Bayesian Hierarchical Modeling of Asymmetric Effect of Autocorrelated Error

***Oloyede, I.**

Department of Statistics, University of Ilorin, Ilorin, Nigeria

Received: November 10, 2015; Revised: December, 24, 2015; Accepted: December 31, 2015.

Abstract

This study investigates the asymmetric effect of autocorrelated error of hierarchical model via Bayesian paradigm. The study employed full Bayesian experiment by considering the marginal conditional posteriors density of the model parameters estimate. The extreme cases of autocorrelated error were considered by selecting -0.99 and 0.99 for rho. The seed was set to 12345; β_s *were set at 2.5, 1.5, 0.5; Xs variables were generated using uniform distribution. The number of replications of our 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 removal of the effect of autocorrelation in Markov Chain Monte Carlo simulation. The study revealed that positive correlation had higher impact than negative correlation when the magnitude is 0.9; whereas at lower correlation, negative correlation had higher impact. The study affirmed improvement in consistency and efficiency on the model parameters estimates.*

Keywords: Asymmetric effect, Autocorrelated error, Bayesian inference, Pooled regression, Shared variance.

1.0 Introduction

Hierarchical linear modeling is an advanced form of ordinary least squares regression estimator which stands as analysis of variances in the outcome variables particularly when regressors are at varying hierarchical levels. Both individual and pooled effects were considered in hierarchical modeling [1]. Assuming data are examined by ordinary least squares, such procedure would lack consideration of shared variance and this will tantamount to loss of standard inferences of model parameter estimates. The wide spread application of hierarchical multi-level data analysis emanates from previous works of some researchers [2-4]. Hierarchical modeling had been applied to varied fields of studies including education, health, business, social works, as well as statistical techniques such as mixed level and effects, random and fixed effects, covariance components-modeling [4]. It is best to examine the relationship between predictors and outcome variable with the premise of shared variances of both levels 1 and 2. The advantage of this approach is premised on the requirement of fewer assumptions to be considered when exploring multilevel data [4].

Hierarchical models and more general; linear and non-linear mixed models are extensively applied in the natural and social sciences, and discussed in detail from [5, 6]. Perry [7] examined hierarchical models via a fast moment-base through comparison of different estimations under simulation. Bottou [8] proposed stochastic gradient descent; while other researchers examined splitting strategies where data are broken into 10 subsets in which separate estimates for each subset are computed followed by combination of the estimates by averaging them [6, 9, 10]. Bates *et al*. [11] examined hierarchical problems with maximum likelihood, adopting gradient-free optimization procedure.

Hierarchical linear models allow for the simultaneous investigation of the relationship within a given hierarchical level, as well as the relationship across levels. Two models were developed in order to achieve this; one that reflects the relationship within lower level units, and a second that models how the relationship within lower level units varies between units, thereby correcting for the violations of aggregating or disaggregating data. This study therefore examined asymmetric effects of autocorrelation error of level 1 in hierarchical modeling.

*** Corresponding Author: Tel: +234(0)8053049890; Email: oloyede.i@unilorin.edu.ng**

2.0 Model Designs

2.1 Equations Underlying Hierarchical Linear Models

In two-level hierarchical models, separate level-1 models (e.g., respondents' status) are developed for each level-2 unit. They take the form of simple regressions developed for each individual *i*:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (1)$$

where:
Y_{ij}= outcome variable for *i*th level-1 unit nested within the *j*th level-2 unit,
X_{ij}= level-1 explanatory variable,
β_{0j}= intercept for the *j*th level-2 unit,
β_{1j}= coefficient associated with *j*th level-2 unit, and
e_{ij}= error associated with *i*th level-1 unit nested within *j*th level-2 unit.

In level-1, the assumption $(e_{ij}) \sim (0, \sigma^2)$ holds. In reverse, the present study violated this assumption by assuming: $\mathrm{E}(e_{ij}) = 0$, $var(e_{ij}) = \rho e_{ij} + e_{ij_{-1}}$ in level 1.

In level-2 models, the level-1 parameters ($\beta_{0j}$and β_{1j}) were used as dependent variables and were related to each of the level-2 explanatory variables. Level-2 models described the variability across multiple groups and were referred to as between-unit models [12]. The study considered the case of a single level-2 predictor using equations 2 and 3.

$$\beta_{oj} = \gamma_{oo} + \gamma_{oj}Z_j + e_{oj} \quad (2)$$

$$\beta_{1j} = \gamma_{1o} + \gamma_{1j}Z_j + e_{1j} \quad (3)$$

where:
β_{oj}= intercept for the *j*th level-2 unit;
β_{1j}= slope for the *j*th level-2 unit;
Z_j= value on the level-2 predictor;
γ_{oo}= overall mean intercept adjusted for *Z;*
γ_{10}= overall mean intercept adjusted for *Z;*
γ_{01}= regression coefficient associated with *Z* relative to level-1 intercept;
γ_{11}= regression coefficient associated with *Z* relative to level-1 slope;
e_{0j}= random effects of the *j*th level-2 unit adjusted for *Z* on the intercept;
e_{1j}= random effects of the *j*th level-2 unit adjusted for *Z* on the slope.

Thus level-2 models have two errors terms (e_{0j} and e_{1j}) that made hierarchical linear model different from regression equation. The study introduced a combined model for the classification of variables and coefficients in terms of the level of hierarchy:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + e_{1i}X_{ij} + e_{0i} + \tau_{ij} \quad (4)$$

The model described in equation (4) is known as mixed model which incorporates the level-1 and level-2 predictors (X_{ij} and Z_j) as deterministic and fixed, a cross-level term (Z_jX_{ij}), but the composite error ($e_{1i}X_{ij} + e_{0i} + \tau_{ij}$) made it stochastic and random.

Let

$$\theta_i = \begin{bmatrix}\alpha_i \\ \beta_i\end{bmatrix} \sim N(\theta_o, \Sigma), \; i = 0, 1, \dots k \quad (5)$$

k is the number of parameters,

where $\theta_o = \begin{bmatrix}\alpha_o \\ \beta_o\end{bmatrix}$ and Σ is a 2×2 positive definite symmetric covariance matrix .

Priors:

$$\sigma^2|a,b \sim IG(a,b)$$

$$\theta_0|\eta,C \sim N(\eta,C)$$

$$\Sigma^{-1}|\rho,R \sim W([\rho R]^{-1},\rho)$$

where W denotes Wishart distribution.

Likelihood: $\frac{1}{[\sigma^2/(1-\rho)]^{N/2}} exp\left(-\frac{1}{2\sigma^2/(1-\rho)}\sum_{i=1}^{n}(y_i - X_i\theta_i)'(y_i - X_i\theta_i)\right)$ (6)

Joint posterior density:

$p(\theta_i,\theta_0,\Sigma^{-1},\sigma^2|y) \propto [\prod_{i=1}^{n} p(y_i|x_i,\theta_i,\sigma^2)\, p(\theta_i|\theta_0,\Sigma^{-1})]p(\theta_0|\eta,C)p(\sigma^2|a,b)p(\Sigma^{-1}|\rho,R)$ (7)

The observations

$$y_i = \begin{bmatrix} y_i 1 \\ y_i 2 \\ . \\ . \\ . \\ y_i n \end{bmatrix} \quad \begin{bmatrix} 1 & x_i 1 \\ 1 & x_i 2 \\ 1 & . \\ 1 & . \\ 1 & x_i n \end{bmatrix} \quad x_i = \begin{bmatrix} 1 & X_i 1 \\ 1 & X_i 2 \\ . & . \\ . & . \\ . & . \\ 1 & X_i n \end{bmatrix} \quad (8)$$

Complete posterior conditional for θ_i

$p(\theta_i|\theta_0,\Sigma^{-1},\sigma^2,y) \propto p(y_i|X_i,\theta_i,\sigma^2)p(\theta_i|\theta_0,\Sigma)$ (9)

Then we have $\theta_i|\theta_0,\Sigma^{-1},\sigma^2.y \sim N\left((X_i'X_i/\sigma^2 + \Sigma^{-1})^{-1}(X_i'y_i/\sigma^2 + \Sigma^{-1}\theta_0), (X_i'X_i/\sigma^2 + \Sigma^{-1})\right)$ (10)

Complete posterior conditional for θ_0

$p(\theta_0|\theta_i,\Sigma^{-1},\sigma^2,y) \propto \prod_{i=1}^{n} p(\theta_i|\theta_0,\Sigma^{-1})\, p(\theta_0|\eta,\mathrm{C})$ (11)

Level 2: $\theta_i|\theta_0,\Sigma^{-1}$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ . \\ . \\ . \\ \theta_n \end{bmatrix} = \begin{bmatrix} 1_1 \\ 1_2 \\ . \\ . \\ . \\ 1_n \end{bmatrix} \theta_0 + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix} \quad (12)$$

Thus we have $\theta_0|\theta_i,\Sigma^{-1},\sigma^2,y \sim N\left(\left(\check{\mathrm{I}}'(I_n \otimes \Sigma^{-1})\check{I} + C^{-1}\right)^{-1} \left(\acute{\mathrm{I}}'(I_n \otimes \Sigma^{-1})\check{\theta} + C^{-1}\eta\right),\right)$ (13)

Complete posterior conditional for σ^2

$p(\sigma^2|\theta_i,\theta_0,\Sigma^{-1},y) \propto \prod_{i=1}^{n} p(y_i|X_i,\theta_i,\sigma^2)\, p(\sigma^2|a,\mathrm{b})$ (14)

$\propto \frac{1}{[\sigma^2]^{N/2}} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i\theta_i)'(y_i - X_i\theta_i)\right) \times \frac{1}{[\sigma^2]^{a+1}} exp\left(-\frac{1}{b\sigma^2}\right)$ (15)

$\propto \frac{1}{[\sigma^2]^{N/2+a+1}} exp\left(-\frac{1}{\sigma^2}[1/2\sum_{i=1}^{n}(y_i - X_i\theta_i)'(y_i - X_i\theta_i) + b^{-1}]^{-1}\right)$ (16)

Thus we have $\sigma^2|\theta_i,\theta_0,\Sigma^{-1},y \sim IG\left(\frac{N}{2} + a, [1/2\sum_{i=1}^{n}(y_i - X_i\theta_i)'(y_i - X_i\theta_i) + b^{-1}]^{-1}\right)$ (17)

where $N = p\,X\,n$

Complete posterior conditional for Σ^{-1}

$$p(\Sigma^{-1}|\theta_i,\theta_0,\sigma^2,y) \propto [\prod_{i=1}^{n} p(\theta_i|\theta_0,\Sigma^{-1})]p(\Sigma^{-1}|\rho,\mathrm{R}) \quad (18)$$

$$\propto |\Sigma^{-1}|^{n/2} exp(-1/2\sum_{i=1}^{n}(\theta_i-\theta_0)'\Sigma^{-1}(\theta_i-\theta_0)) \times |\Sigma^{-1}|^{(\rho-3)/2} exp(-[1/2]tr[\rho R\Sigma^{-1}]) \quad (19)$$

$$= |\Sigma^{-1}|^{(n+\rho-3)/2} exp\left(-1/2([\sum_{i=1}^{n} tr(\theta_i-\theta_0)(\theta_i-\theta_0)'\Sigma^{-1}] + tr[\rho R\Sigma^{-1}])\right) \quad (20)$$

$$= |\Sigma^{-1}|^{(n+\rho-3)/2} exp\left(-1/2tr([\sum_{i=1}^{n}(\theta_i-\theta_0)(\theta_i-\theta_0)' + \rho R]\Sigma^{-1})\right) \quad (21)$$

Thus we have $\Sigma^{-1}|\theta_i,\theta_0,\sigma^2,y \sim \mathbb{W}([\sum_{i=1}^{n}(\theta_i-\theta_0)(\theta_i-\theta_0)' + \rho R]^{-1}, n+\rho)$ (22)

2.2 Bayesian Hierarchical Models

The Bayesian hierarchical modeling of autocorrelated error was presented. The pooled regression represents level one while individual regression denotes sub-group and both level have shared variances that ordinary least squares estimator cannot capture. Both negative and positive autocorrelated errors were considered. Parameters were obtained through the posterior point estimate of Gibbs sampler simulation. Bias and mean square error were computed to measure consistency and efficiency respectively. The levels of convergence of the chains were monitored [5] and graphic analysis was carried out using coda package in R. Multivariate normal and inverse gamma distributions were chosen as priors for parameter estimates $\beta's$ and σ^2 respectively.

3. Results and Discussion

3.1 Pooled Regression Absolute Bias Hierarchical Modeling with Autocorrelated Error

The study observed that when ρ is -0.99 and 0.99, the effect of the negative autocorrelated error on the model parameter of the pooled estimates is less than that of positive autocorrelated error for all the parameters estimates (Table 1). Considering ρ at -0.77 and 0.77, the inverse effects as compared to higher autocorrelated error of 0.99 was observed. The negative autocorrelated error had higher effect on the model parameters estimates compared with positive autocorrelated error for all the parameters as presented in Figs. 1 and 2.

Table 1: Absolute Bias of Bayesian Hierarchical Modeling with Autocorrelated Error

		$firm1$	$firm2$	$firm3$	$firm4$	$firm5$	$firm6$	$pooled$	σ^2
-0.99	β_0	0.0024	0.0018	0.0036	0.0193	0.0403	0.0811	0.0403	22.3807
	β_1	0.0115	0.0036	0.0004	0.0005	0.0078	0.0049	0.0078	
	β_2	0.0093	0.0318	0.0217	0.0091	0.0019	0.0005	0.0019	
	β_3	0.0153	0.0109	0.006	0.001	0.0033	0.0136	0.0033	
0.99	β_0	0.0025	0.0018	0.0039	0.0214	0.0438	0.0831	0.0438	23.49
	β_1	0.0121	0.0035	0.0004	0.0006	0.0085	0.0050	0.0085	
	β_2	0.0098	0.0311	0.0234	0.0102	0.0021	0.0005	0.0021	
	β_3	0.0160	0.0106	0.0064	0.0011	0.0036	0.0139	0.0036	
-0.77	β_0	0.0021	0.0015	0.0035	0.0182	0.0497	0.0773	0.0498	20.1284
	β_1	0.0103	0.003	0.0004	0.0005	0.0096	0.0047	0.0097	
	β_2	0.0083	0.026	0.0213	0.0087	0.0024	0.0004	0.0024	
	β_3	0.0136	0.0089	0.0058	0.0009	0.0040	0.0129	0.0040	
0.77	β_0	0.0018	0.0018	0.0035	0.0191	0.0429	0.0764	0.0429	18.6336
	β_1	0.0087	0.0035	0.0003	0.0005	0.0083	0.0046	0.0083	
	β_2	0.0070	0.0311	0.0210	0.0091	0.0021	0.0004	0.0021	
	β_3	0.0115	0.0106	0.0058	0.0010	0.0035	0.0128	0.0035	

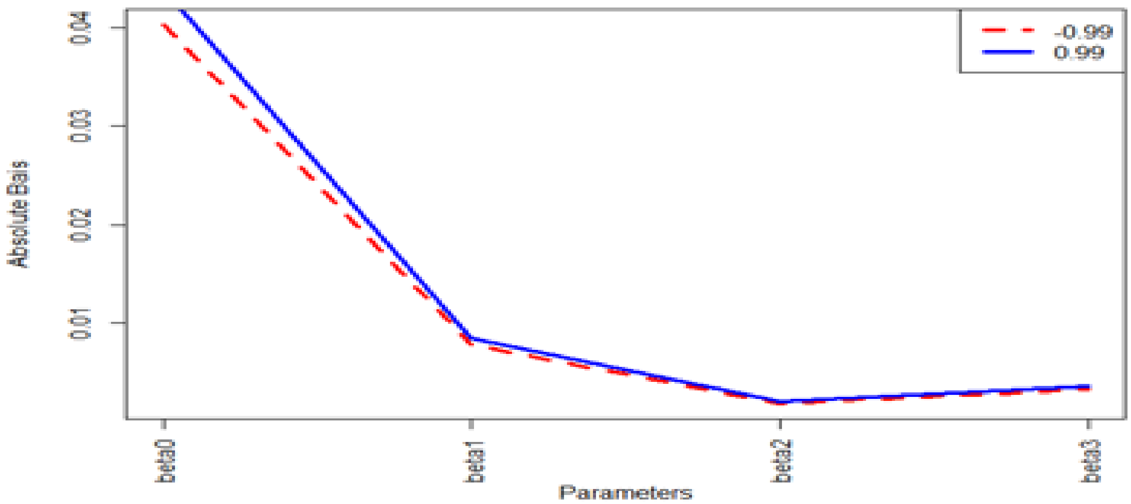

Fig. 1: Bayesian Hierarchical Model with Asymmetric Autocorrelated error at -0.99 and 0.99

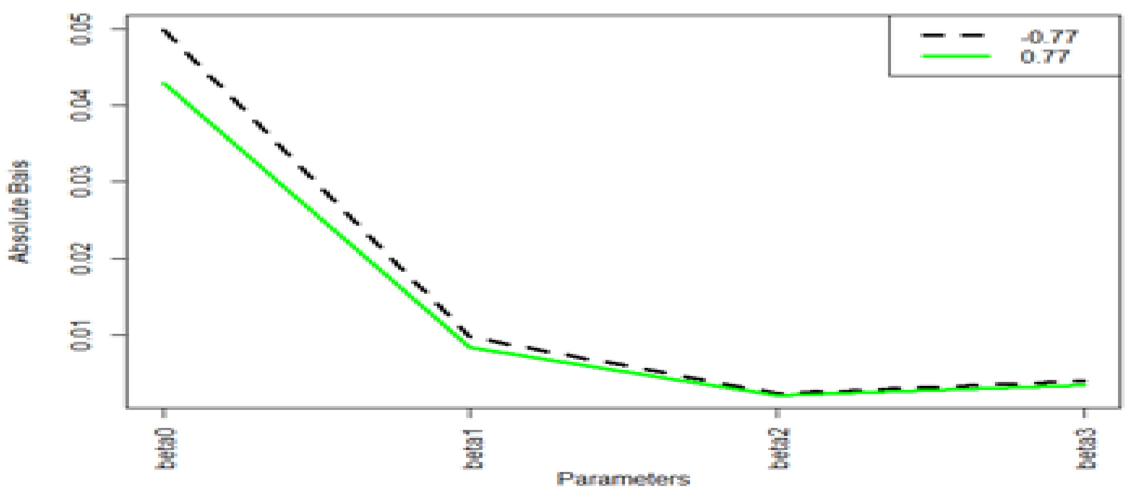

Fig. 2: Bayesian Hierarchical Model with Asymmetric Autocorrelated error at -0.77 and 0.77

3.2 Individual regression Absolute Bias Hierarchical Modeling with Autocorrelated Error

The Mean Squares Error of Bayesian Hierarchical Model with autocorrelated error is presented in Table 2. Considering -0.99 and 0.99 autocorrelated errors for the individual regressions, firm1 firm3, firm4, firm5 and firm6 showed that positive autocorrelated error had higher effects compared with negative autocorrelated error. However, the reverse is the case when considering firm2 in which case the negative autocorrelated error had higher effect for all the model parameters estimates. When considering -0.77 and 0.77 autocorrelated errors for the individual regressions; firm1 firm3, firm4, firm5 and firm6 showed that negative autocorrelated error had higher effects compared with positive autocorrelated error, whereas the reverse is the case when considering firm2 where the positive autocorrelated error had higher effect for all the model parameters estimates.

Table 2: Mean Squares Error of Bayesian Hierarchical Modeling with Autocorrelated Error

ρ		$firm1$	$firm2$	$firm3$	$firm4$	$firm5$	$firm6$	$pooled$
-0.99	β_0	2.3313	2.4395	1.9204	2.5149	1.6294	2.2538	1.6294
	β_1	0.2943	0.2713	0.2331	0.2169	0.1670	0.2225	0.1674
	β_2	0.3539	0.2613	0.229	0.2349	0.1661	0.2935	0.1661
	β_3	0.4669	0.6594	0.3957	0.3275	0.2955	0.3324	0.2955
0.99	β_0	2.5635	2.3302	2.2326	3.1064	1.9252	2.3688	1.9252
	β_1	0.3236	0.2592	0.2709	0.2679	0.1973	0.2338	0.1973
	β_2	0.3892	0.2496	0.2662	0.2902	0.1962	0.3085	0.1962
	β_3	0.5134	0.6299	0.46	0.4045	0.3491	0.3493	0.3491
-0.77	β_0	1.8567	1.6217	1.8473	2.2540	2.4809	2.0483	2.4809
	β_1	0.2344	0.1804	0.2242	0.1944	0.2543	0.2022	0.2543
	β_2	0.2819	0.1737	0.2202	0.2106	0.2529	0.2667	0.2529
	β_3	0.3719	0.4383	0.3806	0.2935	0.4499	0.3021	0.4499
0.77	β_0	1.3210	2.3246	1.8009	2.4835	1.8424	2.0013	1.8424
	β_1	0.1668	0.2585	0.2186	0.2142	0.1888	0.1975	0.1888
	β_2	0.2005	0.2490	0.2147	0.2320	0.1878	0.2606	0.1878
	β_3	0.2646	0.6283	0.3711	0.3234	0.3341	0.2951	0.3341

3.3 Pooled Regression Mean Squares Error Hierarchical Modeling with Autocorrelated Error

From the mean squares error of Bayesian Hierarchical modeling with autocorrelated error, at ρ = -0.99 and ρ = 0.99, the effects of the negative autocorrelated error on the model parameters estimates were lower compared with positive autocorrelated error for all the parameters estimates. Considering ρ =-0.77 and ρ = 0.77, it was observed that the negative autocorrelated error had higher effects on the model parameters estimates compared with positive autocorrelation for all the parameters. This is presented in Figs. 3 and 4.

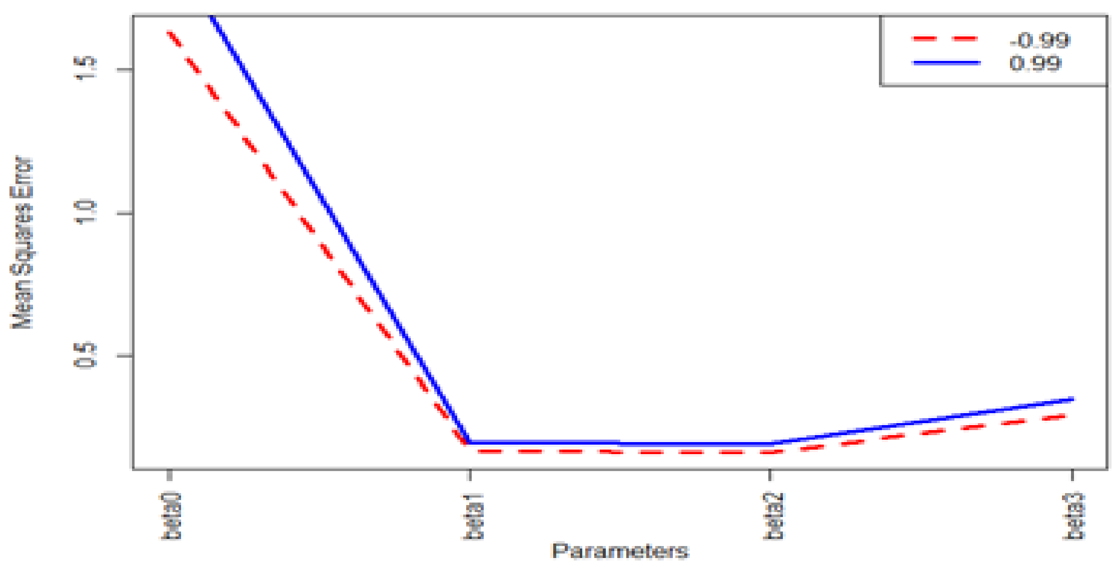

Fig. 3: Mean Squares Error of Bayesian Hierarchical Model with Asymmetric Autocorrelated error at -0.99 and 0.99

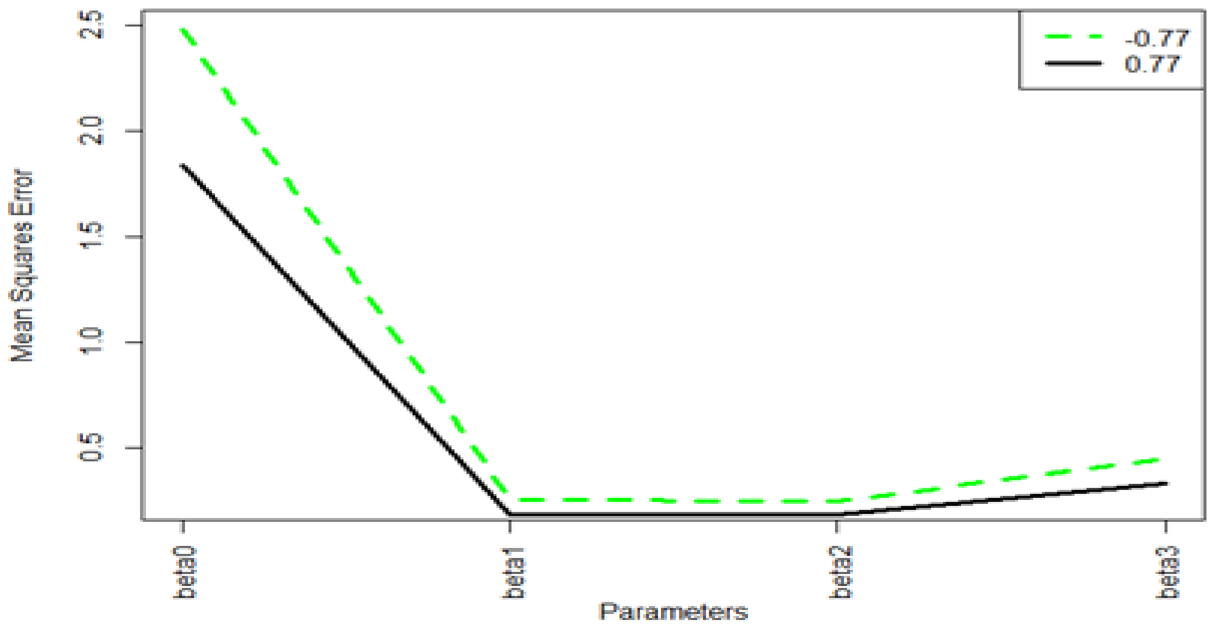

Fig. 4: Mean Squares Error of Bayesian Hierarchical Model with Asymmetric Autocorrelated error at -0.77 and 0.77

3.4 Individual Regression Absolute Bias Hierarchical Modeling with Autocorrelated Error

Considering -0.99 and 0.99 autocorrelated errors for the individual regressions, firm1 firm3, firm4, firm5 and firm6 showed that positive autocorrelated error had higher effects than negative autocorrelated error. The reverse is the case when considering firm2 where the negative autocorrelated error had higher effect for all the model parameters estimates compared with its negative autocorrelated error. When considering -0.77 and 0.77 autocorrelated errors for the individual regressions, firm1 firm3, firm4, firm5 and firm6 showed that negative autocorrelated error had higher effects than positive autocorrelated error. However, the reverse is the case when considering firm2 in which the positive autocorrelated error had higher effect for all the model parameters estimates compared with its negative autocorrelated error.

4.0 Conclusion

This paper presented a simple way of modeling and estimating autocorrelated error Bayesian hierarchical model with autocorrelated error using Gibbs sampler. The study considered asymmetric effects of autocorrelated error observed at higher correlation under absolute bias criterion. Findings revealed that positive autocorrelated error had higher effects on the model parameters estimates compared with negative autocorrelated error. At the lower autocorrelated error, the study established that negative autocorrelated error had higher effects on the model parameters estimates compared with positive autocorrelated error. The study established the same scenario for both absolute bias and mean squares errors criteria. The approach employed in this research can be applied to further studies in the area of linear mixed model, multi-level modeling and other econometrics models.

References

[1] Heather, W., Andrea, F., Christine, J.M. and Meredith, R. (2012). An introduction to Hierarchical Linear Modeling. Tutorials in Quantitative Methods for Psychology, Vol. 8, No.1, pp. 52-69.

[2] Smith, A.F.M. (1973). A General Bayesian Linear Model. Journal of Royal Statistical Society, Series B (Methodological), Vol. 35, No. 1, pp. 67-75.

[3] Dempster, A.P., Rubin, D.B. and Tsutakawa,, R.K. (1981). Estimation in Covariance components Models. Journal of the American Statistical Association, Vol. 76, No. 374, pp. 341-353.

[4] Raudenbush, S.W. and Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd edn., Newbury Park, California: Sage.

[5] Snijders, T.A.B. and Bosker, R.J. (2012). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, 2nd edn., London, Sage Publishers.

[6] Scott, S.L., Blocker, A.W., Bonassi, F.V., Chipman, H.A., George, E.I. and McCulloch, R.E. (2013). Bayes and Big Data: The Consensus Monte Carlo Algorithm http://research.google.com/pubs/pub41849.html (as at 12/12/2015).

[7] Perry, O.O. (2015). Fast Moment-Based Estimation for Hierarchical Models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 3, pp. 1–40.

[8] Bottou, L. (2012). Stochastic Gradient Descent Tricks, In: Montavon, G, Orr, G.B. and Muller, K.R. (eds), Numeral Networks: Tricks of the Trade, 2nd edn., pp. 421-436. Springer.

[9] Gebregziabher, M., Egede, L., Gilbert. G.E., Hunt, K. Knitter, P.J. and Mauldin, P. (2012). Fitting Parametric Random Effects Models in very Large Data Sets with Application to VHA National Data. BMC Medical Research Methodology, Vol. 12, No.163, pp. 1-14.

[10] Khanna, R., Zhang, L., Agarwal, D. and Chen, B.C. (2013). Parallel Matrix Factorization for Binary Response in Big Data, IEEE International Conference, pp. 430.

[11] Bates, D., Maechler, M., Bolker, B. and Walker S. (2013). lme4: Linear Mixed-Effects Models using Eigen and S4. R Package Version 1, pp. 1-7.

[12] Gill, J. (2003). Hierarchical Linear Models, In: Kempf-Leonard, K. (ed), Encyclopedia of Social Measurement. New York: Academic Press.