Original Research Article

BAYESIAN ESTIMATION OF NORMAL LINEAR REGRESSION MODEL WITH HETEROSCEDASTICITY ERROR STRUCTURES

Abstract

Non-constant error variance in Normal Linear Regression Model (NLRM) is an econometric problem generally referred to as heteroscedasticity. Its presence renders statistical inference invalid. Classical approach to its detection, estimation and remediation are widely discussed in the econometric literature. However, estimation of a NLRM using the Bayesian approach when heteroscedasticity problem is present is a major gap in the existing stock of knowledge on this subject. This approach has grown widely in recent times because it combines out-of-sample information with observed data. The study derived Bayesian estimators of the NLRM in the presence of functional forms of heteroscedasticity. Variance was treated as a linear function and as an exponential function of exogenous variables. The estimators are found to be unbiased and consistent and the precision values tend to zero. The estimates obtained from the estimators approximately 95% draws fall within each of the corresponding credible interval. Therefore, the results obtained for the derived Bayesian estimators for different functional forms of heteroscedasticity considered are similar, thus, providing a credible alternative to the existing classical methods which depend solely on the sample information.

Keywords: Asymptotic behaviour; Estimator; Linear function, Exponential function, exogenous variables

1.0 Introduction

Non-constant error variance in Normal Linear Regression Model (NLRM) is an econometric problem generally referred to as heteroscedasticity. Classical approach to its detection, estimation and remediation are widely discussed in the econometric literature (White, 1980; Gujarati, 2003; Cribari-Neto, 2004) amongst others. The consequence of the presence of heteroscedasticity in NLRM renders the classical inference invalid. For instance, the classical Ordinary Least Squares (OLS) estimators of the NLRM parameters are no longer efficient. That is, they are no longer best estimators. In addition, the covariance matrix of the estimated coefficients of the NLRM is no longer consistent and therefore the tests of hypotheses are no longer valid.

These effects cannot be ignored as earlier noted by Geary (1966), White (1980), Pasha (1982), and Hadri and Guermat (1999) amongst others.

The work of White (1980) possibly marked the beginning of investigation into the problem of statistical inference in econometrics. In literature, White (1980) was the most cited article in economics between 1980 and 2005 with 4,318 cites. The paper introduced what is now regarded as a 'revolutionary' idea of inference that is robust to the heteroscedasticity of unknown form. This initial idea has since been extended to other robust inference combining both heteroscedasticity and autocorrelation of unknown forms. Many developments took place rapidly in the frequentist (or classical) literature following the publication of White (1980). Notable ones include: the heteroscedasticity-consistent covariance matrix (HCCM) estimators by MacKinnon and White (1985), Davidson and MacKinnon (1993), Cribari-Neto (2004), the heteroscedasticity and autocorrelation consistent (HAC) covariance estimator include Hansen (1982), White and Domowitz (1984), Newey and West (1987).

In recent times, the application of Bayesian principles in econometrics has witnessed tremendous growth. The principle is based on a degree-of-belief interpretation of probability contrary to the relative-frequency interpretation of the classical methods. The Bayesian principle assumes that coefficients and covariance matrix of the NLRM have prior distributions. This approach is very attractive to applied econometricians because it combines out-of-sample information with observed data. Estimation of a NLRM using the Bayesian approach in the presence of heteroscedasticity is a relatively new area being explored in the econometric literature. Recent papers connected to heteroscedasticity consistent covariance estimators using the Bayesian approach include: Muller (2009), Poirier (2011), Norets (2012), Startz (2013) and Koop (2003).

Sequel to the above progress in the econometric literature, the identifiable gap in the stock of knowledge is the lack of understanding of the nature of Bayesian inference when the structure or form of the heteroscedasticity is known rather than being unknown or assumed in estimating the NLRM. For a NLRM with heteroscedastic errors, the Generalized Least Squares (GLS) of the frequentist is affected and similarly, the mean of the posterior distribution which is the Bayesian equivalent is also affected. To the best of our knowledge, a little work have been carried out on the Bayesian parameters estimation in linear regression model especially when the error variances differ across observation. It is therefore the objective of the paper to examine the behaviour of the spread of the posterior density when the structure of the heteroscedasticity is linear and exponential.

The rest of the paper is structured as follows. Following this introduction, section 2 derives the Bayesian estimators of the parameters of the NLRM with heteroscedastic error term. In section 3, the linear and the exponential error structures in the covariance matrix of the NLRM are formulated. In section 4, a simulation experiment is conducted and the results discussed. Finally, summary and concluding remarks are given in section 5.

2.0 Bayesian Estimation of NLRM with Heteroscedasticity Error Term

We consider a linear regression model;

$$y = X\beta + U \qquad (1)$$

where

$$U \sim MVN(0_N, h^{-1}\Omega)$$

and

$$\Omega = Diag\,(h_i^{-1}) = \begin{bmatrix} h_1^{-1} & 0 & \cdots & 0 \\ 0 & h_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_N^{-1} \end{bmatrix}$$

y is a $(N \times 1)$ vector of the dependent variable, X is a $N \times (k+1)$ matrix of explanatory variables values (including a column of ones for the regression constant), β is an $(k+1) \times 1$ parameters vector and U is an $(N \times N)$ positive definite matrix and h is the precision given as $h_i^{-1} = \sigma_i^2$.

2.1 The Likelihood Function

Once an appropriate model or distribution has been specified to describe the characteristics of a set of data, the immediate issue is one of finding desirable parameter estimates. From a classical perspective the ideal is the Maximum Likelihood Estimator (MLE) which provides a general method for estimating a vector of unknown parameters in a possibly y in a random variable with probability density function $f(y)$ which I characterized by a set of p unknown parameters $\Theta^1 = (\Theta_1, \Theta_2, \cdots, \Theta_p)$. A random sample of T observations $(y_1, y_2, \cdots, y_T)$ is available and the likelihood L, is defined as the joint density of the observations, that is, $L = f(y_1, y_2, \cdots, y_T) = \prod f(y_i; \Theta)$. The attraction of MLE is that subject to fairly minor conditions, it has very desirable properties in large samples (asymptotically).

In this study, using the definition of the Multivariate Normal density, the likelihood of model (1) when the variance differs across observations can be written as;

$$p(y \mid \beta, h, \Omega) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \mid \Omega \mid^{-\frac{1}{2}} \exp\left\{ -\frac{h}{2}(y - X\beta)^{'}\Omega^{-1}(y - X\beta) \right\} \qquad (2)$$

Maximizing the likelihood function in (2) to have

$$\hat{\beta} = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}y \qquad (3)$$

$$\hat{\sigma}^2 = s^2 = \frac{(y - X\beta)^{'}\Omega^{-1}(y - X\beta)}{N - k} \qquad (4)$$

The equations (3) and (4) above represent the Generalized Least Squares (GLS) of the frequentist.

It proves convenient to re-write the likelihood in (2) in a slightly different way. The product $(y - X\beta)^{'}\Omega^{-1}(y - X\beta)$ in (2) can be expressed in terms of the Ordinary Least Squares (OLS) estimator $\hat{\beta}$ of β.

Thus, we then have

$$(y - X\beta)^{'}\Omega^{-1}\ (y - X\beta) = (y - X\beta + X\hat{\beta} - X\hat{\beta})^{'}\Omega^{-1}(y - X\beta + X\hat{\beta} - X\hat{\beta})$$

where $\hat{\beta} = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}y$ and then

$$(y-X\beta)'\Omega^{-1}(y-X\beta)=(y-X\hat{\beta})'\Omega^{-1}(y-X\hat{\beta})+(\hat{\beta}-\beta)'X'\Omega^{-1}X(\hat{\beta}-\beta) \quad (5)$$

From (4), $(N-k)s^2=(y-X\hat{\beta})'\Omega^{-1}(y-X\hat{\beta})$ (6)

Substituting (6) in (5), the likelihood function in (2) then becomes

$$P(y\mid\beta,h,\Omega)=\frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\left\{\exp\left[-\frac{h}{2}(N-k)s^2+(\hat{\beta}-\beta)'X'\Omega^{-1}X(\beta'-\beta)\right]\right\} \quad (7)$$

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

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

The first expression in the curly bracket in (8) resembles the kernel of the multivariate Gaussian density while the second expression also looks like the kernel of the Gamma density. The result simply suggests a Normal-Gamma prior for the likelihood.

3.0 Linear and Exponential Heteroscedasticity Error Structures and the Posterior Densities

The list of the forms of heteroscedasticity structures is not exhaustive, but in this study, two most prevalent forms of heteroscedasticity structures in econometric literature were investigated. The first form of heteroscedasticity structure considered variance is a linear function of exogenous variables is, $w_i=h(z_i',\alpha)$, where $z_i'=(z_{i1},z_{i2},\cdots,z_{ip})$ is a $p\times 1$ vector of observations on a set of exogenous variables related to the regressors and $\alpha'=(\alpha_1,\alpha_2,\cdots,\alpha_p)$ is a $p\times 1$ vector of parameters. The linear model remains $y=X\beta+U$ with

$$E(U)=0$$

and

$$E(UU')=w_i=\begin{bmatrix}(z_1,\alpha)^2 & 0 & \cdots & 0\\ 0 & (z_2,\alpha)^2 & & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & & (z_N,\alpha)^2\end{bmatrix}$$

The justifications of linear function are linearity and additivity of the relationship between dependent and independent variables, statistical independence of errors, homoscedasticity (constant variance) of the errors and normality of the error distribution.

The second form of heteroscedasticity structure by Harvey's (1976) considered variance as an exponential function of exogenous variables. This variance as an exponential function is a very flexible, general model that includes most of the useful formulations as special cases. The general formulation is, $w^*_i = h(\exp(z_i^{'}, \alpha))$, where $z_i^{'}$ and α are as earlier defined.

The specification of the linear model is the same model in (1), with $E(U) = 0$

and

$$E(UU^{'}) = w_i^* = \begin{bmatrix} \exp(z_1^{'}, \alpha)^2 & 0 & \cdots & 0 \\ 0 & \exp(z_2^{'}, \alpha)^2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \exp(z_N^{'}, \alpha)^2 \end{bmatrix}$$

where $h(.)$ is a positive function which depends on parameters α and explanatory variables, z_i. The structures described above were substituted into the likelihood to obtain the likelihood function in (8).

3.1 The Priors and Their Distributions

The most substantial aspect of Bayesian analysis is the specification of appropriate prior distribution for the parameters. In specifying, the following questions should be asked and answered. When should prior come from? How should they be determined and to what extent can they be justified? Probability distributions $p(\phi)$, ideally representing someone's prior information about parameter values are likely to describe the sampling distribution. Priors are meant to reflect any information that researcher has before seeing the data which he wishes to incorporate in the data analysis. Hence, prior can take any form (informative and non informative). There are several ways to choose priors in Bayesian analysis, depending on the available information and the specific form of model (8). For a fully Bayesian analysis, hyper priors for variances are introduced in a further stage. In our study, estimates of hyper priors are available from a previous analysis. We use these estimates, along with expert knowledge of estimation of parameters in NLRM in the presence of heteroscedasticity structures to elicit β_0 and Ω_0. However, it is necessary and common in literature to choose particular classes of priors that are easy to interprete and / or which make computation easier (Gelman, 2006). Hence, natural conjugate priors have both advantages. The conjugate prior is the one which when combined with the likelihood yields a posterior that falls in the same class of distributions (Raifa and Schlaifar, 1961). The likelihood in (8) suggests that Normal-Gamma prior are appropriate for the parameters β and h in this study.

Prior for β condition on h is of the form:

$$p(\beta \mid h) = \frac{h^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}} \mid \Omega_0 \mid^{\frac{1}{2}}} \left\{ \exp\left[-\frac{1}{2}(\hat{\beta} - \beta_0)^{'} (\Omega_0)^{-1} (\hat{\beta} - \beta_0) \right] \right\} \qquad (9)$$

and prior for h is of the form

$$p(h)=\frac{1}{\Gamma(\frac{v_0}{2})(\frac{2s_0^{-2}}{v_0})^{\frac{v_0}{2}}}\left\{h^{\frac{v_0-2}{2}}\exp\left[\frac{hv_0}{2s_0^{-2}}\right]\right\} \qquad (10)$$

Where, β_0 and $\frac{1}{\Gamma(\frac{v_0}{2})(\frac{2s_0^{-2}}{v_0})^{\frac{v_0}{2}}}$ in (9) and (10) are the priors for β and integrating constant respectively. So that the joint prior for β and h then becomes

$$p(\beta,h)=\frac{h^{\frac{v_0+k}{2}-1}}{(2\pi)^{\frac{k}{2}}\,|\Omega_0|^{\frac{1}{2}}\,\Gamma(\frac{v_0}{2})(\frac{2s_0^{-2}}{v_0})^{\frac{v_0}{2}}}\left\{\exp-\frac{1}{2}\left[(\hat{\beta}-\beta_0)'(\Omega_0)^{-1}(\hat{\beta}-\beta_0)+\frac{v_0}{s_0^{-2}}\right]\right\} \qquad (11)$$

The above expression is written in compact form as:

$$p(\beta,h)=f_{NG}(\beta,h\,|\,\beta_0,\Omega_0\;,s_0^{-2},v_0) \qquad (12)$$

We finally specify non-informative uniform prior for Ω, that is, $p(\Omega_0)\propto 1$

3.2 The Posterior Distributions

Combining the prior distributions in (11) and the likelihood function in (8), we can obtain the posterior distribution. Then, from the joint density $p(\beta,h\,|\,y)$ is given by

$$p(\beta,h\,|\,y)\propto p(y\,|\,\beta,h)\,p(\beta,h) \qquad (13)$$

which becomes

$$p(\beta,h,\Omega\,|\,y)\;\propto\;\frac{1}{(2\pi)^{\frac{N}{2}}}\left\{h^{\frac{k}{2}}\exp\left[-\frac{h}{2}(\hat{\beta}-\beta)'X'\Omega^{-1}X(\hat{\beta}-\beta)\right]\right\}\left\{h^{\frac{k}{2}}\exp\left[-\frac{hv}{2s_0^{-2}}\right]\right\}$$

$$\times\frac{h^{\frac{k+v-1}{2}}}{(2\pi)^{\frac{N}{2}}\,|\Omega_0|^{\frac{1}{2}}\,\Gamma(\frac{v_0}{2})(\frac{2s_0^{-2}}{v_0})^{\frac{v_0}{2}}}\left\{\exp-\frac{1}{2}\left[(\hat{\beta}-\beta_0)'\Omega_0^{-1}(\hat{\beta}-\beta_0)+\frac{v_0}{s_0^{-2}}\right]\right\}$$

(14)

From the joint posterior distributions in (14), the following three conditional densities were obtained.

(i) The conditional posterior density of β is;

$$P(\beta \mid h, \Omega, y) = N(\beta_n, \Omega_n) \tag{15}$$

where

$$\beta_n = \Omega_n [\Omega_0^{-1}\beta_0 + hX^{'}\Omega_0^{-1}X\hat{\beta}_{GLS}]$$

$$\Omega_n = [\Omega_0^{-1} + hX^{'}\Omega_0^{-1}X]^{-1}$$

(ii) The conditional posterior density of h is;

$$p(h \mid \beta, \Omega, y) = G[s_n^{-2}, v_n] \tag{16}$$

where

$$s_n^{-2} = \frac{v_n}{(y - X\beta)^{'}\Omega^{-1}(y - X\beta) + v_0 s_0^2};$$

and

$$v_n = N + v_0$$

4.0 Data Generation Process and Discussion of Results

4.1 Data Generation Process

We specify a linear regression model

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U$, where $U \sim N(0, h^{-1}\Omega)$. y could not be determined except values are set for $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 and U, we therefore arbitrarily set 2, 4, 6, 8 and 10 respectively. ε_i was simulated from a unit Gaussian density, i.e. $\varepsilon_i \sim N(0,1)$, so that

$$h_i^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The disturbance terms to be used are generated by specifying the variance-covariance matrix for the error terms, the diagonal $N \times N$ matrix, the squares OLS residuals with robust standard errors are obtained by taking the square root estimated variance-covariance matrix $\Omega = PP^{'}$, since Ω is a symmetric positive definite matrix, we decompose it by a non-singular matrix P such that

$$P = \begin{bmatrix} \sqrt{\sigma_0^2} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\sigma_1^2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\sigma_2^2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\sigma_3^2} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\sigma_4^2} \end{bmatrix}$$

The error term U is then generated by $U = P\varepsilon_1$. We also generate explanatory variables X_1, X_2, X_3 and X_4 from uniformly distributed, that is, $X = U(0,10)$ for $i = 1,2,3,4$. However, the following transformations are made depending on the form of heteroscedasticity to be introduced using $y^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U(h(X_i))$, where $h(X_i)$ is the forms of heteroscedasticity. The following forms of $h(X_i)$ are specified. Given as; $h(z_i^{'}, \alpha)$ and $h(\exp(z_i^{'}))$.

We then resort to the equation $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U$ to determine the values of y.

The following hyper parameters were used in estimating the model parameters.

$v_0 = 5, s_0^{-2} = 4.0 \times 10^{-8}$

$$\beta_0 = \begin{bmatrix} 0 \\ 5 \\ 5 \\ 10 \\ 10 \end{bmatrix}$$

and

$$\Omega_0 = \begin{bmatrix} 2.40 & 0 & 0 & 0 & 0 \\ 0 & 6.0 \times 10^{-7} & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.6 \end{bmatrix}$$

Table 1: Posterior mean for β, h, α $Std..devs$. and $95\% HPDI's$ for n=25

Heteroscedasticity (Linear and Exponential function)

Parameters	Linear function		Exponential function	
	Means(S.D's)	95$ HPDI's	Means (S.D's)	95% HPDI's
β_0	0.06874(0.9947)	$[-1.5637 \quad 1.7142]$	0.06945(1.2548)	$[-1.9787 \quad 2.1174]$
β_1	3.8122(0.0032)	$[3.8070 \quad 3.8175]$	3.8122(0.0032)	$[3.8070 \quad 3.8175]$
β_2	6.1586(0.1219)	$[5.9591 \quad 6.3589]$	6.1593(0.1862)	$[5.8646 \quad 6.4561]$
β_3	8.2626(0.2668)	$[7.8278 \quad 8.7013]$	8.2607(0.2138)	$[7.9041 \quad 8.6178]$
β_4	10.1423(0.1471)	$[9.9055 \quad 10.3782]$	10.1420(0.2331)	$[9.7734 \quad 10.5148]$
h	0.0000(0.0000)	$[0.0000 \quad 0.0000]$	0.0000(0.0000)	$[0.0000 \quad 0.0000]$
α_1	0.2515(0.3838)	$[-0.4545 \quad 0.5893]$	0.1369(0.7004)	$[-0.9761 \quad 1.4044]$
α_2	-0.1333(0.1804)	$[-0.3204 \quad 0.2637]$	-0.0943(0.7686)	$[-1.3567 \quad 1.2973]$
α_3	-0.0469(0.5601)	$[-1.0963 \quad 0.9545]$	-0.0382(1.1085)	$[-1.5507 \quad 1.8521]$
α_4	-0.1866(0.3433)	$[-0.4314 \quad 0.5363]$	-0.0659(0.5138)	$[-0.8294 \quad 0.6567]$

The table above shows the posterior means for $\beta's$, Standard deviation (parentheses) , h and also 95% credible interval.

Table 2: Posterior mean for β, h , α $Std..devs$. and $95\% HPDI's$ for n=50

Heteroscedasticity (Linear and Exponential function)

Parameters	Linear function		Exponential function	
	Means(S.D's)	95$ HPDI's	Means (S.D's)	95% HPDI's
β_0	2.0669(0.3856)	$[1.4352 \quad 2.6977]$	2.0669(0.3851)	$[1.4352 \quad 2.6977]$
β_1	3.9391(0.0032)	$[3.9338 \quad 3.9443]$	3.9390(0.0032)	$[3.9338 \quad 3.8175]$
β_2	5.9772(0.0598)	$[5.8782 \quad 6.0763]$	5.9772(0.0598)	$[5.8783 \quad 6.0763]$
β_3	8.0883(0.0528)	$[8.0010 \quad 8.1751]$	8.0883(0.0528)	$[8.0010 \quad 8.1751]$
β_4	9.9713(0.0587)	$[9.8748 \quad 10.0679]$	9.9713(0.0587)	$[9.8748 \quad 10.0679]$
h	0.0000(0.0000)	$[0.0000 \quad 0.0000]$	0.0000(0.0000)	$[0.0000 \quad 0.0000]$
α_1	0.1117(0.2575)	$[0.0000 \quad 0.7073]$	0.1369(0.7004)	$[-0.9761 \quad 1.4044]$
α_2	0.0020(0.0047)	$[0.0000 \quad 0.0128]$	-0.0943(0.7686)	$[-1.3567 \quad 1.2973]$
α_3	0.0812(0.1848)	$[-0.5062 \quad 0.0000]$	-0.0382(1.1085)	$[-1.5507 \quad 1.8521]$
α_4	-0.0792(0.1825)	$[-0.4999 \quad 0.0000]$	-0.0659(0.5138)	$[-0.8294 \quad 0.6567]$

The table above shows the posterior means for $\beta's$, Standard deviation (parentheses) , h and also 95% credible interval.

Table 3: Posterior mean for β, h, α $Std..devs$. and $95\% HPDI's$ for n=100

Heteroscedasticity (Linear and Exponential function)

Parameters	Linear function Means(S.D's)	95$ HPDI's	Exponential function Means (S.D's)	95% HPDI's
β_0	2.2035(0.6386)	$[1.1533 \quad 3.2461]$	2.2035(0.6386)	$[1.1533 \quad 3.2461]$
β_1	4.0029(0.0032)	$[3.9977 \quad 4.0082]$	4.0029(0.0032)	$[3.9977 \quad 4.0082]$
β_2	5.9749(0.0628)	$[5.8713 \quad 6.0786]$	5.9749(0.0628)	$[5.8713 \quad 6.0786]$
β_3	8.1097(0.0709)	$[7.9932 \quad 8.2263]$	8.1097(0.0709)	$[7.9932 \quad 8.2263]$
β_4	9.8697(0.0761)	$[9.7439 \quad 9.9950]$	9.8697(0.0761)	$[9.7439 \quad 9.9950]$
h	0.0000(0.0000)	$[0.0000 \quad 0.0000]$	0.0000(0.0000)	$[0.0000 \quad 0.0000]$
α_1	0.5045(0.8949)	$[0.8054 \quad 2.2449]$	0.1369(0.7004)	$[-0.9761 \quad 1.4044]$
α_2	0.2355(0.6212)	$[-0.6820 \quad 1.4877]$	−0.0943(0.7686)	$[-1.3567 \quad 1.2973]$
α_3	−0.4838(1.0896)	$[-2.9223 \quad 0.7535]$	−0.0382(1.1085)	$[-1.5507 \quad 1.8521]$
α_4	−0.1949(0.5330)	$[-0.9752 \quad 0.6625]$	−0.0659(0.5138)	$[-0.8294 \quad 0.6567]$

The table above shows the posterior means for $\beta's$, Standard deviation (parentheses), h and also 95% credible interval.

Table 4: Posterior mean for β, h, α $Std..devs$. and $95\% HPDI's$ for n=150

Heteroscedasticity (Linear and Exponential function)

Parameters	Linear function Means(S.D's)	95$ HPDI's	Exponential function Means (S.D's)	95% HPDI's
β_0	1.6212(0.4810)	$[0.8318 \quad 2.4078]$	1.6212(0.4810)	$[0.8318 \quad 2.4078]$
β_1	3.8990(0.0032)	$[3.8937 \quad 3.9042]$	3.8990(0.0032)	$[3.8937 \quad 3.9042]$
β_2	6.0510(0.0549)	$[5.9603 \quad 6.1419]$	6.0051(0.0549)	$[5.9603 \quad 6.1419]$
β_3	7.9805(0.0577)	$[7.8856 \quad 8.0755]$	7.9805(0.0577)	$[7.8856 \quad 8.0755]$
β_4	10.1059(0.0546)	$[10.0157 \quad 10.1961]$	10.1057(0.0546)	$[10.0157 \quad 9.9950]$
h	0.0000(0.0000)	$[0.0000 \quad 0.0000]$	0.0000(0.0000)	$[0.0000 \quad 0.0000]$
α_1	0.0404(0.4514)	$[-0.4041 \quad 0.7180]$	0.1369(0.7004)	$[-0.9761 \quad 1.4044]$
α_2	−0.1369(0.5515)	$[-1.0890 \quad 0.2655]$	−0.0943(0.7686)	$[-1.3567 \quad 1.2973]$
α_3	0.2573(0.4355)	$[-0.3419 \quad 1.0417]$	−0.0382(1.1085)	$[-1.5507 \quad 1.8521]$
α_4	−0.3063(0.4910)	$[-0.9133 \quad 0.8907]$	−0.0659(0.5138)	$[-0.8294 \quad 0.6567]$

The table above shows the posterior means for $\beta's$, Standard deviation (parentheses), h and also 95% credible interval.

Table 5: Posterior mean for β, h , α $Std..devs$. and $95\% HPDI's$ for n=200
Heteroscedasticity (Linear and Exponential function)

Parameters	Linear function		Exponential function	
	Means(S.D's)	95$ HPDI's	Means (S.D's)	95% HPDI's
β_0	2.2945(0.4721)	$[1.5193 \quad 3.0666]$	2.2945(0.4721)	$[1.5193 \quad 3.0666]$
β_1	4.0568(0.0032)	$[4.0516 \quad 4.0621]$	4.0568(0.0032)	$[4.0516 \quad 4.0621]$
β_2	5.9254(0.0524)	$[5.8389 \quad 6.0122]$	5.9254(0.0524)	$[5.8389 \quad 6.0122]$
β_3	7.9463(0.0489)	$[7.8656 \quad 8.0265]$	7.9463(0.0489)	$[7.8656 \quad 8.0265]$
β_4	10.0083(0.0476)	$[9.9297 \quad 10.0868]$	10.0083(0.0476)	$[9.9297 \quad 10.0868]$
h	0.0000(0.0000)	$[0.0000 \quad 0.0000]$	0.0000(0.0000)	$[0.0000 \quad 0.0000]$
α_1	$-0.1997(0.0857)$	$[-0.2365 \quad 0.0000]$	0.1369(0.7004)	$[-0.9761 \quad 1.4044]$
α_2	0.2242(0.0962)	$[0.0000 \quad 0.0000]$	$-0.0943(0.7686)$	$[-1.3567 \quad 1.2973]$
α_3	0.2033(0.0873)	$[0.0000 \quad 0.2408]$	$-0.0382(1.1085)$	$[-1.5507 \quad 1.8521]$
α_4	$-0.3844(0.1650)$	$[-0.4552 \quad 0.0000]$	$-0.0659(0.5138)$	$[-0.8294 \quad 0.6567]$

The table above shows the posterior means for $\beta's$, Standard deviation (parentheses) , h and also 95% credible interval.

4.2 Discussion of Results

In this section, an R code was written for the implementation of the Gibbs sampling and Metropolis-Hasting algorithms for the Bayesian estimation of parameters of a Normal Linear Regression Model with heteroscedasticity structures considered. Normal prior was specified for the coefficients β while Gamma prior was specified for precision h, such that the resulting posterior has a Normal-Gamma density for homoscedasticity version of the model. The derived Bayesian estimators for the homoscedasticity version are in closed forms $p(\beta \mid y, h) \sim N(\beta_n, \Omega_n)$ and $p(h \mid y, \beta) \sim N(s_n^{-2}, v_n)$ .The derived Bayesian estimators for heteroscedasticity of known forms are also in closed forms; $p(\beta \mid h, \Omega, y) \sim N(\beta_n, \Omega_n)$ and $p(h \mid \beta, \Omega, y) \sim G(s_n^{-2}, v_n)$. The Bayesian linear regression models of these two cases were fitted to each data set and parameter estimates yielded by each case are presented in Tables 1 to 5.

The two scenarios described above using derived Bayesian estimators for the normal linear regression model in (1) based on the Monte Carlo experiment were presented. Five different sample sizes $n =$ 25, 50, 100, 150 and 200 using the data generation process presented in section 4.1 were considered.

For the two scenarios, the posterior means for $\beta' s$ are unbiased and consistent for all the sample sizes considered as shown in the tables 1 to 5 above. The value of precision h tends to zero in all cases as expected. The estimated coefficients of $\beta' s$ approximately 95% draws fall within each of the corresponding credible interval. Finally, the difference in the Bayesian estimators derived is noticed in the highest posterior density intervals (HPDI's).

5. Conclusion

This paper has attempted to fill some noticeable gaps in econometric literature. Bayesian estimators of heteroscedasticity structures were derived in normal linear regression model. The estimators are found to be unbiased and consistent with the initial values specified. This confirms the validity of the derived estimators, thus providing a credible alternative to the existing classical methods which depend solely on the sample information.

References

Cribari-Neto, F. (2004). Asymptotic inference under heteroscedasticty of un-known form. *Computational Statistics and Data Analysis* 45, 215-233.

Davidson, R. & Makinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press; *Econometric Theory* 3, 631-635.

Geary, R.C. (1966). A note on residual heterovariance and estimation efficiency in regression. *American Statistician*, 20, 30-31.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*,1(3),515-533.

Gujarati, D.N. (2003). *Basic Econometrics*, 4th Ed., New York: McGraw-Hill.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes, *Econometrica*, 60(4), 967-972.

Hadri, K. & Guermat, C. (1999). Heteroscedasticity in stochasic frontier models. A Monte Carlo Analysis, 1-8.

Harvey, A.C (1976). Estimating Regression Model with Multiplicative Heteroscedasticity. *Ecomonetrica*, 44, 461-465.

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

Mackinnon, J.G. & White, H. (1985). Some heteroscedasticty consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 53-57.

Muller, U. K. (2009). Risk of Bayesian inference in misspecified model andthe Sandwich covariance matrix, working paper, Princeton University.

Newey, W. D. & West K. D. (1987). A simple positive semi-de_nite,heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703-708.

Norets, A. (2012). Bayesian regression with nonparametric heteroskedasticity,working paper, Princeton University.14

Pasha, G.R. (1982). Estimation Methods for Regression Models with Unequal Error Variances, Ph.D thesis, University of Warwick.

Poirier, D.J. (2011). Bayesian interpretations of heteroscedastic Consistent Covariance Estimatiors Using the Informed Bayesian Bootstrap, *Econometric Reviews*, 30,457-468.

Raifa,H & Schlaifer,R. (1961). *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.

Richard, S (2013). Bayesian Heteroscedasticity Robust Standard Errors. Department Economics, University of California, Santa Barbara.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.

White H. & Domowitz, I. 1984. Nonlinear regression with dependent observations, *Econometrica* 52, 143-161.