ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 7 No. 1March 2021*

Comparative Analysis on Estimation of Linear Regression Models in the Presence of Autocorrelated error.

***Adewole, A. I[1]., Bodunwa, O. K.[2], Egunjobi K.A[3]**

[1]Department of Mathematics,
Tai Solarin University of Education
Ijagun Ogun State Nigeria.

[2]Department of Statistics,
Federal University of Technology Akure,
Ondo State Nigeria.

[1]Department of Physics
Tai Solarin University of Education
Ijagun Ogun State Nigeria.

Email: hayorhade2005@gmail.com

Abstract
The research explored practical framework of comparative analysis on estimation of linear regression models in the presence of autocorrelated error. Insight to the investigation of various methods of estimation in the context of autocorrelated error is for proper examination of fitness of a model on a data set that the error is serially correlated. Ordinary Least squares; Quantile Regression and Bayesian Regression methods were employed in estimation of linear regression model with first order autocorrelated error. Simulated data set and empirical case were considered to illustrate how the designed method actually performed in realistic settings that were previously considered intractable from a Bayesian perspective. Bayesian approach gives a robust and less biased estimates when dealing with non-normality and non-constant variance assumption.

Keywords: Regression Analysis, Autocorrelated error, Methods of Estimation,Model fitness.

Introduction
Regression analysis seeks to find the relationship between one or more independent variables and a dependent variable, certain widely used methods of regression have favorable properties if their underlying assumptions are true, but can give wrong inference and misleading decisions

**Author for Correspondence*

if those assumptions are not true. Autocorrelation of the error terms leads to inefficient but sill biased estimates of the coefficients (Rao and Griliches 1969). Gujarati (2003), identified several ways in which autocorrelation may be induced; which include inertia, specification bias, excluded variables,manipulation of data and non-stationarity. The estimation of coefficients in a simple regression model with auto correlated errors is an important problem and has received a great deal of attention in the literature, William *et al.,* (2012), Garba*et al*., (2013), Adenomon*et al.,* (2015), Adewole and Fasoranbaku (2020). Theoretical results established that ordinary least squares regression models could be inadequateif the probability distribution of the observed response variables do not follow a symmetric distribution, Min and Kim (2004). In contrast to mean regression models, according to Cai and Xiao (2012), quantile regression models belong to a robust model family which can give an overall assessment of the covariate effect at different quantiles of the outcome. Several proposals in the literature attempt to obtain a more robust form of quantile regression, among many others, Giloni (2006) and Neykov (2012). MCMC sampling enables exact inference for any sample size without resorting to asymptotic calculations. Bayesian methods do not need to be tested for their sampling properties (Gelman 2008) instead they are concerned with the facts that the correct likelihood and prior are being employed so that Monte Carlo Markov chain (MCMC) methods converge to the implied posterior distribution. However, performing a regression does not automatically give a reliable relationship between the variables but choosing an estimator that gives a precise and reliable parameter estimates. This study extensively compared various approaches of estimating linear regression models with violation of underlying ordinary least square's regression assumptionand investigate the autocorrelation structure across the entire distribution using Monte Carlo simulations and empirical study.

Material and Methods

Consider the following simple linear mean regression model,

$$Y = X^T\beta + \varepsilon, \qquad (1)$$

E(Y/X=x) = $X^T\beta$. The β explains the changes in the mean of the response variable Y due to changes in X. In ordinary least square regression β is estimated by solving

$$\hat{\beta} = \underset{(\beta \epsilon R^p)}{argmin} \sum_{t=1}^{n} (y_t - x_t{}^T)^2. \qquad (2)$$

The quantile regression model is defined as

$$y_t = x_t^T\beta(\tau) + \varepsilon_t(\tau), \qquad (3)$$

where$\beta(\tau)$ is a vector of unknown parameters of interest, estimation of $\beta(\tau)$ proceeds by minimizing

$$\hat{\beta}_\tau = argmin_{\beta \in R^k} \sum_{t=1}^{n} \rho_\tau(y_t - x_t'\beta_\tau), \qquad (4)$$

where the loss function ρ is simplified as

$$\rho_\tau(u) = [\tau - I(u < 0)]u, \qquad (5)$$

and the model's residuals are formulated as an indicator function with

$$\text{I}\,\{u\} = \begin{Bmatrix} 1 & for\ u < 0 \\ 0 & otherwise \end{Bmatrix}. \qquad (6)$$

The quantile $\hat{\beta}_\tau$ is the τ^{th} quantile.

The loss function is not differentiable; solutions to the minimization cannot be derived explicitly. Linear programming method in 'R' was designed to obtain quantile regression estimates for $\widehat{\beta}_\tau$.

Estimation of simple linear regression model using quantile regression method in the presence of autocorrelated error was achieved by considering the regression model as follows;

Let $y_t = x_t^T \beta_\tau + \varepsilon_t$
where, $\boldsymbol{X_t}' = (\boldsymbol{1}, \boldsymbol{x_{t2}}, \ldots, \boldsymbol{x_{tk,}})', \boldsymbol{\beta(\tau)} = \left(\boldsymbol{\beta_{1,1}(\tau)}, \ldots, \boldsymbol{\beta_{1,q_1}(\tau)}, \ldots, \boldsymbol{\beta_{r,q_r}(\tau)}\right)''$**,**and
$\varepsilon_t = \sum_{j=1}^{p} \rho_j \varepsilon_{t-j} + u_t$ (7)
For equation 1 the ε_t follows the process;
$\varepsilon_t = \rho_1 \varepsilon_{t-1} + u_t$ (8)
where$u_t \sim iiid(N0, \sigma^2)$.
$E(u_t) = 0$, $(u_t) = \sigma^2 t$, $Cov(u_t, u_s) \neq 0$,for $t = 1, \ldots, n, X_{it}$ is the q dimensional predictors, $u_t$follows independently identical normal distribution with mean 0 and variance.ρ_j, j =1,…,p is the autocorrelation coefficient of order p which determines the dependency of the error term ε_t
In contrast to the standard linear regression model, the error terms are correlated. Estimating the parameters in the model in equation (3) is to transform it as follows;
$y_t^* = X'^*_t \beta(\tau) + u_t$ (9)
wherey_t^* and X'^*_t represent the following transformed variable:-
$y_t^* = y_t - \rho y_{t-1}$ (10)
for$t = 2,3, \ldots, n$
$X'^*_t = X_t - \rho X_{t-1}$ (11)
for$t = 2,3, \ldots, n$
Where ρ is the autocorrelation coefficient of order1, X'^*_{it} and y_t^* are the $(t-p) \times q$ dimensional matrix and $t-p$ dimensional vector respectively which depend on autocorrelation coefficient ρ.Quantiles estimates were obtained for selected quantiles covering the lower, median and upper quantiles from equation (9) using simplex method of optimization in R.
In Bayesian estimation of parameters of regression model in (9) begins by erecting a likelihood and prior for the parameters involved, the model inference theoretically requires the initial values $(y_0, \ldots, y_{t-p})$ and $(x_1, \ldots, x_{t-p})$, noting that error term u_t are independent normal, the assumption that errors are independent over all individual and time periods implies that the transformed model simply reduces to the standard linear regression framework, the density was expressed as
$f(y_t^* | X'^*_t, \alpha, \beta, \rho) = \frac{1}{(2\pi\sigma^2)^{(t-p)/2}} exp\left[\frac{-(y_t^* - X'^*_t)^T (y_t^* - X'^*_t)}{2\sigma^2}\right]$ (12)
With $\rho = (\rho_1, \rho_2, \ldots, \rho_n)'$ and X^*_t and Y^*_t are the (t-p) X q dimensional matrix and (t-p) dimensional vector respectively,
In view of the above, conjugate prior for $\beta, and\ \sigma^2$ was chosen separately.
Prior of $\beta \sim N(\beta_p, B_p)$ (13)
Where $\beta_p$and $B_p$are the prior mean and the covariance respectively.
all the posterior moment of β exist, the posterior of β still follows normal distribution i. e
$\beta | y, \sigma^2 \sim N(\beta_p, B_p)$ (14)
For the prior on σ^2, inverse gamma distribution $IG(a, b)$ invGamma(shape = $n_{0,}$ scale = $s_{0,}$) was chosen with density
$f(x | n_0, s_0) = \frac{s_0{}^{n_0}}{\Gamma(n_0)} x^{-n_0-1} exp\left(-\frac{s_0}{x}\right)$ (15)
with parameters $a = \frac{n_0}{2}$ and $b = \frac{s_o}{2}$
the posterior distribution for σ follows an inverse Gamma distribution
$\sigma^2 | y, \beta, \sim IG(\frac{n^*}{2} \frac{s^*}{2})$ (16)
With $n^* = n_o + 3n$ and $s^* = s_o + 2\sum_{i=1}^{n} \mu_i v_i + \sum_{i=1}^{n} \left(\frac{(Y_t - X_t' \beta - \mu v_i)^2}{\sigma^2 v_i}\right)$
The full conditional posterior distribution of $\beta\ and\ \sigma^2$ is not of tractable form, so therefore Gibb's sampling method was employed to estimate the posterior.
The Gibb's samplers sampled from
$\sigma^2 |\ \beta, y$

$\beta|\ \sigma^2, y$
Which converges to the joint conditional posterior distribution combining the likelihood function of the data, $L(y_t|\beta, \sigma^2,)$ given by

$$p(\beta, \sigma^2, |y_t) \propto L(y_t|\beta, \sigma^2,).p(\sigma^2|\beta, y_t).\ \ p(\beta|\sigma^2, y_t) \quad (18)$$

This yields the following full conditional posteriors

$$\begin{bmatrix} \beta|y_t, \sigma^2 \sim N(\beta_p, B_p) \\ \sigma^2|y_t, \beta, \sim IG\ \left(\frac{n^*}{2}, \frac{s^*}{2}\right) \\ \end{bmatrix} \quad (19)$$

Based on the conditional posterior densities of βandσ^2 which are not analytically tractable in equation (19), we turn to MCMC computation method using Gibb's sampling to draw samples from the posterior.

The Gibb's sampler is an iterative Monte Carlo scheme designed to extract conditional posterior distribution from intractable joint distribution, from the Gibb' sample, discard an initial number of generates as being unrepresentative of the posterior (the burn in) and average the remaining sample to produce an estimate of the posterior mean in the presence of auto correlated error, which are then used for posterior inferences. The numerical standard error (NSE) derived through the use of a central limit theorem under the weak conditions necessary for Gibb's sampler to converge to a sequence of draws from $P(\theta\ |y)$ was used to evaluate the accuracy of the Monte Carlo approximation of both the posterior means and standard deviation in the simulation study.

To illustrate the efficiency and sampling properties of the MCMC estimator, simulation study was conducted as follows.

Simulation study
In order to estimate β , considering the model

$$y_t^* = X'^*_t\beta(\tau) + u_t \quad (20)$$

Given ρ, generate random numbers of u_t for $t = 1, \ldots, n$ based on the assumption of

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (21)$$

Where$\varepsilon_t \sim N(0, \sigma^2)$, $\sigma^2_{u_t} = 1$ and ρ is an AR(1) parameter autocorrelation coefficient ,all the explanatory variableswere generated as AR(1) variates with first order autoregression such that

$X_t \sim N\left(0, \frac{1}{1-\rho^2}\right), X_t = \rho X_{t-1} + v_tandv_t \sim N(0,1)$.Given β,X_t and u_t for $t = 1, \ldots, n$, a data set for y_t^* from the model in equation (20) was obtained, where β_i, $i = 1, \ldots, 4$ assumed 0.3,0.8,1.02,1.3,0.4 and 1.1 respectively, the research considered 0.2, 0.6 and 0.9 as the weak, medium and autocorrelation coefficient respectively. The simulation study considered the small (n = 25), medium (n = 100) and large sample (n= 500) size to evaluate the effect of autocorrelated error in estimation of Bayesian regression parameters, also using $(y_t^*,\ X'^*_t)$ in equation 10 and 11, $t = 1,2, \ldots, n$,posterior estimates for β were obtained from the conditional posterior densities in equation (19). Gibbs sampling algorithm was thus incorporated to draw the MCMC iterates for each parameter. Gibbs sampler ran for 200,000 replications and discard the first 20,000 as burn in period. MCMC sampling was carried out in R (R development Core Team 2016), the Monte Carlo simulation was implemented by taking random draws from the posterior distribution of

β and then averaging the appropriate functions of these draws. Comparison of the model fit of Bayesian regression estimates with different sample size in the presence of autocorrelated error was done via root mean square error and bias.

Empirical study

To illustrate the comparative analysis of the methods involved in estimating simple linear regression model in the presence of autocorrelated error, the data set from Nigeria CBN bulletin which comprised of Nigeria RGDP growth, as the response variable, the money supply, foreign direct, unemployment rate and exchange rate from the period of 1985-2019 as the explanatory variables was considered.

Let the model

$$y_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_{4t} + u_t \quad (22)$$

where

y_t = RGDP growth, x_{1t} = money supply, x_{2t} = foreign direct, x_{3t} = unemployment rate and x_{4t} = exchange rate .

Given β, X_t and u_t for $t = 1, \dots, n$, from the empirical data set above using the model in equation (3.62), where u_t was generated based on the assumption of $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ N(0, σ^2), $\sigma^2_{u_t}$=1 and ρ is an AR(1) parameter autocorrelation coefficient which was determined through the estimation procedures using Cochrane Orcutt (1949) approach of estimating autocorrelation error in regression models; estimate of ρ obtained was 0.70, which lies between 0 and 1 as it is expected for a ρ when autocorrelation is present.The Gibb's sampling algorithm discussed above was used to obtain the posterior estimates for β_i

Quantile Regression Model in the presence of Autocorrelated Error.

In most cases time series data inherits autocorrelation, this property was verified in quantile regression models with Ljung- Box test, the test was applied to residual from the fitted parametric quantile regression model in equation (20)

The procedures include testing under the null hypothesis of no autocorrelation in the residuals

$H_0 : \hat{\rho}_1(\hat{u}_t) = 0$

The box- pierce test statistic is

$$BP(K) = T \sum_{k=1}^{k} \rho_k \ (\hat{\varepsilon}_t) \quad (23)$$

With T the number of observations, $\rho_k(\hat{\varepsilon}_t)$ is the autocorrelation coefficient of order k of the estimated residuals and k the maximum number of lags. Under the assumptions that x_t is an independent, identically distributed sequence with certain moment conditions. $BP(K)$ is asymptotically a χ^2 random variable with $k - p - q$ degree of freedom.

The Ljung Box test statistic for the model is 165.0825 which has a $p-value$of $1.8e^{-14}$.whichconcluded that the economic data has significant autocorrelation. After fitting a parametric model to the numerical data, frequentist estimate regression coefficient $\beta_0^*(\tau), \beta_1^*(\tau), \dots, \beta_4^*(\tau)$ new model was obtained using a resampling procedure with the auto correlated residuals

Mean square error was used as a criterion of validation to measure the relative effectiveness of ordinary least square, quantile regression and Bayesian method of estimation in the presence of autocorrelated error.

Results and Discussion

Table 1 below present the posterior means, the standard deviation of the posterior means of the estimated parameters in brackets, the numerical standard error of the posterior means in curly brackets and the 95% credible interval, formed by the 2.5th and 97.5th sample of the MCMC iterates in square brackets are shown in bold.

Table 1:**Summary statistics based on the simulation study using Bayesian Regression in the presence of autocorrelated error.**

	AUTO	PARAMETER ESTIMATE						
N	$\boldsymbol{\rho}$	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_4}$	**RMSE**	**BIAS**
25	0.2	0.332(0.19) {0.000409} **[0.025, 0.956]**	0.993(0.01) {0.000032} **[0.325,3.049]**	1.002(0.48) {0.00107} **[0.402,1.938]**	0.455(0.02) {0.000035} **[0.113,1.373]**	1.1511(0.09) {0.000053} **[0.043,1.275]**	0.0017	0.0000
	0.6	0.3052 (0.047) {0.000148} **[0.0391,1.963]**	0.8056(0.041) {0.000069} **[0.025, 2.108]**	1.037(0.58) {0.00234} **[0.083,2.285]**	0.372(0.01) {0.000031} **[0.00,1.263]**	1.132(0.04) {0.00267) **[0.386,2.053]**	0.0151	0.0029
	0.9	0.329 (0.08) {0.000074} **[0.2561,2.418]**	0.992(0.02) {0.000063} **[0.009, 1.428]**	1.0282(0.6) {0.00247} **[0.109,1,857]**	0.4134(0.06) {0.000177} **[0.042,2.751]**	1.2935(0.1) {0.00074} **[1.109,2.996]**	0.0022	0.0009
100	0.2	0.235 (0.02) {0.000077} **[0.0353,3.064]**	0.763(0.07) {0.000221} **[0.208.3.489]**	1.0994(0.08) {0.000025} **[0.0387,3.002]**	0.2037(0.01) {0.000057} **[0.196,3.642]**	1.0664(0.04) {0.000028} **[0.002,1.563]**	0.0085	0.0014
	0.6	0.334 (0.06) {0.000197} **[0.2864,2.432]**	0.8623(0.043) {0.000071} **[0.122, 1.638]**	1.0037(0.02) {0.000075} **[0.285,2.118]**	0.4372(0.55) {0.000387} **[0.002,0.452]**	1.1648(0.029) {0.000008} **[0.557, 2.642]**	0.0169	0.0027
	0.9	0.284(0.016) {0.000051} **[0.345, 1.807]**	0.7874(0.018) {0.000057} **[0.109,1.385]**	1.0557(0.26) {0.00063} **[-0.074,1.038]**	0.6147(0.01) {0.000180} **[0.004,2.581]**	1.0735(0.09) {0.000133} **[0.329,2.176]**	0.0019	0.0006
500	0.2	0.388 (0.024) {0.000075} **[0.180, 1.350]**	0.915(0.031) {0.000098} **[0.310,2.490]**	1.0371(0.19) {00060} **[0.270, 3.116]**	0.2576(0.09) {0.000292} **[0.109.1.802]**	1.4725(0.53) {0.000284} **[0.307,1.983]**	0.0324	0.0000
	0.6	0.4558(0.028) {0.000088} **[0.0835,3.146]**	0.8316(0.009) {0.000037} **[0.620, 2.686]**	1.030(0.51) {0.00227} **[0.136,2.906]**	0.2245(0.07) {0.000096} **[0.185,4.274]**	0.9914(0.01) {0.00078} [0.664,3.118]	0.0031	0.000
	0.9	0.4511(0.019) {0.000060} **[0.0543,1.547]**	0.868(0.013) {0.000031} **[-0.028,2.427]**	1.0864(0.32) {0.00101} **[0.728,3.485]**	0.4391(0.016) {0.000051} **[0.246,1.394]**	1.1056(0.91) {0.000003} **[1.072,2.676]**	0.0252	0.0000

From results in Table 1, posterior estimates are very close to their true values and their NSE is very minimal which shows a good convergence. The estimates are all contained in the credible intervals.The effect of autocorrelation coefficient and large sample size does not significantly affect the accuracy and efficiency of Bayesian regression estimates from the validation criteria with the output root mean square and bias of the estimates.

TABLE 2: Results of the estimation of the three regression models in the presence of autocorrelated error using the empirical data.

ESTIMATION	AUTO	PARAMETER ESTIMATE					MSE
	$\rho = 0.7$	β_0	β_1	β_2	β_3	β_4	
OLS		0.3737(0.35)	0.6243(0.14)	0.8826(0.10)	0.5184(0.2)	1.003(0.62)	0.053
QUANTILE REGRESSION	0.05	0.4094(0.02)	1.0040(0.57)	0.6146(0.12)	0.7826(0.6)	1.4981(0.15)	0.0028
	0.50	0.3294(0.50)	0.9591(0.62)	1.0837(0.29)	0.5710(0.09)	0.9317(0.03)	0.0060
	0.90	0.5556(0.73)	0.8553(0.9)	1.4912(0.1)	0.4857(0.6)	1.4915(0.5)	0.0031
BAYESIAN ESTIMATION		0.3710(0.019)	1.0837(0.014)	0.9624(0.071)	0.8924(0.018)	1.037(0.05)	0.00024

Quantile regression procedures report smaller MSE in all the quantiles compared to ordinaryleast regression when autocorrelated error were considered.Bayesian estimate results provides exact estimation which duly account for parameter uncertainty which are all consistent with the theoretical results.Comparing the frequentist approach with the Bayesian approach using Table 2, it was revealed that Bayesian approach produced minimal MSE and standard error which implies that the Bayesian approach of estimating regression models in the presence of serially correlated error outperformed the frequentist

Conclusion

The workgivesan insight on methods of simple linear regression model estimation andinvestigates a practical frameworkofestimating regression models in the presence of autocorrelated error. It was observed that when dealing with non-normality and non-constant variance assumption, Bayesian approach gives a robust and less biased estimates. It also revealsthat as sample sizes increases, the BIAS of Bayesian regression models with autocorrelated errorreduces. Compared to the frequentist estimate, the Bayesian methods still perform better even when the error distribution assumption is violated.The Bayesian approach methods of estimation gives robust and less biased estimates, the research justified the results of William et al., (2012) that worked on the estimation of Bayesian estimates of autocorrelation in a single case design that concluded that Bayesian estimation reduces the role of sampling error.
Future research will look into other violation of ordinary least squares assumptions such as multicollinearity and investigate the performance of numerous estimation methods.

References

Adenomon, M.O., Michael, V.A. and Evans O.P. (2015). A Simulation Study on the performance of Classical Var and Sims – Zha Bayesian Var Models in the Model in the presence of Autocorrelated Errors. *Open Journal of Modelling and Simulations*,3,146-158.

Adewole Ayoade and Fasoranbaku OIusoga(2020).Bayesian Estimation of Regression quantiles in the presence of Autocorrelated Errors. *Tanzania Journal of Science*, 46(1), 76-84.

Cai, Z. and Xiao, Z., (2012). Semiparametricquantile regression estimation in dynamic models with partially varying coefficient.*Journal of econometrics, 167,413-425.*

Cochrane, D. Orcutt, G. H. (1949). "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms". *Journal of the American Statistical Association.* ***44*** *(245): 32–61.*

Garba, M.K., Oyejola, B.A and Yahya, W.B (2013).Investigation of Certain Estimators for modelling Panel Data under Violations of some basic Assumptions. *Mathematical Theory and Modelling*, 3,47-53.

Gelman A. (2008), Objections to Bayesian Statistics. Rejoinder Bayesian analysis(3), 467-478.

Giloni S. Sehadri (2006). Information sharing in a supply chain under ARMA demand. *International Journal of Production Economics, 103, 401-449.*

Gujarati D.N (2003), Basic econometric Mc Graw- Hill Fourth edition.

Min, I., and Kim I., (2004). A Monte Carlo comparison of parametric and Non parametric quantile regression. *Applied economics letters, 11, 71-74*.

NeykovN.M.(2012). Robust Joint Modelling of Mean and Dispersion. Statistics and DataAnalysis 56, 1757-1770.

Rao, P. and Griliches, Z.(1969). Small sample properties of several two stages of Regression Methods in the context of Autocorrelated Errors. *Journals of American Statistical Association* 64, 253-272.

William R., Rindskopf M., Larry V., Hedges S. and Krilym J. (2012). Bayesian Estimates of Autocorrelation in Single case Design.*Behavior Research methods, ISSN 1554-3528, 45,813–821.*