Tanzania Journal of Science 47(3): 981-987, 2021
ISSN 0856-1761, e-ISSN 2507-7961
© College of Natural and Applied Sciences, University of Dar es Salaam, 2021

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 $\beta = \{2.5,3,1.5,1,0,0,0.5\}$; Xs variables were generated as follow: the design matrix was generated from the multivariate normal distribution with mean > 0 and variance σ_i^2. X_1 and X_2 are truncated with Harvey (1976) heteroscedastic error structure; $X_3, \ldots, X_6$ 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 $\sum|\beta_j|^\gamma \leq t$ 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 $\gamma \geq 0$ 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 $\sum|\beta_j| \leq t$ 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 (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 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 λ_1 and λ_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 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 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 multicolliinearity 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 Ω is a positive definite matrix of order $n \times n$. The model is 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 collinear and heteroscedastic, β as a p-vector parameters and u as an n-vector of heteroscedastic error structures σ_i^2 $i = 1, \ldots, n$.

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

Let X_1 and X_2 be truncated 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 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 θ, where $\theta = (\beta_i, \lambda_j, \gamma, \delta)$ give the sample vector $X_i = (i = 1,2,\ldots,p)'$ and $y = (y_1, y_2, \ldots, y_n)'$ is expressed as $L(\theta,\sigma|X,y) = (2\pi\sigma^2)^{-n/2}\prod_{i=1}^{n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - x\beta]^2\right\}$ (2)

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

Bayesian hetero-elastic net

$$L(\theta,\sigma|X,y) = (2\pi\sigma^2)^{-n/2}\prod_{i=1}^{n}\left|w^{-\lambda/2}\right|\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta)\right\} \quad (3)$$

Incorporating elastic net in to the above model we have

$$L(\theta,\sigma|X,y) = (2\pi\sigma^2)^{-n/2}\prod_{i=1}^{n}\left|w^{-\lambda/2}\right|\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1\sum_{i=1}^{p}|\beta_i| + \lambda_2\sum_{i=1}^{p}|\beta_i|^2\right\} \quad (4)$$

Or

$$\hat{\beta}_{HEN} = \underset{\beta}{argmin}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1\sum_{i=1}^{p}|\beta_i| + \lambda_2\sum_{i=1}^{p}|\beta_i|^2 \quad (5)$$

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 β (Li and Lin 2010) is given as

$$\pi(\beta) \propto exp\left\{-\lambda_1\sum_{i=1}^{p}|\beta_i| - \lambda_2\sum_{i=1}^{p}|\beta_i|^2\right\} \quad (6)$$

Instead we proposed multinomial Gaussian prior for the β_i, gamma prior for tuning parameter λ_i, heteroscedastic δ_i and inverse gamma prior σ_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 σ_i as independent marginal distribution. The study assumed a prior density $\pi(\beta_i,\lambda_i,\delta_i,\sigma_i) = \pi(\beta_i)\pi(\lambda_i)\pi(\delta_i)\pi(\sigma_i)$. Thus, $\pi(\beta) \propto (2\pi\sigma_i^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_i^2}(\beta_i - \mu)^2\right\}, \beta > 0$ (7) $\pi(\lambda_i) \propto (\lambda_i)^{a_1+1}\exp(-b_1/\lambda_i), \lambda_i > 0$ (8)

$\pi(\delta_i) \propto (\delta_i)^{c_1+1}\exp(-d_1/\delta_i), \delta_i > 0$ (9)

$\pi(\sigma_i^2) \propto (\sigma_i^2)^{e_1+1}\exp(-f_1/\sigma_i^2), \sigma^2 > 0$ (10)

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

$$\pi(\beta_i,\lambda_i,\delta_i,\sigma_i|X,y) \propto (2\pi\sigma^2)^{-\frac{n}{2}}\pi(\lambda_i)\pi(\delta_i)\pi(\sigma_i^2)\exp\left\{-\frac{1}{2\sigma^2}(\beta_i - \mu)^2\right\} \quad (11a)$$

$$\prod_{i=1}^{n}\left|w^{-\lambda/2}\right|\exp\{-\frac{1}{\sigma^4}(b_1 + d_1 + f_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1\sum_{i=1}^{p}|\beta_i| + \lambda_2\sum_{i=1}^{p}|\beta_i|^2\} \quad (11b)$$

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 we have joint a posterior distribution for $(\beta_i,\lambda_i,\delta_i)$

$$\pi(\beta_0,\beta_1,\beta_2,\lambda,\sigma|X,y) \propto (2\pi)^{-\frac{n}{2}}\pi(\lambda_i)\pi(\delta_i)\exp\left\{-\frac{1}{2}(\beta_i - \mu)^2\right\}\prod_{i=1}^{n}\left|w^{-\frac{\lambda}{2}}\right| \quad (12)$$

$\exp\{-(b_1 + d_1 + f_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}^{-(a_1+c_1+e_1+n/2)}$

Gibbs algorithm update is performed on the full conditional distribution of $\sigma_i^2 \propto IG(a_1 + \frac{n}{2}, b_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2)$. This yields the following full conditional density of the parameters $\beta_i, \lambda_i, \delta_i$ and σ_i:

$$\pi(\beta_i|\lambda_i, \delta_i, X, y) \propto \exp\left\{-\frac{1}{2}(\beta_i - \mu)^2\right\} \prod_{i=1}^{n}\left|w^{-\lambda/2}\right| \exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}^{-(a_1+c_1+e_1+n/2)} \quad (13)$$

$$\pi(\sigma_i|\beta_i.\lambda_i, \delta_i, X, y) \propto (\sigma_i^2)^{-\left(a_1-1-\frac{n}{2}\right)} \exp\left(-\frac{b_1}{\sigma_i^2}\right) \prod_{i=1}^{n}\left|w^{-\lambda/2}\right| \exp\{-\frac{1}{\sigma_i^2}(b_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}^{-(a_1+c_1+e_1+n/2)} \quad (14)$$

$$\pi(\lambda_i|\beta_i, \delta_i, X, y) \propto (\lambda_i)^{\left(c_1-1-\frac{n}{2}\right)} \exp\left(-\frac{d_1}{\lambda_i}\right)\left|w^{-\frac{\lambda}{2}}\right|(d_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}^{-(a_1+c_1+e_1+n/2)} \quad (15)$$

$$\pi(\delta_i|\beta_i, \lambda_i, X, y) \propto (\lambda_i)^{\left(e_1-1-\frac{n}{2}\right)} \exp\left(-\frac{f_1}{\lambda_i}\right)\left|w^{-\frac{\lambda}{2}}\right|(f_1 + \frac{1}{2}\sum_{i=1}^{n}(y_i - x\beta)'w^{-\lambda}(y_i - x\beta) + \lambda_1 \sum_{i=1}^{p}|\beta_i| + \lambda_2 \sum_{i=1}^{p}|\beta_i|^2\}^{-(a_1+c_1+e_1+n/2)} \quad (16)$$

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 β_i , in this study there is no objection for such claim. Thus we adopted those priors along with the stated priors.

$$\pi(\beta_i) = \frac{1}{2\tau} exp\left(-\frac{|\beta_j|}{\tau}\right) \text{ with } \tau = 1\backslash\lambda. \quad (16)$$

Park and Casella (2008) used empirical Bayes estimates for the penalty parameters λ_1 and λ_2, which are the maximization of the data marginal likelihood. This is accomplished by treating β_i, δ_i ,σ_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; β was set at $\beta = \{2.5,3,1.5,1,0,0,0.5\}$; Xs variables were generated as follows: the design matrix was generated from the multivariate normal distribution with mean > 0 and variance σ_i^2. X_1 and X_2 are truncated with Harvey (1976) heteroscedastic error structure; $X_3, \ldots, X_6$ 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 ,λ_1 , λ_2 and σ^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 $\hat{\beta}_6$ increase algebraically as the scale of heteroscedasticity increases Thus, there exists consistency. 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

Sample	λ_1	λ_2	δ^i	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
	6.27	9.31	0.1	0.1161	0.5053	0.0134	0.0006	0.0335	0.0305
	6.27	6.53	0.2	0.2246	1.0073	0.0244	0.0061	0.0309	0.0262
	6.27	4.38	0.3	0.3281	1.5075	0.0376	0.0121	0.0289	0.0215
	9.40	8.80	0.4	0.4248	2.0045	0.0481	0.0303	0.0579	0.0565
	9.40	6.62	0.5	0.5199	2.4988	0.0614	0.0437	0.0635	0.0617
	9.40	5.08	0.6	0.6106	2.9901	0.0762	0.0588	0.0688	0.0665
	9.40	3.99	0.7	0.6964	3.4780	0.0923	0.0757	0.0737	0.0706
25	9.40	3.20	0.8	0.7775	3.9624	0.1097	0.0943	0.0778	0.0739
	9.40	2.6	0.9	0.8536	4.4427	0.1286	0.1148	0.0812	0.0760
	9.40	2.17	1	0.9245	4.9191	0.1489	0.1369	0.0835	0.0769
	9.40	0.63	2	1.33	9.4061	0.4246	0.4593	0.0090	0.0274
	34.9	5.05	0.1	0.1497	0.6251	0.0347	0.0373	0.0207	0.0167
	34.9	2.08	0.2	0.2984	1.2493	0.0646	0.0703	0.0502	0.0398
	34.9	1.06	0.3	0.4472	1.8725	0.0975	0.1064	0.0828	0.0649
	34.9	0.62	0.4	0.5957	2.4942	0.1332	0.1457	0.1182	0.0917
	34.9	0.41	0.5	0.7440	3.1137	0.1715	0.1877	0.1564	0.1199
	34.9	0.29	0.6	0.8918	3.7305	0.2123	0.2325	0.1971	0.1494
	34.9	0.21	0.7	1.0389	4.3439	0.2553	0.2797	0.2401	0.1799
100	34.9	0.16	0.8	1.1851	4.9534	0.3004	0.3292	0.2852	0.2111
	34.9	0.13	0.9	1.3301	5.5585	0.3473	0.3807	0.3323	0.2429
	34.9	0.1	1	1.4737	6.1585	0.3957	0.4341	0.3811	0.2749
	34.9	0.02	2	2.7947	11.7726	0.9120	1.0190	0.9159	0.5663
	26.69	31.6	0.1	0.1586	0.6049	0.0077	0.0096	0.0083	0.0079
	26.69	20.6	0.2	0.3173	1.2091	0.0152	0.0197	0.0154	0.0137
	26.69	13.1	0.3	0.4755	1.8117	0.0236	0.0311	0.0231	0.0202
	26.69	8.67	0.4	0.6329	2.4121	0.0328	0.0435	0.0316	0.0272
	26.69	6.05	0.5	0.7895	3.0096	0.0429	0.0572	0.0406	0.0346
	26.69	4.43	0.6	0.9449	3.6038	0.0538	0.0719	0.0502	0.0424
1000	26.69	3.37	0.7	1.0990	4.1941	0.0656	0.0878	0.0603	0.0505
	26.69	2.64	0.8	1.2515	4.7798	0.0782	0.1046	0.0706	0.0588
	26.69	2.13	0.9	1.4021	5.3606	0.0915	0.1225	0.0814	0.0673
	26.69	1.7	1	1.5508	5.9357	0.1055	0.1413	0.0924	0.0758
	26.69	0.50	2	2.8908	11.2847	0.2768	0.3691	0.2001	0.1501

Table 2 showed the mean squared error criteria, the mean squares errors for $\hat{\beta}_s$ decrease algebraically as the sample sizes increase irrespective of the scale of heteroscedasticity. Thus, sample size 1000 has the least mean squares error, 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 Guermat (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

Sample	λ_1	λ_2	δ^i	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
	6.27	9.31	0.1	0.0525	0.2902	0.3203	0.3126	0.7168	0.7132
	6.27	6.53	0.2	0.0826	1.0433	0.2648	0.2579	0.5910	0.5871
	6.27	4.38	0.3	0.1341	2.2959	0.2196	0.2128	0.4872	0.4832
	9.40	8.80	0.4	0.2023	4.0370	0.1822	0.1759	0.4043	0.3998
	9.40	6.62	0.5	0.2883	6.2597	0.1523	0.1462	0.3345	0.3303
	9.40	5.08	0.6	0.3876	8.9534	0.1283	0.1224	0.2770	0.2731
	9.40	3.99	0.7	0.4973	12.1072	0.1096	0.1038	0.2298	0.2260
25	9.40	3.20	0.8	0.6146	15.7088	0.0954	0.0897	0.1909	0.1873
	9.40	2.6	0.9	0.7369	19.7454	0.0853	0.0797	0.1588	0.1554
	9.40	2.17	1	0.8615	24.2031	0.0789	0.0736	0.1323	0.1289
	9.40	0.63	2	1.7698	88.4746	0.1885	0.2188	0.0179	0.0181
	34.9	5.05	0.1	0.0313	0.3989	0.1295	0.1332	0.1178	0.1188
	34.9	2.08	0.2	0.0963	1.5675	0.1095	0.1130	0.0988	0.0988
	34.9	1.06	0.3	0.2059	3.5119	0.0959	0.0999	0.0859	0.0839
	34.9	0.62	0.4	0.3598	6.2255	0.0886	0.0938	0.0788	0.0738
100	34.9	0.41	0.5	0.5576	9.6987	0.0876	0.0947	0.0776	0.0680
	34.9	0.29	0.6	0.7986	13.9193	0.0928	0.1027	0.0825	0.0663
	34.9	0.21	0.7	1.0819	18.8719	0.1043	0.1181	0.0934	0.0684
	34.9	0.16	0.8	1.4065	24.5385	0.1223	0.1409	0.1106	0.0741
	34.9	0.13	0.9	1.7708	30.8984	0.1468	0.1716	0.1344	0.0832
	34.9	0.1	1	2.1734	37.9282	0.1781	0.2102	0.1648	0.0954
	34.9	0.02	2	7.8108	138.5934	0.8347	1.0413	0.8415	0.3234
	26.69	31.6	0.1	0.0259	0.36681	0.0111	0.0113	0.0129	0.0128
	26.69	20.6	0.2	0.1013	1.4626	0.0092	0.0096	0.0108	0.0106
	26.69	13.1	0.3	0.2265	3.2827	0.0079	0.0085	0.0092	0.0090
	26.69	8.67	0.4	0.4011	5.8185	0.0072	0.0082	0.0081	0.0078
	26.69	6.05	0.5	0.6237	9.0583	0.0068	0.0084	0.0075	0.0069
	26.69	4.43	0.6	0.8932	12.9881	0.0070	0.0092	0.0073	0.0066
1000	26.69	3.37	0.7	1.2081	17.5909	0.0077	0.0112	0.0076	0.0065
	26.69	2.64	0.8	1.5665	22.8475	0.0088	0.0137	0.0082	0.0066
	26.69	2.13	0.9	1.9662	28.7363	0.0106	0.0173	0.0093	0.0071
	26.69	1.7	1	2.4053	35.2334	0.0129	0.0218	0.0107	0.0078
	26.69	0.50	2	8.3569	127.3451	0.0769	0.1365	0.0403	0.0228

Conclusion

The study observed that modelling hetero-elasticnet 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 $X_3,\ldots,X_6$ 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

Casella G, Ghosh M, Gill J and Kyung M 2010 Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 5(2): 369-411.

Cobb C and Douglas P 1928 A theory of production. *Am. Economic Rev.* 18:139-165.

Daye J, Chen J and Li H 2012 High-dimensional heteroscedastic regression with an application to eQTL data analysis. *Biometrics* 68(1): 316-326.

Efron B, Hastie T, Johnstone I and Tibshirani R 2004 Least angle regression. *Ann. Stat.* 32(2): 407-499.

Fan J and Li R. 2001 Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96: 1348-1360.

Frank LE and Friedman JH 1993 A statistical view of some chemometrics regression tools. *Technometrics* 35: 109-135.

Fu WJ 1998 Penalized regressions: the bridge versus the Lasso. *J. Comput. Graph. Stat.* 7(3): 397-416.

Gelman A and Rubin DB 1992 A single series from the Gibbs sampler provides a false sense of security/ In: Bernado JM, Berger JO, David AP and A. Smith FM (Eds) *Bayesian Statistics 4,* 625-632. Oxford University Press.

Hadri K and Guermat C 1999 Heteroscedasticity in Stochastic Frontier Models: A Monte Carlo Analysis. University of Exeter, USA.

Harvey AC 1976 Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44: 461-465.

Hoerl AE and Kennard RW 1970 Ridge regression: applications to non-orthogonal problems. *Technometrics* 12: 55-68.

Kayanan M and Wijekoon P 2020 Stochastic Restricted Lasso-Type Estimator in the Linear Regression Model. *J. Probab. Stat.* 2020.

Li Q and Lin N 2010 The Bayesian elastic net. *Bayesian Anal.* 5(1): 151-170.

Park T and Casella G 2008 The Bayesian Lasso. *J. Am. Stat. Assoc.* 103: 681-686.

R Core Team 2020 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Robinson PM 1987 Asymptotic efficient estimation in the presence of Heteroscedasticity of unknown form. *Econometrica* 55: 875-891.

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.*

Tibshirani R 1996 Regression shrinkage and selection via the Lasso. *J. Royal Stat. Soc. Ser. B* 58: 267-288.

Tibshirani R, Saunders M, Rosset S, Zhu J and Knight K 2005 Sparsity and smoothness via the fused Lasso. *J. Royal Stat. Soc. Ser. B* 67: 91-108.

White H 1980 A heteroskedasticity consistent covariance matrix and direct test for heteroskedasticity. *Econometrica* 48: 817-838.

Yuan M and Lin Y 2006 Model selection and estimation in regression with grouped variables. *J. Royal Stat. Soc. Ser. B* 68: 49-67.

Zou H and Hastie T 2005 Regularization and Variable Selection via the Elastic Net. *J. Royal Stat. Soc. Ser. B* 67: 301-320.

Zou H 2006 The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.* 101: 1418-1429.