BAYESIAN HETERO-LASSO (A GIBBS SAMPLING APPROACH)

OLOYEDE, I.
Department of Statistics, University of Ilorin, Ilorin, Nigeria.
E-mail: oloyede.i@unilorin.edu.ng **Phone No:** +234-805-304-9890

Abstract
The study investigates the asymptotic consistency and efficiency of Bayesian estimator due to violation of homoscedasticity cum non-multicollinearity properties. Mean Square Error(MSE) and Bias were the performance measuring criteria on twin non-spherical disturbances. The seed was to 12345; β were set at $\beta = \{2.5,1.5,1,0,0,0.5\}$; Xs variables as design matrix was generated from the multivariate normal distribution with $\mu > 0$ and σ_i^2. X_1 and X_2 were contaminated with Harvey (1976) heteroscedastic error structure; X_3,... X_6 were collinear covariate with pairwise correlation of 0.9, the sample sizes were set as 25, 50,70,100,200,500 and 1000. The number of replications of the experiment was set at 11,000 with burn-in of 1000 which specified the draws that were discarded to remove the effect of the initial values. The thinning was set at 5 to ensure the removal of the effect of autocorrelation in our MCMC simulation. In this paper, the study was able to depict the asymptotic consistency and efficiency of the hetero-lasso estimator at large sample sizes, the study affirmed that Bayesian hetero-lasso estimator performed well when the sample size is large. The outcome of the study revealed improved performances of the estimator in the model parameter estimates asymptotically.

Keywords: LASSO, Bayesian Inference, heteroscedasticity, mulcollinearity and Gibbs sampler

Introduction
Collinearity arises from two sources namely model and data based collinearities. Model based collinearity arises when the model are defined with collinearity structures such as lasso type estimator of which some tuning and shrinkage parameters are defined not in ordinary form, these parameters make significant impact on the selection and estimation of the parameters., the later comes from the structure of datasets where two or more predictors variables are pairwise related. More interestingly, the two types of heteroscedasticity which are model and data based heteroscedasticity earlier examined in Oloyede (2013) were incorporated in to the model and data used in the study with four different problems were mixed up. The adverse effects of collinearity manifested primarily in affecting the parameter estimates and their standard errors (Peter, 1999). A regression coefficient does not reflect any inherent effect of the particular predictor variable on the response variable but only a marginal or partial effect, given whatever other correlated predictor variables are included in the model (Neter etal, 1999).

Correlations among covariates in the predictor variables of linear econometric model cum presence of heteroscedasticity affect the precision of the inferences of the parameter estimates. Obviously, the inferences violate efficiency and consistency properties of estimation, though the estimators may be unbiased, the standard error and test of hypothesis computed for the estimators are invalid, Mean Squares Error and Bias may be inflated. 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 & Guermat, 1999; Robinson, 1987; White, 1980).

Hoerl and Kennard (1970) proposed ridge regression which minimises residual sum of squares subject to a constraint $\sum|\beta_j|^{\gamma} \le t$ where the shrinkage parameter $\gamma = 2$. 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 $\gamma > 0$ 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 $\sum|\beta_j| \leq t$ 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. Series of extension of lasso emerged recently in the literatures, adaptive lasso was invented by Zou (2006), the elastic net was introduced by Zou and Hastic (2005) which minimises RSS subject to constraint $\lambda_1 \sum|\beta_j| + \lambda_2 \sum_{j=1}^{p}|\beta_j|^2 \leq t$, where λ_1 and λ_2 denote tuning parameters one and two. Tibshirani (2005) proposed fused lasso while Group lasso was proposed by Yuan and Lin (2006), Smoothly Clipped Absolute Deviation SCAD was introduced by Fan and Li (2001). Daye et al (2011) explored high dimensional heteroscedastic regression. All these estimators did not consider the incorporation of heteroscedastic error structure, this is what this paper intend to examine. Heteroscedasticity, a significant non-spherical disturbance with multicollinearity was recently looking into in the literature. Severien Nkurunziza *et al* (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. Wenjiang (1998) proposed general approach to solve bridge regression for $\gamma \geq 1$. In their studies Qing et al (2010) opined that Bayesian elastic net outperformed elastic net in variable selection for more complicated models, it equally outperformed Bayesian lasso in prediction accuracy for small samples from less sparse modes. The choice of penalty parameters λ_1 and λ_2 is somehow simple by introducing hyper priors on them. This was exemplified by Park and Cassela (2008). Minjung etal (2010) claimed that all the lasso models with the exception of elastic net, the λ and β 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 are several relatively small signals, they opined that dirichlet prior can shrink all of them towards zero.

Since there are two categories of multicollinearity {data and model based multicollinearity} so as also there exist data and model based heteroscedasticity, the objective of this paper is to examine the asymptotic properties of linear regression model when there are presence of both multicollinearity and heteroscedasticity in both data and model.

Model Designs: Bayesian Hetero-Lasso

Let $y = X\beta + u$ with $u \sim N(0, \sigma_i^2 \Omega)$ where Ω is a positive definite matrix of order $n \times n$. A case where $u \sim N(0, \sigma^2 \mathrm{I})$ is a homoscedastic model with constant variance, but when $u \sim N(0, \sigma_i^2 \Omega)$ indicates unequal variances of the diagonal element of $n \times n$ matrix of $E(uu')$ which is regarded as heteroscedastic error structure, $y = X\beta + u$ be truncated with both collinear of different tuning γ and one component heteroscedasticity error structure with δ as the scale; y as an n-vector of random responses; X as an $n \times p$ design matrix of corrupted with collinear and heteroscedastic, β as a p-vector parameters and u as an n-vector of heteroscedastic error structures σ_i^2 $i = 1, \ldots, n$.

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + u_i \qquad 2.1$$

Let X_1 and X_2 be contaminated with multiplicative heteroscedasticity using Harvey (1976) which can be expressed as $\sigma_i^2 = \sigma^2(\beta_0 + \beta_1 X_1 + \beta_2 X_2)^\delta$ where δ is an unknown parameter which determine the degree of heteroscedasticity, thus our variables in X_3 to X_6 are embedded with collinearity. Adopting a full Bayesian inference, we examined the likelihood function, prior distribution for the parameters, hyper-parameters in the model, with MCMC algorithm.

The likelihood function of θ, where $\theta = (\beta_i, \lambda_j, \gamma, \delta)$ given the sample vector $X_i = (i = 1,2,\dots,p)'$ and $y = (y_1, y_2, \dots, y_n)'$ is expressed as

$$L(\theta, \sigma | X, y) = (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - x\beta]^2\right\} \quad 2.2$$

Incorporating multiplicative heteroscedastic-lasso into the likelihood on equation 2.2, thus error u is changed to w

$$L(\beta_i, \lambda_i, \delta_i, \sigma_i | X, y) = \left(2\pi\sigma_i^2\right)^{-n/2} \prod_{i=1}^{n} \left|w^{-\delta_I/2}\right| \exp\left\{-\frac{1}{2\sigma_i^2} \sum_{i=1}^{n} (y_i - x\beta)' w^{-\delta_I} (y_i - x\beta)\right\} \quad 2.3$$

To derive the full Bayesian density, the error density function eq 2.3 is conjugated with Gaussians, and inverse-gamma priors. It is noteworthy that Zou and Hastic(2005) said solving the Lasso problem is just like deriving marginal posterior density mode of $\beta|y$ particular when the prior distribution of β is given as $\pi(\beta) \propto exp\{-\lambda_1 \sum_{i=1}^{p} |\beta_i|\}$ Qing et al(2010). Instead multi-normal Gaussian prior is proposed for β_i, gamma prior for tuning parameter λ_i, heteroscedastic δ_i and inverse gamma prior for σ_i^2. Marginal posterior density is obtained by integrating the joint posterior density with respect to each parameter, thus, expert opinion were 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 $\pi(\beta_i, \lambda_i, \delta_i, \sigma_i) = \pi(\beta_i)\pi(\lambda_i)\pi(\delta_i)\pi(\sigma_i)$ as expressed below:

$$\pi(\beta) \propto \left(2\pi\sigma_i^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_i^2} (\beta_i - \mu)^2\right\}, \beta > 0; \quad 2.4$$

$$\pi(\lambda_i) \propto (\lambda_i)^{a_1+1} \exp(-b_1/\lambda_i), \lambda_i > 0 \quad 2.5$$

$$\pi(\delta_i) \propto (\delta_i)^{c_1+1} \exp(-d_1/\delta_i), \delta_i > 0 \quad 2.6$$

$$\pi(\sigma_i^2) \propto \left(\sigma_i^2\right)^{e_1+1} \exp\left(-f_1/\sigma_i^2\right), \sigma^2 > 0 \quad 2.7$$

The posterior distribution of $\theta = (\beta_i, \lambda_i, \delta_i, \sigma_i)$. Considering independence among the parameters is given by :

$$\pi(\beta_i, \lambda_i, \delta_i, \sigma_i | X, y) \propto$$
$$(2\pi\sigma^2)^{-\frac{n}{2}} \pi(\lambda_i)\pi(\delta_i)\pi(\sigma_i^2) \exp\left\{-\frac{1}{2\sigma^2} (\beta_i - \mu)^2\right\} \prod_{i=1}^{n} \left|w^{-\lambda/2}\right| \exp\{-\frac{1}{\sigma^4}(b_1 + d_1 + f_1 + \frac{1}{2}\sum_{i=1}^{n} (y_i - x\beta)' w^{-\lambda} (y_i - x\beta) + \lambda_1 \sum_{i=1}^{p} |\beta_i|\} \quad 2.8$$

where a_1, b_1, c_1, d_1, e_1, f_1 are the hyper-parameters for the gamma and inverse-gamma priors. Hyper-parameters are excluded for β_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 $\pi(\beta_i, \lambda_i, \delta_i, \sigma_i | X, y)$ with respect to σ_i, thus the joint posterior distribution for $(\beta_i, \lambda_i, \delta_i)$ is obtained as:

$$\pi(\beta_0, \beta_1, \beta_2, \lambda, \sigma | X, y) \propto$$
$$(2\pi)^{-\frac{n}{2}} \pi(\lambda_i)\pi(\delta_i) \exp\left\{-\frac{1}{2} (\beta_i - \mu)^2\right\} \prod_{i=1}^{n} \left|w^{-\frac{\lambda}{2}}\right| \exp\{-(b_1 + d_1 + f_1 + \frac{1}{2}\sum_{i=1}^{n} (y_i - x\beta)' w^{-\lambda} (y_i - x\beta) + \lambda_1 \sum_{i=1}^{p} |\beta_i|\}^{-(a_1 + c_1 + e_1 + n/2)} \quad 2.9$$

Gibbs Algorithm update is performed on the full conditional distribution of $\sigma_i^2 \propto IG(a_1 + \frac{n}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{x\beta})'\mathrm{w}^{-\lambda}(\mathrm{y_i} - \mathrm{x\beta}) + \lambda_1 \sum_{\mathrm{i=1}}^{\mathrm{p}}|\beta_\mathrm{i}|)$. This yielded the following full conditional density of the parameters $\beta_i, \lambda_i, \delta_i$ and σ_i:

$$\pi(\beta_i|\lambda_i, \delta_i, X, y) \propto \exp\left\{-\frac{1}{2}(\beta_i - \mu)^2\right\}\prod_{i=1}^{n}|w^{-\lambda/2}| \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{x\beta})'\mathrm{w}^{-\lambda}(\mathrm{y_i} - \mathrm{x\beta})+\lambda_1 \sum_{\mathrm{i=1}}^{\mathrm{p}}|\beta_\mathrm{i}|\}^{-(a_1+c_1+e_1+n/2)} \quad 2.10$$

$$\pi(\sigma_i|\beta_i.\lambda_i, \delta_i, X, y) \propto (\sigma_i^2)^{-(\mathrm{a_1}-1-\mathrm{n}/2)}\exp(-b_1/\sigma_i^2)\prod_{i=1}^{n}|w^{-\lambda/2}| \exp\{-\frac{1}{\sigma_i^2}(b_1 + \frac{1}{2}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{x\beta})'\mathrm{w}^{-\lambda}(\mathrm{y_i} - \mathrm{x\beta}) + \lambda_1 \sum_{\mathrm{i=1}}^{\mathrm{p}}|\beta_\mathrm{i}|\}^{-(a_1+c_1+e_1+n/2)} \quad 2.11$$

$$\pi(\lambda_i|\beta_i, \delta_i, X, y) \propto (\lambda_i)^{(\mathrm{c_1}-1-\mathrm{n}/2)}\exp(-d_1/\lambda_i)|w^{-\lambda/2}|(d_1 + \frac{1}{2}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{x\beta})'\mathrm{w}^{-\lambda}(\mathrm{y_i} - \mathrm{x\beta}) + \lambda_1 \sum_{\mathrm{i=1}}^{\mathrm{p}}|\beta_\mathrm{i}|\}^{-(a_1+c_1+e_1+n/2)} \quad 2.12$$

$$\pi(\delta_i|\beta_i, \lambda_i, X, y) \propto (\lambda_i)^{(\mathrm{e_1}-1-\mathrm{n}/2)}\exp(-f_1/\lambda_i)|w^{-\lambda/2}|(f_1 + \frac{1}{2}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{x\beta})'\mathrm{w}^{-\lambda}(\mathrm{y_i} - \mathrm{x\beta}) + \lambda_1 \sum_{\mathrm{i=1}}^{\mathrm{p}}|\beta_\mathrm{i}|\}^{-(a_1+c_1+e_1+n/2)} \quad 2.13$$

Results

In this study, Bayesian Hetero-lasso was presented with multiplicative heteroscedasticity structure and collinear covariates. Parameters were obtained through the posterior point estimate of Gibbs sampler simulation. The level 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 package.

Table 1: Performance of Bhetlasso based on Absolute Bias @ Scale of Heteroscedasticity with sample sizes

Sample	λ_1	δ^i	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
	1.6338	0.1	0.0038	0.0038	0.0073	0.0052	0.0081	0.0202
	1.8416	0.6	0.0339	0.021	0.0462	0.0209	0.0948	0.0839
25	1.8416	0.9	0.0505	0.0318	0.0644	0.038	0.1581	0.1087
	1.8416	2	0.1056	0.0763	0.1156	0.1235	0.4617	0.127
	1.4209	0.1	0.0269	0.0091	0.0033	0.0012	0.0184	0.0163
	1.4209	0.6	0.1582	0.0688	0.0192	0.0144	0.1046	0.1146
50	1.4209	0.9	0.2368	0.1043	0.0297	0.0216	0.1514	0.1788
	1.4209	2	0.5226	0.2335	0.0702	0.0463	0.3235	0.4141
	0.3309	0.1	0.0068	0.0043	0.0682	0.0267	0.0014	0.0129
	0.3309	0.6	0.0072	0.0493	0.4146	0.1532	0.0001	0.1419
70	0.3309	0.9	0.007	0.077	0.6248	0.2273	0.0009	0.2184
	0.3309	2	0.0042	0.182	1.4005	0.4974	0.0189	0.4785
	2.638	0.1	0.012	0.0016	0.0097	0.0232	0.0046	0.0099
	2.638	0.6	0.0774	0.0064	0.0466	0.1549	0.0201	0.0653
100	2.638	0.9	0.1174	0.0091	0.0568	0.2451	0.0231	0.1054
	2.638	2	0.2674	0.0176	0.0634	0.6008	0.0124	0.2796
	0.0286	0.1	0.0174	0.0013	0.0295	0.0491	0.0326	0.0609
	0.0286	0.6	0.0053	0.0014	0.0275	0.0604	0.0562	0.102
200	0.0286	0.9	0.0016	0.0026	0.0282	0.0694	0.0681	0.1295
	0.0286	2	0.0249	0.0046	0.044	0.1172	0.0938	0.2531
	2.0532	0.1	0.0017	0.0078	0.0041	0.0057	0.0091	0.0003
	2.0532	0.6	0.0081	0.0457	0.0209	0.0429	0.0543	0.0024
500	2.0532	0.9	0.0121	0.0686	0.0294	0.0666	0.0837	0.0056
	2.0532	2	0.0283	0.1532	0.0528	0.161	0.2027	0.0284
	2.7683	0.1	0.0004	0.0009	0.0055	0.0029	0.0008	0.0003
	2.7683	0.6	0.0033	0.0057	0.0411	0.0144	0.0037	0.0037
1000	2.7683	0.9	0.005	0.0084	0.0631	0.0209	0.006	0.0055
	2.7683	2	0.0117	0.0171	0.1456	0.045	0.0177	0.0103

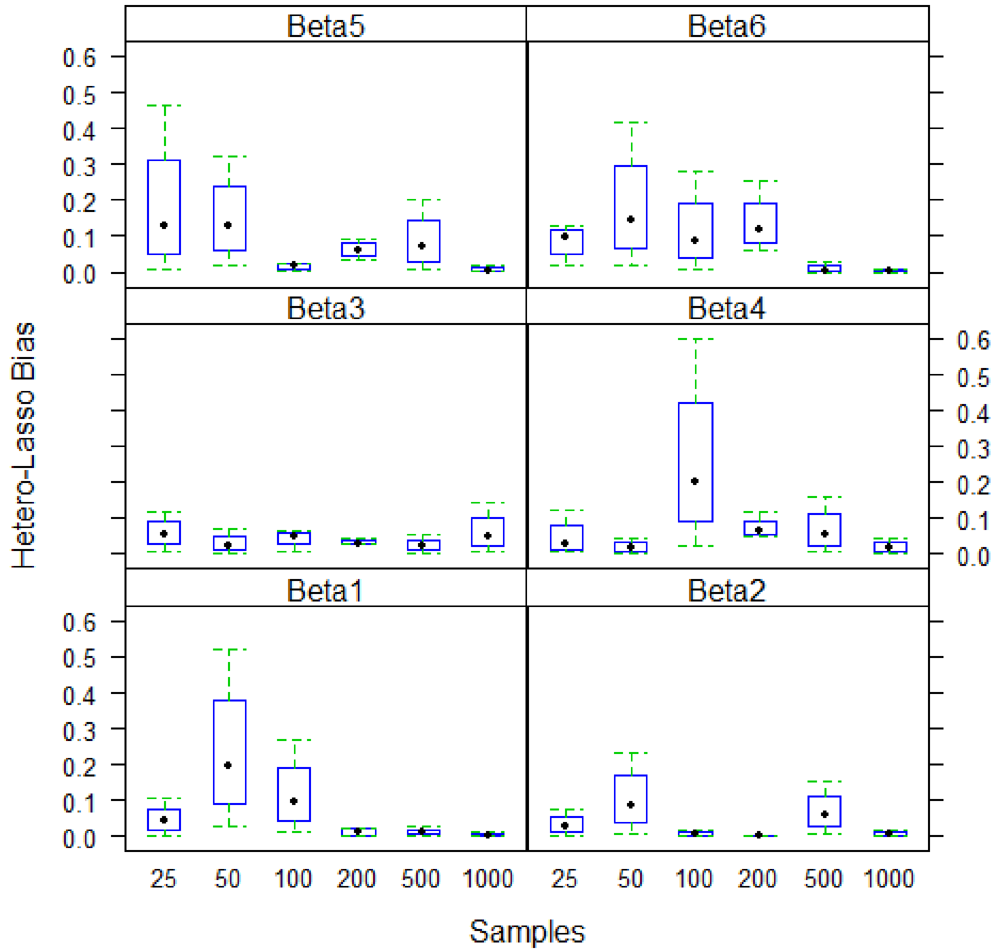

Figure 1: Depicting Performance of Bhetlasso based on Absolute Bias @ Scale of Heteroscedasticity with sample sizes

Table and figure 1 revealed the outcome of the estimation of Bayesian hetero-lasso based on the Absolute Bias Performances with different Scales of Heteroscedasticity and $\rho = 0.9$. The study observed that biases for $\beta_1, \beta_2, \beta_3$, β_4 β_5, and β_6 were consistent, as the scale of heteroscedasticity increased so also the biases were absolutely increased algebraically. At sample size 70, the biases were absolutely interchangeable for β_1 and β_5 which brought about inconsistency while for $\beta_2, \beta_3, \beta_4, \beta_6$, the study observed consistency since increase in scale of heteroscedasticity brought about absolute increase in the biases. Likewise, sample size 200 brought about inconsistency for β_1 and β_2, while other parameters were consistent.

Table 2: Performance of Bhetlasso based on mean squares error @ scale of Heteroscedasticity with sample sizes

Sample	δ^i	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
	0.1	0.0051	0.0001	0.0009	0.0081	0.0048	0.0025
25	0.6	0.0014	0.0006	0.0037	0.0039	0.0131	0.0097
	0.9	0.0028	0.0012	0.0058	0.005	0.029	0.0143
	2	0.0113	0.0059	0.0146	0.0187	0.2158	0.0176
	0.1	0.0015	0.0176	0.0001	0.0026	0.0061	0.0017
50	0.6	0.0257	0.0226	0.0006	0.003	0.0174	0.0148
	0.9	0.0567	0.029	0.0012	0.0034	0.0297	0.0338
	2	0.2735	0.0734	0.0055	0.0053	0.1116	0.173
	0.1	0.0853	0.0369	0.0048	0.0009	0.0145	0.304
70	0.6	0.0853	0.0397	0.1724	0.024	0.0153	0.3281
	0.9	0.0854	0.0435	0.3909	0.0523	0.0156	0.3578
	2	0.0853	0.0716	1.9622	0.2477	0.0153	0.5449
	0.1	0.0011	0.0006	0.0014	0.0006	0.0003	0.001
100	0.6	0.0069	0.0006	0.0037	0.0243	0.0009	0.0052
	0.9	0.0148	0.0007	0.0049	0.0604	0.0011	0.012
	2	0.0727	0.0009	0.0056	0.3614	0.0005	0.0787
	0.1	0.7984	0.0074	1.8383	4.504	1.5227	5.6926
200	0.6	0.7987	0.0074	1.8385	4.5043	1.5237	5.695
	0.9	0.7989	0.0074	1.8385	4.5046	1.5245	5.6985
	2	0.8005	0.0074	1.8386	4.5089	1.5271	5.7325
	0.1	0.0004	0.0002	0.0001	0.003	0.0013	0.0005
500	0.6	0.0005	0.0022	0.0006	0.0049	0.0042	0.0006
	0.9	0.0006	0.0048	0.0011	0.0076	0.0082	0.0006
	2	0.0012	0.0235	0.0029	0.0293	0.042	0.0013
	0.1	0.0001	0	0.0045	0.0005	0.0001	0.0003
1000	0.6	0.0001	0.0001	0.0063	0.0007	0.0002	0.0003
	0.9	0.0001	0.0001	0.0087	0.0009	0.0002	0.0004
	2	0.0003	0.0003	0.0262	0.0025	0.0005	0.0004

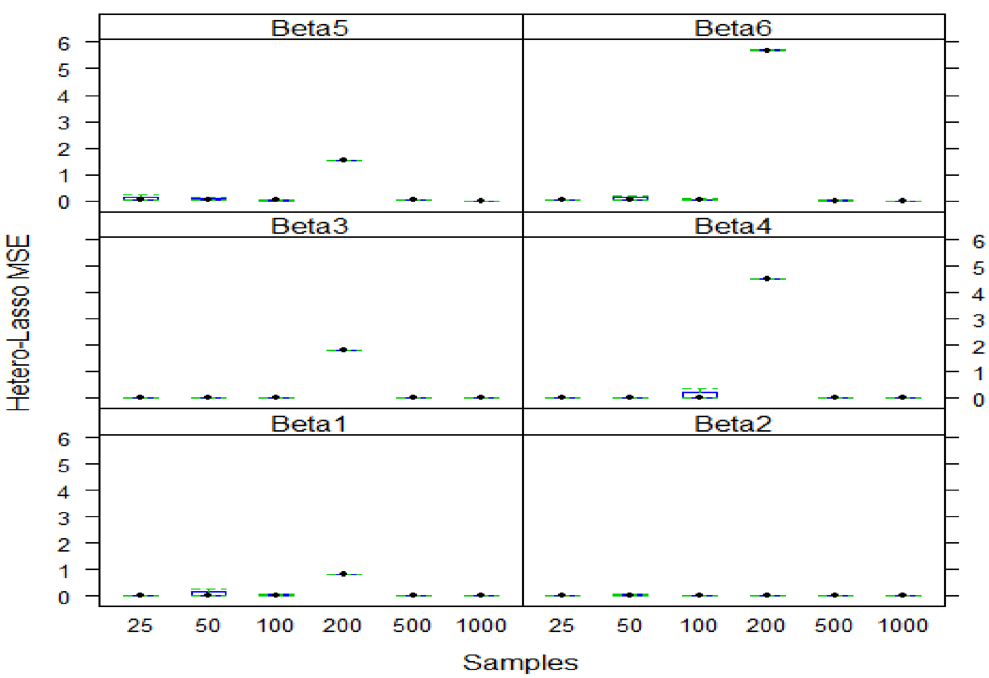

Figure 3.2: Showing the performance of Bhetlasso based on mean squares error @ scale of Heteroscedasticity with sample sizes

Table and figure 2 revealed that mean squares error for β_1, β_2 β_3, β_4 , β_5 and β_6 were all depicting efficiency across all sample sizes considered in the study, as the scale of heteroscedasticity increased, so also the mean squares error for all parameters increased algebraically.

Table 3: Performance of Bhetlasso based on Bias @ sample sizes with Scale of Heteroscedasticity

Samples	δ^i	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
25	0.1	0.0038	0.0038	0.0073	0.0052	0.0081	0.0202
50	0.1	0.0269	0.0091	0.0033	0.0012	0.0184	0.0163
70	0.1	0.0068	0.0043	0.0682	0.0267	0.0014	0.0129
100	0.1	0.012	0.0016	0.0097	0.0232	0.0046	0.0099
200	0.1	0.0174	0.0013	0.0295	0.0491	0.0326	0.0609
500	0.1	0.0017	0.0078	0.0041	0.0057	0.0091	0.0003
1000	0.1	0.0004	0.0009	0.0055	0.0029	0.0008	0.0003
25	0.6	0.0339	0.021	0.0462	0.0209	0.0948	0.0839
50	0.6	0.1582	0.0688	0.0192	0.0144	0.1046	0.1146
70	0.6	0.0072	0.0493	0.4146	0.1532	0.001	0.1419
100	0.6	0.0774	0.0064	0.0466	0.1549	0.0201	0.0653
200	0.6	0.0053	0.0014	0.0275	0.0604	0.0562	0.102
500	0.6	0.0081	0.0457	0.0209	0.0429	0.0543	0.0024
1000	0.6	0.0033	0.0057	0.0411	0.0144	0.0037	0.0037
25	0.9	0.0505	0.0318	0.0644	0.038	0.1581	0.1087
50	0.9	0.2368	0.1043	0.0297	0.0216	0.1514	0.1788
70	0.9	0.007	0.077	0.6248	0.2273	0.0009	0.2184
100	0.9	0.1174	0.0091	0.0568	0.2451	0.0231	0.1054
200	0.9	0.0016	0.0026	0.0282	0.0694	0.0681	0.1295
500	0.9	0.0121	0.0686	0.0294	0.0666	0.0837	0.0056
1000	0.9	0.005	0.0084	0.0631	0.0209	0.006	0.0055
25	2	0.1056	0.0763	0.1156	0.1235	0.4617	0.127
50	2	0.5226	0.2335	0.0702	0.0463	0.3235	0.4141
70	2	0.0042	0.182	1.4005	0.4974	0.0189	0.4785
100	2	0.2674	0.0176	0.0634	0.6008	0.0124	0.2796
200	2	0.0249	0.0046	0.044	0.1172	0.0938	0.2531
500	2	0.0283	0.1532	0.0528	0.161	0.2027	0.0284
1000	2	0.0117	0.0171	0.1456	0.045	0.0177	0.0103

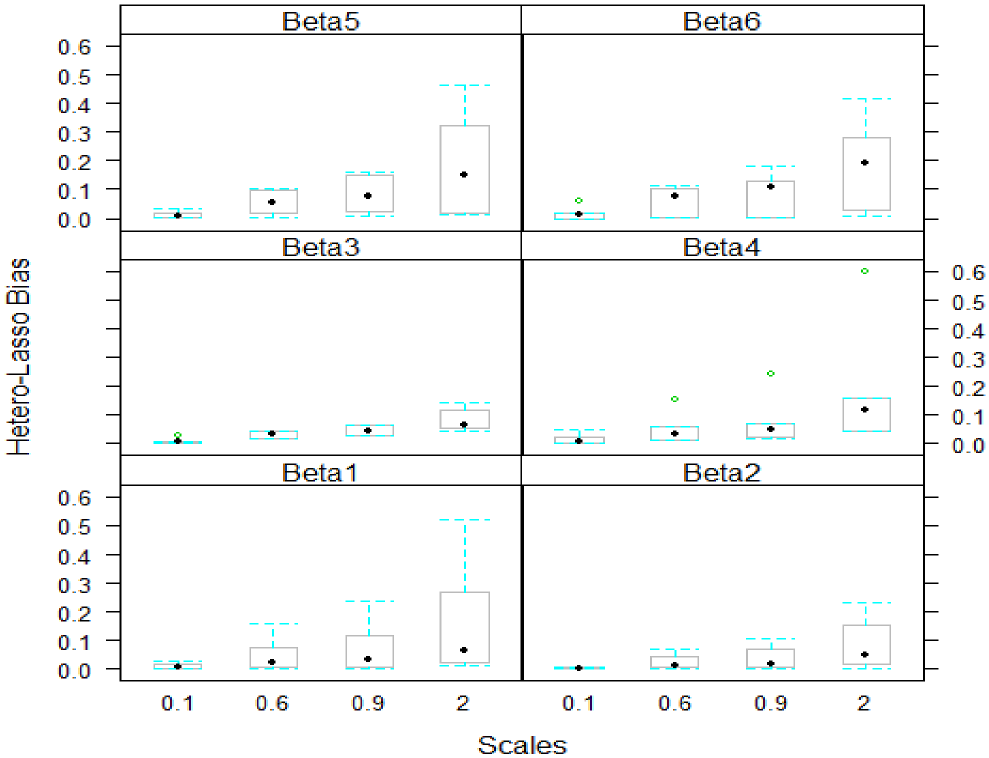

Figure 3: Depicting performance of bhetlasso based on bias @ sample sizes with scale of heteroscedasticity

Table and figure 3 revealed the inconsistency from sample sizes 25 to 200 for the parameter estimates β_1, β_2, β_3 , β_4 , β_5 and β_6 across the scale of heteroscedasticity, this is so since an increase in sample sizes brought about interchangeable in the biases obtained for all the parameters. Interestingly at large sample sizes of 500 and 1000, the study observed consistency since an increase in sample size brought about decrease in the bias.

Table 4: Performance of bhlasso based on mean squares error @ sample sizes with scale of heteroscedasticity

Samples	δ^i	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
25	0.1	0.0051	0.0001	0.0009	0.0081	0.0048	0.0025
50	0.1	0.0015	0.0176	0.0001	0.0026	0.0061	0.0017
70	0.1	0.0853	0.0369	0.0048	0.0009	0.0145	0.304
100	0.1	0.0011	0.0006	0.0014	0.0006	0.0003	0.001
200	0.1	0.7984	0.0074	1.8383	4.504	1.5227	5.6926
500	0.1	0.0004	0.0002	0.0001	0.003	0.0013	0.0005
1000	0.1	0.0001	0	0.0045	0.0005	0.0001	0.0003
25	0.6	0.0014	0.0006	0.0037	0.0039	0.0131	0.0097
50	0.6	0.0257	0.0226	0.0006	0.003	0.0174	0.0148
70	0.6	0.0853	0.0397	0.1724	0.024	0.0153	0.3281
100	0.6	0.0069	0.0006	0.0037	0.0243	0.0009	0.0052
200	0.6	0.7987	0.0074	1.8385	4.5043	1.5237	5.695
500	0.6	0.0005	0.0022	0.0006	0.0049	0.0042	0.0006
1000	0.6	0.0001	0.0001	0.0063	0.0007	0.0002	0.0003
25	0.9	0.0028	0.0012	0.0058	0.005	0.029	0.0143
50	0.9	0.0567	0.029	0.0012	0.0034	0.0297	0.0338
70	0.9	0.0854	0.0435	0.3909	0.0523	0.0156	0.3578
100	0.9	0.0148	0.0007	0.0049	0.0604	0.0011	0.012

200	0.9	0.7989	0.0074	1.8385	4.5046	1.5245	5.6985
500	0.9	0.0006	0.0048	0.0011	0.0076	0.0082	0.0006
1000	0.9	0.0001	0.0001	0.0087	0.0009	0.0002	0.0004
25	2	0.0113	0.0059	0.0146	0.0187	0.2158	0.0176
50	2	0.2735	0.0734	0.0055	0.0053	0.1116	0.173
70	2	0.0853	0.0716	1.9622	0.2477	0.0153	0.5449
100	2	0.0727	0.0009	0.0056	0.3614	0.0005	0.0787
200	2	0.8005	0.0074	1.8386	4.5089	1.5271	5.7325
500	2	0.0012	0.0235	0.0029	0.0293	0.042	0.0013
1000	2	0.0003	0.0003	0.0262	0.0025	0.0005	0.0004

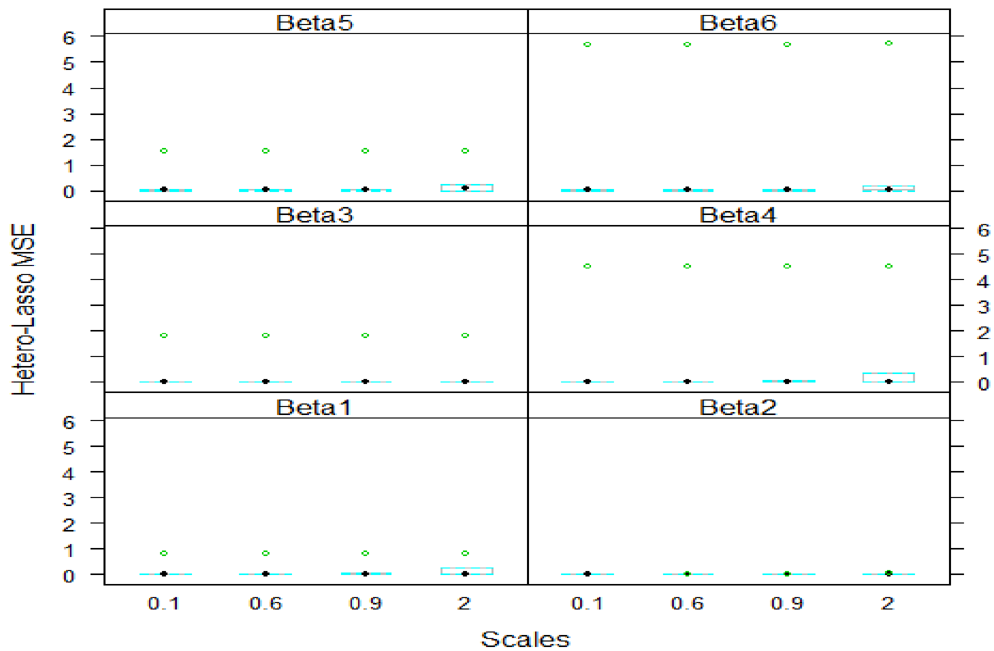

Figure 4: Performance of bhlasso based on mean squares error @ sample sizes with scale of heteroscedasticity

Table and figure 4 revealed similar patterns for all the scale of heteroscedasticity from scale 0.1 to 2, the Mean Squares Error depicted inefficiency for sample from sample sizes 25 to 200 but brought about efficiency for large samples from 500 sample size upward. Asymptotically, the study opined that the effects of the two non-spherical disturbances were reduce at larger sample sizes.

Conclusion

In this paper, a simple way of modelling and estimating heteroscedastic-collinear model under simulation approach (MCMC) was presented by incorporating it into the celebrated lasso estimator. The study observed that modelling hetero-lasso in a full Bayesian improved the precision of the inferences of the estimates at larger samples. The study found that $_{1}$ and $_{2}$ were affected as the scale of heteroscedasticity increased while $_{3}$- $_{6}$ behaved in different way. The study concluded that asymptotically, there exist consistency and efficiency in the estimation. The approach can be applied to further studies in the area of simultaneous equation and other econometric models and non-econometric models.

References

Cobb, C., & Douglas, P. (1928). A theory of production. *American Economic Review*, *18,* 139-250.

Daye, J., Chen, J., & Li, H. (2011). High-dimensional heteroscedastic regression with an application to eQTL. *Data Analysis Biometrics*, *68*(1), 316-326.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32,* 407-451.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96,* 1348-1360.

Frank, I. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics, 35,* 109-135.

Hadri, K., & Guermat, C. (1999). *Heteroscedasticity in stochastic frontier models: A monte carlo analysis.* University of Exeter, USA., P. 18.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica, 44,* 461-465.

Hoerl, A. E., & Kennard, R. W (1970). Ridge regression: Applications to non-orthogonal problems. *Technometrics, 12,* 55-68.

Minjung, K., Jeff, G., Malay, G., & George, C. (2010). Penalized regression, standard errors, and *Bayesian Lassos, 5*(2), 369-412.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association, 103,* 681-686.

Peter, L. B. (1999). The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharmaceutical Research*, 6(5).

Qing, L., & Nan, L. (2010). The Bayesian elastic net. *International Society for Bayesian Analysis, 5*(1), 151-170.

R-Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Robinson, P. M. (1987). Asymptotic efficient estimation in the presence of Heteroscedasticity of unknown form. *Econometrica, 55,* 875 -891.

Severien, N., & Eric, Y. L. (2012). Shrinkage and LASSO strategies in high dimensional heteroscedastic models. Department of Mathematics and Statistics, University of Windsor, Windsor Ontario Canada N9B 3P4.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society. Series B* 67, 91-108.

Wenjiang, J. F. (1998). Penalized regression: The bridge versus the Lasso. American Statistical Association, Institute of Mathematical Statistics and interface Foundation of North America. *Journal of Computational and Graphical Statistics, 7*(3), 397-416.

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

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* 68, 49-67.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67, 301-320.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101,* 1418-1429.