

On Seemingly Unrelated Regression and Single Equation Estimators Under Heteroscedastic Error and Non-Gaussian Responses

Razaq B. Afolayan, Alabi W. Banjoko, Mohammed K. Garba and *Waheed B. Yahya

Department of Statistics, University of Ilorin, Ilorin, Nigeria

afolayan.rb@unilorin.edu.ng | {banjokoalabi | garbamk | dr.yah2009}@gmail.com

Received: 20-JAN-2020; Reviewed: 04-MAR-2020; Accepted: 06-JUN-2020

<http://dx.doi.org/10.46792/fuoyejt.v5i2.469>

Abstract- This study investigated the efficiency of Seemingly Unrelated Regression (SUR) estimator of Feasible Generalized Least Square (FGLS) compared to robust MM-BISQ, M-Huber, and Ordinary Least Squares (OLS) estimators when the variances of the error terms are non-constant and the distribution of the response variables is not Gaussian. The finite properties and relative performance of these other estimators to OLS were examined under four forms of heteroscedasticity of the error terms, levels of Contemporaneous Correlation (Cc) with gamma responses. The efficiency of four estimation techniques for the SUR model was examined using the Root Mean Square Error (RMSE) criterion to determine the best estimator(s) under different conditions at various sample sizes. The simulation results revealed that the SUR estimator (FGLS) showed superior performance in the small sample situations when the contemporaneous correlation (ρ) is almost perfect ($\rho=0.95$) with the gamma response model while MM-BISQ was the best under low contemporaneous correlation. The relative efficiencies of MM-BISQ, M-Huber and FGLS estimators over the OLS are respectively 89%, 71%, and 14% in a small sample ($n \leq 30$) and 49%, 32% and 1% in large sample sizes ($n > 30$) under gamma response model. The study concluded that MM-BISQ and M-Huber estimators are the most efficient estimators for modeling systems of simultaneous equations with non-Gaussian responses under either homoscedastic or multiplicative heteroscedastic error terms irrespective of the sample size.

Keywords- Contemporaneous correlation, Feasible Generalized Least Square, Heteroscedasticity, Homoscedasticity, Seemingly unrelated Regression

1 INTRODUCTION

Seemingly Unrelated Regression (SUR) equation is a system of multiple regression equations in which each equation has a single dependent and one or more independent or exogenous variables as in the standard regression model. The set of equations in SUR has no link or relationship with one another except that their disturbances are said to be correlated. Zellner (1962) introduced the SUR estimation procedure for systems of regression equation as against the general estimation method of Ordinary Least Square estimator (OLS). Before the introduction of the SUR estimation procedure, OLS has been the common method for estimating the parameters of the regression models.

The OLS is a single-equation estimation method that does not account for the interactions that may exist among the different regression equations as in SUR. The OLS is unbiased, consistent, and remains the most efficient estimator when all the assumptions of the classical linear regression model are satisfied. The single by single equation estimation of the SUR model without using the information from error terms across equations by OLS is still unbiased and consistent but seizes to be the best linear unbiased estimator (BLUE) (Yahya et al, 2008). Given this, the inferences about the parameters of the model using the statistic from the OLS estimator become invalid. Zellner (1962) viewed that the disturbance terms of these equations are likely to be contemporaneously correlated because of unobservable factors that influence the disturbance term in one equation may affect the disturbance terms in other equations. Estimating these equations separately without accounting for the non-zero covariance structure of the errors leads to inefficient parameter estimates.

However, the joint estimation procedure of SUR using the information from the correlation among the errors of different equations is more efficient than the separate equation estimation procedure of the ordinary least square and the gain in efficiency is achieved if the contemporaneous correlation between the disturbances across equation is very high and other assumptions of the classical regression model are satisfied (See Judge et al, 1988; Zellner, 1962 & 1963; Zellner & Theil, 1962; Yahya et al, 2008). The assumption of normally distributed random variables is rather a myth than reality in real-life applications. For instance, the concentration of elements and their radioactivity in geology and mineral sciences usually follow a lognormal distribution (Malanca et.al, 1996). In medicine, the latent periods of infectious disease do not follow a normal distribution. The distributions of particles, chemicals, and organisms in the environment are often lognormal (Sartwell, 1966; Kondo, 1977; Biondini, 1976). In the field of Social and economic sciences, the distribution of the variables such as age at marriage, family size, farm size, income, etc., deviate from the normal distribution (Preston, 1981).

The violation of each assumption has attracted the attention of many researchers especially in the regression model but the violation of homoscedasticity or normality in the system of regression model has not been investigated. When the responses are non-normal and error variances are non-constant, the consequences of SUR estimator or single equation estimator-based statistics can be more misleading than when any of the assumptions is independently violated. More so, since many real-life data do not satisfy the normality assumption there is a need for further investigation of the estimation methods that fit the non-normal regression model with different error variance structures. Therefore, the study examines the performances of some alternative estimators of the SUR model under non-normal continuous responses with heteroscedastic error terms.

*Corresponding Author

2 LITERATURE REVIEW

After the Proposed Aitken's Generalized Least Squares (AGLS) estimator also referred to as Feasible Generalized Least Square (FGLS) by Zellner (1962), much extensive theoretical and empirical applications of the work has been recorded in econometrics, statistics and other areas in recent time. For instance, finite sample properties of SUR estimator (SURE) have been studied in the literature (e.g. Zellner, 1962 & 1963; Kakawani, 1967, Zellner & Huang, 1962; Zellner & Theil, 1962) by asymptotic expansions (e.g. Srivastava and Maekawa, 1995), or by simulation (e.g. Kmenta & Gilbert, 1968).

Several other researchers also worked on the SUR model under the violation of basic model assumptions involving different types of data such as time series, Panel, Cross-sectional data and so on. Kmenta and Gilbert (1970), considered the problem of estimating a system of regression equations in which the disturbances are both serially and contemporaneously correlated. They developed an estimator that is consistent and more efficient than SURE in the situation in which each of the disturbance follows a first-order autoregressive scheme. Other works within the SUR frame-work with autocorrelated disturbances and error component can be found in (Guilkey and Schmidt, 1973; Avery, 1977; Walter Kramer, 1980; Baltagi, 1980; Messemmer and parks, 2004; Alaba, 2010).

Takada et al (1995), presented methods of resolving the problem of non-singularity in the covariance matrix of the errors in the SUR model and proposed an efficient procedure for estimation. The empirical study of the estimator was investigated by studying the diffusion processes of videocassette recorders across different geographic regions in the US, which exhibits a singular covariance matrix. The empirical results show that the procedure is efficient in tackling the problem and provide plausible estimation results. The efficiency of the SUR model when exogenous variables across-equations are correlated was investigated by Yahya et al. (2008). In their work, it was established that for a large sample size ($n \geq 100$), the SURE would still be efficient if a correlation exists among the exogenous variable in SUR model. However, under small and moderate sample sizes, they recommended a Tolerable non-orthogonal correlation point (TNCP) of ± 0.2 under which the SUR estimator would still be efficient.

Olamide and Adepoju (2013), considered a situation in which the assumption of independence of error terms in the SUR model is violated. The performances of FGLS, Iterative Ordinary Least Squares (IOLS), and Ordinary Least Squares (OLS) were compared in the SUR model with first-order autoregressive error terms. Olanrewaju (2013) noted that one of the problems of time series data is the autocorrelation of the error terms and extended the work of Yahya et al. (2008) to situations where disturbances in different equations follow the first-order autocorrelation. The effect of multicollinearity, autocorrelation, and correlation between the error terms on some methods of estimation of the system of the simultaneous equation were investigated through Monte Carlo experiment. The performances of Ordinary Least

Squares (OLS), Three-Stage Least Squares (3SLS), Feasible Generalized Least Squares (FGLS), Maximum Likelihood (ML), Full Information Maximum Likelihood (FML) and Multivariate Regression (MR) were investigated extensively at different levels of multicollinearity, autocorrelation and correlation between the error terms. The results of the study showed that the ML estimator is preferred when there is a presence of autocorrelation and multicollinearity in the model. However, when there is a correlation between the error term, SUR and 3SLS should be preferred.

SUR model also gained an appreciable application in econometrics and applied sciences. For instance, Sparks (2004) developed a SUR procedure that is applicable to environmental situations especially when missing and censored data are inevitable. Singh and Ullah (1974) extended Zellner's (1962) SUR model to credibility regression model with random coefficient and proposed estimators that are asymptotically more efficient than Zellner's estimator. In share equation systems with a random coefficient, Mandy & Martins-Filho (1993) proposed a consistent and asymptotically efficient estimator for SUR systems that have an additive heteroscedastic contemporaneous correlation. They followed Amemiya (1977) by using Generalized Least Squares to estimate parameter of covariance matrix.

Recently, Afolayan and Adeleke (2018) investigated the performances of some estimators of the SUR model with heteroscedastic error terms using the Monte Carlo Simulation approach. The study revealed that the OLS estimator performs best in a small sample size ($n < 30$) situation with low contemporaneous correlation (0.2) except when there is quadratic heteroscedasticity where Robust Huber-M estimator was the best. Meanwhile, FGLS performs best in medium and large samples under low contemporaneous correlation and the gain in efficiency is magnified at high contemporaneous correlations. The study, however, does not consider situations where the distribution of the response variables may be skewed as evidenced in many applications. Hence, the need to investigate the efficiency of SUR estimators under skewed (non-normal) responses with different heteroscedastic structures in a system of regression models as is being undertaken in this work.

3 THE SUR FRAMEWORK AND ESTIMATION PROCEDURE

The SUR model is given by

$$Y_i = X_i\beta_i + \varepsilon_i, \quad i = 1, \dots, M \quad (1)$$

where Y_i is $n \times 1$ vector denoting observations for the i^{th} equation and X_i is a $n \times k$ matrix of non-stochastic regressors, β_i is a $k \times 1$ vector of parameters while ε_i is a $n \times 1$ vector of disturbances.

The equation (1) could be written in compact form as;

$$Y = X\beta + \tilde{\varepsilon} \quad (2)$$

and in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{mn \times 1} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{bmatrix}_{mn \times \sum_{i=1}^m K_i} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}_{\sum_{i=1}^m K_i \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{mn \times 1} \quad (3)$$

3.1 FEASIBLE GENERALIZED LEAST SQUARE ESTIMATOR

The assumption regarding the disturbances in equation (1) is that the disturbances are contemporaneously correlated with mean, $E(\varepsilon) = 0$ and variance-covariance of the disturbances denoted by Ω is $E(\varepsilon\varepsilon') = \text{Var}(\varepsilon) = \Omega = \Sigma \otimes I$ where $\otimes$ is the Kronecker product operator, I is an identity matrix of order $N \times N$ and $\Sigma = ((\sigma_{ij}))$ is an $M \times M$ symmetric and positive definite matrix such that

$$E(\varepsilon_i \varepsilon_j') = \begin{cases} \sigma_{ii} I & \text{for } i = j, \\ \sigma_{ij} I & \text{for } i \neq j, \end{cases} \quad \begin{matrix} j = 1, \dots, m \\ i = 1, \dots, m \end{matrix} \quad (4)$$

where Ω is an $m \times m$ matrix of the form:

$$E(\varepsilon_i \varepsilon_j') = \text{var.}(\varepsilon_i) = \Omega = \begin{pmatrix} \sigma_{11} I & \sigma_{12} I & \cdots & \sigma_{1m} I \\ \sigma_{21} I & \sigma_{22} I & \cdots & \sigma_{2m} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} I & \sigma_{m2} I & \cdots & \sigma_{mm} I \end{pmatrix} \quad (5)$$

Estimating $k \times 1$ vector of parameters in the equation (2) is equivalent to applying OLS to each equation since the covariance matrix is no longer scalar of the form $\sigma^2 I$. The OLS estimator for β in each equation in equation (2) is given as:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (6)$$

The estimator of equation (2) is unbiased and consistent but less efficient and its variance-covariance matrix is given by

$$V(\hat{\beta}) = (X'X)^{-1} X'\Omega(X'X)^{-1} \quad (7)$$

Aitken (1934) suggested an infeasible estimator called the Generalized Least Square (GLS) estimator for a known m by m positive definite variance-covariance matrix Ω using a Weighted Least Square (WLS) approach. The idea is to transform the model in (2) such that the disturbances covariance matrix in (3.6) becomes $\Omega = \sigma^2 I$. That is, transform the joint model in (2) by a square and invertible but non-diagonal weighting matrix A , such that:

$$A\Omega A' = I_n \Leftrightarrow \Omega = (A'A)^{-1} \quad (8)$$

Multiplying the compact SUR model in (3.1) by A to obtain

$$AY = AX\beta + A\varepsilon \quad (9)$$

with $Y^* = AY$, $X^* = AX$, $\varepsilon^* = A\varepsilon$, then, model (2) becomes

$$Y^* = X^*\beta + \varepsilon^* \quad (10)$$

The $E(\varepsilon^* \varepsilon^{*'}) = E(A\varepsilon\varepsilon'A') = AE(\varepsilon\varepsilon')A' = A\Omega A' = I_m$, thus the transformed model now has its $E(\varepsilon^*) = 0$ and $\text{Var}(\varepsilon^*) = \sigma^2 I_{nm}$.

The GLS estimator of β in (3.9) is given by

$$\widehat{\beta}_{GLS} = (X^*X^*)^{-1} X^{*'} Y^* = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y \quad (11)$$

And its variance- covariance matrix is

$$V(\beta_{GLS}) = \sigma^2 (X'\Omega^{-1}X)^{-1} \quad (12)$$

However, in practice, the value of the covariance matrix Ω is generally unknown. Zellner (1962) suggested replacing Ω by its consistent estimator $\hat{\Omega}$ and estimating the parameters in the SUR model in two steps called the Feasible GLS (FGLS) estimator and is given as

$$\widehat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}Y \quad (13)$$

Where $\hat{\Omega} = S \otimes I$ is an $m \times m$ matrix based on the single equation of OLS disturbances and $S = (s_{ij})$, where $s_{ij} = \varepsilon_i' \varepsilon_j / (n - k)$, see (Yahya et.al, 2008).

3.2 HUBER M-REGRESSION ESTIMATOR

The M-estimators for regression which was introduced by Huber (1973), is a generalization of the OLS estimation procedure by minimizing the sum of a less rapidly increasing function (objective function) of the residuals instead of minimizing the sum of squared residuals as;

$$\hat{\beta} = \min \sum_{i=1}^n \rho(y_i - X_i' \beta) \quad (14)$$

Equivalent to the M-Estimator of location, the robustness of the estimator is determined by the choice of weight function. The solution is not scale-equivariance, and thus the residuals must be standardized by a robust scale estimator $\hat{\sigma}_e$ usually the median absolute deviation (MAD). Differentiating the objective function in (14) and setting the partial derivative to zero gives the score function:

$$\hat{\beta} = \min \sum_{i=1}^n \psi((y_i - X_i' \beta) / \hat{\sigma}_e) X_i' = 0 \quad (15)$$

where ψ is the derivative of the objective function ρ . Equation (3.13) is a P system of normal equations. Then ψ is replaced by appropriate weights that decrease as the size of the residual increases, defined by $w(\varepsilon) = \frac{\psi(\varepsilon)}{\varepsilon}$ and $w_i = w(\varepsilon)$. Hence (15) becomes

$$\hat{\beta} = \min \sum_{i=1}^n \omega_i ((y_i - X_i' \beta) / \hat{\sigma}_e) X_i' = 0 \quad (16)$$

In order to solve the equation in (16) with respect to β , an iterative procedure called Iterative Re-Weighted Least Squares (IRWLS) is employed as follows;

1. Set the iteration counter $t=0$ and select the initial estimates $\beta^{(0)}$ from initial OLS estimates.
2. Calculate the residuals from OLS in (1) and set as $\varepsilon^{(0)}$

3. Select any weight function of choice and apply to the initial OLS residuals to create an initial weight, $w(\epsilon^{(0)})$
4. The first iteration, $t=1$, uses weighted least squares (WLS) to minimize $\sum \omega_i^1 \epsilon_i^2$ and obtain $\beta^{(1)}$. In matrix form, where W represents an $n \times n$ diagonal matrix of individual weights, the solution is
$$\hat{\beta}^{(1)} = (X'WX)^{-1} X'WY$$
5. At each iteration t , calculate the residual $\epsilon^{(t-1)}$ and associated weights $w(\epsilon^{(t-1)})$ from the previous iteration
6. Solve for new WLS estimates $\hat{\beta}^{(t)} = (X'W^{(t-1)}X)^{-1} X'W^{(t-1)}Y$, where $W^{(t-1)} = \text{diag}(W_i^{(t-1)})$, the current weight matrix and X is the model design matrix, with X' as transpose.
7. Steps 4-6 are repeated until the estimated coefficients converge (see Maronna *et al.*, 2006).

M-estimates based on Huber's ψ function were used in this study with Objective and bi-weight function defined respectively as:

$$\rho(\epsilon)_{\text{huber}} = \begin{cases} \frac{1}{2}\epsilon^2 & \text{for } |\epsilon| \leq k \\ k|\epsilon| - \frac{1}{2}k^2 & \text{for } |\epsilon| > k \end{cases} \quad (17)$$

$$w_{\text{huber}} = \begin{cases} 1 & \text{for } |\epsilon| \leq k \\ \frac{k}{|\epsilon|} & \text{for } |\epsilon| > k \end{cases} \quad (18)$$

Huber's ψ functions have computational advantages but sensitive to leverage points; see Maronne *et al.* (2006) for more details. The choice of $c = 1.345$ recommended by Huber (1981) produced a relative efficiency of approximately 95% when the error density is normal.

3.3 TURKEY BI-SQUARES MM-REGRESSION ESTIMATOR

The Turkey bi-squares MM-estimator is a robust estimator that combines the high breakdown point of S-estimator and high efficiency of M-estimator. The breakdown point is a measure of the proportion of outliers that can be addressed before these observations affect the model. The S-estimator proposed by Rousseeuw and Yohai (1984) minimizes the dispersion of the residuals and it is given as:

$$\min \frac{1}{n} \sum_i^n \rho\left(\frac{\epsilon_i}{\hat{\sigma}_e}\right) \quad (19)$$

where $\hat{\sigma}_e$ is the estimate of residual scale and ρ is the weight function.

MM-estimator requires three stages. The first stage finds the regression parameter using S estimator which is consistent and has a high breakdown point of 50% but not necessarily efficient. In the second stage, the M-estimator of the residuals is calculated using the residuals obtained from the initial S estimator. The M-estimator of the regression parameter as described in subsection 3.2 is then used in the third stage to estimate the regression parameter that is consistent, robust to outlier with high efficiency and asymptotically normally distributed

(Maronna *et al.* (2006)). The MM-estimator uses Turkey's objective function given as:

$$w_{\text{bisquare}} = \begin{cases} \frac{c^2}{6} \{1 - [1 - (\epsilon/c)^3]\} & \text{for } |\epsilon| \leq k \\ \frac{c^2}{6} & \text{for } |\epsilon| > k \end{cases} \quad (20)$$

and Turkey bi-weight functions stated below:

$$w_{\text{bisquare}} = \begin{cases} \{[1 - (\epsilon/c)^2]^2\} & \text{for } |\epsilon| \leq c \\ (0) & \text{for } |\epsilon| > c \end{cases} \quad (21)$$

where the constant $c = 4.685$ as determined by Susanti *et al.* (2014) which produced 95% relative efficiency.

4 DATA GENERATING PROCEDURE

To examine the performances of OLS, Huber-M, and Turkey-MM relative to FGLS (SUR) estimator, the simulation study considers a system of SUR equations containing two distinct linear regression equations and multiplicative heteroscedasticity error terms as suggested by Harvey (1976) with the assumption that the variance of the error term varies with the mean response. That is $\text{Var}(\epsilon'_i) = \sigma^2 e^{E[Y_i]}$.

The multiplicative heteroscedasticity henceforth referred to as the exponential heteroscedasticity is given as;

$$h_{i1} = \sigma^2 \exp(\alpha_1 + \gamma_{20}X_{i1} + \gamma_{21}X_{i2}), \quad i = 1, 2,$$

where α_1, γ_{20} and γ_{21} are arbitrarily fixed constants. For instance, we fixed $\alpha_1 = 0.6$, $\gamma_{20} = 0.8$, $\gamma_{21} = 0.4$, and $\sigma^2 = 1$.

To assess the asymptotic and small sample properties of the various estimators under the violation of normality of the responses, homoscedasticity of the error terms and, the correlation among errors across equations, the entire simulation experiment was performed for various sample sizes (n): $n = 500, 250, 100, 50, 30, 20$ with 1000 replicates in each case. However, the estimation method, model specification, and form of heteroscedasticity as given by Afolayan and Adeleke (2018) were used in the study for easy comparison. The simulation study was generated given the SUR model in equation(1.0) as:

$$Y_1 = X_1\beta_1 + \epsilon_1$$

$$Y_2 = X_2\beta_2 + \epsilon_2$$

with exogenous variables, error terms and parameters defined as follows:

$$X_1 = (1, X_{11}, X_{12}); \beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})$$

$$X_2 = (1, X_{21}, X_{22}); \beta_2 = (\beta_{20}, \beta_{21}, \beta_{22}) \text{ and } \epsilon = (\epsilon_1, \epsilon_2)$$

The values of the model parameters were set as follows; $\beta_{10} = -20$, $\beta_{11} = 15$, $\beta_{12} = 13$, $\beta_{20} = 11.5$, $\beta_{21} = 10.0$, and $\beta_{22} = 6.5$

The contemporaneous correlations between the errors from the two equations were specified as:

$\rho = 0.0, 0.2, 0.5, 0.7, \text{ and } 0.95$.

The SUR model therefore becomes:

$$y_{i1} = -20 + 15X_{i1} + 13X_{i2} + \varepsilon_{i1} \quad h_{ij}; j=1, 2, 3, 4. \quad (22)$$

$$y_{i2} = 11.5 + 10X_{i1} + 6.5X_{i2} + \varepsilon_{i2} \quad h_{ij}; \text{ for } i=1, \dots, n. \quad (23)$$

The simulations of observations for the two regression equation model of (22 & 23) were carried out as follows:

1. Each of the exogenous variables in $[X_1, X_2]$ was generated from uniform distribution $(-0.5, 0.5)$ i.e. $X_c \sim U(-0.5, 0.5)$, $r = 1, 2$, for various sample sizes.
2. The error terms $[\varepsilon_{i1}, \varepsilon_{i2}]$ in equation (22 & 23) were drawn from multivariate gamma distribution (MVG) with parameters scale=4, shape=2 and variance-covariance = Σ_ε with diagonal=1, off-diagonal= ρ_ε and correlation coefficient $\rho_\varepsilon = 0.0, 0.2, 0.5, 0.7$ and 0.95 i.e. $MVG(4, 1; \Sigma_\varepsilon)$.

The heteroscedastic structure for error variance, h_{ij} as given in section 4 were generated using the specified values: $\alpha_1 = 0.6$, $\gamma_{20}=0.8$, $\gamma_{21} = 0.4$, $\sigma^2=1$ based on generated exogenous variables $[X_1, X_2]$.

To study the effect of small and large samples on the properties of estimators of β we considered samples of sizes 20 & 30; 50 & 100 and 250 & 500 for small, medium, and large samples respectively.

The simulation was repeated 1000 times and for each, we got 1000 estimates for each replication, the value of the parameters $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})$ and $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})$ were estimated for the following estimators:

- i. Feasible Generalized Least Squares (FGLS)
- ii. Ordinary Least Squares (OLS)
- iii. Huber M-Estimator (Huber)
- iv. Turkey Bi-Squares MM- Estimator (BISQ)

The performances of these estimators at different heteroscedastic structures, contemporaneous correlations, and sample sizes were evaluated using the Root Mean Square Error (RMSE) and their Relative Efficiency (RE) to OLS. In other words, the relative efficiency of an estimator, in this case, is measured by the degree to which the estimator performs similar to the OLS estimator.

For two unbiased estimators θ_1 and θ_2 , the relative efficiency of θ_2 to θ_1 is given by:

$$RE = \frac{RMSE(\theta_1)}{RMSE(\theta_2)}$$

Here, θ_1 is the OLS estimator and θ_2 is any other estimator. If the relative efficiency is 1, then it means that the estimator is as efficient as the OLS, if the error distribution of the data is normal. The relative efficiency of 1.2 for

instance, implies the estimator θ_2 is 20% more efficient than the OLS estimator.

The RMSE of the regression parameters estimator $\hat{\beta}_{ij}$ of parameter β_{ij} in $\beta = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22})$ is calculated as:

$$RMSE = \frac{1}{K} \sqrt{\sum_{k=1}^K \sum_{i=1}^2 \sum_{j=0}^2 (\beta_{ijk} - \hat{\beta}_{ijk})^2}, \text{ with } K = 1000, \text{ the number of replications.}$$

5 RESULTS AND DISCUSSION

The results for exponential and homoscedastic error terms are discussed under the gamma response model. Meanwhile, the results on Quadratic, Linear, and Square root heteroscedasticity under the gamma response model were similar to that of Exponential heteroscedasticity, therefore their results are not presented here due to space. Again, results on exponential and Square root were also not presented under the Lognormal response model for the same reason given earlier. Table 1 shows the results for the case of exponential heteroscedasticity and gamma response at various levels of contemporaneous correlations.

At small sample sizes $n \leq 30$ when the correlation is absent (i.e. at $\rho = 0.0$) between the error terms in different equations, the performances of OLS and FGLS estimators are similar based on RMSE criterion while the Huber and Bi-squares performed better than the OLS and FGLS estimators. The relative efficiency of Huber and Bi-squares estimators to OLS is 23% and 30% respectively. However, as the level of correlation increases ($0.2 \leq \rho \leq 0.7$), the performances of Huber and Bi-squares estimators decreased as evident from their estimated relative efficiencies which decreases from 21% to 14% and 27% to 16% respectively. However, both estimators still performed better than the OLS and FGLS.

Nevertheless, the performance of FGLS has improved tremendously and better than OLS. In summary, in terms of RMSE, the most efficient estimator under the exponential heteroscedastic error terms in the Gamma response model at small sample sizes $n \leq 30$ with the absence of correlation or moderately high correlation ($\rho = 0.0$ or $0.2 \leq \rho \leq 0.7$) is Bi-squares estimator. When the correlation is almost perfect ($\rho = 0.95$) at sample sizes $n \leq 50$, FGLS shows better efficiency over other estimators. The graphical presentations of the various performances of all these estimators are provided by Fig 1 in the Appendix which showed the plot of RMSE yielded by each estimator against the different sample sizes.

As sample sizes increased, specifically at $n \geq 50$ in the absence of correlation, OLS and FGLS were similar in performance and were consistent. Huber and Bi-squares continuously maintained better performance over OLS and FGLS. The relative efficiency of Huber and Bi-squares over OLS is between 20% to 23% and 29% to 43% respectively.

Table 1. Estimated RMSE and RE for different Estimators at Various Levels of Contemporaneous Correlations under the Exponential Heteroscedastic Gamma Response Model

Sample Size (n)	Estimator	Cc $\rho=0.0$		Cc $\rho=0.2$		Cc $\rho=0.5$		Cc $\rho=0.7$		Cc $\rho=0.95$	
		RMSE	RE	RMSE	RE	RMSE	RE	RMSE	RE	RMSE	RE
20	OLS	48.9384	1.00	53.2834	1.00	56.4428	1.00	18.1692	1.00	55.7569	1.00
	FGLS	49.4507	0.99	52.2794	1.02	52.7690	1.07	16.5943	1.09	38.9281	1.43
	HUBER	39.7426	1.23	44.0038	1.21	48.0720	1.17	15.9618	1.14	49.4964	1.13
	BI-SQ	37.7426	1.30	42.0488	1.27	46.6847	1.21	15.7030	1.16	48.3997	1.15
30	OLS	41.6984	1.00	40.2076	1.00	43.1733	1.00	14.5887	1.00	45.5751	1.00
	FGLS	42.6573	0.98	40.3472	1.00	42.3686	1.02	13.9337	1.05	33.8621	1.35
	HUBER	34.8722	1.20	34.9920	1.15	38.6882	1.12	13.1156	1.11	40.5911	1.12
	BI-SQ	32.8739	1.27	34.0380	1.18	37.8853	1.14	12.7619	1.14	39.3795	1.16
50	OLS	35.6911	1.00	37.5678	1.00	40.5786	1.00	13.6563	1.00	41.8095	1.00
	FGLS	35.9675	0.99	37.3849	1.00	39.4048	1.03	12.9898	1.05	33.1375	1.26
	HUBER	29.6377	1.20	31.4056	1.20	35.2072	1.15	11.6860	1.17	36.3063	1.15
	BI-SQ	27.6269	1.29	29.4435	1.28	33.5609	1.21	11.1159	1.23	34.9453	1.20
100	OLS	30.2379	1.00	31.7156	1.00	35.0625	1.00	11.2856	1.00	37.0484	1.00
	FGLS	30.4133	0.99	31.6426	1.00	34.4978	1.02	10.8856	1.04	31.8901	1.16
	HUBER	24.8286	1.22	26.5312	1.20	30.5373	1.15	9.7781	1.15	32.0775	1.15
	BI-SQ	22.4362	1.35	24.5602	1.29	28.8979	1.21	9.2712	1.22	30.5703	1.21
250	OLS	26.5354	1.00	28.7715	1.00	32.4383	1.00	10.2531	1.00	34.2529	1.00
	FGLS	26.5522	1.00	28.6184	1.01	31.9222	1.02	9.8907	1.04	31.6721	1.08
	HUBER	21.8165	1.22	23.7330	1.21	27.8342	1.17	8.7382	1.17	29.7591	1.15
	BI-SQ	19.3993	1.37	21.2960	1.35	25.8080	1.26	8.0430	1.27	27.9974	1.22
500	OLS	25.0425	1.00	27.1047	1.00	30.8832	1.00	9.8702	1.00	32.9058	1.00
	FGLS	25.0446	1.00	26.9882	1.00	30.5509	1.01	9.7188	1.02	31.7810	1.04
	HUBER	20.3845	1.23	22.5394	1.20	26.6674	1.16	8.4001	1.18	28.7197	1.15
	BI-SQ	17.5187	1.43	19.9602	1.36	24.5184	1.26	7.6486	1.29	26.7832	1.23

For sample sizes $n > 50$, the increase in the levels of contemporaneous correlation could not improve the efficiency of FGLS over Huber and Bi-squares, even at ($\rho = 0.95$). This is to reaffirm that FGLS or OLS is not the best estimator for modeling a non-normal response model regardless of the levels of contemporaneous correlation between the error terms in different models. The robust Bi-square is the most efficient estimator for modeling exponential heteroscedastic error terms with Gamma response.

Table 2 presents RMSE for model parameter estimates when the error term is homoscedastic at various levels of contemporaneous correlation for different sample sizes. When there is an absence of correlation ($\rho = 0.0$) between the error terms in different equations and error terms is homoscedastic using the RMSE criterion, the best estimator is Bi-squares at different sample sizes ($n \geq 20$). The Huber estimator performed better than OLS and FGLS but below the Bi-squares estimator. At this point, it shows clearly that when the response is non-normal, OLS or FGLS is less efficient. Graphical presentations of all these results are presented by Fig 2 in the Appendix.

When the correlation is ($0.2 \leq \rho \leq 0.7$) at various sample sizes $n \leq 500$ and the error term is homoscedastic using the RMSE criterion, Bi-squares is the best estimator followed by Huber estimator. The performances of FGLS continue to improve as the correlation increases but perform below Bi-squares and Huber estimators. The relative efficiency of Bi-squares, Huber and FGLS to OLS reduces under homoscedastic error term than when the error is heteroscedastic (13-29%, 11-20% and 2-9% respectively).

In summary, the best estimator for modeling homoscedastic Gamma response at different levels of contemporaneous correlation notwithstanding the size of the sample using RMSE criterion is the Bi-squares estimator. This same inference was obtained under other forms of heteroscedasticity of the error terms in a system of simultaneous equations.

Table 2. Estimated RMSE and RE for different Estimators at Various Levels of Contemporaneous Correlations under Homoscedastic Gamma Response Model

Sample Size (n)	Estimator	Cc $\rho=0.0$		Cc $\rho=0.2$		Cc $\rho=0.5$		Cc $\rho=0.7$		Cc $\rho=0.95$	
		RMSE	RE	RMSE	RE	RMSE	RE	RMSE	RE	RMSE	RE
20	OLS	22.7647	1.00	24.3944	1.00	25.8178	1.00	23.8179	1.00	25.6066	1.00
	FGLS	22.8307	1.00	24.1065	1.01	24.2715	1.06	21.8774	1.09	16.8463	1.52
	HUBER	18.8108	1.21	20.5622	1.19	22.4032	1.15	21.3922	1.11	22.9194	1.12
	BI-SQ	17.9699	1.27	19.6576	1.24	21.7162	1.19	21.0180	1.13	22.4364	1.14
30	OLS	19.1940	1.00	18.5527	1.00	19.8854	1.00	19.0428	1.00	21.3412	1.00
	FGLS	19.5614	0.98	18.5333	1.00	19.3577	1.03	17.8327	1.07	14.6193	1.46
	HUBER	16.2468	1.18	16.2024	1.15	17.9091	1.11	17.0774	1.12	19.1322	1.12
	BI-SQ	15.3560	1.25	15.7288	1.18	17.4842	1.14	16.5405	1.15	18.6055	1.15
50	OLS	16.6502	1.00	8.6736	1.00	18.7456	1.00	17.5877	1.00	19.5068	1.00
	FGLS	16.7676	0.99	8.7207	0.99	18.0749	1.04	16.4477	1.07	13.6843	1.43
	HUBER	13.9511	1.19	7.5122	1.15	16.5335	1.13	15.4502	1.14	16.9501	1.15
	BI-SQ	13.1377	1.27	6.8585	1.26	15.8402	1.18	14.8094	1.19	16.4477	1.19
100	OLS	13.6673	1.00	14.3013	1.00	15.5410	1.00	14.0942	1.00	16.3450	1.00
	FGLS	13.7130	1.00	14.2533	1.00	15.1872	1.02	13.4207	1.05	12.4152	1.32
	HUBER	11.3899	1.20	11.9373	1.20	13.5278	1.15	12.4344	1.13	14.2139	1.15
	BI-SQ	10.4954	1.30	11.1793	1.28	12.9387	1.20	11.9938	1.18	13.7115	1.19
250	OLS	11.7760	1.00	12.4805	1.00	13.7955	1.00	11.9508	1.00	14.5030	1.00
	FGLS	11.7945	1.00	12.4412	1.00	13.5271	1.02	11.3328	1.05	11.7934	1.23
	HUBER	9.9413	1.18	10.5096	1.19	12.0408	1.15	10.4224	1.15	12.7663	1.14
	BI-SQ	8.9876	1.31	9.6541	1.29	11.3971	1.21	9.8569	1.21	12.2149	1.19
500	OLS	10.5230	1.00	11.2493	1.00	12.5674	1.00	10.8907	1.00	13.1193	1.00
	FGLS	10.5230	1.00	11.2191	1.00	12.3788	1.02	10.6087	1.03	11.5541	1.14
	HUBER	8.8316	1.19	9.5866	1.17	11.0783	1.13	9.5153	1.14	11.7642	1.12
	BI-SQ	7.8623	1.34	8.7151	1.29	10.4116	1.21	8.9211	1.22	11.2537	1.17

6 CONCLUSION

In this paper, we compared the FGLS estimator with OLS, M-Huber, and MM-BISQ estimators in two seemingly unrelated regression models when responses are non-Gaussian under different heteroscedastic error structures. The OLS estimator lost its BLUE property in a small sample under the non-normal response model. Also, the FGLS was consistent with an increase in contemporaneous correlation but not efficient even at correlations between $0.5 \leq \rho \leq 0.7$. The MM-BISQ and M-Huber estimators showed lower RMSE than the OLS or FGLS in both small and large sample sizes regardless of the heteroscedastic structures. The FGLS only outperforms other estimators in small samples when the contemporaneous correlation between the error terms in the two equations is almost perfect ($\rho = 0.95$) and the error terms are homoscedastic, Linear or square root heteroscedastic.

The robust MM-BISQ and M-Huber estimators are more efficient than the OLS or FGLS estimator in estimating the parameters in two SUR model with gamma or lognormal responses. The efficiency gain for using either BISQ or Huber is magnified in small samples ($n \leq 30$) and reduced as sample size increases. This study concludes that the single equation estimation procedure of MM-BISQ or M-Huber showed higher efficiency in modeling non-

normal response equations than the joint estimation procedure of FGLS.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers and journal editors for their comments and observations all of which have improved the original draft of this paper.

REFERENCES

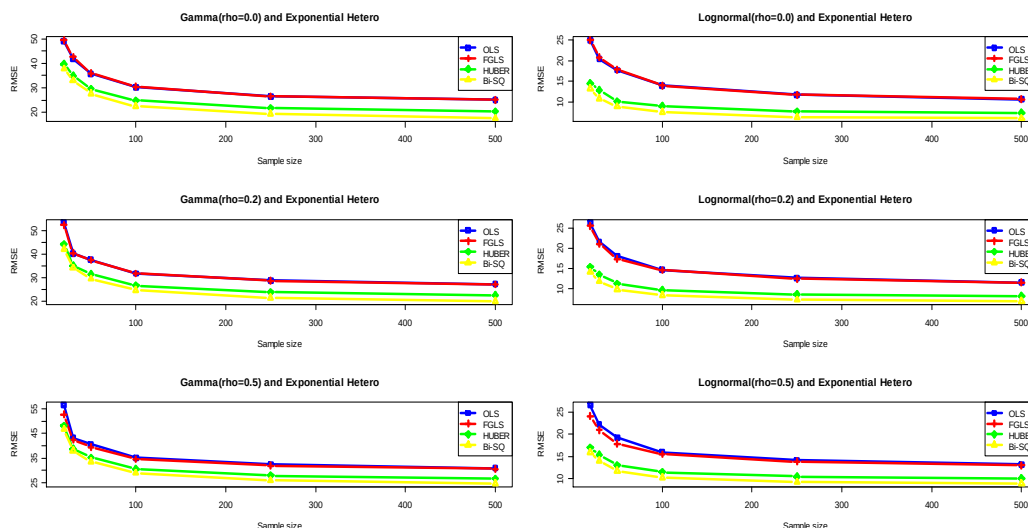
- Afolayan, R. B. & Adeleke, B. L. (2018). On the Efficiency of Some Estimators for Modeling Seemingly Unrelated Regression with Heteroscedastic Disturbances, *IOSR Journal of Mathematics (IOSR-JM)*, 14(4), 01-13. <http://www.iosrjournals.org>
- Aitken, A. C. (1934). On Least-Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh*. 55, 42 - 48.
- Amemiya, T. (1977). A Note on a Heteroscedastic Model. *Journal of Econometrics*, 6, 365 - 370.
- Avery, R. B. (1977). Error Components and Seemingly unrelated regressions. *Econometrica* 45, 199-209.
- Baltagi, B. H. (1980). On seemingly unrelated regressions with error Components. *Econometrica*, 48(6), 1547-1551.
- Biondini, R. (1976). Cloud motion and rainfall statistics. *Journal of Applied Meteorology* 15: 205-224
- Carroll, R. J. (1982). Adapting for Heteroscedasticity in Linear models. *Annals of Statistics*, 10(4), 1224-1233.
- Guilkey, D. K. & P. Schmidt. (1973) Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors, *Journal of the American Statistical Association*. 68, 642-647.

- Harvey, A. C. (1976). Estimating Regression Model with Multiplicative heteroscedasticity, *Econometrica*, 44(3), 461-465.
- Huber, P. J. (1981). Robust Statistics, Wiley, New York.
- Huber, P.J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo, *Annals of Statistics*, 1(5); 799-821.
- Judge, G. G., Hill, R.C., Griffiths, W. E. & Lutkepohl, H. (1988). Introduction to theory and practice of Econometrics, 2nd ed. New York, Wiley, 444-468.
- Kakwani, N. C. (1967). The unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators. *Journal of the American Statistical Association*, 62: 141 – 142.
- Kmenta, J. and Gilbert, R. F., (1970). Estimation of seemingly unrelated regressions with autoregressive disturbances. *Journal of the American Statistical Association*, 65, 186-197.
- Kmenta, J. and Gilbert, R. F., (1968). Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regression. *Journal of American Statistical Association*, 63, 1180-1200.
- Kondo, K. (1977). The log-normal distribution of the incubation time of exogenous diseases. *Japanese Journal of Human Genetics* 21:217-237.
- Kramer, W. (1980). Finite Sample Efficiency of Ordinary Least Squares in the Linear Regression Model with Authocorrelated Errors. *Journal of Ameriacan Statistical Association*, 75(372): 1005-1009.
- Malanca, A., Gaidolfi, L., Pessina, V. & Dallara, G. (1996). Distribution of 226-Ra, 232-Th and 40-K in soils of Rio Grande do Norte (Brazil). *Journal of Environmental Radioactivity* 30:55-67.
- Mandy, D. M. & Martins-Filho, C., (1993). Seemingly Unrelated Regressions Under Additive Heteroscedasticity: Theory and share equation applications. *Journal of Econometrics*, 58, 315 – 346.
- Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). Robust Statistics: Theory and Methods. John Wiley and Sons Ltd, England.
- Messemer, C. and Parks, R. W., (2004). Bootstrap Methods for Inference in a SUR model with Autocorrelated Disturbances. University of Washington Economics Working Paper No. UWEC-2004-24
- Olamide, E. I. and Adepoju, A. A. (2013). Estimating Seemingly Unrelated Regressions with First Order Autoregressive Disturbances. *Studies in Mathematical Sciences*, 6(2), 40-57
- Olanrewaju, S.O. (2013). Effect of Multicollinearity and Autocorrelation on some Estimators of System of Regression equations. Unpublished PhD thesis Submitted to Department of Statistics, University of Ilorin.
- Preston, F.W. (1981). Pseudo-log-normal distributions. *Ecology* 62:355-364.
- R Development Core Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rousseeuw, P. J. and Yohai, V. J. (1984). Robust Regression by Mean of S Estimators, Robust and Nonlinear Time Series Analysis, New York, 256-274, doi: 10.1007/978-1-4615-7821-5-15.
- Sartwell, P.E. (1966): The incubation period and the dynamics of infectious disease. *American Journal of Epidemiology* 83: 204-216.
- Singh, B. and Ullah, A. (1974). Estimation of seemingly unrelated regressions with random coefficients. *Journal of the American Statistical Association*, 69, 191-195.
- Sparks, R. (2004). SUR Models Applied to an Environmental situation with missing Data and Censored Values. *Journal of Applied mathematics and Decision Sciences*. 8, 15-32.
- Srivastava, V. K. & Maekawa, K. (1995). Efficiency Properties of Feasible Generalized Least Squares Estimators in SUR Models under Non-normal. *Journal of Econometrics*, 66: 99-121. [https://doi.org/10.1016/0304-4076\(94\)01609-4](https://doi.org/10.1016/0304-4076(94)01609-4).
- Takada, H., Ullah, A. and Chen, Yu-Min (1995). Estimation of seemingly unrelated regression model when the error covariance matrix is singular. *Journal of the Applied Statistics*, 22(4), 517-530
- Yahya W. B., Adebayo, S. B., Jolayemi, E.T., Oyejola, B. A. and Sanni, O.O.M. (2008). Effects of non-orthogonally on the efficiency of seemingly unrelated regression (SUR) models. *InterStat Journal*, <http://interstat.statjournals.net/.29>.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57, 348-368.
- Zellner, A. (1963). Estimators for seemingly unrelated regression equations: some exact finite sample results. *Journal of the American Statistical Association* 58: 977-992.
- Zellner, A. and Theil, H. (1962). Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrical*, 30, 54 – 78.
- Zellner, A. and Huang, D.S. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. *International Economic Review* 3(3): 300-313.

APPENDICES

THE GRAPHICAL DESCRIPTION OF THE PERFORMANCES OF THE FOUR ESTIMATORS CONSIDERED.

A. EXPONENTIAL HETEROSCEDASTICITY



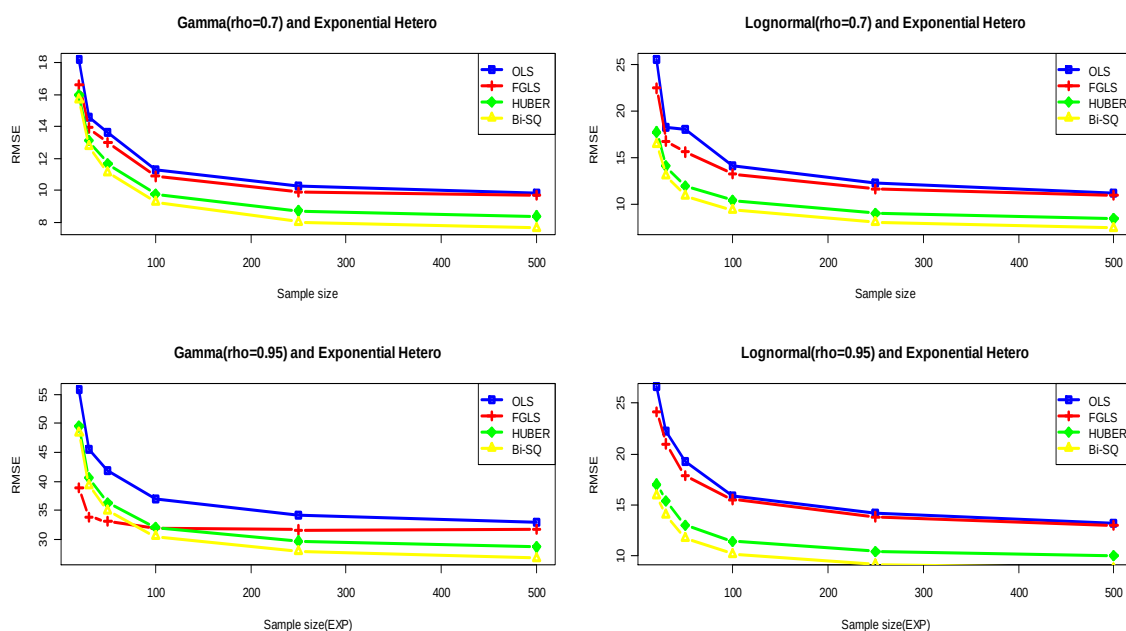


Fig. A1: Plots of RMSE of the estimators under exponential Heteroscedasticity with Gamma and Lognormal responses at various Contemporaneous Correlations ($\rho = \rho$).

B. HOMOSCEDASTIC ERROR

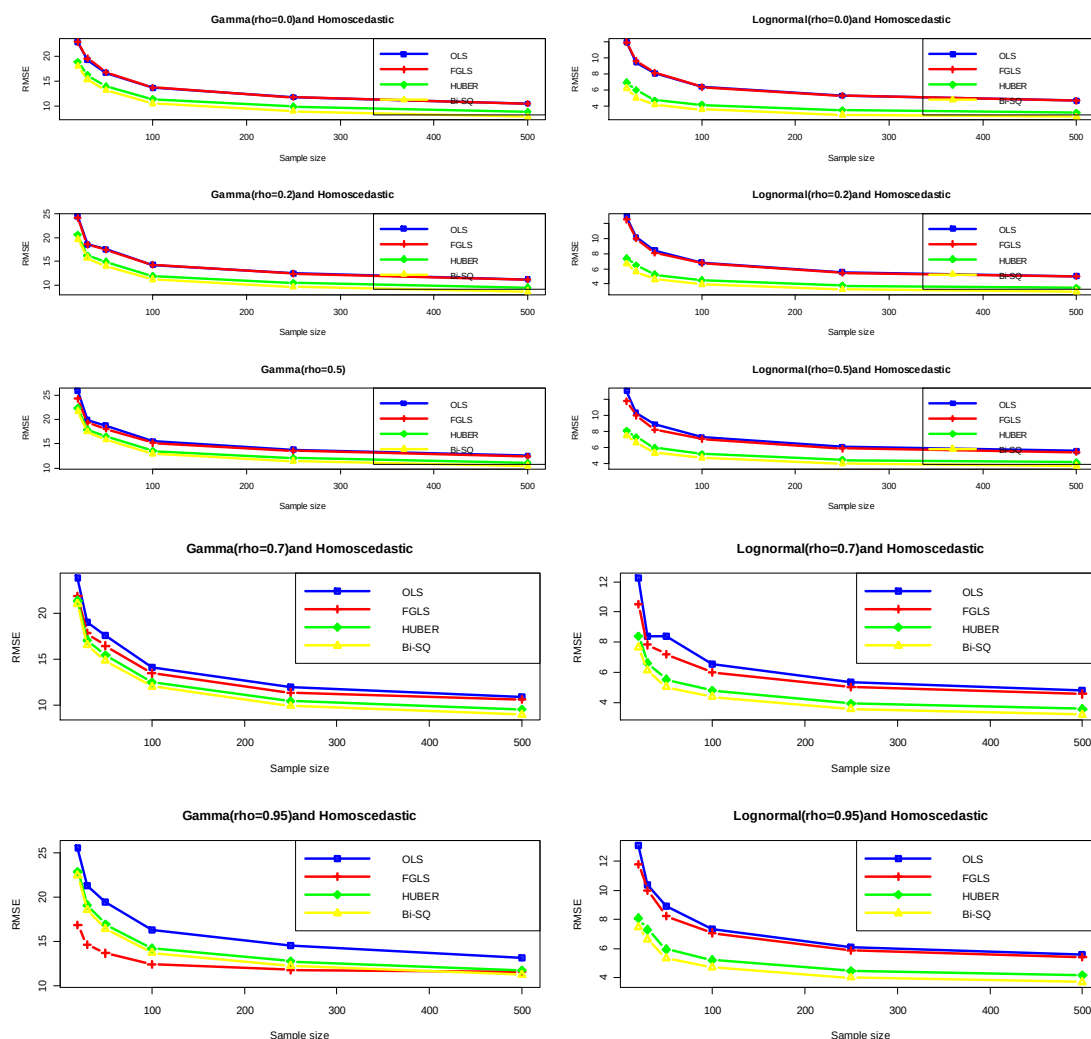


Fig. A2: Plots of RMSE of the estimators under homoscedastic error with Gamma and Lognormal responses at various Contemporaneous Correlations ($\rho = \rho$).