Websites:
http://www.sciencepub.net
http://www.sciencepub.net/report

Emails:
editor@sciencepub.net
reportopinion@gmail.com

Report and Opinion

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

Missing observations is one of the problems often encountered in data analysis

Yem Funky

yemfunky@ymail.com

Abstract: Missing observations is one of the problems often encountered in data analysis. The remaining observations are either used or the missing ones omitted before the analysis is done. In the research, four estimators of linear regression model (OLSE, PRWE, COCR and HILU) with autocorrelated error terms of order one (AR(1)) were studied and the effect of missing observations on them was determined. The mid-range measure of location was adapted to estimating values and its performances were compared with the existing ones namely; Amelia, Hmisc, Arithemetic Mean, Median, Mice and Mid-Range. Monte Carlo experiments were conducted 1000 times on a linear regression model with three explanatory variables having autocorrelation values of 0.8, 0.9, 0.95, 0.99, 0.999 and error variance of 4, 25 and 100 at five sample sizes namely; 15, 20, 30, 40 and 50. The study considered situation where there 1, 2, 3 and 4 missing observations on the dependent variable. The estimators were compared using the bias and mean square error criteria. Results show that the PWRE and COCR estimator are more efficient when there are missing observations or no missing observation. Replacing the missing observations with the estimated ones, the research identified the arithmetic mean, median and mid-range methods with PWRE and COCR estimator as better estimators. The results from real life data set also support these findings.

[Yem Funky. Idrus. **Missing observations is one of the problems often encountered in data analysis**. *Rep Opinion* 2022;14(3):24-36]. ISSN1553-9873 (print);ISSN 2375-7205 (online). http://www.sciencepub.net/report. 4. doi:10.7537/marsroj140322.04.

Keywords: observation; problem; encountered; data; analysis

I. INTRODUCTION

One of the serious problems faced in data collection and analysis is missing observations. Nature does not always provide a complete data set, hence, the mechanism for observing time series may be imperfect, equipment failure, and human error or the disregarding of inaccurate measurements can introduce missing values. In statistics, analysis is expected to be carried out with complete observations. Where any observation is missing either by natural or human error, the missing values are either eliminated or estimated before the analysis is carried out.

Analysis of time series data constitutes an important area of statistics especially in identifying the nature of the phenomenon represented by the sequence. However, missing observations in time series data are very common (David, 2006). This happens when an observation may not be made at a particular time, due to faulty equipment, lost records or a mistake, which cannot be rectified until later. When this happens, it is necessary to obtain estimates of the missing value for better understanding of the nature of the data and make possible a more accurate forecast (Howell, 2007).

Most estimators can only be derived in a "large sample" context (asymptotic properties).For example, one estimation procedure may be chosen over another because it is known to provide consistent and asymptotically reliable parameter estimates in certain stochastic environments. When sample sizes are small and asymptotic formulae do not accurately represent sampling behavior, such reliance on asymptotic theory can lead to serious bias problems and low levels of inferential accuracy.

Most practitioners had access to ordinary least square (OLS) software packages for estimating regression parameters in the 1960s, 1970s and early 1980s, but not to nonlinear packages. This prompted researchers to devise clever analytical and /or iterative procedures that would allow practitioners to obtain results similar to those obtained by nonlinear procedures using OLS. As a result, there are numerous methods for estimating time series regression parameters. The use of each depends on assumptions existing in the model. For example, the error terms in time series analysis follow autoregressive of order one (AR (1)). Also, Cochrane Orcutt and PraisWinsten are among the estimators that become relevant.

In all domains of quantitative research, such as economics, medical, environment, life sciences and social sciences, the existence of missing values is unavoidable (Takahashi & Ito 2013). It has been proven that ignoring missing values can lead to biased estimates and incorrect conclusions (Guan & Yusoff, 2011). In other words, an insufficient approach to handling missing data in a statistical analysis will result in incorrect estimates and inferences. In general, there are ways of dealing with missing values known as traditional or modern approaches. Some commonly used traditional ways are listwise deletion and pairwise deletion. For imputation methods: mean imputation, hot-deck imputation, and stochastic imputation are some of the most commonly used (George et al. 2015). The modern approaches include those based on maximum likelihood and multiple imputations (Acock 2005). Studies on handling missing values are largely for univariate or regression model data and the regression model parameters are the most popular.

The most common method for estimating regression model parameters is to use the Ordinary Least Squares (OLS) estimator. For example, in a model where the observations occur at different times, the OLS estimates are far from their true values. Multicollinearity, auto correlated error terms, outliers, or their combined presence caused the OLS estimator to produce inaccurate estimates. Various authors have concentrated on the specific effects of each of these issues. The combined effect of multicolinearity and auto correlated error terms, for example, has gotten a lot of attention. Both developed a ridge estimator based on GLS and demonstrated its effectiveness. In the presence of AR (1) errors, we found an expression for the mean squared error (MSE) of ridge estimators. Collinearity and autocorrelation interact to inflate the MSE of OLS and ridge estimators, as shown in this expression. Provided adaptive ridge estimators based on generalized least squares (GLS) for regression problems with collinear independent variables and auto correlated errors, which were compared to the traditional OLS estimator. When the independent variables are related and serially correlated, GLS-based methods are the best. For multiple linear models that suffer from both autocorrelation AR (1) and multicollinearity, the two Stages Ridge Estimator was proposed.

The autoregressive (AR) models are simple and widely used for modeling the generation of time series. It assumes that each observation in the time series is a noisy linear combination of some previous observations plus a constant shift. The simplest form is autoregressive of order one (AR (1)) in which the error term at present is related to the previous. This often common in econometric time series data. Consequently, it intended to estimate missing value in this study and determine its effects on estimator of regression model with AR (1).

In the presence of first-order autoregressive disturbances, AR(1), and irrelevant measurement artifacts, we discuss the robust estimation of a linear trend from a noisy time series $Y_1, Y_2, \ldots, Y_n$ of moderate size n. (outliers). The model is as follows:

$$Y_t = \mu + \beta(t - m - 1) + \varepsilon_t$$
$$\varepsilon_t = \varphi\varepsilon_{t-1} + \delta_t \qquad (1.2)$$

where $t = 1, \ldots, n;\quad n = 2m + 1$ and δ_t is the number of innovations produced by a white noise process with a mean of zero and a variance of $\sigma^2 > 0$. We interpret as the central level by centering time and assuming stationary errors, $|\varphi| < 1$. In econometrics, the estimation of linear trends in the presence of AR(1) errors has gotten a lot of attention. A number of papers, for example, compare the efficiencies of the ordinary least squares (OLS), generalized least squares (GLS), first differences (FD), Cochrane-Orcutt (CO), and Prais-Winsten (PW) estimators for estimating the slope β, often under the idealized assumption that φ is known (see Kramer, 1980, 1982, Steman and Trenkler, 2000, and the references cited therein). Methods based on least squares, on the other hand, are extremely susceptible to outlier contamination. Simple, reliable alternatives become more appealing as a result of this. Davies, Fried, and Gather investigated the robust fitting of linear trends to data within a moving time window of moderate length n recently (2004). The central level μ is of primary importance in this context. They find Siegel's (1982) repeated median (RM) to be very suitable for trend estimation due to its robustness, stability, and computational tractability, based on a comparison of robust regression techniques. The ordinary repeated median, on the other hand, treats the data as independent, despite the fact that autocorrelations can produce monotonic data patterns that resemble time-varying trends.

In this paper, we look at whether the repeated median can be improved when there is AR(1) noise. Simultaneous estimation of the autoregressive and trend parameters using robust regression is one option. Rousseeuw and Leroy (1987) and Meintanis and Donatos (1988) proposed robust regression techniques for fitting AR models to data with a constant level (1999). Preliminary estimation of followed by trend estimation from transformed data is another approach.

METHODS.

In this chapter, the process of analyzing the data to provide insight into the phenomenon of using another approach to determine effect of missing value estimation methods on estimators of AR (1) model were discussed in line with the aim and objectives of the study.

3.1 The proposed method of estimating missing observations

Mid-range Approach: Mid-range determines the number that is halfway between the minimum and maximum numbers of a data set. It is a statistical tool that identifies a measure of center like mean, media or mode. The formula is given as:

$$Mid - range = \frac{X_{max} + X_{min}}{2} \qquad (3.1)$$

3.2 Simulation Study

The linear regression model with three explanatory variables and error terms following AR (1) is used for simulation study. This is given as:

$$y_i = \beta_0 + \beta_{1i} x_i + \beta_{2i} x_i + \beta_{3i} x_i + \mu_i \quad (3.2)$$

where $\mu_i = \square\mu_{i-1} + \square_i$

The ε_i was generated to follow normal distribution as $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma^2 = 4, 25, 100$ and the error terms of the model was generated following the idea of Ayinde (2008) and Ayinde and Olaomi (2008) assuming the process start from infinite past and continue to operate. That is,

$$\mu_1 = \frac{\varepsilon 1}{\sqrt{1-\rho 2}} \quad (3.3)$$

and $\mu_1 = \square\mu_{i-1} + \varepsilon_i$, i=2, 3, …, n (=15, 20, 30, 40, 50)

Each of the regressors was generated as: $x_i \sim N(0, 1)$ and the regression coefficients were taken as (0, 0.1, 0.5, 0.9). The autocorrelation values were be taken as 0.8, 0.9, 0.95, 0.99, and0.999. The value of the dependent variables were then determined and the experiments perform 1000 times.

Missing values (1 missing, 2 missing, 3 missing and 4 missing) were created in the dependent variable for the five sample sizes as shown in Table 3.

Table 3.1: Missing observations at various sample size in the study

s/n	Sample Size (n)				
	15	20	30	40	50
1					
2					
3	Missing				
4		Missing			
5					
6	Missing		Missing		
7					
8		Missing		Missing	
9	Missing				
10					
11					
12	Missing	Missing	Missing		Missing
13					
14					
15					
16		Missing		Missing	
17					
18			Missing		
19					
20					
21					
22					
23					
24			Missing	Missing	Missing
25					
26					
27					
28					
29					
SS30					
31					
32				Missing	
33					
34					
35					
36					Missing
37					
38					

39					
40					
41					
42					
43					
44					
45					
46					
47					
48					Missing
49					
50					

3.3 Methods of Missing value used in the Study

i. Amelia
ii. Hmisc
iii. Arithmetic mean:
iv. Median
v. Mice
vi. Mid-range

3.4 Estimators Used in the Study

The following estimators of AR (1) were used in the study

i. Ordinary Least Square Estimator by Friedrich Gauss (1795)
ii. Cochrane Orcutt Estimator by D.Cochrane and G.orcutt (1949)
iii. Prais Winsten Estimator by Winsten (1954)
iv. Hildreth Lu Estimator by John Lu(1960)

The above estimation methods were be applied to the simulated data when there was no missing value, when there was 1 missing value, 2 missing values, 3 missing values and 4 missing values; and furthermore, when the missing values were estimated with the techniques outlined in 3.3.

3.5 Criteria for comparison of estimators

Bias $(\hat{\beta}) = \left| \frac{1}{1000}\sum_{j=1}^{1000} \sum_{i=1}^{2}(\hat{\beta}_{ij} - \beta_i) \right|$,(Bias closest to zero)

MSE $(\hat{\beta}) = \frac{1}{1000}\sum_{j-1}^{1000} \sum_{i=1}^{2}(\hat{\beta}_{ij} - \beta_i)^2$

Having estimated the criteria above for each of the estimators, then, at each level of autocorrelation, error variance and sample size, all the estimators were ranked. The number of times each estimator is best (rank first) was then counted over the levels of autocorrelation (5) and error terms (3). Thus, a total of fifteen (15) counts are expected. An estimator with the highest frequent is considered to

3.6 Application to real life Data

A data set with autocorrelation problem was be used. Missing values were created following the idea in the simulation study. The estimators were compared using the Adjusted co-efficient of determination.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.0 Introduction

In this chapter, the results of our investigation and findings are discussed. In section 4.1, results showing performances of the estimator at different levels of autocorrelation when there is no missing value are presented and discussed. In section 4.2, the performances of the estimators with missing values are presented and discussed. In section 4. 3, the missing values were estimated and replaced and the estimators were then used for estimations and the results compared.

4.1: Results When There No Missing Value

The sample of simulated results when n= 15 is presented in Table 4.1 From Table 4.1, it can be seen that the biases of the estimators are very close to zero except for the OLS estimator when the autocorrelation level is very close to 1 and error variance is large. Moreover, the OLS estimator is known to unbiased in the present of autocorrelation.

Furthermore, it can be seen that the most efficient estimator is either the PRWE or COCR. Having counted the number of time each estimator has closest bias to zero and minimum MSE over the levels of autocorrelation and error variance, Table 4.2 was obtained.

From Table 4.2, it becomes clearer that the most efficient estimator is either PRWE or COCR estimator. Moreover; the HILU estimator does compete in term of bias the PRWE and COCR estimator.

Table 4.1: Bias and Mean Square Errors Of The Estimators When Sample Size (N) = 15

		4		25		100	
AUTOCORRELATION VALUE	ESTIMATORS	BIAS	MSE	BIAS	MSE	BIAS	MSE
0.8	OLS	0.0126466	2.6245	0.0316165	16.4034	0.063233	65.6137
	PRW	0.0286749	1.0194	0.0716873	6.3711	0.1433746	25.4842
	COC	0.0233685	1.0445	0.0458075	6.0569	0.0756956	25.926
	HIL	0.0056333	6.1119	0.0140832	38.1992	0.0281663	152.7966
0.9	OLS	0.0851593	4.076	0.2128983	25.475	0.4257966	101.9001
	PRW	0.0128821	0.9409	0.0322052	5.8809	0.0644105	23.5237
	COC	0.0075995	0.9694	0.0298416	6.0374	0.068413	24.1419
	HIL	0.1535423	11.3091	0.3838557	70.6816	0.7677114	282.7263
0.95	OLS	0.0518346	6.4468	0.1893135	43.6869	0.378627	174.7475
	PRW	0.0226275	0.8271	0.0113946	5.7376	0.0227891	22.9504
	COC	0.0354493	0.84	0.0264328	6.6534	0.0402514	23.1802
	HIL	0.003002	19.6856	0.0330642	133.0802	0.0661284	532.3209
0.99	OLS	0.0539855	32.2796	0.1349638	201.7475	0.2699275	806.9898
	PRW	0.01324	0.8251	0.0331	5.1571	0.0662001	20.6283
	COC	0.0168035	0.8301	0.0455327	5.1737	0.0938757	20.6846
	HIL	0.0468575	105.8797	0.1171439	661.7479	0.2342877	2646.9916
0.999	OLS	0.7696498	309.1603	1.9764535	1899.6917	3.3353271	6786.7315
	PRW	0.0126483	0.843	0.0228359	5.1729	0.0076086	18.9007
	COC	0.0100016	0.8384	0.0168484	5.139	0.0050401	18.7633
	HIL	0.0601007	1023.8692	0.1586344	6288.2485	0.4787131	22466.7809

TABLE 4.2: Frequency of Estimators with minimum Bias and MSE When counted Over levels of Autocorrelation and Error Variance: No missing value

Sample Size	Estimator	Bias	MSE
15	OLSE	0	0
	PRWE	6	11
	COCR	5	4
	HILU	4	0
20	OLSE	4	0
	PRWE	7	10
	COCR	4	5
	HILU	0	0
30	OLSE	2	0
	PRWE	4	5
	COCR	6	10
	HILU	3	0
40	OLSE	0	0
	PRWE	3	6
	COCR	7	9
	HILU	5	0
50	OLSE	3	0
	PRWE	2	15
	COCR	10	0
	HILU	0	0

4.2: Results When There are Missing Values

The summary of the performances of the estimators at different sample sizes and the number of missing observation is presented in Table 4.3. This is further expressed graphically in Figure 4.1A and 4.1B.

From these, it can be observed that even though the estimators competes in term of bias the performances of the OLSE estimator gets better with increased number of missing values. The PRWE and COCR estimator are generally more efficient than the OLSE and HILU.

TABLE 4.3: Frequency of Estimators with minimum Bias and MSE When counted Over levels of Autocorrelation and Error Variance: Missing values

		BIAS					MSE				
		0	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
15	OLSE	0	0	2	1	7	0	0	0	0	0
	PRWE	6	2	0	8	3	11	10	1	1	10
	COCR	5	7	8	6	4	4	5	14	14	5
	HILU	4	6	4	0	1	0	0	0	0	0
20	OLSE	4	8	0	3	3	0	0	0	0	0
	PRWE	7	6	8	8	7	10	7	14	9	0
	COCR	4	0	7	4	2	5	8	1	6	15
	HILU	0	1	0	0	3	0	0	0	0	0
30	OLSE	2	2	0	1	4	0	0	0	0	0
	PRWE	4	4	5	4	5	5	11	9	11	10
	COCR	6	9	10	10	6	10	4	6	4	5
	HILU	3	0	0	0	0	0	0	0	0	0
40	OLSE	0	3	1	0	3	0	0	0	0	0
	PRWE	3	4	9	11	3	6	11	4	12	8
	COCR	7	6	5	4	7	9	4	11	3	7
	HILU	5	2	0	0	2	0	0	0	0	0
50	OLSE	3	3	1	2	2	0	0	0	0	0
	PRWE	2	2	6	4	4	15	15	7	10	10
	COCR	10	10	6	9	8	0	0	8	5	5
	HILU	0	0	2	0	1	0	0	0	0	0

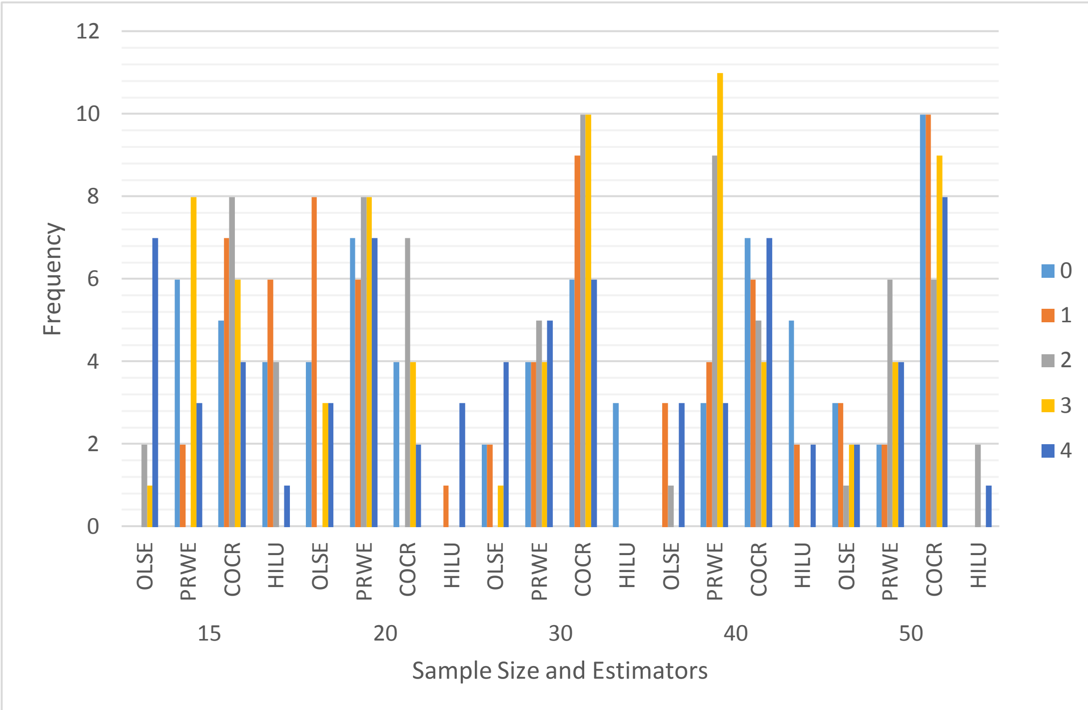

FIGURE 4.1A: Frequency of Estimators with minimum Bias When counted Over levels of Autocorrelation and Error Variance: Missing values

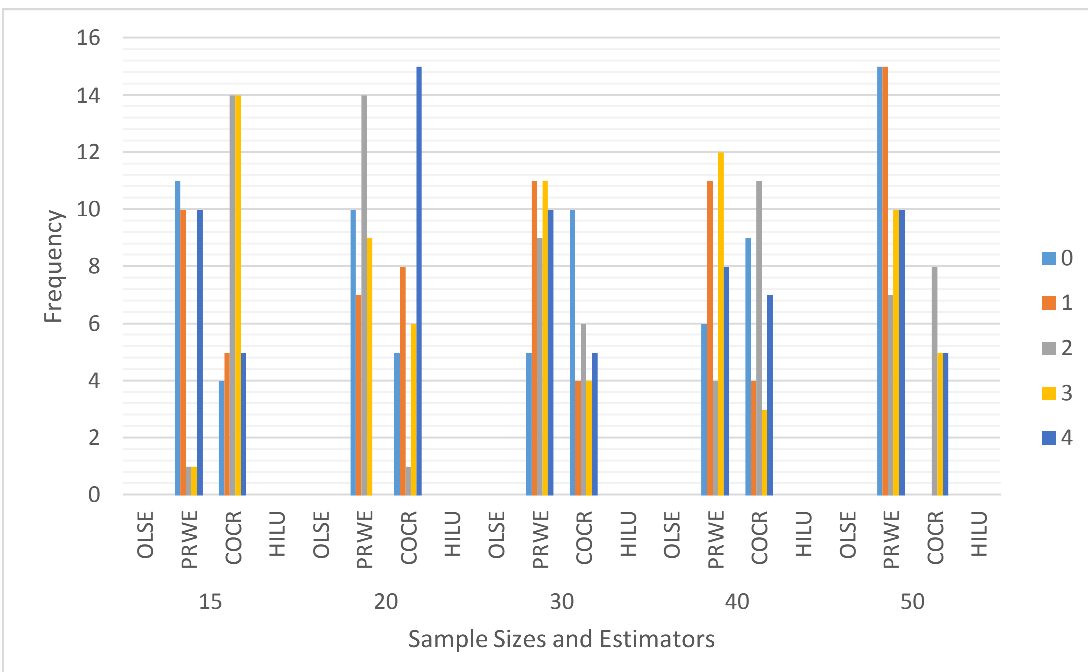

FIGURE 4.1B: Frequency of Estimators with minimum MSE When counted Over levels of Autocorrelation and Error Variance: Missing values

4.3: Results When Missing Values are replaced

The results of the performances of the estimators and the methods of estimating missing values are presented in Figure 4.1, 4.2, 4.3 and 4.4

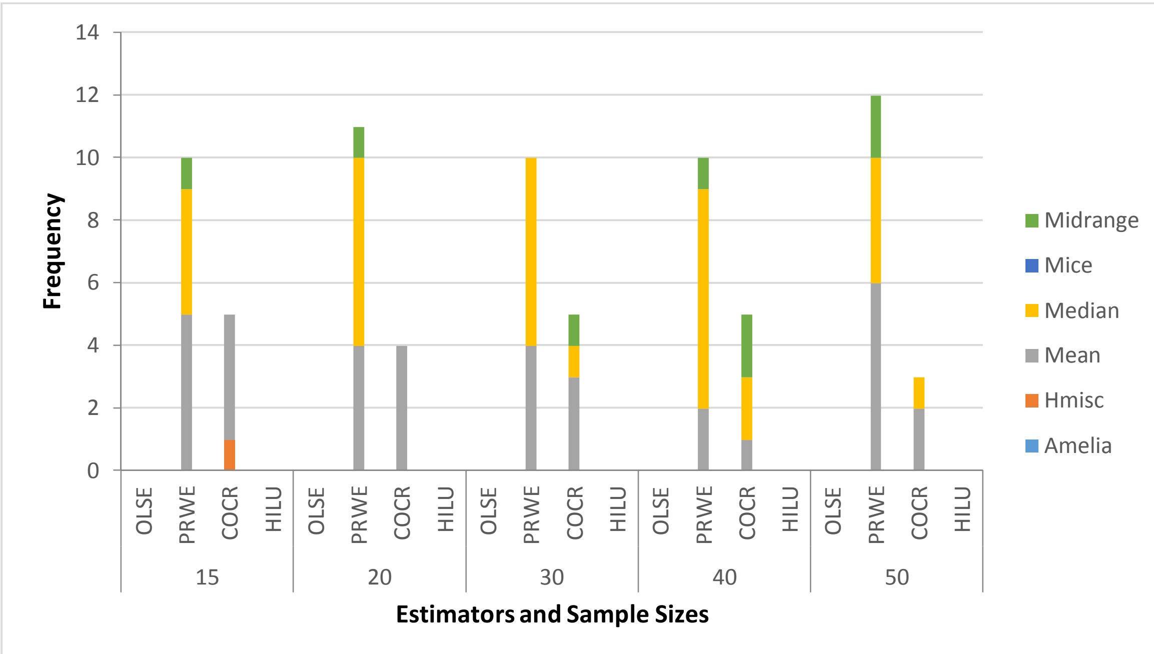

FIGURE 4.2A: Frequency of minimum MSE of the Estimators with Methods of Missing value estimation When counted Over levels of Autocorrelation and Error Variance: One Missing Value

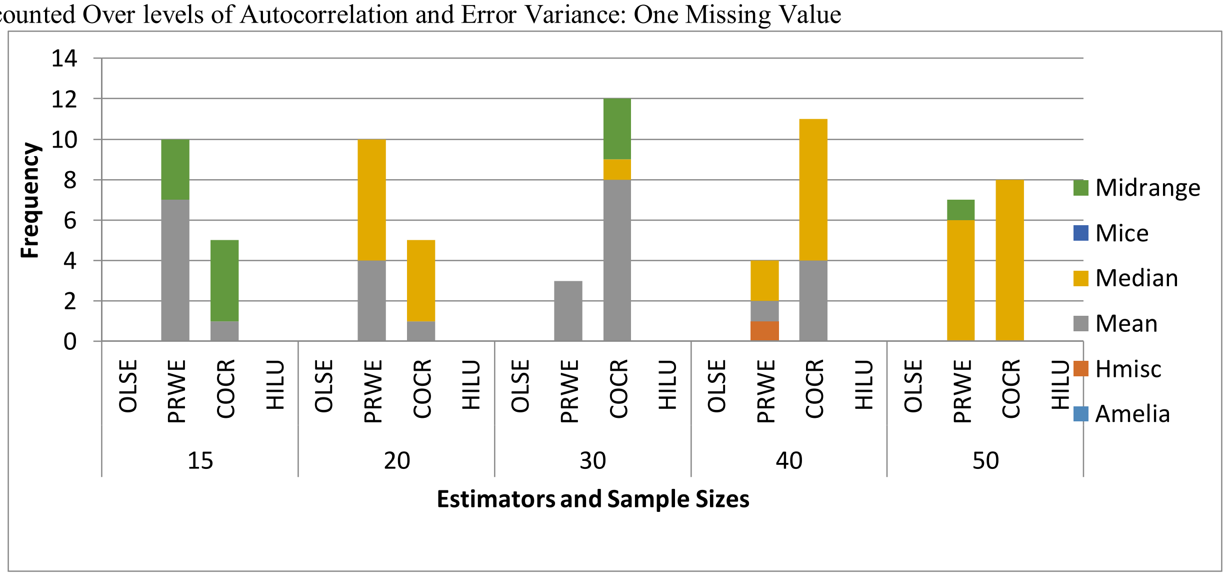

FIGURE 4.2B: Frequency of minimum MSE of the Estimators with Methods of Missing value estimation When counted Over levels of Autocorrelation and Error Variance: Two Missing Values

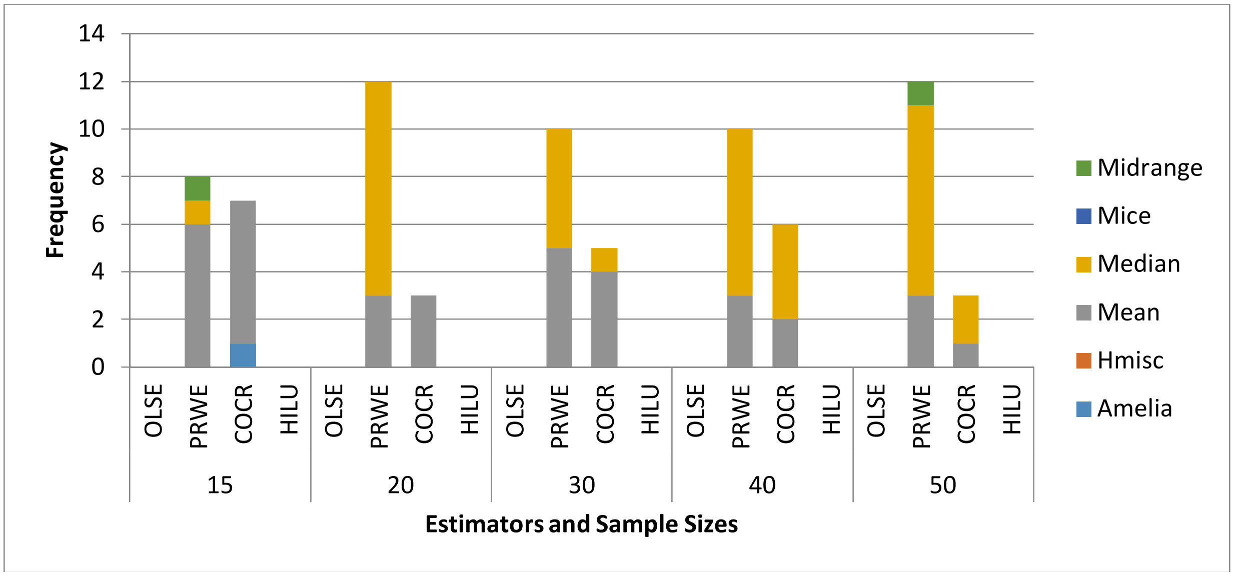

FIGURE 4.2C: Frequency of minimum MSE of the Estimators with Methods of Missing value estimation When counted Over levels of Autocorrelation and Error Variance: Three Missing Values

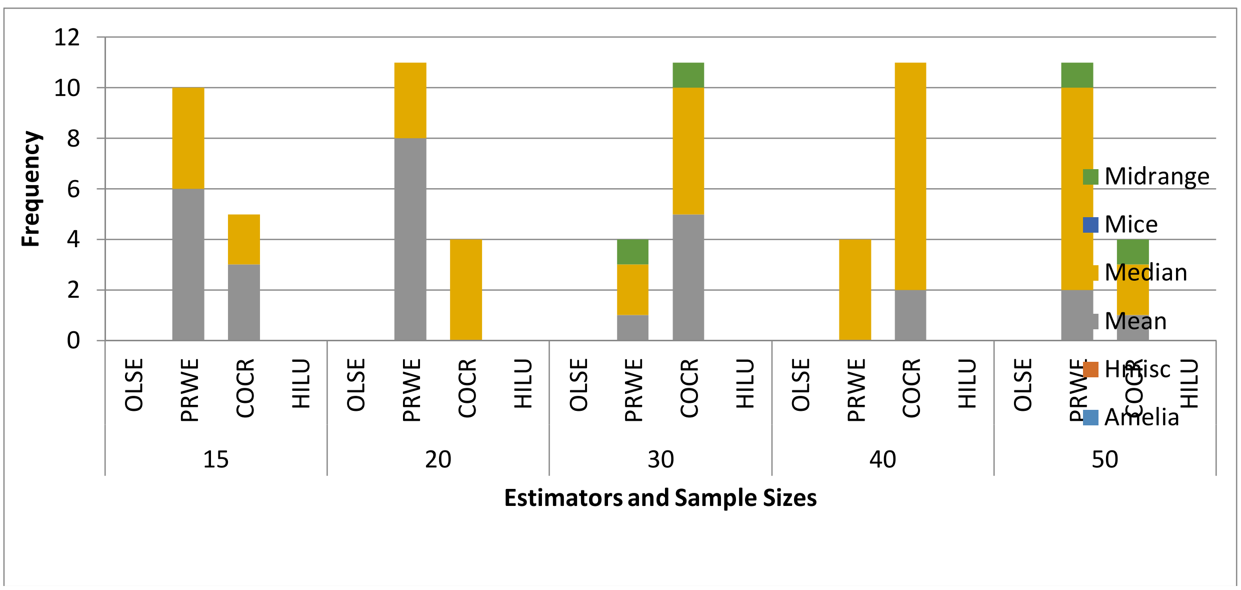

FIGURE 4.2D: Frequency of minimum MSE of the Estimators with Methods of Missing value estimation When counted Over levels of Autocorrelation and Error Variance: Four Missing Values

The figures clearly suggest the use of Arithmetic mean, median and less frequently, the proposed mid-range approach with PRWE or COCR estimator for estimation after replacement.

4.4: Application to Real Life Data Set

This dataset covered the products in the manufacturing sector of Iraq in the period of 1960 to 1990. It has been used by Hussein and Abdalla (2012) and it covered variables on product value in the manufacturing sector**(Y),** value of imported intermediate (X_1), imported capital commodities (X_2) and value of imported raw materials (X_3). The results of the analysis with the estimators when there is no missing observation and when the two missing observations invoked and replaced after their estimation are presented in Table 4.4. From the table, it can be summarily observed that using median to estimate missing value and using the PRWE is best.

TABLE 4.4: Results from real life data analysis

No. of Missing value	Estimators	C	X1	X2	X3	Other Statistics
0	OLSE	208.885 (42.9851)	0.612954 (0.654327)	1.25626 (0.271925)	-1.22126 (1.50334)	Adj. R2= 0.988484 DW= 0.904749
	PRWE	333.597 (143.786)	0.501528 (0.389629)	1.15766 (0.231033)	-1.43049 (0.940812)	Adj. R2= 0.992865 DW= 2.043961
	CORC	603.138 (248.741)	0.437879 (0.373264)	1.12239 (0.22204)	-1.4042 (0.898111)	Adj. R2= 0.993194 DW= 2.146318
	HILU	610.048 (254.096)	0.43704 (0.372892)	1.12124 (0.221882)	-1.40271 (0.897195)	Adj. R2= 0.993194 DW= 2.148401
2 replaced With AM	OLSE	336.517 (97.2732)	0.358215 (1.48071)	1.11515 (0.615352)	0.0975355 (3.40199)	Adj.R2= 0.938122 DW= 2.027536
	PRWE	337.957 (94.5313)	0.335769 (1.49471)	1.13920 (0.603578)	0.013522 (3.4228)	Adj.R2= 0.938228 DW= 1.953999
	CORC	356.776 (97.5827)	0.252294 (1.50587)	1.17665 (0.608323)	-0.0291947 (3.44130)	Adj.R2= 0.937483 DW= 2.009507
	HILU	356.776 (97.5827)	0.252294 (1.50587)	1.17664 (0.608323)	-0.0291932 (3.44130)	Adj.R2= 0.937483 DW= 2.009508
2 replaced With Median	OLSE	290.668 (69.8220)	0.459775 (1.06284)	1.17415 (0.441695)	-0.462568 (2.44192)	Adj.R2= 0.968322 DW= 1.817586
	PRWE	289.883 (72.2049)	0.481429 (1.04815)	1.15398 (0.450599)	-0.406537 (2.41809)	Adj.R2= 0.968395 DW= 1.906611
	CORC	308.345 (74.6763)	0.410028 (1.05039)	1.18648 (0.452302)	-0.453849 (2.41835)	Adj.R2= 0.968388 DW= 1.982154
	HILU	308.345 (74.6774)	0.410036 (1.05039)	1.18647 (0.452305)	0.45383 (2.41834)	Adj.R2= 0.968388 DW= 1.982187
2 replaced With Mid-range	OLSE	346.010 (129.042)	0.695165 (1.96430)	1.39905 (0.816323)	-2.86220 (4.51306)	Adj.R2= 0.895662 DW= 2.166209
	PRWE	347.477 (120.398)	0.695596 (2.00087)	1.43418 (0.776501)	-3.0832 (4.56211)	Adj.R2= 0.896721 DW= 1.991374
	CORC	365.096 (124.625)	0.611666 (2.02609)	1.47082 (0.786911)	-3.11549 (4.61084)	Adj.R2= 0.894514 DW= 2.025356
	HILU	365.096 (124.625)	0.611666 (2.02609)	1.47082 (0.786911)	-3.11549 (4.61084)	Adj.R2= 0.894514 DW= 2.025356

CHAPTER FIVE
RESULTS AND DISCUSSION
CONCLUSION AND RECOMMENDATION
Conclusion and Recommendations

Out of the four estimators of linear regression model with autocorrelation error terms considered, PWRE and COCR estimator are more efficient. Replacing the missing observations with the estimated ones by the method of arithmetic mean, median and mid-range with PWRE and COCR estimator provide better results.

REFERENCES

1. A.F. Siegel. Robust regression using repeated medians. Biometrika. 69:242-244, 1982.
2. Acock, A.C.2005. Working with missing values. Journal of marriage and Family 67:1012- 1028

3. Adewale F. Lukman , Kayode Ayinde, Sek Siok Kun,and Emmanuel T. Adewuyi (2019). A Modified New Two-Parameter Estimator in a Linear Regression Model, https://doi.org/10.1155/2019/6342702
4. Almed, M.R. & Al-Khazaleh, A.M.H. (2008) Estimation of Missing Data by Using the Filtering Process in a Time Series Modeling.
5. Ali, D. A and Midi, H. (2020), On the Robust Parameter Estimation Method for Linear Model with Autocorrelated Errors in the Presence of High Leverage Points and Outliers in the Y-Direction Malaysian Journal of Mathematical Sciences 14(3): 505-517.
6. Ayinde, K. and Ipinyomi, R.A (2007): A Comparative study of the OLS and some GLS estimators when normally distributed Regressors are stochastic. Trends in Applied Science Research 2 (4):354-359.ISSN 1819-3579.
7. Ayinde, K. (2008): Performances of Some Estimators of Linear Model when Stochastic Regressors are correlated with Autocorrelated Error Terms. European Journal of Scientific Research, 20 (3):558-571.
8. Ayinde, K. and Olaomi, J. O. (2008): A Study of Robustness of Some Estimators of Linear Model with Autocorrelated Error Terms When Stochastic Regressors Are Normally Distributed. Journal of Modern Applied Statistical Methods, 7(1):246- 252.
9. Ayinde, K. and Iyaniwura, J. O. (2008): A Comparative Studyof the Performances of Some Estimators of Linear Model with Fixed and Stochastic Regressors. Global Journal of Pure and Applied Sciences Vol 14, No3, 2008: 363-369
10. Ayinde, K. Emmanuel O. Apata , Oluwayemisi O. Alaba (2012): Estimators of Linear Regression Model and Prediction under Some Assumptions Violation. Open Journal of Statistics, 2012, 2, 534-546
11. Bar-Joseph, Z. & Gerber, G. K., Gifford, D. K., Jaakkola, T. S., & Simon, I. (2003). Continuous representations of time-series gene expression data. Journal of Computational Biology, 10(3-4):341–356, 2003.
12. Box, G.E., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). Time series analysis: forecasting and control. John Wiley & Sons. ISBN 978-1-118-67502-1
13. Burke, S. P., Godfrey, L. G. and Tremayne, A. R. (1987). Testing AR (l) Against MA (l) Disturbances in the Linear Regression Model: An Alternative Procedure, Review of Economic Studies (1990) 57, 135-145.
14. Daniel, P. and George, C. T. (1991) A Note on Likelihood Estimation of Missing Values in Time Series, The American Statistician, 45:3, 212-213, http://dx.doi.org/10.1080/00031305.1991.10475804
15. Daniel, L. T. (1987): A Note on the Efficiency of the Cochrane-Orcutt Estimator of the AR (1) Regression Model. Journal of Econometrics 36 (1987) 369-376. North-Holland
16. David, S.C.F. (2006) Methods for the Estimation of Missing Values in Time Series. Cowan University Press, Western Australia.
17. Davies, P.L. Fried, R. and Gather, U. (2004). Robust signal extraction for on-line monitoring data. J. Statistical Planning and Inference. 122:65-78.
18. Formby, B., Hill, R.C. and Johnson, S. R.(1984). “Advance Econometric Methods,” Springer-Verlag, New York, Berlin, Heidelberg, London, Paris and Tokyo, 1984.
19. Fried, R. and Gather, U. (2004). Robust Trend Estimation for AR (1) Disturbances Technical Report, No. 2004, 64
20. Godfrey, L. G. (1987), "Discriminating Between Autocorrelation and Misspecification in Regression Analysis: An Alternative Strategy", Review of Economics and Statistics, 69, 128-134.
21. George, N.I., Bowyer, J.F., Crabtree, N.M. & Chang, C.W. 2015. An iterative leave-one- out approach to outlier detection in RNA-Seq data. Plos One 10(6): e0125224. Doi: 10.1371/ journal. pone.0125224G.
22. Gottman, J. M. (1981). Time-series analysis: A comprehensive introduction for social scientists, volume 400.Cambridge University Press Cambridge.
23. Guan, N.C. & Yusoff, N.S.B. 2011. Missing values in data anaalysis: Ignore or Impute? Education in Medicin Journal 3(1): 6-11.
24. Howell, D.C. (2007). The Analysis of Missing Data. In: Outhwaite, W. and Turner, S., Eds., Handbook of Social Science Methodology, Sage, London. https://doi.org/10.4135/9781848607958.n11
25. Housila,P.S. and Suryal, K.P. (2016): A Modified Procedure for Estimating the Population Mean in Two-occasion Successive Samplings. Afrika Statistika Vol. 12 (3), 2017, pages 1347–1365. DOI: http://dx.doi.org/10.16929/as/2017.1347.108

26. Hyun, K. (2013) The prevention and handling of the missing data. Korean journal of Anesthesiology, 64(5)
27. Ifederu, A. (2006).Estimation of the parameters of Single Linear Regression Models with autocorrelated error terms which are also correlated with the trended regressor. Unpublished M.sc Thesis, University of Ibadan, Nigeria.
28. Ibrahim, J. G. (1990), "Incomplete data in generalized linear models". J. Am. Statist. Ass. 85, 765-769
29. Ibrahim, J. G. and Lipsitz, S. R. (1996), "Parameter estimation from incomplete data in binomial regression when the missing data mechanism is non-ignorable". Biometrics, 52, 1071-1078
30. Kramer, W. (1980). Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. J. Amer. Statist. Assoc. 75:1005- 1009.
31. Kramer,W.(1982). Note on estimating linear trend when residuals are autocorrelated. Econometrica 50:1065-1067.
32. Kramer. W. and F.Marnol, 2002. OLS- based asymptotic inference in linear regression models with trending regressors and AR (P) Disturbances.Commun.Stat.Theory Methods, 31:261-270
33. Kramer, W., 1998.Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares with trending Regressors and stationary Autoregression disturbances In: Econometric in Theory and Practices. Galata Kutchenhoff (Eds.).festschriff for Han schneewei β, pp: 137-140
34. Little, R. J. A. and Rubin, D. B. (2002). Statistical analysis with missing data, 2nd Ed. Hoboken, NJ: Wiley.
35. Luengo, J. (2011). Missing values in data mining (online) *Available at: http://sci2s.ugr.es/MVDM/index.php*
36. Lukman, A. A. and Ayinde, K. (2019). Developing a new estimators in linear regression Model 3rd International Conference on Science and Sustainable Development (ICSSD 2019) Journal of Physics: Conf. Series 1299 (2019) 012128 IOP. Publishing doi:10.1088/1742-6596/1299/1/012128
37. Maeshiro, A, 1976. Autoregressive transformations, trended independent variable and autocorrelated disturbance terms. Rev.Econ.Stat, 58: 497-500.
38. Mizon, G. E. And Hendry, D. F. (1980), "An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification", Review of Economic Studies, 47, 21-45.
39. Maddala, G. S. (2002). "Introduction to Econometrics," 3rd Edition, John Willey and Sons Limited, Hoboken.
40. Mcgiffin, P. B. & Murthy, D. N. P. (1980) Parameter estimation for auto-regressive systems with missing observations, International Journal of Systems Science, 11:9, 1021-1034, DOI: 10.1080/00207728008967071
41. Meintanis, S.G. and Donatos, G.S.(1999). Finite-sample performance of alternative estimators for autoregressive models in the presence of outliers. Computational Statistics & Data Analysis 31:323-339, 1999.
42. Murthy, D.N.P. (1977). Parameter estimation in AR models with missing observations, unpublished paper presented at SIAM conference, Philadelphia
43. Nwabueze, J.C. (2005). Performance of estimators of linear model with autocorrelated error terms with exponential independent variable.J.Nig.ASssoc.Math.Phys. 9: 385-388.
44. Neter, J. and Wasserman,W. (1974). "Applied Linear Model," Richard D. Irwin Inc..Olaomi, J.O. and Adepoju, A.A. (2009): Efficiency in Linear Model with AR (1) and Correlated Error-Regressor (Pp. 46-61). An International Multi-Disciplinary Journal, Ethiopia Vol. 3 (3), April, 2009 ISSN 1994-9057 (Print) ISSN 2070-0083 (Online).
45. Rousseeuw, P.J. and Leroy, A.M.(1987) Robust Regression and Outlier Detection. Wiley, New York, 1987.
46. Rubin, D. B. (1976). Inference and missing data. Biometrika, 63, 581−592.
47. Steman, D. and Trenkler, G. Some further results on the efficiency of the Cochrane-Orcutt estimator. J. Statistical Planning and Inference. 88:205-214, 2000.
48. Suganthi, D. and Dheenathayalan, K. (2008), A Comparison of Statistical Packages in R Tool to Impute Missing Values IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278- 8727
49. Sargan, J.D., and Drettakis, E.G., 1974, Int.Econ.Rev.15, 39.
50. Sara J., Abbas B., Mohammad M. S., Behshid G. and Mohammad Reza B. (2021): Evaluation of Four Multiple T. Imputation Methods for Handling Missing Binary Outcome Data in the Presence of an Interaction between a Dummy and a Continuous Variable. Journal of Probability and Statistics Volume 2021, Article ID

6668822, 14 pages https://doi.org/10.1155/2021/6668822
51. Takahashi, M. and Ito, T. (2013). Multiple imputation of missing values in economic surveys: Comparison of competing algorithms. Proceedings 59th ISI World Statistics Congress, Hong Kong, August 25-30th.
52. Tanja Krone. Casper J. Alber . Marieke, E. Timmerman (2017). A comparative simulation study of AR (1) estimators in short time series Published online: 9 December 2015. doi :10.1007/s11135-015-0290-1
53. Taylor, S. J. (2007) Modelling financial time series. Wecker, W.E., 1978, Stoch.Proc.Applic. 8,153.
54. Zhongheng Z. (2015): Multiple imputation with multivariate imputation by chained sequation (MICE) package. doi: 10.3978/j.issn.2305-5839.2015.12.63.

3/22/2022