European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

BAYESIAN ESTIMATION OF UNKNOWN FORM OF HETEROSCEDASTICITY STRUCTURE

Bolanle, Azeez Oseni

Department of Mathematics and Statistics, The Polytechnic, Ibadan, Nigeria

ABSTRACT: *This paper examines the Bayesian estimation of parameters of linear regression model when the assumption of normality is not tenable. Outlier observations have been traced and identified as one of the factors causing departures from normality assumptions. Thus, the Ordinary Least Squares (OLS) estimates are unbiased but the variances are no longer minimum which can hinder the validity of the inferences to be made about the parameters. Nigerian Stock Exchange and Simulated data were used for illustrations. The finding shows that the posterior mean is unbiased, consistent and similar to the results obtained in homoscedasticity version and the degrees-of-freedom obtained are relatively small and the existence of fat tail is confirmed.*

KEYWORDS: conditional posterior density; non-hierarchical prior; gibbs-sampling; outlier; metropolis-hasting

INTRODUCTION

It is generally accepted that the normal distribution is sensitive to departures from the assumptions, because of its 'thin' tails. Outlier observations have been traced and identified as one of the factors causing departures from normal assumptions. Outlier observations can have a marked impact on inferences. Many alternative 'robust' methods have been developed in the studies of Hamsel et al (1986), Rogers and Tukey (1993), Zellner (1976), Geweke (1993) and Koop (2003) amongst others.

One of the possibilities that have been suggested for addressing departure from normality consists of replacing the Normal process with thick-tail distribution, such as Student-t, either in its univariate or multivariate forms. The Student-t distribution is similar to the Normal distribution, but has fatter tails and is more flexible. In fact, the Normal distribution is a special case of the Student-t which occurs as 'v' degree-of-freedom tends to infinity $(v \to \infty)$. Thus, we have a model that allows for a more flexible error distribution without leaving our familiar Normal linear regression model framework. In Bayesian literature, few authors have worked on the Bayesian analysis of linear regression with Student-t disturbances. Among these is the use of Student-t as the distributional assumption has been an important tool, to Jeffreys (1941) for the case of mean estimation. Fraser (1976, 1979) used this distribution in a linear models, and Maronna (1976) discussed maximum likelihood estimation of the mean and covariance matrix in the same situation. Bayesian analysis with independent Student-t linear model was proposed by Geweke (1993) using a Gibbs sampler to sample when the degree-of -freedom of Student-t disturbance is unknown. Geweke (1993) and Fernandez and Steel (201b) specified non-informative prior for β and precision, σ^2 and prior specification for λ and v_λ as gamma and chi-square distributions respectively.

Toshiaki (2001) used the Metropolis-Hasting acceptance-rejection algorithm proposed by Tierney (1994) to sample the degrees-of-freedom from conditional probability. Koop (2003) in his work specified hierarchical prior for degrees-of-freedom v. However, in literature the posterior odds ratio favours Student-t linear models degrees-of-freedom in the range 3 to 7 over normal linear model.

Pati and Dunson (2014) consider the problem of robust Bayesian inference on the mean regression function allowing the residual density to change flexibility with predictors. The proposed class of the model is based on a Gaussian for the collection of residual densities indexed by predictors with heteroscedasticity arising as a special case. Different priors were considered such as the special case in which the residual distribution follow a homoscedastic Student-t distribution with unknown degrees-of-freedom by placing hyper prior on the degrees-of-freedom $V_{\sigma}(a_v, b_v)$, with $G(a \ ,b)$ denoting the gamma distribution with mean a/b, one can obtain a data adaptive approach to down-weighting outliers in estimating the mean regression function with the Student-t low degrees of freedom is heavy-tailed, outliers are allowed.

Andrew et al. (2008) propose a new proper prior distribution that produces stable, regularised estimates while still being vague enough to be used as a default in routine applied work. Their procedure can be seen as a generalisation of the scale prior of Raftery (1996) to the t-case, with the additional innovation that the prior scale parameter is given a direct interpretation in terms of logistic regression parameters. Andrew et al (2008) are motivated to consider the t-family because flat-tailed distribution allows for robust inference (see, Berger and Berliner (1986); Lange et al (1989) because it allows easy and stable computation in logistic regression by placing iteratively weighted least squares within an approximate Expectation Maximization (EM)algorithm.

The independence Jeffreys prior is widely used in scale mixtures of normals contain some important distributions such as normal, Student-t with v degrees of freedom, logistic, Laplace, Cauchy and exponential power family with power $1 \le q \le 2$. Thus, for this wide and practically important class of distributions the two-piece model with the independence Jeffrey prior leads to valid inference in any sample of two or more observations (Rubio and Steel, 2011). Yeojin et al (2000) recommend a class of weakly informative prior densities for Ω that go to zero in the boundary as Ω becomes degenerate, thus ensuring that the posterior mode (i.e. the maximum penalized likelihood estimate) is always non-degenerate. They recommend a class of Wishart priors with a default choice of hyper-parameters. The degrees of freedom is the number of varying coefficients plus two and the scale matrix is the identity matrix multiplied by a large enough number.

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

Another study worthy of mentioning is the work of Geweke (1993) and Koop (2003) on treatment of Student-t in linear regression model. Geweke (1993) specified $\chi^2_{(v)}$ for hyper-hyper-parameter v the degrees of freedom while Koop (2003) specified hierarchical priors for degrees of freedom v. This study seeks to extend the work of Koop (2003) by specifying non-hierarchical priors for degrees of freedom as equivalent to hierarchical priors adopted in Koop (2003).

The structures of the paper are as follows. In section 2 we describe an overview of Ordinary Least Squares (OLS) estimation of parameters and Bayesian estimation approach when the error terms are heteroscedasticity of unknown form. Inference in the presence of heteroscedasticity of unknown form are developed in section 3. Data generation, model estimation and discussion of results are in section 4. Concluding remarks are in section 5.

Overview of Ordinary Least Squares (OLS)

The model specification is given by

$$y = X\beta + U; \qquad U \sim MVN(0, \sigma^2 I_n) \tag{1}$$

Where y is vector of responses arranged in a vector $y_{n\times 1} = (y_1, y_2, \cdots, y_n)'$ and a matrix $X_{n\times k} \equiv (X_1, X_2, \cdots, X_n)'$ and the rank of X is k.

The least squares estimator of the linear regression model in Eq. (2) seeks to minimize the residual sum of squares in the model in Eq. (1)

$$SSE = \sum_{i=1}^{n} (y_i - X_i\hat{\beta})^2 \tag{2}$$

The estimated vector $\hat{\beta}$ that minimizes β is obtained by differentiating Eq. (2) with respect to β and equating to zero to have

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3}$$

and the estimated value of s^2 is computed by

$$s^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n-k} \qquad ; v = n - k \tag{4}$$

Bayesian Estimation

The Likelihood

Consider the model in Eq. (1) using the definition of the multivariate normal density, we can write likelihood function as:

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

$$P(y|\beta,h)=\frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\left\{\exp\left[-\frac{h}{2}(y-X\beta)^{T}(y-X\beta)\right]\right\} \qquad (5)$$

For derivation, it proves convenient to re-write the likelihood in a slightly different way. From Eq. (3) we have

$$(y-X\beta)^{T}(y-X\beta)=vs^{2}+(\beta-\hat{\beta})^{T}X^{T}X(\beta-\hat{\beta}) \qquad (6)$$

For many technical derivations, it is easier and convenient to work with error precision defined as $h=\sigma^{-2}$ rather than variance. Using the result in (6), the likelihood in Eq. (5) can be written as

$$P(y|\beta,h)=\frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\left\{\exp\left[-\frac{h}{2}\left\{vs^{2}+(\beta-\hat{\beta})^{T}X^{T}X(\beta-\hat{\beta})\right\}\right]\right\} \qquad (7)$$

Eq. (7) can be separated into two by setting $n=v+k$ which leads to

$$p(y|\beta,h)=\frac{1}{(2\pi)^{\frac{N}{2}}}\left\{h^{\frac{k}{2}}\exp\left[-\frac{h}{2}(\beta-\hat{\beta})^{T}X^{T}X(\beta-\hat{\beta})\right]\right\}\left\{h^{\frac{v}{2}}\exp\left[\frac{hv}{2s^{-2}}\right]\right\} \qquad (8)$$

The quantity $\left\{h^{\frac{k}{2}}\exp\left[-\frac{h}{2}(\beta-\hat{\beta})^{'}X^{'}X(\beta-\hat{\beta})\right]\right\}$ in Eq. (8) resembles the kernel of the multivariate normal density and $\left\{h^{\frac{v}{2}}\exp\left[-\frac{hv}{2s^{-2}}\right]\right\}$ also looks like the kernel of the gamma density.

A Scale Mixture of Normal

Suppose a random vector has the conditional multivariate normal distribution with probability density function:

$$y|X\beta,\Sigma,\omega \sim N\left(y|X\beta,\frac{\Sigma}{\omega}\right) \qquad (9)$$

Where ω, in turn, is a scalar random variable following a

$$\omega|v \sim G\left(\frac{v}{2},\frac{v}{2}\right) \qquad (10)$$

Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

process; here $v > 0$ is a parameter. The density of the joint distribution of y and ω is then

$$p(y,\omega \mid X\beta,\Sigma,v) = \left| 2\pi(\frac{\Sigma}{\omega}) \right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - X\beta)^T \Sigma^{-1}(y - X\beta)\right\}$$

$$\frac{(v/2)^{\frac{v}{2}}}{\Gamma(v/2)} \omega^{\frac{v}{2}-1} \exp\left\{-\frac{v\omega}{2}\right\} \qquad (11)$$

The joint density was integrated with respect to ω to hsave

$$p(y,\omega \mid X\beta,\Sigma,v) \propto \frac{(v/2)^{\frac{v}{2}}}{\Gamma(v/2)} \omega^{\frac{v}{2}-1} \exp\left\{-\frac{v\omega}{2}\right\} \int_0^{\infty} \omega^{\frac{v}{2}-1} \exp\left\{-\omega \frac{(y - X\beta)^T \Sigma^{-1}(y - X\beta)}{2}\right\} \partial\omega$$

(12)

Eq. (12) indicates that the integrand is the kernel of the density

$$G\left(\omega \mid \frac{N+v}{2}, \frac{(y - X\beta)^T \Sigma^{-1}(y - X\beta) + v}{2}\right)$$

Hence, the integral in Eq. (12) is equal to the reciprocal of the integration constant of the corresponding distribution, that is

$$\int_0^{\infty} \omega^{\frac{v}{2}-1} \exp\left\{-\omega \frac{(y - X\beta)^T \Sigma^{-1}(y - X\beta) + v}{2}\right\} \partial\omega$$

$$= \frac{\Gamma(\frac{N+v}{2})}{\left[\frac{(y - X\beta)^T \Sigma^{-1}(y - X\beta)}{2}\right]^{\frac{n+v}{2}}} \qquad (13)$$

Employing Eq. (12) and Eq. (13), and rearranging we then obtain

$$p(y \mid X\beta,\Sigma,v) = \frac{(v)^{\frac{v}{2}} \Gamma(\frac{N+v}{2})}{\Gamma(\frac{v}{2}) \mid v\pi\Sigma \mid^{\frac{1}{2}}} \left[(y - X\beta)^T \Sigma^{-1}(y - X\beta)\right]^{\frac{N+v}{2}}$$

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

$$\frac{\Gamma(\frac{N+v}{2})}{\Gamma(\frac{v}{2})\,|v\pi\Sigma|^{\frac{1}{2}}}\left[1+\frac{(y-X\beta)^T\Sigma^{-1}(y-X\beta)}{v}\right]^{\frac{N+v}{2}} \quad (14)$$

This is the density of n-dimensional multivariate t-distribution with mean vector $X\beta$, scale matrix Σ, and degrees of freedom parameter v.

Student-t linear regression model

$$y_i \mid X \sim t(X_i\beta, h^{-1}, v) \quad (15)$$

Where v is the degree of freedom parameter and v is part of the specification of the model. The model can be expressed as;

$$f(y\mid\beta,h,v)=\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\Gamma v\pi h^{-1}}\left[1+\frac{(y-X_i^T\beta)^2}{vh^{-1}}\right]^{-\frac{(N+v)}{2}} \quad (16)$$

We can write the Student-t likelihood function as

$$p(y\mid\beta,h,v \sim \prod_{i=1}^{N} t(y_i, X_i^T\beta, h, v) \quad (17)$$

$$p(y\mid\beta,h,v)=\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\Gamma v\pi h^{-1}}\prod_{i=1}^{N}\left[1+\frac{(y-X_i^T\beta)^2}{vh^{-1}}\right]^{-\frac{(N+v)}{2}}$$

The full likelihood of the model can be expressed as

$$p(y\mid\beta,h,v)=\frac{\Gamma((v+1)/2)}{\Gamma(v/2)\Gamma v\pi}h^{\frac{N}{2}}\prod_{i=1}^{N}\left(1+\frac{(y-X_i^T\beta)}{vh^{-1}}\right)^{-\frac{(v+N)}{2}} \quad (18)$$

An equivalent specification of Eq. (9) is

$$y_i \mid X \sim N(X\beta, h^{-1}\omega_i) \quad (19)$$

Where, $\beta_{k\times1}\equiv(\beta_{1,}\beta_2,\cdots,\beta_\kappa)'$ is a vector of unknown parameters, $\omega_{k\times1}\equiv(\omega_{1,}\omega_2,\cdots,\omega_\kappa)'$ is a vector of unknown parameters and h is unknown parameter.

$$y_i = X_i^T\beta + U_i; \quad U_i \sim N(0, h^{-1}\omega_i) \quad (i=1,2,\cdots,N)$$

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

It is more convenient to work with error precision rather than variances and, hence, we define;

$$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_N)' \equiv (\omega_1, \omega_2, \cdots, \omega_N)'$$

$$\text{var}(U) = h^{-1}\omega_i; \qquad \Sigma \equiv diag(\omega_1, \omega_2, \cdots, \omega_N)$$

The likelihood function of the Eq.(19) can be expressed as;

$$p(y \mid \beta, h, \lambda) = h^{\frac{N}{2}} \prod_{i=1}^{N} \lambda_i^{\frac{1}{2}} \exp\left[-\frac{\lambda_i (y_i - X_i^T \beta)^2}{2h^{-1}} \right] \tag{20}$$

The likelihood function in Eq. (7), can be re- written as;

$$p(y \mid \beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{k}{2}} \exp\left[-\frac{h}{2} (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta}) \right] \right\} \left\{ h^{\frac{v}{2}} \exp\left[\frac{hv}{2s^{-2}} \right] \right\} \tag{21}$$

The right hand side and left hand side of Eq. (21) follows Normal and Gamma densities respectively.

The Priors and their distributions

The likelihood in Eq.(21) suggests that Normal-Gamma prior could be used for the parameter β and h.

Prior for β condition on h is of the form:

$$\beta \mid h \sim N(\underline{\beta}, h^{-1}\underline{\Sigma})$$

Prior for h is of the form

$$h \sim G(\underline{s}^{-2}, \frac{2\underline{s}^{-2}}{\underline{v}})$$

$$p(\beta, h) = \frac{h^{\frac{\underline{v}+k}{2}-1}}{(2\pi)^{\frac{k}{2}} \mid \underline{\Sigma} \mid^{\frac{1}{2}} \Gamma(\frac{\underline{v}}{2})(\frac{2\underline{s}^{-2}}{\underline{v}})^{\frac{\underline{v}}{2}}} \left\{ \exp - \frac{h}{2} \left[(\beta - \underline{\beta})^T (\underline{\Sigma})^{-1} (\beta - \underline{\beta}) + \frac{\underline{v}}{\underline{s}^{-2}} \right] \right\} \tag{22}$$

Eq.(22) is the normal-Gamma prior, where, $\underline{\beta}$ and $\frac{1}{\Gamma(\frac{\underline{v}}{2})(\frac{2\underline{s}^{-2}}{\underline{v}})^{\frac{\underline{v}}{2}}}$ are the prior for β and integrating constant respectively.

The above Eq.(22) is the Normal-Gamma prior defined as;

$$p(\beta,h) = f_{NG}(\beta,h \mid \underline{\beta},\underline{\Sigma},\underline{s}^{-2},\underline{v})$$

We specify prior for degrees-of-freedom in two steps. The first step is $p(v_\lambda)$ defined by

$$p(\lambda) = \prod_{i=1}^{N} f_G(\lambda_i \mid 1, v_\lambda) \tag{23}$$

with density function given as;

$$p(\lambda) = \prod_{i=1}^{N} cG^{-1} \lambda_i^{\frac{v_\lambda - 2}{2}} \exp\left(-\frac{\lambda_i v_\lambda}{2}\right) \tag{24}$$

where the integrating constant given by $cG^{-1} = \left(\frac{v_\lambda}{2}\right)^{\frac{v_\lambda}{2}} \Gamma\left(\frac{v_\lambda}{2}\right)^{-N}$

while the second step is $p(v_\lambda)$ given by

$$p(v_\lambda) = f_G(v_\lambda \mid v_{0\lambda}, 2) \tag{25}$$

with density function given as

$$p(v_\lambda) = cG^{-1} \exp\left(\frac{v_\lambda}{v_{0\lambda}}\right) \tag{26}$$

where the integrating constant is cG^{-1}. Exponential distribution which is simply the Gamma with 2 degrees-of-freedom was specified.

The Posterior distributions

The posterior is proportional to the product of likelihood and the prior is of the form

$$p(\beta,h,\Sigma \mid y) \propto \frac{h^{\frac{\underline{v}+k}{2}-1}}{(2\pi)^{\frac{k}{2}} |\underline{\Sigma}|^{\frac{1}{2}} \Gamma\left(\frac{\underline{v}}{2}\right)\left(\frac{2\underline{s}^{-2}}{\underline{v}}\right)^{\frac{\underline{v}}{2}}} \left\{\exp - \frac{h}{2}\left[(\beta-\underline{\beta})^T(\underline{\Sigma})^{-1}(\beta-\underline{\beta}) + \frac{\underline{v}}{\underline{s}^{-2}}\right]\right\}$$

$$\frac{1}{(2\pi)^{\frac{N}{2}}}\left\{h^{\frac{k}{2}} \exp\left[-\frac{h}{2}(\beta-\hat{\beta})^T X^T \Sigma^{-1} X(\beta-\hat{\beta})\right]\right\}\left\{h^{\frac{\underline{v}}{2}} \exp\left[\frac{h\underline{v}}{2\underline{s}^{-2}}\right]\right\} \tag{27}$$

The posterior density of β is

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

$$P(\beta \mid h,\Sigma,y) \propto \exp\left[-\frac{h}{2}\left\{vs^2(\Sigma)+(\beta-\hat{\beta}(\Sigma))^T X^T\Sigma^{-1}X(\beta-\hat{\beta}(\Sigma))\right\}\right]\exp\left[-\frac{h}{2}(\beta-\underline{\beta})^T(\underline{\Sigma})^{-1}(\beta-\underline{\beta})\right]$$

Where

$$\beta_n=[\underline{\Sigma}^{-1}+hX^T\Sigma^{-1}X]^{-1}[\underline{\Sigma}^{-1}\underline{\beta}+hX^T\Sigma^{-1}X\hat{\beta}]$$

and

$$\Sigma_n=[\underline{\Sigma}^{-1}\underline{\beta}+hX^T\Sigma^{-1}X]^{-1}$$

$$\beta_n=\Sigma_n[\underline{\Sigma}^{-1}\underline{\beta}+hX^T\Sigma^{-1}X\hat{\beta}]^{-1}$$

So that β is sampled from

$$\beta \mid h,\Sigma,y \sim N(\beta_n(\Sigma),\Sigma_n)$$

The posterior density of h is

$$p(h \mid \beta,\Sigma,y) \propto h^{\frac{N+v_0}{2}-1}\exp\left\{-\frac{h}{2}\left[(y-X\beta)^T\Sigma^{-1}(y-X\beta)+\underline{v}\underline{s}^2\right]\right\} \quad (28)$$

$$v_n=N+\underline{v}$$

$$v_n s_n^2=(y-X\beta)^T\Sigma^{-1}(y-X\beta)+\underline{v}\underline{s}^2$$

$$s_n^2=\frac{(y-X\hat{\beta})^T\Sigma^{-1}(y-X\hat{\beta})+\underline{v}\underline{s}^2}{v_n}$$

then

$$s_n^{-2}=\frac{v_n}{(y-X\hat{\beta})^T\Sigma^{-1}(y-X\hat{\beta})+\underline{v}\underline{s}^2} \quad (29)$$

Therefore h is sampled from

$$h \mid \beta,\Sigma,y \sim G[s_n^{-2},v_n]$$

The posterior density of λ is

$$p(\lambda_i \mid \beta, h) = \prod_{i=1}^{N} \lambda_i^{\frac{1}{2}} \exp\left[-\frac{\lambda_i (y_i - X_i^T \beta)^2}{2h^{-1}} \right] \prod_{i=1}^{N} \lambda_i^{\frac{v_\lambda - 2}{2}} \exp\left\{ -\frac{\lambda_i v_\lambda}{2} \right\}$$

$$p(\lambda_i \mid \beta, h) = \prod_{i=1}^{N} \lambda_i^{\frac{v-1}{2}} \exp\left[-\frac{\lambda_i (y_i - X_i^T \beta)^2}{2h^{-1}} + \frac{v_\lambda}{2} \right]$$

$$p(\lambda_i \mid \beta, h) = \lambda_i^{\frac{v-1}{2}} \exp\left[-\frac{\lambda_i \sum_{i=1}^{N} (y_i - X_i^T \beta)^2}{2h^{-1}} + \frac{v_\lambda}{2} \right] \tag{30}$$

$$\lambda_i = G\left[-\frac{v+1}{2} + \frac{hU_i^2}{2} + \frac{v_\lambda}{2} \right] \qquad where \quad U_i^2 = \sum_{i=1}^{n} (y_i - X_i^T \beta)^2$$

Establishing $p(v_\lambda \mid y, \beta, h, \lambda)$ is relatively straightforward, since v_λ does not enter the likelihood and it can be confirmed that $p(v_\lambda \mid y, \beta, h, \lambda) = p(v_\lambda \mid \lambda)$ via Bayes theorem:

$$p(v_\lambda \mid \lambda) \propto \prod_{i=1}^{N} cG^{-1} \lambda_i^{\frac{v_\lambda - 2}{2}} \exp\left(-\frac{\lambda_i v_\lambda}{2} \right) \exp\left(\frac{v_\lambda}{\underline{v}_\lambda} \right)$$

Hence, the posterior density of v_λ is given as

$$p(v_\lambda \mid y, \beta, h, \lambda) \propto \left(\frac{v_\lambda}{2} \right)^{N v_\lambda} \Gamma\left(\frac{v_\lambda}{2} \right)^{-N} \exp(-\eta v_\lambda); \tag{31}$$

$$\eta = \frac{1}{v_{0\lambda}} + \frac{1}{2} \sum_{i=1}^{N} \left\{ (\log \lambda_i^{-1} + \lambda_i) \right\}$$

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

Inferences of Heteroscedasticity of Unknown Form

To carry out Bayesian inference in the presence of heteroscedasticity of unknown form described above, a posterior simulator known as Gibbs sampling and Metropolis-Hasting are required. The posterior for β , h, λ and v_λ are simply those derived in Eq.(27), (28) and (31).

i. $p(\beta \mid h, \Omega, y)$ is sampled from Normal distribution

$$\beta \mid h, \Sigma \sim N(\beta_n(\Sigma), \Sigma_n)$$

ii. $p(h \mid \beta, \Omega, y)$ is sampled from Gamma distribution

$$h \mid y, \beta \sim G[s_n^{-2}, v_n]$$

iii. $p(\lambda_i \mid \beta, h, y)$ which depends on α is sampled from density below

$$\lambda_i = G\left[\frac{v+1}{2}, \frac{hU_i^2}{2} + \frac{v_\lambda}{2}\right] \qquad ; U_i^2 = \sum_{i=1}^{N}(y_i - X_i^T\beta)^2$$

$p(v_\lambda \mid \beta, h, y)$ does not take the form of any convenient density. Nevertheless, a

Metropolis-Hasting Algorithm was employed to have a complete posterior simulator.

$$p(v_\lambda \mid y, \beta, h, \lambda) \propto \left(\frac{v_\lambda}{2}\right)^{Nv_\lambda} \Gamma\left(\frac{v_\lambda}{2}\right)^{-N} \exp\left(-\eta v_\lambda\right) \;\; ;$$

$$\eta = \frac{1}{\underline{v}_\lambda} + \frac{1}{2}\sum_{i=1}^{N}\left\{(\log \lambda_i^{-1} + \lambda_i)\right\}$$

Data Generation Process and Model Estimation

In an attempt to estimate the parameters of linear regression model in presence of heteroscedasticity of known form, we adopted Markov-Chain Monte Carlo (MCMC) experiment. The Monte-Carlo experiment is carried out as follows;

Empirical Illustration

The data used for empirical application is a daily data operation of the Nigeria Stock Exchange market and comprises: Nigeria Stock Exchange All Share Index (ASEASI), Earning Per Share

(EPS), Return on Asset (ROA) and Earnings before Interest Taxes and Management (EBITM). The dependent variable is NSEASI, the explanatory variables X_1, X_2, X_3 *and* X_4 are EPS, ROE, ROA and EBITM respectively and the number of observation is 545.

We set $s^2 = 4x10^{-8}$ and $v = 5$ for hyper parameters and prior for the regression coefficient β and co-variance matrix Σ

are

$$\beta = \begin{bmatrix} 0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 2.40 & 0 & 0 & 0 & 0 \\ 0 & 6.0x10^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 & 0 \\ 0 & 0 & 0 & 0.60 & 0 \\ 0 & 0 & 0 & 0 & 0.6 \end{bmatrix}$$

. Finally, we employed a Metropolis within Gibbs algorithm to draw posterior results for β and α . Inferences on parameters are based on 25000 draws with 5000 burn-in are discarded.

Simulated Illustration

i. Consider the model

$$y = X\beta + U_i ; \quad U \sim N(0, \sigma_i^2)$$

Where y is the dependent variable and X is the explanatory variables, β is the co-efficient and U_i are normally and independently distributed with $E(U_i) = 0$ and $E(U_i^2) = \sigma_i^2$

ii. We generate error term ε_i which are normally and independently distributed with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma_i^2$ for $i = 1, 2, ..., N$ i.e. $\varepsilon_i \sim N(0,1)$

iii. We specify the variance-covariance matrix for the error terms: diagonal $N \times N$ matrix, with the squared OLS residuals, with 'robust standard errors' are obtained by taking the square root estimated variance-covariance matrix. $\Sigma = PP'$.
Since Σ is a symmetric positive definite matrix, we decompose it by a non-singular matrix P such that:

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

$$P = \begin{bmatrix} \sqrt{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_N^2} \end{bmatrix}$$

iv. The error term U is generated by $U = P\varepsilon$. The error term has a scalar covariance matrix property.

v. The coefficients are set at $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (2, 4, 6, 8, 10)$, we also generate explanatory variables X_1, X_2, X_3, X_4 from uniformly distributed from 0 to 10 i.e. $X_i \sim U(0,10)$ for $i = 1, 2, 3, 4$.

vi. Given $\beta's$, (X_1, X_2, X_3, X_4) and U, we then obtain y.

vii. Finally, we apply Bayesian method to the model for different sample sizes $n = 25, 50, 100, 150$ and 200.

The focus of the paper is on the estimation of parameters of linear regression model in the presence of heteroscedasticity of unknown form. However, the interest is on the following conditional distributions of β, h and α. Finally, we employed a Metropolis within Gibbs algorithm to draw posterior results for β and v. Inferences on parameters are based on 25000 draws with 5000 burn-in are discarded.

Table 4.1:Posterior mean for or β, and v_λ, Std. devs. and HPDI's for Nigeria Stock Exchange Data

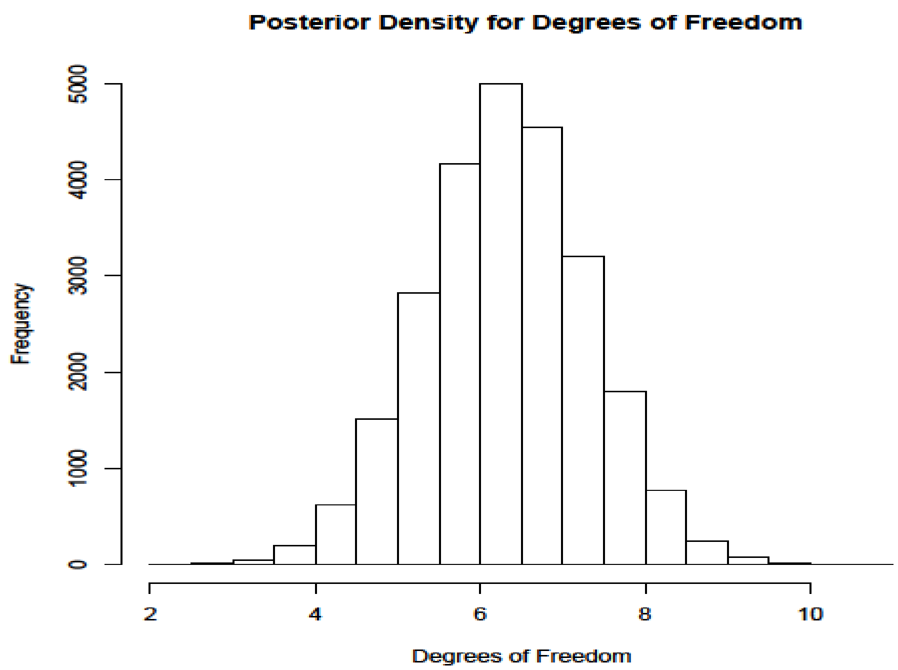
Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	22423.49	559.1550	[21524.31	23332.55]
β_1	-0.35961	0.09998	[-0.52451	-0.195729]
β_2	-4108.37	316.0074	[-4632.937	-3591.609]
β_3	23361.13	1849.283	[20330.87	26421.498]
β_4	440.8911	63.6679	[338.9292	541.96880]
h	$1.668e^{-11}$	$2.638e^{-09}$	[0.000000	0.000000]

The Table above shows the posterior means for β, h and standard deviation and 95% credibility intervals

Table 4.2: Posterior mean for or β, and v_λ, Std. devs. and HPDI's for Nigeria Stock Exchange Daata

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	22425.48	554.08060	[21511.55	23343.98]
β_1	-0.3592	0.10160	[-0.526800	-0.19310]
β_2	-4106.66	318.1138	[-4634.366	-3584.652]
β_3	23343.230	1861.685	[20282.09	26420.85]
β_4	440.5800	62.2382	[338.0367	542.942]
v_λ	5.0100	1.0110	[3.348500	6.66360]

The Table above shows the posterior means for β, v_λ , standard deviation and 95% credibility intervals

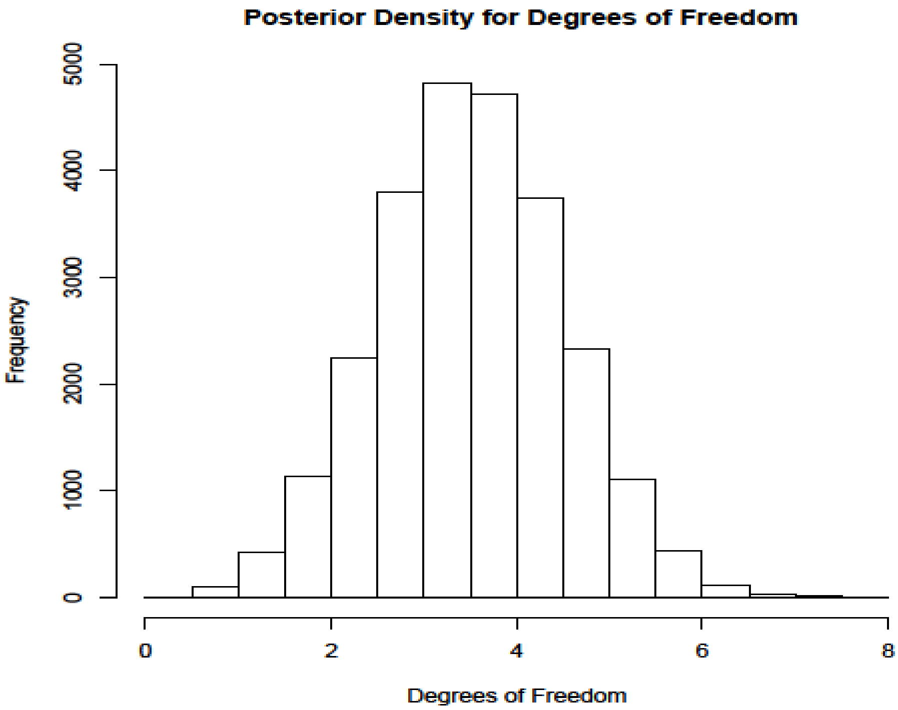


Posterior Density for Degree-of-freedom

Table 4.3: Posterior mean for or β, and v_λ, Std. devs. and HPDI's for simulated Daata n=25

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	0.0690	0.72140	[0.741300	4.238800]
β_1	3.8123	0.0718	[3.849800	4.17420]
β_2	6.1602	0.0679	[5.764100	6.0721]
β_3	8.2597	0.0759	[7.651900	7.99220]
β_4	10.1422	0.0759	[9.751500	10.0860]
v_λ	3.5062	0.9959	[1.448200	4.64250]

Posterior mean for or β, and v_λ, Std. devs. and HPDI's for simulated Daata n=25

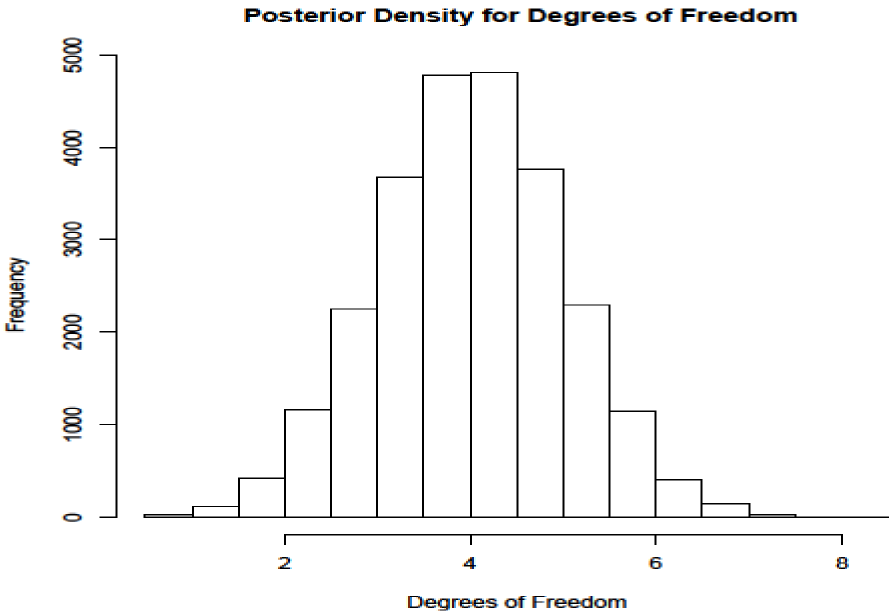


Posterior Density for Degree-of-freedom

Table 4.4: Posterior mean for or β, and v_{λ}, Std. Devs. and HPDI's for simulated Daata n=50

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	2.4641	0.5092	[1.621200	3.2960]
β_1	4.0299	0.1294	[3.817500	4.25180]
β_2	5.9304	0.1287	[5.722100	6.15450]
β_3	7.8328	0.1604	[7.527700	8.1172]
β_4	9.9315	0.1266	[9.724100	10.1506]
v_{λ}	3.5062	0.9886	[1.882800	5.1`542]

The Table above shows the posterior means for β, v_{λ} standard deviation and 95% credibility intervals



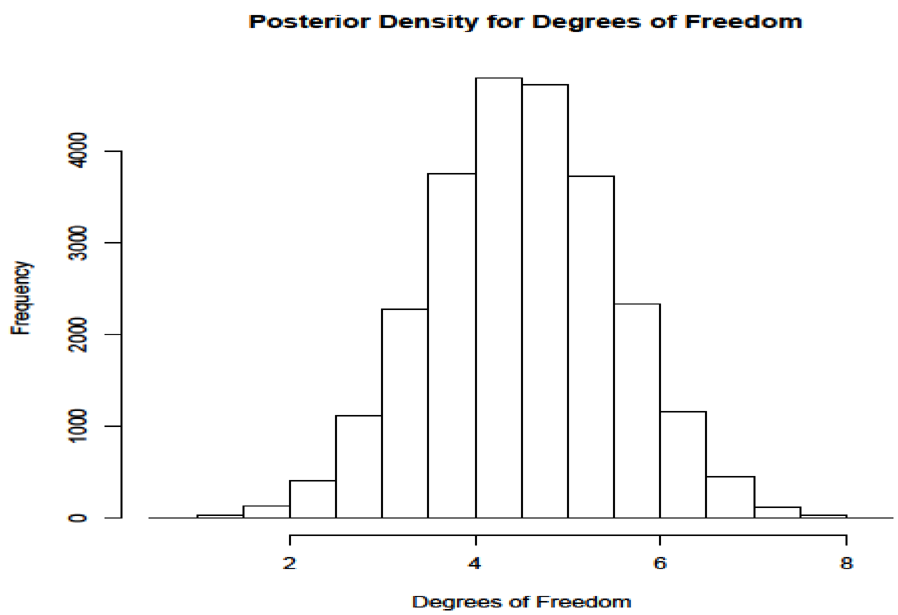
Posterior Density for Degree-of-freedo

Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

Table 4.5: Posterior mean for or β, and v_λ, Std. devs. and HPDI's for Simulated Daata n=100

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	2.206	0.4788	[1.122000	3.812200]
β_1	4.0029	0.0531	[3.847600	4.192800]
β_2	5.9743	0.0469	[5.753500	6.08300]
β_3	8.1095	0.0492	[7.660000	7.99090]
β_4	9.8708	0.0530	[9.756100	10.0889]
v_λ	4.5069	1.0049	[2.347200	5.63930]

The Table above shows the posterior means for β, v_λ, standard deviation and 95% credibility intervals

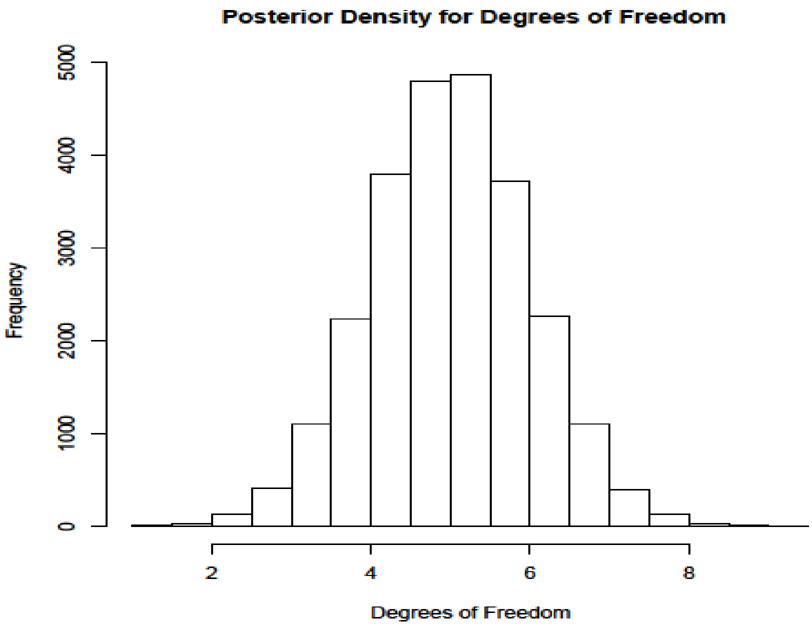


Posterior Density for Degree-of-freedom

Table 4.6: Posterior mean for or β, and v_λ, Std. devs. and HPDI's for Simulated Daata n=150

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	1.6222	0.3104	[1.586200	3.31670]
β_1	3.8983	0.0333	[3.900700	4.4177]
β_2	6.0513	0.0333	[5.783300	6.31240]
β_3	7.9800	0.0346	[7.677400	8.19390]
β_4	10.1063	0.0333	[9.789600	10.29480]
v_λ	4.9987	0.9978	[2.866700	6.166100]

The Table above shows the posterior means for β, v_λ ,standard deviation and 95% credibility intervals



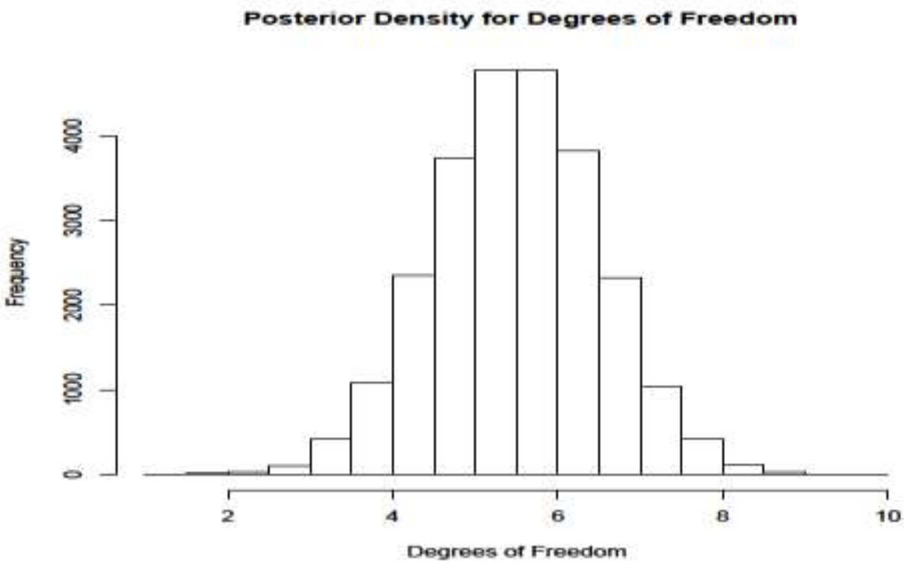
Posterior Density for Degree-of-freedom

European Journal of Statistics and Probability
Vol.8, No.3, pp, 1-21, October 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

Table 4.7: Table 6: Posterior mean for or β, and v_λ, Std. devs. and HPDI's for Nigeria Stock Exchange Daata n=200

Parameters	Means	Std. Dev.	95%HPDI 2.5%	97.5%
β_0	2.2934	0.3272	[1.536500	3.3918]
β_1	4.0567	0.0359	[3.875300	4.2817]
β_2	5.9257	0.0360	[5.777900	6.17760]
β_3	7.9462	0.0351	[7.676800	8.06510]
β_4	10.0082	0.0317	[9.689800	10.2804]
v_λ	5.4976	0.9941	[3.332800	6.63870]

The Table above shows the posterior means for β, v_λ, standard deviation and 95% credibility intervals



Posterior Density for Degrees-of-freedom

DISCUSSION OF RESULTS

This study employs method for Bayesian inference in a linear regression model when the assumption of normality is not tenable. The normal process was replaced with mixture of normal- Student-t linear regression model which is equivalent to treatment of heteroscedasticity of unknown form considered in Geweke (1993) and Koop (2003). The computations are based on Gibbs sampling and Metropolis-Hasting algorithms. We adopt Normal-Gamma prior for parameter β and precision, h as suggested by the likelihood itself. It was shown in this methodology how to use non-hierarchical prior as against hierarchical prior for degrees-of-freedom parameter of Student-t linear model.

The conditional posterior distribution of β, h, λ and v_λ were obtained using Gibbs sampling and Metropolis-Hasting algorithms as a relief to the difficulties of sampling directly from the intractable joint posterior distribution. More importantly, this methodology was applied to Nigeria Stock Exchange data and simulated data using different sample sizes $n = 25, 50, 100, 150$ and $200.$ We specify the initial values for parameter β to be $2, 4, 6, 8$ and 10.

The posterior mean for β in tables 4.3 to 4.7 are similar and consistent with the initial values specified and with the homoscedasticity version in tables 4.1 to 4.2 The estimated coefficients of the parameters approximately 95% draws fall within each of the corresponding credible intervals. The histogram description of the posterior for degrees-of-freedom indicates the error exhibit substantial deviations from normality and all the posterior values obtained for degrees of freedom the parameter of Student-t linear model is in line with values in the ranges 3 to 7 as specified in the literature.

CONCLUSION

Generally, we inferred from the various results obtained that Bayesian estimation of parameters of a linear regression model when the assumption of normality is not tenable using mixture of normals which is also equivalent to treatment of heteroscedasticity of unknown form performed creditably well like its counterpart classical frequentist technique for modelling the linear regression. However, we also conclude that Bayesian estimation of parameters of linear regression model in the presence of heteroscedasticity of unknown form improve the precision of the estimates amongst others.

References

Andrews, G., Aleks, J., Maria, G.P and Yu-Sung SU (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics.* Vol 2. No. 4, 1360-1383

Berger, J.O. and Berliner, L.M. (1986). Robust Bayes analysis with epsilon-contaminated priors. *Ann. Statist.* 14. 461-486. MR0840509.

Fernandez, C.E. Ley and M.F.J. Steel. (2001b). "Benchmark priors for Bayesian model Averaging", *Journal of Econometrics* 100, 381-427.

Fraser, D.A.S. (1976). “Necessary analysis and adaptive inference”, (with discussion). Journal of the American Statistical Association, 71, 99-113.

Fraser, D.A.S. (1979). Inference and Linear Models. McGraw-Hill Publishing Co.

Gelfand, A.E and Smith, A.F.M.(1990). Sampling Based Approaches to calculating Marginal Densities, *Journal of American Statistical Association*,85, 398-409.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics,* Vol. 8,No.1. pp. 19-40

Hampel, F.R.,E.M. Ronchetts, P.G. Rousseeuw and W.A. Stahel(1986). Robust Statistics. Wiley

Jeffreys, H. (1941). An invariant form for the prior probability is estimation problems. *Proceedings of the Royal Society of London.* Series A, 183: 453-461.,51-67.

Koop, G. (2003). Bayesian Econometrics. John Wiley & Sons Ltd,UK.

Lange, K. L., Little, R.J.A and Taylor, J.M.G (1986). Robust Statistical modelling using the t-distribution. J. Statist. Assoc.84: 881-896. MR1134486.

Maronna, R.A. (1976). “Robust M-estimators of multivariate location and scatter”. *Annals of Statistics*,4

Pati Debdeep and Dunson David B. (2014). ‘Bayesian nonparametric regression with varying residual density’: *Annals of the Institute of Statistical Mathematics* 66 (1): 1-31. Doi.10.1007/s10463-013-0415-z.

Raftery, A.E.(1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* 83: 251-266. MR1439782.

Rubio, J. and Steel, M.F.J. (2011). Inference in Two-Piece Location- Scale Models with Jeffreys priors: //www.2. *warwick.ac.uk/fac/sci/statistics/crism/research*/2011/paper,11-13.

Rogers, W.H. and Tukey, J.W.(1993)."Understanding some long-tailed distributions. *Statistica Neerlandica* 26,211-226.

Tierney, L.(1994). Exploring posterior distributions using markov-chains. In Computing Science and Statistic: pro: 23rd Sym Interface (Keramidas E, ed.) *Am Stat. Assoc, Alexandria*, VA, 8p.)

Yeojin Chung, Andrew Gelma, Sophia Rabe-Hesketh, Jinchen Liu and Vincent Dorie (2000). Weakly Prior for Point Estimation of Covariance Matrices in Hierarchical Models. *Journal of Educational and Behavioural Statistics*

Toshiaki Watanabe (2001). Price volatility trading volume,and market depth: evidence from the Japanese stock index futures market. *Applied Financial Economics*,Vol.11 issue 6, page651-658.

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