Journal of Engineering and Applied Sciences 14 (19): 7072-7076, 2019
ISSN: 1816-949X

Comparison of Estimate Methods of Multiple Linear Regression Model with Auto-Correlated Errors when the Error Distributed with General Logistic

Ebtisam K. Abdulah, Ahmed D. Ahmed and Baydaa I. Aboulwahhab
Department of Statistics, College of Administration and Economics, Baghdad, Iraq,
ebtisamsa@yahoo.com

Abstract: In this research, we studied the multiple linear regression models for two variables in the presence of the autocorrelation problem for the error term observations and when the error is distributed with general logistic distribution. The auto regression model is involved in the studying and analyzing of the relationship between the variables and through this relationship, the forecasting is completed with the variables as values. A simulation technique is used for comparison methods depending on the mean square error criteria in where the estimation methods that were used are (generalized least squares, M robust and Laplace) and for different sizes of samples (20, 40, 60, 80, 100, 120). The M robust method is demonstrated the best method for all values of correlation coefficients as (ϕ = -0.9, -0.5, 0.5, 0.9). So, we applied it to the data that was obtained from the Ministry of Planning in Iraq/Central Organization for Statistics which represents the consumer price index for the years 2004-2016. So, we confirmed that the dollar exchange rate is directly affected by the increase in annual inflation rates and the ratio of currency to the money supply.

Key words: Autocorrelation, generalized least squares method, Laplace robust method, logistic distribution, M robust method, multiple linear regression

INTRODUCTION

The price indexes for consumers are considered to be one of the important indicators in the process of the planning and economic developments. Throughout the prices, we can observe the changes that occur in the market and attempt to control its directions in order to protect the consumer. Multiple linear regression analysis method is considered as one of the researches important tools that is used in the study and analysis of the relationship between the dependent variable and the independent variables. With the presence of a mathematical equation and the use of this relation, a prediction can be made with one variable among the other variables. One of the classical assumptions of linear regression is that it does not have an autocorrelation problem between the values of random error variables. However, one of the causes of this problem in the linear regression model is the existence of economic fluctuations that affect the time series or the personal estimates of the data or the failure to formulate a precise linear relationship of the mathematical model. Therefore, both Eakambaram and Elangovan used a set of methods when there was a problem of the autocorrelation of errors and found that the least absolute error was the best (Eakambaram and Elangovan, 2010). Dietz studied the robust estimation for a simple linear regression model when the random error is distributed as normal and a couple of methods are used (Theil method and unweighted average method) (Dietz, 1986). Ahmed discussed the simple linear regression with the autocorrelation problem when the error distributed with a normal distribution. He used three methods (least squares, unweighted average and Theil method) and made a comparison between them by using mean squares error criteria (MSE) and the simulation method. Then he applied these methods to wheat production data in Iraq (Ahmed, 2015). The researcher Tanizaki made a comparison between the maximum likelihood method and the Bayesian method for the linear regression model with the presence of the auto correlated error. He obtained that the Bayesian method is more efficient (Tanizaki, 1989). Chib (1993) developed the practical framework for high order autoregressive processes of Gaussian and Student-t regression models. This research aims to compare between the methods which are: Generalized Least Squares, M Robust and Laplace by using the simulation of the linear regression model with two variables and with the autocorrelation of the first order when the random error is distributed as a general logistic distribution.

Corresponding Author: Ebtisam K. Abdulah, Department of Statistics, College of Administration and Economics, Baghdad, Iraq, ebtisamsa@yahoo.com

MATERIALS AND METHODS

To illustrate these three methods, we first begin by addressing the problem of autocorrelation by assuming that the linear regression model with two variables is as follows:

$$Y_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + e_i, \quad i = 1, 2, ..., n \tag{1}$$

Where:

- Y_i = The dependent variable
- X_{1i}, Y_{2i} = The independent variables
- γ_0, γ_1, γ_2 = The regression parameters
- e_i = The random error and the autocorrelation of first order by Ayinde *et al.* (2015):

$$e_i = \phi e_{i-1} + \delta_i, \quad i = 1, 2, .., n$$

Where:

- ϕ = The simple autocorrelation coefficient between random errors where ($-1 \leq \phi \leq 1$)
- δ = The independent errors and distributed general logistic distribution with probability density function:

$$f(\delta) = \frac{\frac{b}{\alpha}\exp\left(\frac{-\delta_i}{\alpha}\right)}{\left[1+\exp\left(\frac{-\delta_i}{\alpha}\right)\right]^{b+1}}, \quad -\infty < \delta < \infty$$

Where:

- b = The shape parameter
- α = The scale parameter
- δ = The random error and the cumulative density function for this distribution is:

$$F(\delta) = \left[1 + \exp\frac{-\delta_i}{\alpha}\right]^{-b}$$

The skewed of the distribution is positive when (b>1) and negative when (b<1) and symmetric when (b = 1). As for solving the autocorrelation problem we used the transformation method and our model which is Eq. 1 in the following form:

$$Y_i - \phi Y_{i-1} = (1-\phi)\gamma_0 + \gamma_1 (X_{1i} + \phi X_{1i-1}) + \gamma_2 (X_{2i} - \phi X_{2i-1})$$
$$Y_i^* = \gamma_0^* + \gamma_2 X_{1i}^* + \gamma_2 X_{2i}^* + \delta$$
$$\gamma_0^* = (1-\phi)\gamma_0$$

Generalized Least Squares Method (GLSM): In order to obtain the General Least Squares (GLS) method, we minimize the sum squares of errors as:

$$\sum \delta_i^2 = \sum \left(Y_I^* - \gamma_0^* - \gamma_1 X_{1i}^* - \gamma_2 X_{2i}^*\right)^2$$

As for estimation form by using matrices is as follows:

$$\tilde{\gamma} = \left(x^{*'} x^*\right)^{-1} \left(x^{*'} y^*\right)$$

M robust method: This method requires to minimize Eq.:

$$\sum_{i=1}^{n} \rho\left(y_i^* - x_i^{*'} \gamma\right)$$

where (ρ) is a convex function as well as it's symmetric. The formula for estimating the parameters of the M Robust method is as follows:

$$\hat{\gamma} = \left(x^{*'} w x^*\right)^{-1} \left(x^{*'} w y^*\right)$$

where, the elements of diagonal weighted matrix are:

$$w_i = \psi\left(\frac{y_i^* - x_i^{*'}\gamma_a}{\hat{\sigma}}\right) \Big/ \left(\frac{y_i^* - x_i^{*'}\gamma_a}{\hat{\sigma}}\right)$$

Where:

- ψ = The partial derivative for the parameter vector
- γ = The function (ρ)
- γ_a = The vector of minor parameter and it is estimated by GLSM
- $\hat{\sigma}$ = The scale parameter is estimated as the form Huber:

$$\hat{\sigma} = 1.438\left[(\text{median})\left|\delta - \text{median}(\delta)\right|\right]$$

The use of Huber's function is as follows:

$$\rho(\delta) = \begin{cases} \delta/2 \\ f\,|\delta| - f^2/2 \end{cases}$$

$$\psi(\delta) = \begin{cases} \delta & |\delta| \leq f \\ f\,\text{sign}(\delta) & |\delta| \prec f \end{cases}$$

Whereas, f = 1.345

Laplace method (LP): The base of this method depends on minimizing the absolute values of the sum of residuals as follows:

$$\text{Min}\sum_{i=1}^{n} |\delta_i|^g \qquad 1 \leq g \leq 2$$

This method is known as least absolute error, when the value of g = 1. As for when g = 2 it is called ordinary least squares. So, the parameters are represented by (γ_0^L, γ_1^L, γ_2^L).

RESULTS AND DISCUSION

The experiment has been repeated (1000) times and for different sample sizes (n = 20, 40, 60, 80, 100, 120) by using the MATLAB program with correlation coefficients (ϕ = -0.9, -0.5, 0.5, 0.9) and the initial values of parameters are (γ_0 = 0.5, γ_1 = 1, γ_2 = 1). The supposed values of parameters distribution are (b = 1, α = 1 and 2). As for the random variables, they are generated by using this form:

$$\delta = -\text{Ln}\left(u^{-\frac{1}{b}} - 1\right)\alpha$$

where, the values of variables (Y_0) (X_1, X_2) and (X_{10}, X_{20}) are generated with the following forms (Tiku *et al*., 2000; Beach and Machinnon, 1978):

- X_1~Rand
- X_2~Rand
- $X_{10} \sim \text{Rand}/\sqrt{1-\phi^2}$
- $X_{20} \sim \text{Rand}/\sqrt{1-\phi^2}$
- $Y_0 \sim \delta_0/\sqrt{1-\phi^2}$
- (Y_0) and (X_{10}, X_{20}) represent the first observation of variables that has been preserved

Table 1-6, show the values of mean squares error of parameter and for all the methods when the error is distributed as general logistic distribution.

Throughout these table, we notice that for all sample sizes, the increasing of autocorrelation value with shape parameter (b = 1) of distribution. The M method is the best depending on the MSE for parameters when the error is distributed as a general logistic. In second place comes LP and then follows it GLS.

Applied data: The data was obtained from the Ministry of Planning in Iraq/Central Organization for Statistics/Department of Statistics which represents the consumer price index for the years (2004-2016) where, Y is the exchange rate of the dollar, X_1 represents annual inflation rates and X_2 is the ratio of currency to money supply and the data is as follows:

- X_1 = [26.8, 37.1, 53.1, 30.9, 12.7, 8.3, 2.5, 5.6, 6.1, 1.9, 2.2, 1.4, 0.1]
- X_2 = [80.357, 78, 74.833, 70, 69.833, 64, 51, 47.417, 48.667, 48.083, 49.25, 45.167, 56.417]

Table 1: Values of mean square error for parameters and sample size (n = 20, b = 1)

	n = 20			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.8058	0.2795	0.0417	0.0174
γ_1	0.1959	0.1641	0.1459	0.1688
γ_2	0.2036	0.1642	0.1442	0.1701
M				
γ_0	0.7946	0.2765	0.0409	0.0173
γ_1	0.1900	0.1613	0.1426	0.1647
γ_2	0.1987	0.1611	0.1398	0.1652
LP				
γ_0	0.7960	0.2767	0.0410	0.0178
γ_1	0.1890	0.1625	0.1441	0.1645
γ_2	0.1991	0.1614	0.1411	0.1658
2				
GLS				
γ_0	1.8564	0.8847	0.1030	0.0254
γ_1	0.7321	0.6399	0.5765	0.6464
γ_2	0.7646	0.6426	0.5676	0.6518
M				
γ_0	1.8402	0.8759	0.1003	0.0249
γ_1	0.7139	0.6303	0.5634	0.6311
γ_2	0.7492	0.6302	0.5507	0.6347
LP				
γ_0	1.8437	0.8791	0.1014	0.0267
γ_1	0.7108	0.6411	0.5694	0.6363
γ_2	0.7538	0.6279	0.5567	0.6394

Table 2: Values of mean square error for parameters and sample size (n = 40, b = 1)

	n = 40			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.2409	0.0758	0.0173	0.0086
γ_1	0.0443	0.0363	0.0350	0.0415
γ_2	0.0492	0.0399	0.0374	0.0428
M				
γ_0	0.2355	0.0744	0.0171	0.0087
γ_1	0.0432	0.0354	0.0342	0.0403
γ_2	0.0481	0.0392	0.0365	0.0416
LP				
γ_0	0.2375	0.0752	0.0172	0.0088
γ_1	0.0434	0.0359	0.0346	0.0405
γ_2	0.0485	0.0396	0.0370	0.0424
2				
GLS				
γ_0	0.4824	0.2108	0.0328	0.0099
γ_1	0.1632	0.1412	0.1372	0.1564
γ_2	0.1852	0.1559	0.1452	0.1601
M				
γ_0	0.4705	0.2064	0.0322	0.0099
γ_2	0.1591	0.1379	0.1338	0.1519
γ_3	0.1810	0.1530	0.1418	0.1553
LP				
γ_0	0.4755	0.2108	0.0327	0.0102
γ_1	0.1603	0.1404	0.1356	0.1528
γ_2	0.1828	0.1555	0.1435	0.1579

- Y = [1454, 1473, 1477, 1266, 1206, 1183, 1187, 1199, 1234, 1233, 1218, 1251, 1281]

Table 3: Values of mean square error for parameters and sample size (n = 60, b = 1)

	n = 60			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.1384	0.0414	0.0101	0.0059
γ_2	0.0214	0.0167	0.0151	0.0188
γ_2	0.0175	0.0136	0.0150	0.0189
M				
γ_0	0.1357	0.0409	0.0100	0.0060
γ_1	0.0211	0.0164	0.0146	0.0184
γ_2	0.0170	0.0132	0.0145	0.0184
LP				
γ_0	0.1361	0.0409	0.0101	0.0060
γ_1	0.0213	0.0166	0.0149	0.0186
γ_2	0.0172	0.0133	0.0147	0.0183
2				
GLS				
γ_0	0.2502	0.0998	0.0163	0.0064
γ_1	0.0785	0.0651	0.0588	0.0689
γ_2	0.0634	0.0525	0.0582	0.0689
M				
γ_0	0.2442	0.0978	0.0161	0.0064
γ_1	0.0773	0.0638	0.0568	0.0671
γ_2	0.0614	0.0510	0.0564	0.0669
LP				
γ_0	0.2453	0.0977	0.0162	0.0066
γ_1	0.0779	0.0639	0.0581	0.0682
γ_2	0.0621	0.0517	0.0569	0.0672

Table 4: Values of mean square error for parameters and sample size (n = 80, b = 1)

	n = 80			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.0872	0.0254	0.0078	0.0046
γ_1	0.0118	0.0091	0.0089	0.0115
γ_2	0.0117	0.0089	0.0087	0.0112
M				
γ_0	0.0863	0.0252	0.0077	0.0047
γ_1	0.0115	0.0088	0.0086	0.0113
γ_2	0.0115	0.0087	0.0084	0.0108
LP				
γ_0	0.0867	0.0253	0.0077	0.0047
γ_1	0.0115	0.0089	0.0086	0.0114
γ_2	0.0115	0.0087	0.0086	0.0109
2				
GLS				
γ_0	0.1421	0.0565	0.0118	0.0050
γ_1	0.0415	0.0352	0.0339	0.0401
γ_2	0.0413	0.0342	0.0334	0.0394
M				
γ_0	0.1399	0.0554	0.0115	0.0050
γ_1	0.0401	0.0340	0.0329	0.0392
γ_2	0.0403	0.0334	0.0324	0.0379
LP				
γ_0	0.1408	0.0555	0.0117	0.0051
γ_1	0.0405	0.0342	0.0331	0.0397
γ_2	0.0405	0.0334	0.0330	0.0380

Table 5: Values of mean square error for parameters and sample size (n = 100, b = 1)

	n = 100			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.0659	0.0185	0.0059	0.0037
γ_1	0.0074	0.0055	0.0054	0.0070
γ_2	0.0080	0.0059	0.0056	0.0074
M				
γ_0	0.0648	0.0183	0.0059	0.0037
γ_1	0.0073	0.0053	0.0052	0.0069
γ_2	0.0078	0.0057	0.0055	0.0073
LP				
γ_0	0.0649	0.0184	0.0059	0.0037
γ_1	0.0073	0.0054	0.0053	0.0069
γ_2	0.0078	0.0058	0.0055	0.0073
2				
GLS				
γ_0	0.1024	0.0384	0.0082	0.0038
γ_1	0.0251	0.0211	0.0204	0.0241
γ_2	0.0268	0.0221	0.0211	0.0255
M				
γ_0	0.1003	0.0377	0.0082	0.0039
γ_1	0.0244	0.0203	0.0198	0.0234
γ_2	0.0261	0.0216	0.0207	0.0248
LP				
γ_0	0.1006	0.0382	0.0082	0.0039
γ_1	0.0246	0.0204	0.0199	0.0235
γ_2	0.0264	0.0220	0.0211	0.0251

Table 6: Values of mean square error for parameters and sample size (n = 120, b = 1)

	n = 120			
	Φ			
α/methods	-0.9	-0.5	0.5	0.9
1				
GLS				
γ_0	0.0542	0.0151	0.0047	0.0031
γ_1	0.0056	0.0039	0.0039	0.0052
γ_2	0.0059	0.0042	0.0043	0.0056
M				
γ_0	0.0535	0.0149	0.0047	0.0031
γ_1	0.0054	0.0038	0.0038	0.0052
γ_2	0.0057	0.0040	0.0041	0.0055
LP				
γ_0	0.0536	0.0149	0.0047	0.0031
γ_1	0.0055	0.0038	0.0039	0.0052
γ_2	0.0057	0.0041	0.0042	0.0055
2				
GLS				
γ_0	0.0814	0.0295	0.0062	0.0032
γ_1	0.0183	0.0149	0.0147	0.0174
γ_2	0.0196	0.0159	0.0161	0.0190
M				
γ_0	0.0799	0.0288	0.0061	0.0032
γ_1	0.0177	0.0144	0.0144	0.0170
γ_2	0.0187	0.0151	0.0157	0.0185
LP				
γ_0	0.0803	0.0290	0.0062	0.0033
γ_1	0.0178	0.0144	0.0147	0.0172
γ_2	0.0189	0.0154	0.0158	0.0185

The data has been tested after taking the standard degree (Z-score) for the dependent variable Y and it seems that has a general logistic distribution where the value of (D = 0.272), the tabulated value with (0.05), significance level is (0.3772)and from the Durbin-Watson table with sample size (n = 13), we find the lower values for the testing (d_L = 0.861) and the higher value was (d_U =1.562) and since, the obtained value for testing (D.W = 0.5) which falls under the interval (0<D.W<d_L), so, this indicates that there is an autocorrelation problem. Therefore, we estimated the correlation coefficient equal to (0.75) where zero was given for the first observations (Y_0) and (X_{10}, X_{20}). The values of parameters of multiple linear regression by using M Robust with the presence of autocorrelation $\hat{\gamma}_0 = 0.0977, \hat{\gamma}_1 = 0.8793, \hat{\gamma}_2 = 16.2061$ knowing that the data was divided by (1000) for simplification.

CONCLUSION

In this study, we have shown the comparison of estimate methods of multiple linear regression models with auto correlated errors and when the error is distributed with general logistic. Table, show that M Robust Method is the best when the auto correlated coefficient increases with the shape parameter (b = 1) and for different sample sizes when the error term distributes general logistic. Also, we indicate from the value of MSE for the parameter decreases when the sample size increases. The value of MSE for the parameters decreases when the value of positive auto correlation coefficient increases (ϕ = 0.5, 0.9) with respect to the parameters of the variables X_1, X_2. Moreover, the value of MSE for the parameters increases when the value of negative auto correlation coefficient increases for different sample sizes and for all values of (α). Finally, as a real application we have concluded that each of the annual inflation rates and the ratio of currency to the money supply has a direct impact on the dollar exchange rates.

REFRENCES

Ahmed, A., 2015. Comparison between some of estimation methods of multiple linear regression model with auto correlated errors. J. Econ. Administrative Sci., 21: 675-686.

Ayinde, K., A.F. Lukman and O.T. Arowolo, 2015. Combined parameters estimation methods of linear regression model with multicollinearity and autocorrelation. J. Asian Sci. Res., 5: 243-250.

Beach, C.M. and J.S. Mackinnon, 1978. A maximum likelihood procedure for regression with autocorrelated errors. Econometrica, 46: 51-58.

Chib, S., 1993. Bayes regression with autoregressive errors: A Gibbs sampling approach. J. Econ., 58: 275-294.

Dietz, J., 1986. A comparison of robust estimators in simple linear regression. Master Thesis, Institute of Statistics Mimeograph Series, North Caroline State University, Raleigh, North Carolina.

Tiku, M.L., W.K. Wong, D.C. Vaughan and G. Bian, 2000. Time series models in non-normal situations: Symmetric innovations. J. Time Ser. Anal., 21: 571-596.