World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

On the efficiency of convoluted weighted method for missing values in household surveys

Faweya Olanrewaju *, Akinyemi Oluwadare and Ayeni Taiwo Michael

Department of Statistics, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria.

World Journal of Advanced Research and Reviews, 2023, 18(01), 815–832

Publication history: Received on 10 March 2023; revised on 17 April 2023; accepted on 19 April 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.1.0696

Abstract

Household surveys collect information on social and demographic characteristics in which the household constitutes the sampling units. They are effective means of obtaining variety of data needed for informed policy formulation and monitoring of national development. Despite their increasing recognition, household surveys suffer from non-response which creates bias estimates leading to wrong inference if not properly addressed. Existing estimators for handling missing values focused on the quality of missing value estimates without considering the resulting effects on location and scale parameters with regard to the nature of missingness and number of missing values. The convoluted weighted method was more efficient than the existing estimators with increasing number and nature of missingness. Its use will enhanced the level of precision of the estimated population parameter in the presence of missing values in household surveys.

Keywords: Household survey; Non-response; Bias; Missing Value; Demographic characteristics

1. Introduction

The effect of missing value from non-response on household survey data has generated a strong reaction from the researchers in the last decades. The main purpose of most survey analysis is to estimate various population parameters. When missing value occurs, estimate of any type of parameter are subject to certain effects of missing value which may be measured, minimized or otherwise tolerated by the researcher.

Household surveys are more than often affected by missing value from non-response. In addition, the process of obtaining data is precisely never free of missing value from non-response. Due to the effects of missing value on survey estimates, taking strategies of prevention and the handling of missing value in household surveys is clearly necessary[2].

Considering the fact that non-response which resulted to missing values is of two types:

Item and unit non-response, the handling strategies differ on the types of non-response.

1.1. Types of Non-response

Non-response can be manifested either as unit or item non-response

1.1.1. Unit Non-response

This refers to outright failure of a sampled subject to participate in a study. Its effects is that the realized sample is smaller than planned, making estimates less precise. Unit non-response in household surveys arises because of refusal

* Corresponding author: Ayeni Taiwo Micahel

to participate non-at-homes, units closed. Away in holiday, unit vacant or demolished, not locatable and language barriers or physical impairment.

1.1.2. Item non-response

It occurs in any kind of multivariate study in which a sampled subject responds to some but not all survey items. Its impacts on a statistic are exactly the same as that of unit non-response, but the damage is limited to statistic produced using data from the affected items.

The causes of item nonresponse are different from those unit non-responses. Whereas unit non-response occurs from a decision based on a brief description of the survey, item non-response arises after the measurement has been fully revealed. It arises because of item refusals “don’t knows”, omissions and answers deleted.

1.2. Nature of missing data (Missing Data Mechanism)

Knowledge of the nature of missing data is a central element in choosing an appropriate statistical technique to deal with missing data. According to [4], there exist three natures of missing data and they are missing completely at random (MCRAR), Missing at random (MAR) and Missing not at random(MNAR).

1.3. Missing completely at random (MCAR)

This is the assumption that missing data occur totally at random without any relation to the other observed or unobserved data. Here, the distribution of missing values R is thus assumed to be independent of both the forget variable Y and auxiliary variable X. in MCAR, response behaviour R and auxiliary variable are unrelated. Hence if there is a strong relationship between Y and X, then variable Y and response behaviour R has no relationship. Thus,

$$P\left(R/Y,X\right)=P(\mathrm{R}).$$

The most important feature of data which are MCAR is that the analysis remains unbiased.

Thus, data on family income would be considered MCAR if people with lower incomes were more likely to report their family income than people with high incomes, missingness would be correlated with income level.

1.4. Missing at Random (MAR)

In general, MAR occurs when there is no direct relationship between the targeted variable Y and the response behaviour R and the same time there is a relationship between the auxiliary variable and the response behaviour R. here, estimate may be biased and the problem can solved by applying a weighting technique using auxiliary variable. This is expressed as;

$$P(R/Y,\mathrm{X})=\mathrm{P(R/\,Y;X)}\ .$$

For example, people who are depressed might be less inclined to report their income, and thus reported income will be related to depression.

1.5. Missing not at Random (MNAR)

Here, missing values are assigned to be related to the uunobserved dependant variable vector Y_i^M, in addition to the remaining observered values. Hence there is a direct relationship between the target variable Y and the response behaviour Rand this relationship cannot be explained by an auxiliary variable, here estimate are biased and correlation techniques biased on auxiliary variables may be able to reduced bias. This is expressed as;

$$P(R/Y,X)=P(R/Y^{M},Y^{0},X)\ .$$

1.6. Consequences of MCAR, MAR and MNAR

The main consequence of MCAR is loss of statistical power. The good thing about MCAR is that analyses yield unbiased parameter estimates. MAR missingness also yields unbiased parameter estimates but MNAR yields biased parameter estimates.

1.7. Some existing Techniques to improve Response Rate

In limiting non-response in survey, many techniques have been found relevant in minimizing the proportion of non-response or its effect. Among them are discussed below:

1.8. Improvements in the Data Collection Strategies

These methods appear to be the most obvious remedy to minimize non-response.

Some of attempts which can lead to improvement in response rate are:

- Assurance of confidentiality which helps to alleviate fear respondents may have about the use of their responses for purpose other than those stipulated for the survey.
- Developing a rapport with the communities or the respondents through social engagement. Sensitize people about upcoming surveys through radio jingles, print media and television.
- Contacting and educating community heads and chiefs in the nature and benefits of a survey enhance cooperation.
- Questionnaire must be clear and concise. Any terms should be clearly defined.
- The survey format must be unambiguous and consistent. Instructions should be as explicit as possible.
- Good outlook (mode of dressing) of the interviewer.
- Motivation of the respondent to cooperate by using incentives either financial or materials.
- Use of locals with knowledge of the terrain and good command of the local language and with necessary academic qualification by the interviewer will improve cooperation on the part of the respondents.

1.9. Call-Backs and Reminders Strategies

Call-backs is the most common and successful way of reducing the percentage of non-response especially the \not-at-homes". It can be described as deliberate new attempts to obtain response from the non-respondents.

Also, in a mail survey, those who do not respond to the initial mailing may be sent a reminder (and a new copy of the questionnaire).

1.10. Substitution/Replacement Strategy

New sample members in this technique are substituted for unit non-respondents as a means to maintain the intended sample size but the bias from non-respondents will not usually be reduced.

1.11. Appropriate/Correct Questioning Pattern

A design with correct questioning format usually promotes high response rate. In [3] satisfying theory, there are three factors that affect the process of answering questions, these are:

- Motivation of the respondent to perform the task.
- Difficulty of the task.
- Respondent's cognitive ability to perform the task.

This theory explains why some respondents perform the cognitive task of answering questions better than others. The theory built on the question answering process model of [5].

2. Methods

Convoluted weighted method: this estimator was proposed by [1] providing the following predictions;

$$\hat{Y}^*_{CWM} = \hat{\alpha} Y^*_{LS} + (1-\alpha)\hat{Y}^*_{SR}$$

where

$$\hat{\alpha} = \frac{\left[\left(1-\frac{KR_c}{(t_c-k+2)b_c^{'}X_c^{'}b_cX_c}\right)X_*b_c\right]^2 - \left[\left(1-\frac{KR_c}{(t_c-k+2)b_c^{'}X_c^{'}b_cX_c}\right)\left(X_*b_c\right)^2\right]}{\left[\left(1-\frac{KR_c}{(t_c-k+2)b_c^{'}X_c^{'}b_cX_c}\right)X_*b_c - X_*b_c\right]^2} \quad (1)$$

$$Y_{LS}^* = X_*b_c$$

$$Y_{SR}^* = \left(1-\frac{KR_c}{(t_c-k+2)b_c^{'}X_c^{'}b_cX_c}\right)X_*b_c \quad (2)$$

$$b_c = \left(X_c^{'}X_c\right)^{-1}X_{cy_{obs}}^{'}$$

$$R_c = (y_c - X_cb_c)^{'}(y_c - X_cb_c)$$

$t_c =$ indicates the number of observed classed

$k =$ Number of explanatory variables (which is appositive scalar)

2.1. Efficiency Comparison

If the data are complete, then $S^2 = \frac{\sum_{t=1}^{T}(y_t-\hat{y}_t)^2}{T-k}$ is the corresponding estimator of variance (σ^2). If T-t_c cases are incomplete, that is, observation y_{mis} are missing in the model, then the variance σ^2 can be estimated using the complete case estimator as:

$$\sigma_c^2 = \frac{\sum_{t=1}^{t_c}(yt-\hat{y}t)^2}{t_c-k} \quad (3)$$

If the missing data are imputed using Least Squares (Yates) method, then we have the estimator

$$\hat{\sigma}_{LS}^2 = \frac{1}{T-k}\left[\sum_{t=1}^{t_c}(yt-\hat{y}t)^2 + \sum_{t=t_c+1}^{T}(y_t-\hat{y}_{LS})^2\right]$$

$$\hat{\sigma}_{LS}^2 = \frac{\sum_{t=1}^{t_c}(y_t-\hat{y}_t)^2}{T-k}$$

$$\hat{\sigma}_{LS}^2 = \frac{\sum_{t=1}^{t_c}(y_t-X*b_c)^2}{T-k} \quad (4)$$

Which makes use of t_c observations but has T-K instead of $t_c - k$ degrees of freedom. As

$$\hat{\sigma}_{LS}^2 = \hat{\sigma}_C^2\,\frac{t_c-k}{T-k} < \hat{\sigma}_c^2 \quad (5)$$

If the missing data are imputed using Stein Rule approach, then we have the estimate

$$\hat{\sigma}_{SR}^2 = \frac{\sum_{t=1}^{t_c}\left[y_t-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right]^2}{T-k} \quad (6)$$

If the missing data are imputed using the alternative Convoluted Weighted Estimator (CWE), then we have the estimate of variance as

$$\hat{\sigma}^2_{CWM} = \frac{\sum_{t=1}^{\mathrm{t_c}}\left[\mathrm{y_t}-\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\}\right]^2}{\mathrm{T-k}} \quad (7)$$

2.2. Efficiency Comparison: Alternative Estimator Using Convoluted Weighted Technique versus Least Squares

The proposed convoluted weighted method will be more efficient than the Least Squares method

if and only if

$$\frac{\hat{\sigma}^2_{LS}}{\hat{\sigma}^2_{CWM}} > 1 \quad (9)$$

Equation (8) implies $\hat{\sigma}^2_{LS} > \hat{\sigma}^2_{CWM}$ which also implies

$$\hat{\sigma}^2_{LS} - \hat{\sigma}^2_{CWM} > 0 \quad (10)$$

Substituting (4) and (7) in (9), we have

$$\frac{\sum_{t=1}^{\mathrm{T}}(\mathrm{y_t}-X*b_c)^2}{\mathrm{T-k}} - \frac{\sum_{t=1}^{\mathrm{t_c}}\left[\mathrm{y_t}-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2}{\mathrm{T-k}} > 0$$

multiplying through by *T-k*, we have

$$\sum_{t=1}^{\mathrm{T}}(\mathrm{y_t}-X*b_c)^2 - \frac{\sum_{t=1}^{\mathrm{t_c}}\left[\mathrm{y_t}-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2}{\mathrm{T-k}} > 0$$

Since the series span through the same index, the above inequality is true if and only if

$$(\mathrm{y_t}-X*b_c)^2 - \left[\mathrm{y_t}-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2 > 0$$

$$(\mathrm{y_t}-X*b_c)^2 > \left[\mathrm{y_t}-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2$$

Raising both sides of the above inequality to power $\frac{1}{2}$, we get

$$(\mathrm{y_t}-X*b_c) > \left[\mathrm{y_t}-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]$$

$$-X*b_c > -\tilde{\alpha}X*b_c-(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c$$

$$-X*b_c+\tilde{\alpha}X*b_c > -(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c$$

$$-(1-\tilde{\alpha})X*b_c > -(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c$$

Dividing through by $-(1-\tilde{\alpha})$

$$X*b_c < \left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c$$

Multiplying through by $(X*b_c)^{-1}$

$$1 < 1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}$$

$$1-1 < -\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}$$

$$0 < -\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}$$

$$-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c} > 0$$

Multiply through by - 1 and recall that negative multiplication reverse order

$$\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c} < 0 \tag{11}$$

Hence, the alternative convoluted weighted method will be more efficient than Least Squares method if equation (11) holds. That is, if

$$\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c} < 0$$

2.3. Efficiency Comparison: Proposed Technique versus Stein Rule Techniques

The convoluted weighted method (CWM) will be more efficient than Stein's rule only if

$$\frac{\hat{\sigma}^2_{SR}}{\hat{\sigma}^2_{CWM}} > 1 \tag{12}$$

Equation (12) implies

$$\hat{\sigma}^2_{SR} > \hat{\sigma}^2_{CWM}$$

$$\Rightarrow \hat{\sigma}^2_{SR} - \hat{\sigma}^2_{CWM} > 0 \tag{13}$$

Substituting (6) and (7) in (13), we have

$$\frac{\sum_{i=1}^{t_c}\left[y_t-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right]^2}{T-k} - \frac{\sum_{i=1}^{t_c}\left[y_t\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2}{T-k} > 0$$

Multiplying through by *T*-*k*, we have

$$\sum_{i=1}^{t_c}\left[y_t-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right]^2 - \sum_{i=1}^{t_c}\left[y_t\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2$$

Since the series span through the same index, the above inequality will hold if and only if

$$\left[y_t-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right]^2 - \left[y_t\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}\right]^2$$

Raising both sides of the inequality to power of $\frac{1}{2}$, we have

$$y_t-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c > y_t-\left\{\tilde{\alpha}X*b_c+(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c\right\}$$

$$-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c > -\tilde{\alpha}X*b_c-(1-\tilde{\alpha})\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\,b_c}\right)X*b_c$$

$$-\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c > -\tilde{\alpha}X*b_c - \left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c + \tilde{\alpha}\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c$$

Collecting like terms

$$0 > -\tilde{\alpha}X*b_c + \tilde{\alpha}\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c$$

$$\tilde{\alpha}X*b_c > \tilde{\alpha}\left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c$$

Since $\tilde{\alpha}$ is a scaler, dividing above inequality through by $\tilde{\alpha}$ yields

$$X*b_c > \left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)X*b_c$$

Also, multiplying through by $(X*b_c)^{-1}$

$$1 > \left(1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c}\right)$$

Therefore, the above inequality or more succinctly, our proposed convoluted weighted method is more efficient than the steins rule if and only if

$$1-\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c} < 1$$

$$\frac{kR_c}{(t_c-k+2)b_c'X_c'X_c\ b_c} > 0$$

3. Data Analysis

3.1. Numerical Illustration with Survey Data

In order to investigate the performance of the developed technique, we carried out a 2-Stage Stratification Design on Akure North Local Government Area Iju/ Ita Ogbolu in Ondo-State which consist of urban and rural area. A Simple Random Sampling of 15 Enumeration Areas (EAs) was selected from a total of 33(5 Urban and 28 Rural) EAs in Akure North. In each of selected EAs, a list of households was prepared and a Simple Random sampling (SRS) of 100 households were selected from the list. Informations on household Income (Y), age(X_1) and year of schooling(X_2) from each selected household were observed. The data from this survey are available in appendix A to actualize the following:

- To determine the effects of missingness on descriptive and inferential statistics when different proportions of data are missing.
- To evaluate the performance of the proposed model with some existing techniques of handling missing data under different percentage of missing data.

Three demographic variables; Y(income N'000), Age (x_2) and year of schooling (x_1) were considered.

Then, differing amounts were deleted at random causing MCAR data which had 0, 1,5,12,23 and 44% missing data.

In MAR situation y becomes missing as follows: 0% for complete data set, 5% when $x_1 < 5$, 12% when $x_2 < 55$, 23% when $x_1 < 6$ or $x_2 < 50$.

Sorting according to the actual y values deleting the cases to give 6 different rates of missing data.

Table 1 Performance of Some Missing Data Techniques For Parameter Estimates when Different Percentage of Data are Missing Under MAR Assumption of Missingness Using Survey Data.

EST.	% of	MISSING DATA TECHNIQUES				
PAR.	MISSINGNESS	MI	LS	SR	Proposed	LW
					CWM	LW
MEAN ($\bar{y}$)	0%	13.814	13.814	13.814	13.814	13.814
	5%	14.3309	13.73923	13.7395	13.73949	14.40389
	12%	13.79838	13.9359	13.93581	13.93581	13.83693
	23%	16.0668	13.60591	13.6062	13.60622	16.2539
	44%	16.1926	13.74462	13.90653	13.80945	16.38911
COR (ρ)	0%	0.946	0.946	0.946	0.946	0.946
	5%	0.688	0.967	0.967	0.967	0.967
	12%	0.657	0.956	0.956	0.956	0.956
	23%	0.5	0.968	0.968	0.968	0.968
	44%	0.419	0.969	0.871	0.937	0.871
VAR ($\hat{\sigma}^2$)	0%	46.577	46.577	46.577	46.577	46.577
	5%	40.08759	47.4471	47.44181	47.44195	42.04409
	12%	38.13801	47.24294	47.24176	47.24176	42.82502
	23%	25.06315	50.00722	50.00148	50.00124	32.26894
	44%	13.14075	51.63002	51.23706	50.40304	23.41699
SKEW (S_k)	0%	0.292	0.292	0.292	0.292	0.292
	5%	0.332896	0.217973	0.218127	0.21812	0.293745
	12%	0.273841	0.247429	0.247397	0.247397	0.243485
	23%	0.508641	0.263444	0.26358	0.263582	0.358312
	44%	0.919367	0.217336	0.210038	0.259461	0.582377
KURT (K)	0%	2.616	2.616	2.616	2.616	2.616
	5%	2.79315	2.609991	2.610098	2.610102	2.659034
	12%	3.138895	2.624818	2.624781	2.624781	2.797235
	23%	3.854819	2.504469	2.504602	2.504612	2.979674
	44%	5.552385	2.576848	2.596357	2.594897	3.075064
CV	0%	49.62	49.62	49.62	49.62	49.62
	5%	44.18059	50.13518	50.13139	50.13149	45.01658
	12%	44.75595	49.32117	49.3209	49.3209	47.29432
	23%	31.15935	51.97433	51.97021	51.97003	34.94901
	44%	22.38688	52.03251	52.65302	51.2723	29.52638

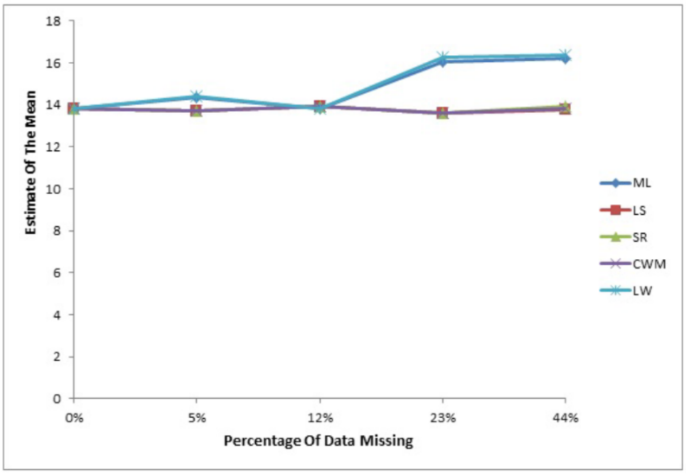

Figure 1 Graph of mean for MAR imputed data by Percentage of the Data; Missing using Survey Data

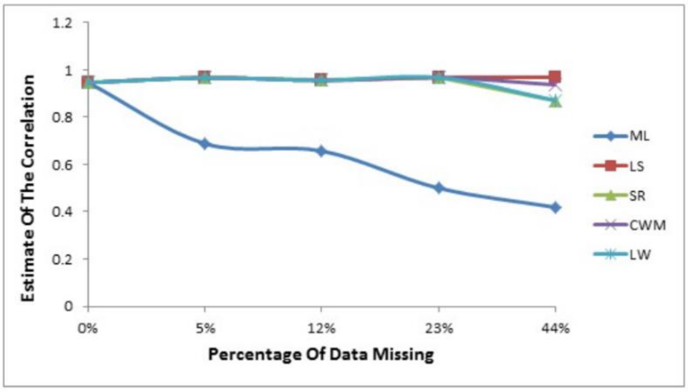

Figure 2 Graph of correlation for MAR imputed data by Percentage of Data; Missing using Survey Data

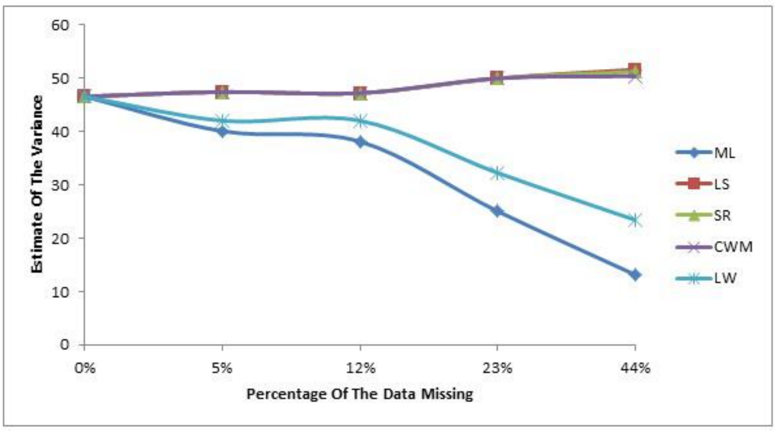

Figure 3 Graph of variance for MAR imputed data by Percentage of the DataMissing using Survey Data

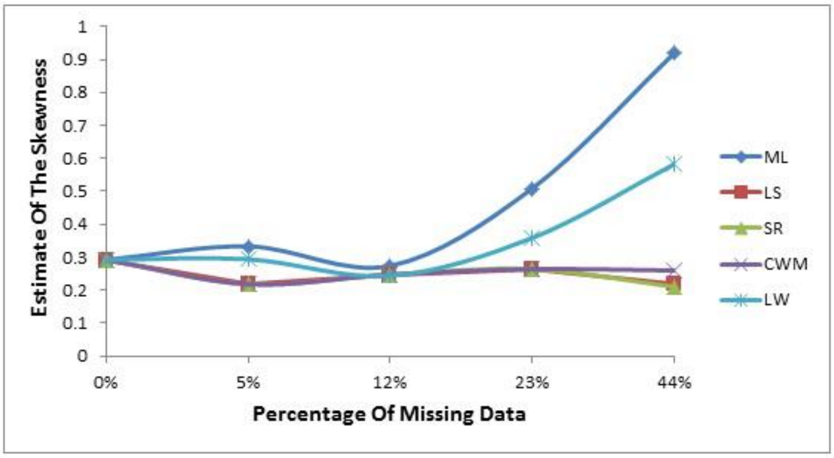

Figure 4 Graph of skewness for MAR imputed data by Percentage of the Data; Missing using Survey Data

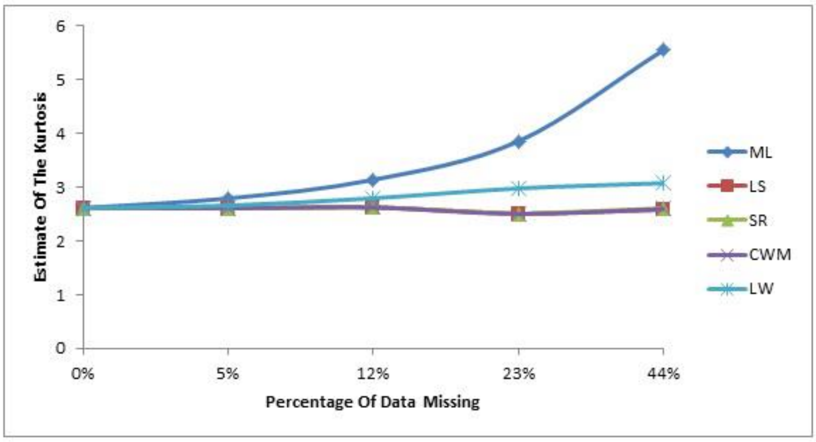

Figure 5 Graph of kurtosis for MAR imputed data by Percentage of the DataMissing using Survey Data

Figure 6 Graph of coefficient of variation for MAR imputed data by Percentage of the Data; Missing using Survey Data

Table 2 Performance of Some Missing Data Techniques For Parameter Estimates when Different Percentage of Data are Missing Under MNAR Assumption of Missingness Using Survey Data.

EST.	**% of**	**MISSING DATA TECHNIQUES**				
PAR.	**MISSINGNESS**	**MI**	**LS**	**SR**	**Proposed**	**LW**
					CWM	**LW**
MEAN ($\bar{y}$)	0%	13.814	13.814	13.814	13.814	13.814
	5%	14.49056	13.97812	13.84395	13.84639	13.88946
	12%	12.17703	13.95554	13.95534	13.95546	12.36552
	23%	12.23131	14.04591	14.01833	13.96311	11.48885
	44%	11.061	14.4385	14.5921	13.91216	9.637288
COR (ρ)	0%	0.946	0.946	0.946	0.946	0.946
	5%	0.606	0.967	0.967	0.967	0.979
	12%	0.588	0.966	0.966	0.978	0.966
	23%	0.478	0.92	0.863	0.975	0.695
	44%	0.42	0.897	0.829	0.974	0.705
VAR ($\hat{\sigma}^2$)	0%	46.577	46.577	46.577	46.577	46.577
	5%	26.26837	46.08873	47.4183	47.4967	34.36302
	12%	38.86101	49.97552	49.97043	49.97311	29.30968
	23%	27.12435	51.65988	48.20559	50.15867	25.17617
	44%	18.89883	48.62315	50.6775	50.64028	17.85335
SKEW (S_k)	0%	0.292	0.292	0.292	0.292	0.292
	5%	-0.05984	0.287699	0.256216	0.260343	-0.0691
	12%	0.083983	0.321818	0.32165	0.321715	-0.20054
	23%	-0.23919	0.337992	0.343632	0.327701	-0.16146
	44%	-0.27929	0.279842	0.333821	0.296285	-0.08753
KURT (K)	0%	2.616	2.616	2.616	2.616	2.616
	5%	2.185599	2.712855	2.702257	2.711416	2.16819
	12%	2.370098	2.730782	2.730542	2.730597	2.092673
	23%	2.396741	2.678539	2.699595	2.740979	2.234413
	44%	2.812341	2.721712	2.731868	2.672794	2.428488
CV	0%	49.62	49.62	49.62	49.62	49.62
	5%	35.36971	48.56783	49.74082	49.77317	42.20465
	12%	51.19359	50.65614	50.65427	50.65519	43.78177
	23%	42.58009	51.17135	50.72134	49.5282	43.67355
	44%	39.30276	50.69062	50.87755	50.12182	43.84349

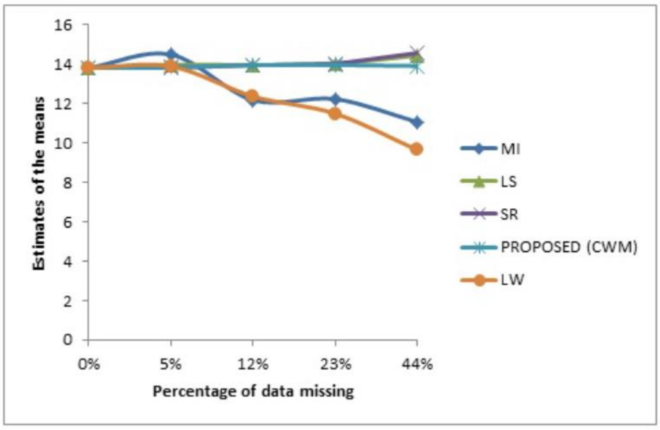

Figure 7 Graph of mean for MNAR imputed data by Percentage of the Data; Missing using Survey Data

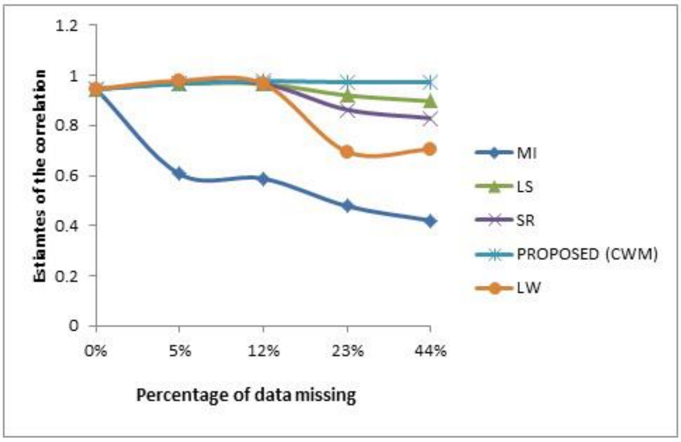

Figure 8 Graph of correlation for MNAR imputed data by Percentage of the Data; Missing using Survey Data

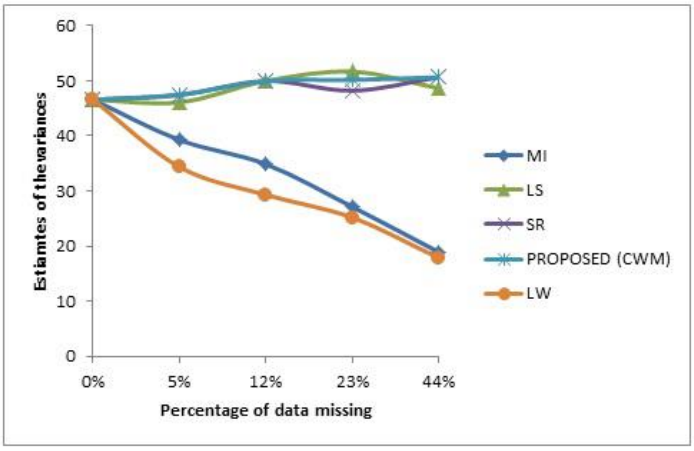

Figure 9 Graph of variance for MNAR imputed data by Percentage of the Data; Missing using Survey Data

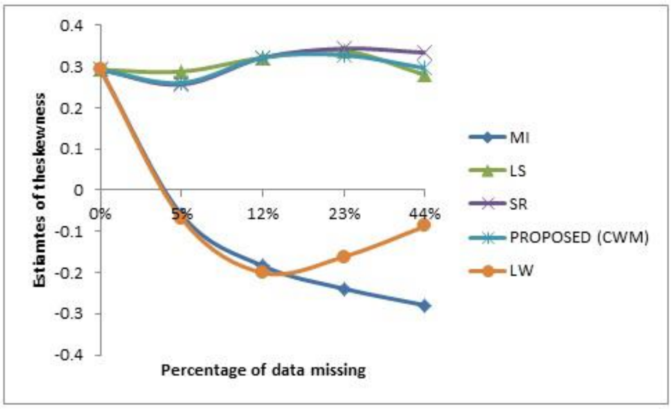

Figure 10 Graph of skewness for MNAR imputed data by Percentage of the Data; Missing using Survey Data

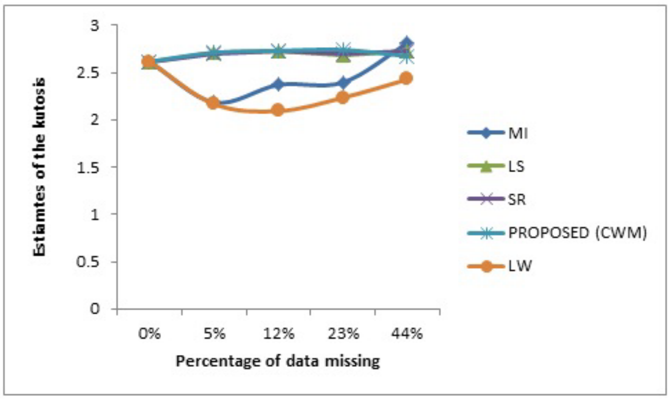

Figure 11 Graph of kurtosis for MNAR imputed data by Percentage of the Data; Missing using Survey Data

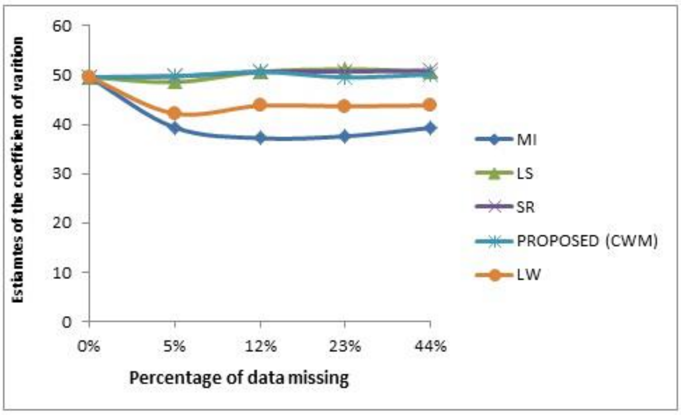

Figure 12 Graph of coefficient of variation for MNAR imputed data by Percentage of the Data; Missing using Survey Data

Table 3 Performance of missing techniques for parameter estimates under MAR (with RMSE in parenthesis) Using Survey Data.

COMPLETE	**MAR**				
PARAMETER	**MI**	**LS**	**SR**	**Proposed CWM**	**LW**
mean = 13.814	15.097	13.773	13.797	13.757	15.221
	(1.213)	(0.138)	(0.154)	(0.136)	(1.293)
VAR = 46.577	29.107	49.082	48.981	48.772	35.139
	(12.562)	(2.114)	(1.960)	(1.662)	(9.172)
STDEV = 6.8247	5.285	7.005	6.998	6.983	5.887
	(1.255)	(0.151)	(0.140)	(0.119)	(0.802)
SKEW = 0.217	0.509	0.217	0.210	0.212	0.369
	(0.291)	(0.038)	(0.035)	(0.035)	(0.150)
E KURT = 0	0.835	0.421	0.416	0.416	0.122
	(1.227)	(0.054)	(0.054)	(0.054)	(0.186)
KURT = 2.616	3.835	2.579	2.584	2.584	2.878
	(1.227)	(0.054)	(0.054)	(0.054)	(0.186)
COV = 49.62	35.621	50.866	50.724	50.769	39.197
	(10.828)	(1.355)	(1.255)	(1.216)	(8.387)
COR = 0.946	0.566	0.965	0.941	0.941	0.957
	(0.128)	(0.047)	(0.047)	(0.006)	(0.014)

Source: Analysis Result from Table 4.3 using MATLAB-software code.

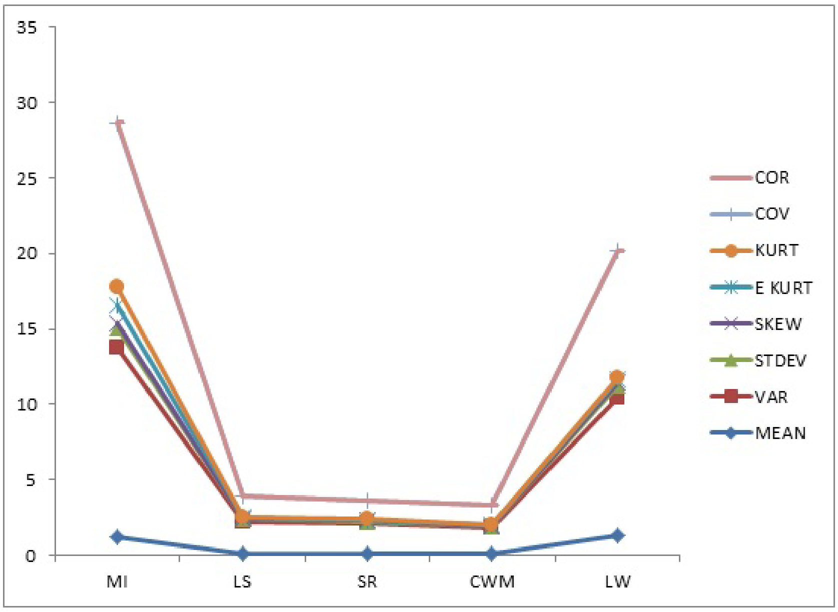

Figure 13 Graph of Root Mean Square Error of some missing data techniques under MAR missingness assumption using Survey Data

Table 4 Performance of Missing Techniques for Parameter Estimates under MNAR (with RMSE in parenthesis) Using Survey Data

COMPLETE	**MAR**				
PARAMETER	**MI**	**LS**	**SR**	**Proposed CWM**	**LW**
mean = 13.814	12.49	13.973	13.952	13.951	11.845
	(1.439)	(0.079)	(0.077)	(0.056)	(1.775)
VAR = 46.577	27.788	49.087	49.068	49.567	26.676
	(8.254)	(2.353)	(1.513)	(1.409)	(6.979)
STDEV = 6.8247	5.229	7.005	7.004	7.04	5.13
	(0.774)	(0.168)	(0.108)	(0.101)	(0.695)
SKEW = 0.217	0.124	0.307	0.314	0.302	0.13
	(0.168)	(0.031)	(0.039)	(0.028)	(0.062)
E KURT = 0	0.559	0.289	0.284	0.286	0.769
	(0.265)	(0.023)	(0.018)	(0.03)	(0.144)
KURT = 2.616	2.441	2.711	2.716	2.714	2.231
	(0.265)	(0.023)	(0.018)	(0.03)	(0.144)
COV = 49.62	42.112	50.129	50.2	50.46	43.376
	(6.734)	(1.126)	(0.665)	(0.459)	(0.784)
COR = 0.946	0.523	0.938	0.936	0.833	0.977
	(0.089)	(0.035)	(0.05)	(0.002)	(0.154)

Source: Analysis Result from Table 4.4 using MATLAB-software code.

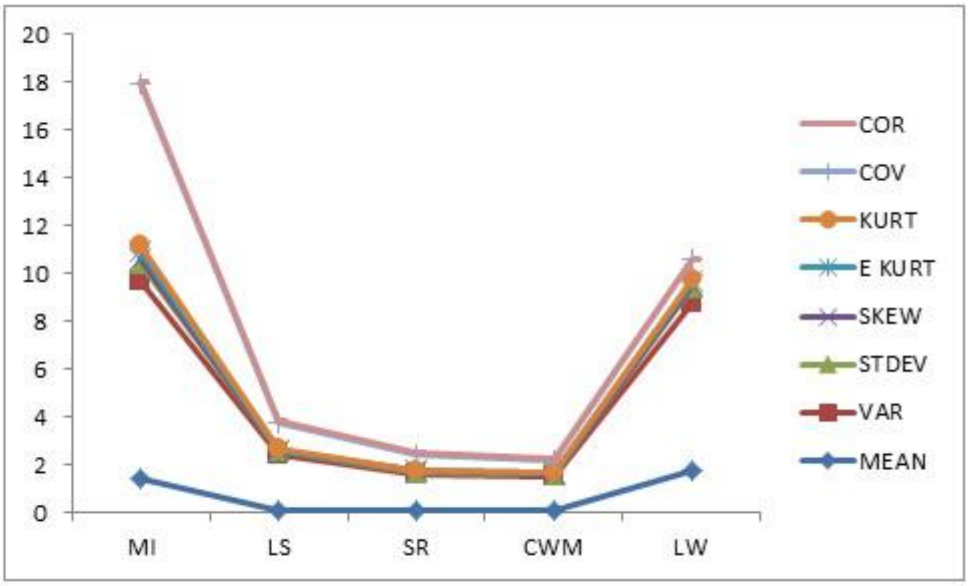

Figure 14 Graph of Root Mean Square Error of some missing data techniques under MNAR missingness assumption using Survey Data

Table 5 Summary of the Results from the figure 1-14 of Performance of some Missing Data Techniques under MAR as Percentage of Missing Value Increases using Real Life Data

Est. Par.	MI	LS	SR	Proposed CWM	LW
Mean	Little percentage of discrepancy	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	Little percentage of discrepancy
Cor	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)
Var	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
Skew	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
Kurt	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	Little percentage of discrepancy
CV	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases

Table 6 Summary of the Results from the figures 1-14 of Performance of some Missing Data Techniques under MNAR as Percentage of Missing Value Increases using Real Life Data.

Est. par.	**MI**	**LS**	**SR**	**Proposed CWM**	**LW**
Mean	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
Correlation	High percentage of discrepancy in the true value as percentage of missingness increases	Little percentage of discrepancy	Little percentage of discrepancy	Approximately constant (within target value)	Approximately constant (within target value)
Variance	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Little percentage of discrepancy	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
Skewness	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value) when the percentage missing is low	Approximately constant (within target value) when the percentage missing is low	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
Kurtosis	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	High percentage of discrepancy in the true value as percentage of missingness increases
CV	High percentage of discrepancy in the true value as percentage of missingness increases	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)	Approximately constant (within target value)

4. Conclusion

From the above findings, we conclude that sometimes, missing data introduce systematic distortion in survey estimates and bias flows from missing data when the causes of the missing data are linked to the survey statistics measure. In addition, the alternative estimator using Convoluted Weighted Method (CWM) performed better regardless of the percentage of the missing data and nature of missingness. It proved to have precise estimate of population parameter in the presence of missing values in household surveys.

Compliance with ethical standards

Acknowledgments

We acknowledged the anonymous reviewers.

Disclosure of conflict of interest

The authors has No conflict of interest.

References

[1] Faweya O., Amahia G.N., Adeniran A.T.(2017) Estimation of Missing Data Using Convoluted Weighted Method in Nigeria Household Survey. Science Journal of Applied Mathematics and Statistics; 5(2): 70-77 http://www.sciencepublishinggroup.com/j/sjams doi: 10.11648/j.sjams.20170502.12 ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

[2] Faweya Olanrewaju and G. N. Amahia(2016)Effect of Missingness Mechanism on Household Survey Estimates in Nigeria. Global Journal of Science Frontier Research: F-Mathematics and Decision Sciences 16(5) Online ISSN: 2249-4626 & Print ISSN: 0975-5896

[3] Krosnic, J. A. 1991. Response strategies for coping with cognitive demands of attitude measures in surveys, Applied cognitive psychology, 5(3); 213-236.

[4] Little, R. J. A. and Rubin, D. B. 1987. Statistical Analysis with Missing Data, Wiley, New York.

[5] Tourangeau, R. and Rasinski, K. 1988. Cognitive process Underlying Context Effects in Attitude Measurement, Psychological Bulletin, 103, 209-314.