sum_{i=1}^{n} x_i x_i'}{(1-\varepsilon^2)} & \frac{-np}{(1-\varepsilon^2)} \\ 0 & \frac{-np}{(1-\varepsilon^2)} & \frac{n(p+1)}{1-\varepsilon^2} \end{bmatrix}$$

Here, Ψ'(.) is the trigamma function.

Proof: This follows as a consequence of proposition 1 of Salazar et al. (2009), and proposition 5 Berger et al. (2009).

We derived some reference priors for the ESGE regression model. We begin with the derivation of the Jeffrey's general rule prior $\pi^{GR}(p, \beta, \varepsilon)$ whichis proportional to the positive square root of the determinant $I(p, \beta, \varepsilon)$. Thus

$$\pi^{GR}(p, \beta, \varepsilon) \propto \sqrt{I_{\beta\beta} I_{\varepsilon\varepsilon} - I_{\beta\varepsilon}^2} \sqrt{\det(I_{PP})}$$

Where

$$\det(I_{PP}) \propto [\left(\frac{P+1}{P^4}\right)\Psi'\left(1+\frac{1}{P}\right)]$$

Then the Jeffrey's general rule is given by

$$\pi^{GR}(p,\beta,\varepsilon) \propto \frac{1}{1-\varepsilon^2}[(p+1)\Gamma\left(\frac{1}{P}\right)\Gamma\left(2-\frac{1}{P}\right)-P^2]^{\frac{1}{2}}[\left(\frac{P+1}{P^4}\right)\Psi'\left(1+\frac{1}{P}\right)]^{\frac{1}{2}}$$

In Jeffrey's general rule prior all the parameters are treated as equally important.

Next we derived two groups $\theta_{(1)} = p$ and $\theta_{(2)} = (\beta, \varepsilon)$ reference priors following Berger and Bernardo (1992), Datta and Ghosh (1996) algorithms. Here p is the parameter of interest and(β, ε) is the vector of nuisance parameters. Then the Fisher information matrix for this grouping is same as first group, and following is reference prior

$$\pi^{2R}(p,(\beta,\varepsilon)) \propto \frac{1}{1-\varepsilon^2}[(p+1)\Gamma\left(\frac{1}{P}\right)\Gamma\left(2-\frac{1}{P}\right)-P^2]^{\frac{1}{2}}[\left(\frac{P+1}{P^4}\right)\Psi'\left(1+\frac{1}{P}\right)]^{\frac{1}{2}}$$

Now, from Tibshirani (1989), Berger and Bernardo (1992), Datta and Ghosh (1996) algorithms, the reference prior is proportional to the square root of the information element for the parameter of interest p times an arbitrary function of the nuisance parameters . Let $\theta_{(1)}=p$, $\theta_{(2)}=\varepsilon$ and $\theta_{(3)}=\beta$. Then, the Fisher information matrix for this grouping and the reference prior are as following

$$I^*(p,\varepsilon,\beta) = \begin{bmatrix} \frac{n(p+1)}{p^4}\acute{\Psi}(1+\frac{1}{p}) & 0 & 0 \\ 0 & \frac{n(p+1)}{1-\varepsilon^2} & \frac{-np}{(1-\varepsilon^2)} \\ 0 & \frac{-np}{(1-\varepsilon^2)} & \frac{\Gamma(\frac{1}{p})\Gamma(2-\frac{1}{p})\sum_{i=1}^{n} x_i x_i'}{(1-\varepsilon^2)} \end{bmatrix}$$

$$\pi^{3G}((p),(\varepsilon),(\beta) \propto [\left(\frac{p+1}{p^4}\right)\Psi'(1+\frac{1}{p}]^{1/2} g(p,\varepsilon)$$

Where $g(p,\varepsilon) = [\frac{1}{1-\varepsilon^2}]^{\frac{1+q}{2}}[\Gamma\left(\frac{1}{p}\right)\Gamma\left(2-\frac{1}{p}\right)]^{\frac{q}{2}}$ is an arbitrary function.

Based on Datta and Ghosh (1996) algorithm, the Jeffrey's general rule prior $\pi^{3G}(p,\beta,\varepsilon)$ and both reference priors $\pi^{2G}(p,(\beta,\varepsilon)), \pi^{3G}((p),(\varepsilon),(\beta))$belong to the class of improper prior distribution given by

$$\pi(\theta) \propto \frac{\pi(p)}{(1-\varepsilon^2)^a}$$

Here $a \in \Re$ is a hyper-parameter and $p(\pi)$ is the marginal prior of p.

4- Simulation Study

This section focused on the simulation study of the Bayesian estimation of the ESGE distribution parameters. Bayesian reference priors have coded by using the R packages. We generate the observation of the interest variable from ESGE distribution with four parameters (β_1,ε, p, σ) based sample size equals to 250 observation , MCMC algorithm have used with 12000 iteration have generate . The MSE criteria have used to judge the parameter estimate accurate. In addition to that 2000 iteration have burn-in also it clearly that we use the following linear model to simulate our data :

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i$$

the initial values of $\beta_1 = 1.2$, $\beta_0 = 1$ and the residuals e_i are generated from $N(0,1)$, $0.75N(0,3) + 0.25N(0,1)$. The following table shows the MSE and its relative efficiency for different residuals distribution.

Table 1 : MSE and the relative efficiency

Model	$\boldsymbol{\beta_1}$ MSE (eff.)	$\boldsymbol{p}$ MSE (eff.)	$\boldsymbol{\varepsilon}$ MSE (eff.)	$\boldsymbol{\sigma}$ MSE (eff.)
$\boldsymbol{\epsilon \sim N(0,1)}$ **ESGE(proposed)** **MLE**	0.01466 (2.15065)	0.033468 (2.1839)	0.01216 (2.0656)	0.0571 (2.0984)
$\boldsymbol{\epsilon \sim 0.75N(0,3)}$ $\boldsymbol{+ 0.25N(0,1)}$ **ESGE(proposed)** **MLE**	0.02558 (2.81213)	0.00986 (2.7349)	0.00654 (1.8933)	0.0422 (1.7986)

The following figures are the trace plot which that tools of MCMC algorithm convergence. The trace plots indicate that the MCMC samples show good convergence to the stationary distribution (the initial value of the true parameters. Also the plots shows there are no flat bits and there are no slow mixing in the trace plot.

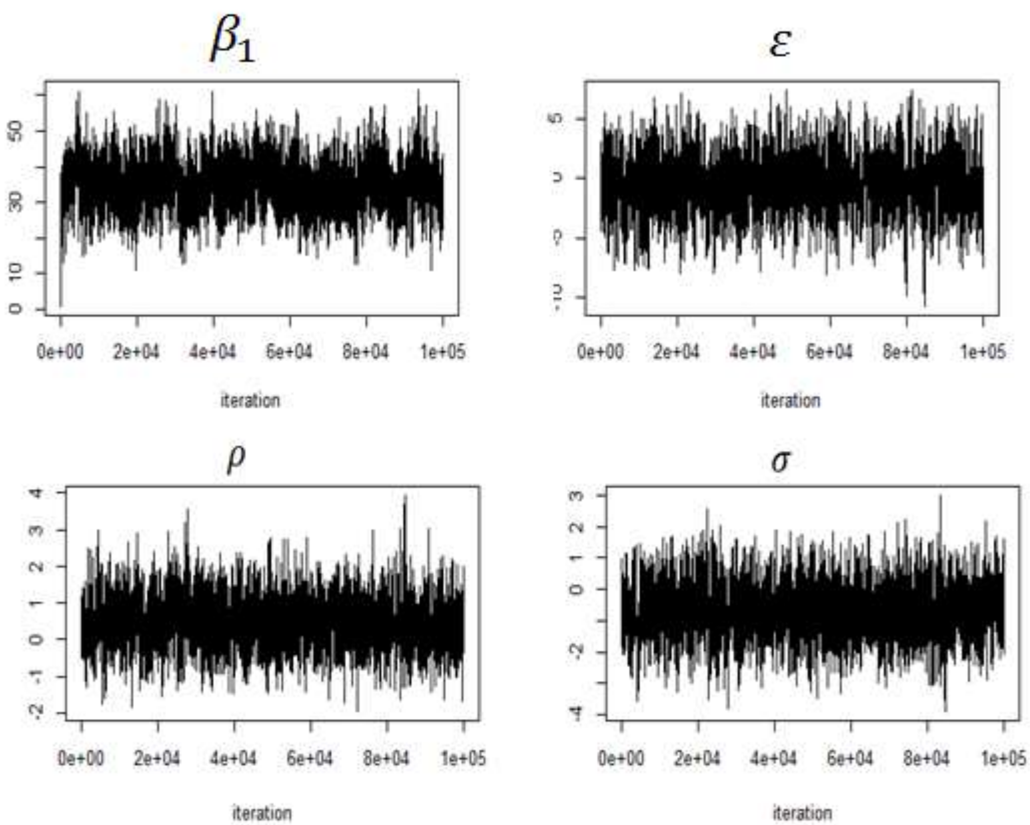

Figure 1. Trace plots of the estimated parameters

Findlay for comparison purpose, we use the BIC criterion for model selection. We compare the proposed model ESGE model with the Laplace distribution to test the goodness of fits for the generated data. The results indicated that the ESGE model fit the data better than the Laplace distribution (the less value of BIC, the better performance).

Table 2 : BIC values of the models

Model	BIC values	Decision
ESGE(proposed)	17.3	Better
Laplace distribution	22.27	

5- Conclusions

The Epsilon skew generalized distribution has studied from the Bayesian prosecution. The Bayesian estimation for the parameters of the proposed model have analyzed by using reference priors to find the posterior distribution mean. The MCMC algorithm has used with R packages and compared the results with Laplace distribution. The BIC shows that the ESGE model performed better.

6- REFERENCES

Azzalini, A., Further results on a class of distributions which includes the normal ones, Statistica, 46,199–208,1986.

Bernardo, J. M., Reference posterior distributions for Bayesian inference, Journal of the Royal Statistical Society B. 41, 113-147, 1979.

Berger,J.O., and Bernardo,J.M., Estimating a product of means: Bayesian analysis with reference priors, JASA,84,200-207,1989

Berger, J. O. and Bernardo, J. M., On the development of reference priors. In Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting. (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith eds.). Clarendon Press: Oxford.35-60,1992a.

Berger, J. O. and Bernardo, J. M., Order group reference priors with application to the multinomial problem, Biometrika.79, 25-37,1992b.

Berger,J.O. , Bernardo,J.M. and Sun,D. The formal definition of reference priors. Ann.. Stat. 37 , 905 -938, 2009 .

Berger,J.O. , Bernardo,J.M. and Sun,D. objective priors for discrete parameter spaces. J.Am. Stat. Assoc. 107 , 636 -648 , 2012 .

Berger,J.O. , Bernardo,J.M. and Sun,D. Overail objective priors (wit discussion).Bayesion Anal. 10(1) 189-221 , 2015 .

Box, G. E. P. and Tiao, G. C., A further look at robustness via Bayes's theorem, Biometrika, 49, 419-432,1962.

Box, G. E. P. and Tiao, G. C., Bayesian Inference in Statistical Analysis. WileyInterscience,1992.

Datta,G.S.,andGhosh,M., On the invariance of noninformative priors, Ann.Stat.,24,141- 159,1996.

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

Fonseca, T.C.O., Ferreira, M.A., and Migon, H., Objective Bayesian analysis for the Student-t regression model,Biometrika .95, 2, 325–333, 2008.

Jeffreys,H., Theory of probability. Oxford University Press.London, 1961. Jeong,J.,andJeong,D., A Test on the Normality in the Tobit Model.(online).2010.

Laplace, P.,Theorie analytique des probabilites. Paris: Courcier, 1812. Mashtare Jr., T., and Huston,A., Utilizing the Flexibility of the Epsilon-Skew-Normal Distribution for Tobit Regression Problems.(online),2008.

Mudholkar, G. S. and Hutson, A. D., The epsilon-skew-normal distribution for analyzing near-normal data, J. Statist.Plan. Infer. 83:291–309,2000.

Mudholkar,G.S., Srivastava, D.K., Exponentiated Weibull family for analyzing bathtub failure data, IEEE Trans.Rel.42, 299-302.1993.

Salazar,E., Ferreira, M.A.,and Migon,H.,Objective Bayesian analysis for exponential power regression models,(online),2009.

Subbotin, M.T., On the Law of Frequency of Error, Mathematicheskii Sbornik, 296–301, 1923. Swets, J. A. and Pickett, R. M., Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press, New York, 1982

Tibshirani , R. Noninformative priors for one for one parameter of many . Biometrika76(3) , (604-608) , 1989 .