Bayesian Hetero-Elasticnet (A Gibbs Sampler Approach)

Isiaka Oloyede
Department of Statistics, University of Ilorin, P.M.B. 1515 Ilorin, Nigeria
E-mail: oloyede.i@unilorin.edu.ng
Received 17 Dec 2020, Revised 30 Jun 2021, Accepted 20 Jul 2021, Published Aug 2021
DOI: https://dx.doi.org/10.4314/tjs.v47i3.9

Abstract
Combined heteroscedasticity and multicollinearity as dual non-spherical disturbances were experimented asymptotically. A Gibbs Sampler technique was used to investigate the asymptotic properties of hetero-elasticnet estimator with mean squares error (MSE) and bias as performance metrics. The seed was set to 12345; β is set at β = {2.5,3,1.5,1,0,0,5}; Xs variables were generated as follow: the design matrix was generated from the multivariate normal distribution with mean > 0 and variance σ². X₁ and X₂ are truncated with Harvey (1976) heteroscedastic error structure; X₃, ..., X₆ are collinear covariate with pairwise correlation between 0.6 and 0.9, the sample sizes were 25, 100 and 1000. The number of replications of the experiment was set at 10,000 with burn-in of 1000 which specified the draws that were discarded to remove the effects of the initial values. The thinning was set at 5 to ensure the removal of the effects of autocorrelation in the MCMC simulation. The study found that there is consistency of estimator asymptotically as the sample sizes increases from 25 to 50 so also to 1000, the larger sample size depicted least bias. The estimator exhibited efficiency asymptotically as larger sample sizes depicted least mean squares error. The study therefore recommended Bayesian hetero-elasticnet when data exhibit both heteroscedasticity and multicollinearity.

Keywords: Elasticnet, Bayesian Inference and Gibbs sampler

Introduction
Correlations among covariates in the explanatory variables of linear regression vis a vis presence of heteroscedasticity affect the precision of the inferences of the parameter estimates. Obviously, the non-spherical disturbances in the data or model usually led to inefficient and inconsistent estimation, though the estimator is unbiased, the standard error and test of hypothesis computed for the estimator are invalid. Thus, the presence of both non-spherical disturbances poses serious threats to the appropriateness of the inferences of the parameter estimates.

Meanwhile, mean squares error (MSE) and bias may be inflated owing to the presence of non-spherical disturbances. From the previous study of heteroscedasticity in the literature, ordinary least squares becomes inefficient and inconsistent when heteroscedasticity is present in the data and or model (Hadri and Guermat 1999, Robinson 1987, White 1980). An example is naturally heteroscedastic model is the popular Cobb-Douglas (1928) production function which had been receiving series of criticisms and modifications since 1928 when the model was formulated.

Hoerl and Kennard (1970) proposed ridge regression which minimises residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of squares subject to a constraint ∑|β_j|^γ ≤ t where the shrinkage parameter γ = 2. Frank and Friedman (1993) introduced bridge regression which minimizes residual sum of
squares (RSS) subject to a constraint $\sum |\beta_j|^\gamma \leq t$ where $\gamma \geq 0$ with a special case of 0. Tibshirani (1996) formulated least absolute shrinkage and selection operator popularly tagged as Lasso with tuning parameter, this off course minimizes residual sum of squares (RSS) subject to a constraint $\sum |\beta_j|^\gamma \leq t$ with $\gamma \geq 0$ which is more or less bridge regression when tuning parameter $\gamma = 1$. Lasso is a special case of penalized least squares which penalizes the parameter estimates and shrink some of the estimates to zero. This is a way to compensate for the presence of multicollinearity in the data and or model, of which if not penalized may make the covariates of explanatory variables to have zero determinant, Severien and Eric (2012). Lasso is a good selection operator which showcases the uninfected estimates after series of iterative algorithms. If the tuning parameter $\gamma = 2$, bridge regression becomes ridge regression. Series of extension of Lasso emerged recently in the literatures, adaptive Lasso was invented by Zou and Hastic (2005) which minimises RSS subject to constraint $\lambda_1 \sum |\beta_j| + \lambda_2 (\sum_{j=1}^p |\beta_j|^2)^\gamma \leq t$, where $\lambda_1$ and $\lambda_2$ denote tuning parameters one and two. Tibshirani et al. (2005) proposed fused Lasso, while Group Lasso was proposed by Yuan and Lin (2006), and Smoothly Clipped Absolute Deviation (SCAD) was introduced by Fan and Li (2001). Daye et al. (2012) explored high dimensional heteroscedastic regression. Heteroscedasticity, a significant non-spherical disturbance with multicollinearity was recently examined in the literature. Severien and Eric (2012) examined shrinkage and Lasso in high dimensional heteroscedasticity models. Due to nonlinearity of the model, the bridge model does not always perform the best in estimation and prediction compared to other shrinkage models, Fu (1998). In their studies Li and Lin (2010) opined that Bayesian elastic net outperformed elastic net in variable selection for more complicated models, it equally outperforms Bayesian Lasso in prediction accuracy for small samples from less sparse modes. The choice of penalty parameters $\lambda_1$ and $\lambda_2$ can be done by introducing hyper priors on them. This was exemplified by Park and Cassela (2008). Cassela et al. (2010) claimed that all the Lasso models with the exception of elastic net, the $\lambda$ and $\beta$ parameters are conditionally independent given the $\gamma$’s shrinkage parameter leading to a straightforward Gibbs sampler. Anirban et al. (2013) proposed Dirichlet prior and compared it with Bayesian Lasso prior, thus concluded that their proposed prior outperformed Bayesian Lasso prior due to its strong concentration around the origin. Should there be several relatively small signals, they opined that dirichlet prior can shrink all of them towards zero. Kayanan and Wijekoon (2020) affirmed that elasticnet performed better in high multicollinearity where Lasso regularization failed.

Since there are two categories of multicollinearity {data and model based multicollinearity} so as also there exist data and model based heteroscedasticity, this paper examines the oracle properties of linear regression model when there are presence of both multicollinearity and heteroscedasticity in both data and model asymptotically using Bayesian hetero-elasticnet. The study observed that most generalised estimators only deemed one non-spherical disturbance, the study therefore fill the gap of accommodating dual non-spherical disturbances both in the data and or model.

**Model Designs**

Let $y = X\beta + u$ with $u \sim N(0, \sigma_i^2\Omega)$ where $\Omega$ is a positive definite matrix of order $n \times n$. The model is truncated with both collinear of different tuning $\gamma$ and one component heteroscedasticity error structure with $\delta$ as the scale; $y$ as an n-vector of random responses; $X$ as an $n \times p$ design matrix of corrupted collinear and heteroscedastic, $\beta$ as a $p$-vector parameters and $u$ as an n-vector of heteroscedastic error structures $\sigma_i^2$ $i = 1, ..., n$. $y = \beta_0 + \sum_{i=1}^p \beta_i X_i + u_i$ (1)
Let $X_1$ and $X_2$ be truncated with multiplicative heteroscedasticity as described by Harvey (1976) which can be expressed as $\sigma_i^2 = \sigma^2(\beta_0 + \beta_1X_1 + \beta_2X_2)\delta$ where $\delta$ is an unknown parameter which determines the degree of heteroscedasticity, some of variables in $Xs$ are embedded with collinearity. Adopting a full Bayesian inference which incorporates the likelihood function, prior distribution for the parameters, and hyper-parameters in the model with MCMC algorithm we have: The likelihood function of $\theta$, where $\theta = (\beta_0, \lambda_i, \gamma, \delta)$ give the sample vector $X_t = (i = 1, 2, ..., p)'$ and $y = (y_1, y_2, ..., y_n)'$ is expressed as

$$L(\theta | \sigma, X, y) = \frac{(2\pi \sigma^2)^{-n/2} \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - x\beta] \right\}}{(\sum_{i=1}^{n} [y_i - x\beta])^2}$$

Incorporating multiplicative hetero-elastic-net into likelihood function above, the study derived conditional likelihood of heteroelastic-net from the product of the error density function. Thus error $u$ is changed to $w$ for easy comprehension.

**Bayesian hetero-elastic net**

$$L(\theta, \sigma | X, y) = \frac{(2\pi \sigma^2)^{-n/2} \prod_{i=1}^{n} |w|^{-\lambda/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - x\beta]'w^{-\lambda}(y_i - x\beta) \right\}}{(\sum_{i=1}^{n} [y_i - x\beta])^2}$$

Incorporating elastic net in to the above model as

$$L(\theta, \sigma | X, y) = \frac{(2\pi \sigma^2)^{-n/2} \prod_{i=1}^{n} |w|^{-\lambda/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - x\beta]'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p} |\beta_i|^2 \right\}}{(\sum_{i=1}^{n} [y_i - x\beta])^2}$$

Or

$$\hat{\beta}_{HEN} = \arg\min\beta \left( y_i - x\beta \right)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p} |\beta_i|^2$$

To derive the full Bayesian density, we truncate the error density function Equation (3) with Gaussians, Laplacian and inverse-gamma priors. It is noteworthy that Zou and Hastie (2005) said solving the Elastic net problem is just like deriving marginal posterior density mode of $\beta | y$ particular when the prior distribution of $\beta$ (Li and Lin 2010) is given as

$$\pi(\beta) \propto \exp\left\{ -\lambda_1 \sum_{i=1}^{p} |\beta_i| - \lambda_2 \sum_{i=1}^{p} |\beta_i|^2 \right\}$$

Instead we proposed multinomial Gaussian prior for the $\beta_i$, gamma prior for tuning parameter $\lambda_i$, heteroscedastic $\delta_i$ and inverse gamma prior $\sigma_i^2$. Marginal posterior density is obtained by integrating the joint posterior density with respect to each parameter, thus, expert opinion can be adopted by assuming the set of parameters $\beta_i, \lambda_i, \delta_i$ and $\sigma_i$ as independent marginal distribution. The study assumed a prior density $\pi(\beta_0, \lambda_i, \delta_i, \sigma_i) = \pi(\beta_0)\pi(\lambda_i)\pi(\delta_i)\pi(\sigma_i)$. Thus, $\pi(\beta) \propto (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_i - \mu)^2 \right\}, \beta > 0$ and $\pi(\lambda_i) \propto (\lambda_i)\delta^{-1} \exp(-b_i/\lambda_i), \lambda_i > 0$ and $\pi(\delta_i) \propto (\delta_i)^{\gamma+1} \exp(-d_i/\delta_i), \delta_i > 0$ and $\pi(\sigma_i^2) \propto (\sigma_i^2)^{\delta+1} \exp(-f_i/\sigma_i^2), \sigma_i^2 > 0$. The posterior distribution of $\theta = (\beta_i, \lambda_i, \delta_i, \sigma_i)$. Considering independence among the parameters is given by:

$$\pi(\beta_0, \lambda_i, \delta_i, \sigma_i | X, y) \propto$$

$$\prod_{i=1}^{n} |w|^{-\lambda/2} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_i - \mu)^2 \right\} \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_i - \mu)^2 \right\}$$

where $a_1, b_i, c_i, d_i, e_i, f_i$ are the hyper-parameters for the gamma and inverse-gamma priors. Hyper-parameters are excluded for $\beta_i$-parameters since they would be estimated from the data and may be arbitrarily small leading to problems which may eventually affect the inferences. Integrating the posterior $\pi(\beta_0, \lambda_i, \delta_i, \sigma_i | X, y)$ with respect to $\sigma_i$, we have we joint a posterior distribution for $\pi(\beta_0, \lambda_i, \delta_i)$ and $\pi(\mu, \beta_0, \lambda_i, \delta_i, \sigma_i | X, y) \propto$

$$\prod_{i=1}^{n} |w|^{-\lambda/2} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_i - \mu)^2 \right\} \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2\sigma^2} (\beta_i - \mu)^2 \right\}$$. 

983
\[
\exp\{-\left(b_1 + d_1 + f_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)^{'}w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}\}
\]

Gibbs algorithm update is performed on the full conditional distribution of \(\sigma^2 \propto IG(\alpha_{1} + \frac{n}{2}, b_{1} + \frac{1}{2}\sum_{i=1}^{n}(y_{i} - x\beta)^{'}w^{-\lambda}(y_{i} - x\beta) + \lambda_{1} \sum_{i=1}^{p}|\beta_{i}| + \lambda_{2} \sum_{i=1}^{p}|\beta_{i}|^2)\). This yields the following full conditional density of the parameters \(\beta_1, \lambda_1, \delta_1\) and \(\alpha_{1}\):

\[
\pi(\beta_i | \lambda_1, \delta_1, X, y) \propto \exp\left\{-\frac{1}{2} (\beta_i - \mu)^2\right\}
\]

\[
\Pi_{i=1}^{n} w^{-\lambda/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n}(y_{i} - x\beta)'w^{-\lambda}(y_{i} - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_{i}| + \lambda_2 \sum_{i=1}^{p}|\beta_{i}|^2\right\} \quad (13)
\]

\[
\pi(\sigma_i^2 | \beta_i, \lambda_i, \delta_i, X, y) \propto \left(\sigma_i^2\right)^{-\left(a_{11} - \frac{1}{2}\right)} \exp \left\{-\frac{b_{1}}{\sigma_i^2}\right\}
\]

\[
\Pi_{i=1}^{n} w^{-\lambda/2} \exp\left\{\frac{1}{2} b_{1} + \frac{1}{2} \sum_{i=1}^{n}(y_i - x\beta)^{'}w^{-\lambda}(y_i - x\beta) - \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\right\} \quad (14)
\]

\[
\pi(\lambda_i | \beta_i, \lambda_i, \delta_i, X, y) \propto \left(\lambda_i\right)^{\left(c_{11} - \frac{1}{2}\right)} \exp \left\{-\frac{f_{1}}{\lambda_i}\right\} \left|w^{-\lambda}\right| \quad (15)
\]

\[
\pi(\delta_i | \beta_i, \lambda_i, X, y) \propto \left(\delta_i\right)^{\left(e_{11} - \frac{1}{2}\right)} \exp \left\{-\frac{f_{1}}{\delta_i}\right\} \left|w^{-\lambda}\right| \quad (16)
\]

**General posterior gibbs sampler procedures for Lasso-type bayes estimates**

It is generally asserted from the expert opinions that Lasso-type are the priors for \(\beta_i\), in this study there is no objection for such claim. Thus we adopted those priors along with the stated priors.

\[
\pi(\beta_i) = \frac{1}{2\tau} \exp\left\{-\frac{|\beta_i|}{\tau}\right\} \text{ with } \tau = 1^{t} \lambda. \quad (16)
\]

Park and Casella (2008) used empirical Bayes estimates for the penalty parameters \(\lambda_1\) and \(\lambda_2\), which are the maximization of the data marginal likelihood. This is accomplished by treating \(\beta_i, \delta_i, \sigma_i^2\). Penalized regression approaches have been used in cases where \(p < n\) and or where \(n < p\). Efron et al. (2004) proposed Least Angle Regression Selection (LARS) for a model selection algorithm. Readers are advised to see Casella et al. (2010) for details.

**Simulation**

A Gibbs sampler technique was adopted to measure the performance of each estimator described in the previous section using MSE and bias. Seed was set to 12345; \(\beta\) was set at \(\beta = \{2.5, 3, 1.5, 1, 0, 0.5\}\): Xs variables were generated as follows: the design matrix was generated from the multivariate normal distribution with mean > 0 and variance \(\sigma^2\). \(X_1\) and \(X_2\) are truncated with Harvey (1976) heteroscedastic error structure; \(X_3, ..., X_6\) are collinear covariate with pairwise correlations between 0.6 and 0.9. The sample sizes were 25, 100 and 1000. The number of replications of the experiments was set at 10,000 with burn-in of 1000 which specified the draws that were discarded to remove the effects of the initial values. The thinning was set at 5 to ensure the removal of the effects of autocorrelation in the MCMC simulation. For the Bayesian experiment, a Gibbs sampler algorithm was developed to simulate the heteroscedastic-elasticnet based models. This was invoked in R Statistical software, R Core Team (2020).

**Results and Discussion**

In this study, a Bayesian Hetero-Elasticnet truncated linear model is presented, with multiplicative heteroscedasticity structure and collinear covariates. Parameters obtained through the posterior point estimate of Gibbs sampler were used to compute bias (measure of consistency) and mean squared error criterion (measure of efficiency). The levels of convergence of the chains were monitored using the method proposed by Gelman and Rubin (1992) and graphic analysis was carried out using coda package in R. Multivariate normal, gamma and inverse gamma
distributions were chosen as priors for parameter estimates \( \lambda_1 \), \( \lambda_2 \) and \( \sigma^2 \), respectively.

Table 1 showed the outcomes of the estimations of Bayesian hetero-elasticnet based on the absolute bias performances with different scales of heteroscedasticity and sample sizes ranging from 25 to 1000. At sample size 25, it was observed that biases for \( \hat{\beta}_1 \) to \( \hat{\beta}_4 \) were consistent with scale of heteroscedasticity as we increased the scale of heteroscedasticity, and also the biases were increasing. But the biases for \( \hat{\beta}_5 \) and \( \hat{\beta}_6 \) depicted inconsistence, the biases decrease up to the scale of 0.3, at the scale of 0.4 they increase as the scale is increased. Surprisingly, the bias decreases at the scale of 2. The biases for sample sizes 100, 200 and 1000 all depicted consistence for all the parameters. Similar patterns were observed with the study of Hadri and Guermat, (1999). It was observed that the bias for \( \hat{\beta}_1 \) to \( \hat{\beta}_4 \) increase algebraically as the scale of heteroscedasticity increases Thus, there exists consistence. The lambda one and two were randomly generated and were used to shrink the problematic parameters towards zero.

**Table 1: Benet based on absolute bias @ scale of heteroscedasticity with sample sizes**

<table>
<thead>
<tr>
<th>Sample</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \hat{\beta}_2 )</th>
<th>( \hat{\beta}_3 )</th>
<th>( \hat{\beta}_4 )</th>
<th>( \hat{\beta}_5 )</th>
<th>( \hat{\beta}_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>6.27</td>
<td>6.53</td>
<td>6.38</td>
<td>8.80</td>
<td>6.62</td>
<td>5.08</td>
<td>3.99</td>
<td>3.20</td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.2</td>
<td>0.38</td>
<td>0.4</td>
<td>0.52</td>
<td>0.6</td>
<td>0.79</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>0.1161</td>
<td>0.2246</td>
<td>0.3281</td>
<td>0.4248</td>
<td>0.5999</td>
<td>0.6106</td>
<td>0.6964</td>
<td>0.7775</td>
</tr>
<tr>
<td></td>
<td>0.5053</td>
<td>1.0073</td>
<td>1.5075</td>
<td>2.0045</td>
<td>2.4988</td>
<td>2.9901</td>
<td>3.4780</td>
<td>3.9624</td>
</tr>
<tr>
<td></td>
<td>0.0134</td>
<td>0.0244</td>
<td>0.0376</td>
<td>0.0481</td>
<td>0.0614</td>
<td>0.0762</td>
<td>0.0923</td>
<td>0.1097</td>
</tr>
<tr>
<td></td>
<td>0.0006</td>
<td>0.0061</td>
<td>0.0121</td>
<td>0.0303</td>
<td>0.0437</td>
<td>0.0588</td>
<td>0.0757</td>
<td>0.0943</td>
</tr>
</tbody>
</table>

Table 2 showed the mean squared error criteria, the mean squares errors for \( \hat{\beta}_5 \) decrease algebraically as the sample sizes increase irrespective of the scale of heteroscedasticity. Thus, sample size 1000 has the least mean squares error,
Oloyede - Bayesian Hetero-Elasticnet (A Gibbs Sampler Approach)

asymptotically; larger sample sizes bring about improvements in the estimation and reduce the effects of the error on the inferences. The study outcome is analogous to that of Hadri and Guerret (1999). Moreover, the mean squares errors have asymptotic efficiency since the MSE decrease as the sample size increases.

Considering the scale of heteroscedasticity, the study revealed that the mean squared errors increase as the scale of heteroscedasticity increases for posterior mean of $\hat{\beta}_s$. The outcome of the study is analogous with the Kayanan and Wijekoon (2020).

**Table 2**: Benet based on MSE @ scale of heteroscedasticity with sample sizes

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\delta^1$</th>
<th>$\beta_1^1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>6.27</td>
<td>9.31</td>
<td>0.1</td>
<td>0.0525</td>
<td>0.0092</td>
<td>0.3203</td>
<td>0.3126</td>
<td>0.7168</td>
<td>0.7132</td>
</tr>
<tr>
<td></td>
<td>6.27</td>
<td>6.53</td>
<td>0.2</td>
<td>0.0826</td>
<td>1.0433</td>
<td>0.2648</td>
<td>0.2579</td>
<td>0.5910</td>
<td>0.5871</td>
</tr>
<tr>
<td></td>
<td>6.27</td>
<td>4.38</td>
<td>0.3</td>
<td>0.1341</td>
<td>2.2959</td>
<td>0.2196</td>
<td>0.2128</td>
<td>0.4872</td>
<td>0.4832</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>8.80</td>
<td>0.4</td>
<td>0.2023</td>
<td>4.0370</td>
<td>0.1822</td>
<td>0.1759</td>
<td>0.4043</td>
<td>0.3998</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>6.62</td>
<td>0.5</td>
<td>0.2883</td>
<td>6.2597</td>
<td>0.1523</td>
<td>0.1462</td>
<td>0.3345</td>
<td>0.3303</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>5.08</td>
<td>0.6</td>
<td>0.3876</td>
<td>8.9534</td>
<td>0.1283</td>
<td>0.1224</td>
<td>0.2770</td>
<td>0.2731</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>3.99</td>
<td>0.7</td>
<td>0.4973</td>
<td>12.1072</td>
<td>0.1096</td>
<td>0.1038</td>
<td>0.2298</td>
<td>0.2260</td>
</tr>
<tr>
<td>100</td>
<td>9.40</td>
<td>3.20</td>
<td>0.8</td>
<td>0.6146</td>
<td>15.7088</td>
<td>0.0954</td>
<td>0.0897</td>
<td>0.1909</td>
<td>0.1873</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>2.6</td>
<td>0.9</td>
<td>0.7369</td>
<td>19.7454</td>
<td>0.0853</td>
<td>0.0797</td>
<td>0.1588</td>
<td>0.1554</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>2.17</td>
<td>1</td>
<td>0.8615</td>
<td>24.2031</td>
<td>0.0789</td>
<td>0.0736</td>
<td>0.1323</td>
<td>0.1289</td>
</tr>
<tr>
<td></td>
<td>9.40</td>
<td>0.63</td>
<td>2</td>
<td>1.7698</td>
<td>88.4746</td>
<td>0.1885</td>
<td>0.2188</td>
<td>0.0179</td>
<td>0.0181</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>5.05</td>
<td>0.1</td>
<td>0.0313</td>
<td>0.3989</td>
<td>0.1295</td>
<td>0.1332</td>
<td>0.1178</td>
<td>0.1188</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>2.08</td>
<td>0.2</td>
<td>0.0963</td>
<td>1.5675</td>
<td>0.1095</td>
<td>0.1130</td>
<td>0.0988</td>
<td>0.0988</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>1.06</td>
<td>0.3</td>
<td>0.2059</td>
<td>3.5119</td>
<td>0.0959</td>
<td>0.0999</td>
<td>0.0859</td>
<td>0.0839</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.62</td>
<td>0.4</td>
<td>0.3598</td>
<td>6.2255</td>
<td>0.0886</td>
<td>0.0938</td>
<td>0.0788</td>
<td>0.0738</td>
</tr>
<tr>
<td>1000</td>
<td>34.9</td>
<td>0.41</td>
<td>0.5</td>
<td>0.5576</td>
<td>9.6987</td>
<td>0.0876</td>
<td>0.0947</td>
<td>0.0776</td>
<td>0.0680</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.29</td>
<td>0.6</td>
<td>0.7986</td>
<td>13.9193</td>
<td>0.0928</td>
<td>0.1027</td>
<td>0.0825</td>
<td>0.0863</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.21</td>
<td>0.7</td>
<td>1.0819</td>
<td>18.8719</td>
<td>0.1043</td>
<td>0.1181</td>
<td>0.0934</td>
<td>0.0684</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.16</td>
<td>0.8</td>
<td>1.4065</td>
<td>24.5385</td>
<td>0.1223</td>
<td>0.1409</td>
<td>0.1106</td>
<td>0.0741</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.13</td>
<td>0.9</td>
<td>1.7708</td>
<td>30.8984</td>
<td>0.1468</td>
<td>0.1716</td>
<td>0.1344</td>
<td>0.0832</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.1</td>
<td>1</td>
<td>2.1734</td>
<td>37.9282</td>
<td>0.1781</td>
<td>0.2102</td>
<td>0.1648</td>
<td>0.0954</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>0.02</td>
<td>2</td>
<td>7.8108</td>
<td>138.5934</td>
<td>0.8347</td>
<td>1.0413</td>
<td>0.8415</td>
<td>0.3234</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>31.6</td>
<td>0.1</td>
<td>0.0259</td>
<td>0.36681</td>
<td>0.0111</td>
<td>0.0113</td>
<td>0.0129</td>
<td>0.0128</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>20.6</td>
<td>0.2</td>
<td>0.1013</td>
<td>1.4626</td>
<td>0.0092</td>
<td>0.0096</td>
<td>0.0108</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>13.1</td>
<td>0.3</td>
<td>0.2265</td>
<td>3.2827</td>
<td>0.0079</td>
<td>0.0085</td>
<td>0.0092</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>8.67</td>
<td>0.4</td>
<td>0.4011</td>
<td>5.8185</td>
<td>0.0072</td>
<td>0.0082</td>
<td>0.0081</td>
<td>0.0078</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>6.05</td>
<td>0.5</td>
<td>0.6237</td>
<td>9.0583</td>
<td>0.0068</td>
<td>0.0084</td>
<td>0.0075</td>
<td>0.0069</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>4.43</td>
<td>0.6</td>
<td>0.8932</td>
<td>12.9881</td>
<td>0.0070</td>
<td>0.0092</td>
<td>0.0073</td>
<td>0.0066</td>
</tr>
<tr>
<td>1000</td>
<td>26.69</td>
<td>3.37</td>
<td>0.7</td>
<td>1.2081</td>
<td>17.5909</td>
<td>0.0077</td>
<td>0.0112</td>
<td>0.0076</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>2.64</td>
<td>0.8</td>
<td>1.5665</td>
<td>22.8475</td>
<td>0.0088</td>
<td>0.0137</td>
<td>0.0082</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>2.13</td>
<td>0.9</td>
<td>1.9662</td>
<td>28.7363</td>
<td>0.0106</td>
<td>0.0173</td>
<td>0.0093</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>1.7</td>
<td>1</td>
<td>2.4053</td>
<td>35.2334</td>
<td>0.0129</td>
<td>0.0218</td>
<td>0.0107</td>
<td>0.0078</td>
</tr>
<tr>
<td></td>
<td>26.69</td>
<td>0.50</td>
<td>2</td>
<td>8.3569</td>
<td>127.3451</td>
<td>0.0769</td>
<td>0.1365</td>
<td>0.0403</td>
<td>0.0228</td>
</tr>
</tbody>
</table>

Conclusion
The study observed that modelling heteroelasticnet in a full Bayesian improves the precision of the inferences of the estimates. The study found that $X_1$ and $X_2$ were affected as the scale of heteroscedasticity was increased while $X_3, ..., X_6$ behave in different way. The
effects of heteroscedasticity on the parameters $X_1$ and $X_2$ asymptotically are in line with the findings of Hadri and Guermat (1999). The study concludes that asymptotically, there exist consistency and efficiency in the estimations. The approach can be applied to further studies in the areas of simultaneous equations and other econometric models.

References
Severien N and Eric YL 2012 Shrinkage and LASSO Strategies in high dimensional Heteroscedastic models. *Department of Mathematics and Statistics, University of Windsor, Windsor Ontario Canada N9B 3P4*.