19- Steffen, L.L. (1995), " The EM Algorithm for Graphical Association Models With Missin Data ", Computational Statistics and Data Analysis, 19, 2, 191-201.
20- Thompson, J. R. (1968), " Some Shrinkage Techinques For Estimating The Mean",JASA, 63, 113-122.
21- Yuan, (2003), " Multiple Imputation Strategies for Missing Data ", SAS Institute.

....
....
....

Sociological Methodology 1987, ed. C. C. Clogg Washington, D.C.: American Sociological Association, 71-103.
3- Cooman, G. de, and Zaffalon. M. (2004), " Updating Beliefs With Incomplete Observations ", Artificial Intelligence, 159, 75-125.
4- Glesser, M. (1964), " Linear Regression Analysis With Missing Observations Among The Independent Variables ", JASA, 59, 834-844.
5- Huang, X. and Zhu, Q. (2002), " A Pseudo-Nearest-Neighbor Approach for Missing Data Recovery on Gaussian Random Data Sets ", Pattern Recognition Letters, Vol. 23, Issue 13.
6- Lee, S.Y. (1986), " Estimation for Structural Equation Models With Missing Data", ",Psychometrika,51,93-99.
7- Little, R. J. A. (1992), " Regression with missing X's", JASA, 87, 1227-1237.
8- Little, R. J. A. and Rubin, D.B. (1987), " Statistical Analysis With Missing Data ", John Wiley, New York.
9- Mehta, J. S. and Srinivasan, R. (1971), " Estimation of The Mean by Shrinkage to a Point ", JASA, 66, 86-90.
10- Muthen, B., Kapplan, D., and Hollis, M. (1987), " On The Structural Equatio Modeling With Data That Are Note Missing at Random ",Psychometrika, 52, 431-462.
11- Nielsen, S. F. (2000), " Reletive Coarsening at Random ", Statistica Neerlandia, 54, 79-
12- Nielsen, S. F. (2001), " Robust Learning With Missing Data ", Klumer Academic Publishers, USA, 45, 147-170.
13- Pandy, B. N. (1979), " On Shrikage Estimation Of Normal Population Variance",Communication in Statistics. Theory, Mathematical. A8(4), 359-365.
14- Ramoni, M. and Sebastiani, P. (2003), " Bayesia Methods ", Intelligent Data Analysis, Springer Verlag. New York.
15- Rahim, M. A. (1976), " Pretest Estimation of Simple Regression Model Using Prior Knowledge of One of The Parameters ", Technometrics, 18, 349-445.
16- Rubin, D. B. (1976a), " Comparing Regression When Some Predictor Variables Are Missing ", Technometrics, VOL 18, 201-205.
17- Rubin, D. B. (1976b), " Inference and Missing Data ", Biometrica, 63, 581-592
18- Schafer, J. (1997), " Analysis of Incomplete Multivariate Data ",Chapman and Hall.

Table (4): Describes the values of the constant a for the SLRM's parameters by SH method

Missing data percentage	10%		20%		30%		40%	
	α	β	α	β	α	β	α	β
$n = 10$	0.1074	0.1243	0.0008	0.0010	0.0823	0.1092	0.0013	0.0017
$n = 30$	0.0141	0.0189	0.0221	0.0297	0.0084	0.0139	0.0252	0.0394
$n = 50$	0.2835	0.3494	0.1721	0.2334	0.1037	0.1664	0.0411	0.0750
$n = 100$	0.3075	0.4058	0.2365	0.3246	0.2316	0.3488	0.2555	0.3820

6. CONCLUSIONS

1. It is found that the OLS estimator and the CC estimator and the AC estimator, are all unbiased, whereas the shrinkage estimator is biased for all sample sizes and for all missing data percentages considered.

2. The proposed shrinkage estimator for estimating the SLRM's parameters was the best and better than the estimators of the OLS and CC and AC methods, because it has less MSE compared with the other mentioned methods.

3. The R.E for the proposed shrinkage estimator was very large for all sample sizes and for all missing percentages considered for all methods.

4. The values of a was between 0 and 1 for all missing percentages which coincide with the condition which was put on the proposed estimator in the theoretical side.

REFERENCES

1- Affifi, A. A. and Elashoff, R. M. (1969), " Missing Observations in Multivariate Statistics, A Note on Simple Linear Regression ", JASA, 64, 358-365.

2- Allison, P.D. (1987), " Estimation of Linear Models With Incomplete Data ", in

4- The R.E. for the parameter β by the CC method has the same properties as for the parameter α by the same method.
5- The R.E for the parameters α and β by the AC method increase when the sample size increase.
6- At sample size $n = 10$, we note that the R.E for the parameter α by the SH method directly proportional with the missing data percentages (10%, 20%, 40%). Whereas for the missing data percentage 30%, we note decreases in R.E for the parameter α.
7- At $n = 30$, we note fluctuations in R.E for the parameter α by SM method for all missing data percentage.
8- At $n = 50.100$, we note that the R.E for the parameter α by SH method directly proportional with all missing data percentage.
9- The R.E for the parameter β has the same properties as for the R.E. of the parameter α.
10- At all sample sizes, we note that the R.E by the SH method is the largest followed by the CC method then the AC method for all missing data percentage.

Table (3): Describes the R.E for the SLRM's parameters

Missing data percentage		10%		20%		30%		40%	
		α	β	α	β	α	β	α	β
$n = 10$	CC	2.2372	1.4240	3.3631	2.7958	13.0307	7.2472	3.4958	2.6205
	AC	1.1086	0.6447	0.9924	0.6390	2.1968	1.0099	0.8774	0.5247
	SH	11.9290	6.0038	1428.5	771.375	30.7204	10.6857	761.8666	346.0373
$n = 30$	CC	1.2005	1.1557	1.3156	1.1358	4.9665	2.8092	1.7475	1.3959
	AC	0.6499	0.5933	0.6345	0.4875	1.3437	0.6037	0.6591	1.3959
	SH	52.7096	35.9062	41.8974	18.7745	204.25	49.1025	30.2592	11.9687
$n = 50$	CC	1.0281	0.9727	1.3599	1.0774	2.2558	1.2240	2.5147	1.6051
	AC	0.6489	0.5659	0.6434	1.0161	1.0593	0.4782	0.8265	0.4346
	SH	2.7102	1.9258	4.7345	2.4716	11.8	3.3459	23.2424	6.7356
$n = 100$	CC	0.6360	1.1504	1.3524	1.0284	1.7389	1.0710	1.8677	1.1036
	AC	0.6303	0.5934	0.6524	0.4338	0.8114	0.4189	0.7233	0.5568
	SH	2.9663	1.5830	3.2385	1.7431	6.2841	1.9323	3.3619	1.0412

From table (4), we note that all values of a by SH method is between zero and one.

(11) It is noticed that the MSE of the parameter β by the SH method has the same properties as the parameter α has.

(12) For all sample sizes $n = 10,30,50,100$, we notice that the MSE for the parameters α and β by the SH method is the least followed by the CC method and then the AC method for all missing data percentage.

Table(2): Describes the MSE for the SLRM's parameters

Missing data percentage		10%		20%		30%		40%	
		α	β	α	β	α	β	α	β
$n = 10$	OLS	1.1428	3.7026						
	CC	0.5108	2.6001	0.3398	1.3243	0.0877	0.5109	0.3260	1.4129
	AC	1.0308	5.7428	1.1515	5.7943	0.5202	3.6662	1.3024	7.0561
	SH	0.0958	0.6167	0.0008	0.0048	0.0372	0.3465	0.0015	0.0107
$n = 30$	OLS	0.1634	0.5745						
	CC	0.1361	0.4971	0.1242	0.5058	0.0329	0.2045	0.0935	0.4109
	AC	0.2514	0.9683	0.2575	1.1784	0.1216	0.9516	0.2479	103959
	SH	0.0031	0.0160	0.0039	0.0306	0.0008	0.0117	0.0054	0.0480
$n = 50$	OLS	0.0767	0.2573						
	CC	0.0746	0.2645	0.0564	0.2388	0.0340	0.2102	0.0305	0.1603
	AC	0.1182	0.4546	0.1192	0.5232	0.0724	0.5180	0.0928	0.5920
	SH	0.0283	0.1336	0.0162	0.1041	0.0065	0.0769	0.0033	0.0382
$n = 100$	OLS	0.0353	0.1086						
	CC	0.0555	0.0944	0.0261	0.1056	0.0203	0.1014	0.0189	0.0984
	AC	0.0560	0.1830	0.0541	0.2503	0.0435	0.2592	0.0488	0.3242
	SH	0.0119	0.0686	0.0109	0.0623	0.0085	0.0562	0.0105	0.1043

From the results that appear in table (3) which represent the R.E for the estimated SLRM's parameters, we deduce the following:

1- For the samples of size $n = 10,30$, we notice that the R.E for the parameter α by the CC method directly proportional with the missing data percentage (10%, 20%, 30%). Whereas for the missing percentage 40%, we notice decreasing in R.E. for the parameter α.

2- For the samples of size $n = 50,100$, we notice that the R.E for the parameter α by the CC method directly proportional with the all missing data percentages.

3- The R.E for the parameter α increases when the sample size increase for all missing percentage chosen.

Table (1): Describes the absolute value for the bias of the SLMR's parameters

Missing data percentage	10%		20%		30%		40%	
	α	β	α	β	α	β	α	β
n=10	0.2949	0.7422	0.0290	0.0700	0.1859	0.1859	0.0386	0.1034
n=30	0.0554	0.1256	0.0699	0.1728	0.0298	0.1071	0.0731	0.2155
n=50	0.1469	0.3071	0.1178	0.2887	0.0770	0.2568	0.0566	0.1910
n=100	0.0830	0.2027	0.0933	0.2231	0.0911	0.2318	0.0940	0.2658

From the results that appear in table (2) below which represent the MSEs for the estimated SLRM's parameters, we deduce that

(1) For the samples of size $n = 10,30$, we note that the MSE for the parameter α by CC method is inversely proportional with the missing data percentage.

(2) It has been noted that the CC method had responded to the sample size where the MSE for the parameter α decrease when the sample size increase for all missing data percentage except for the sample of size $n = 50$ and for missing percentage 30%.

(3) The MSE for the parameter β by CC method has the same properties of the MSE of the parameter α.

(4) For all sample sizes, it is noted that the MSE criterion for the parameter α by the AC method fluctuate for all missing data percentage.

(5) It has been noted that the AC method responded for the sample size where the MSE for the parameter α decrease when the sample size increase.

(6) The MSE for the parameter β by the AC method has the same properties of the MSE of the parameter α by the same method.

(7) For the sample of size $n = 10$, we noticed that the MSE for the parameter α by SH method is inversely proportional with all missing data percentage.

(8) At sample size $n = 30$, it is noticed that the MSE for the parameter α by the SH method fluctuates for all missing percentages.

(9) At sample sizes $n = 50,100$, the MSE for the parameter α by the SH method is inversely proportional with the missing percentages.

(10) For all sample sizes, we notice that the MSE criterion by the SH method for the parameter α fluctuates.

each of the sample size and of the missing observations percentage have been studied. The samples of size (n = 10, 30, 50, 100) have been generated, for all sample sizes used in this work. The observations have been chosen completely at random and the suggested percentage of the missing observations of the study where (10%, 20%, 30%, 40%). Then, the bias, MSE for the estimated parameters have been computed using the following methods:

1. Analysis of complete observations (CC) method
2. Analysis of available observations (AC) method
3. Shrinkage estimators (SH) method

Then a comparison between the estimators of these methods with the estimators of the OLS method using the RE have been made.

5. SIMULATION RESULTS

A SLRM will be used with $\alpha = 1$, $\beta = 2$ and the value of the constant in the weight function is b = 0.5. Then the SLRM will be

$$y_i = 1 + 2x_i + \varepsilon_i$$

From the results that appear in table (1) below which represent the absolute value for the bias of the estimators of model's parameters, we deduce the following:

(1) For the sample of size $n = 10$, it is observed that the bias for α inversely proportional with the missing data percentage.

(2) For samples of size $n = 30,50$, it is observed that the bias for the parameter α fluctuate (unstable) for all missing data percentages.

(3) For sample of size $n = 100$, the bias for α directly proportional with the missing data percentage.

(4) For all sample sizes chosen, we note that the bias for the parameter α fluctuate.

(5) The bias for the parameter β has the same properties of the bias of the parameter α .

$$=\left\{\frac{a^2}{(4b+1)^{3/2}}e^{\frac{2b\lambda^2}{4b+1}}\left[1+\frac{\lambda^2}{4b+1}\right]+\lambda^2\left[1-\frac{2a}{(2b+1)^{3/2}}e^{\frac{-b\lambda^2}{2b+1}}\right]\right\}^{-1} \quad (28)$$

In a similar way, the RE ($\tilde{\beta}$) using (27) is

$$RE(\tilde{\beta})=\left\{\frac{a^2}{(4b+1)^{3/2}}e^{\frac{-2b\lambda^2}{4b+1}}\left[1+\frac{\lambda^2}{4b+1}\right]+\lambda^2\left[1-\frac{2a}{(2b+1)^{3/2}}e^{\frac{-b\lambda^2}{2b+1}}\right]\right\}^{-1} \quad (29)$$

The value of a that minimizes $MSE(\tilde{\alpha})$ of (26) is gotten by equating $\frac{\partial MSE(\tilde{\alpha})}{\partial a}$ to zero, and solving the resulting equation we get

$$a=\left[\lambda^2(4b+1)^{5/2}e^{\frac{b\lambda^2}{(2b+1)(4b+1)}}\right]/\left[(2b+1)^{3/2}(4b+1+\lambda^2)\right] \quad (30)$$

Similarly, the value of a that minimizes the $MSE(\tilde{\beta})$ of (27) is

$$a=\left[\lambda^2(4b+1)^{5/2}e^{\frac{b\lambda^2}{(2b+1)(4b+1)}}\right]/\left[2b+1)^{3/2}(4b+1+\lambda^2)\right] \quad (31)$$

4. THE EXPEREMINTAL SIDE

A Q-basic program has been used to simulate experiments considered in this study. The computer program has generated data for the explanatory variable which follow the uniform distribution in the interval (0,1) and generating data for the random error variable which follow the normal distribution. Then values have been assumed for the parameters α and β of the SLRM to do comparisons between the estimators of different methods used in this work. Then, the effect of

where A_1 is defined in (18) and λ is defined in (19). Thus, what remains to calculate is just C_1. Now from (24a),

$$C_1 = a^2 \int_{-\infty}^{\infty} (\hat{\alpha} - \alpha_0)^2 e^{\frac{-2b(\hat{\alpha}-\alpha)^2}{V(\hat{\alpha})}} \frac{1}{\sqrt{2\pi V(\hat{\alpha})}} e^{-\frac{(\hat{\alpha}-\alpha)^2}{2V(\hat{\alpha})}} d\hat{\alpha}$$

$$= a^2 e^{\frac{-4b\alpha_0^2 - \alpha^2}{2V(\hat{\alpha})}} e^{\frac{(\alpha+4b\alpha_0)^2}{2(4b+1)V(\hat{\alpha})}} \int_{-\infty}^{\infty} \frac{(\hat{\alpha} - \alpha_0)}{\sqrt{2\pi V(\hat{\alpha})}} e^{\frac{-(4b+1)}{2V(\hat{\alpha})}\left[\hat{\alpha} - \frac{\alpha+4b\alpha_0}{4b+1}\right]^2} d\hat{\alpha}$$

after making change of variables, integrating by parts, using (14) and simplifying the results, we get

$$C_1 = \frac{a^2\left[(4b+1)+\lambda^2\right]V(\hat{\alpha})}{(4b+1)^{5/2}} e^{\frac{-2b\lambda^2}{4b+1}} \qquad (25)$$

Now substituting (24b), (24c) and (25) in (23), we get

$$MSE(\tilde{\alpha}) = V(\hat{\alpha})\left\{\frac{a^2}{4(b+1)^{3/2}} e^{\frac{-2b\lambda^2}{4b+1}}\left[1+\frac{\lambda^2}{4b+1}\right] + \lambda^2\left[1 - \frac{2a}{(2b+1)^{3/2}} e^{\frac{-b\lambda^2}{2b+1}}\right]\right\} \qquad (26)$$

In a similar way, we find the MSE for $\tilde{\beta}$ and the result is

$$MSE(\tilde{\beta}) = V(\hat{\beta})\left\{\frac{a^2}{(4b+1)^{3/2}} e^{\frac{-2b\lambda^2}{4b+1}}\left[1+\frac{\lambda^2}{4b+1}\right] + \lambda^2\left[1 - \frac{2a}{(2b+1)^{3/2}} e^{\frac{-b\lambda_2}{2b+1}}\right]\right\} \qquad (27)$$

Now, using the definition of the relative efficiency (R.E), then the R.E ($\tilde{\alpha}$) using (26) is

$$R.E(\tilde{\alpha}) = \frac{MSE(\hat{\alpha})}{MSE(\tilde{\alpha})}$$

$$A_2 = E(\alpha - \alpha_0) = \alpha - \alpha_0) = \lambda\sqrt{V(\hat{\alpha})} \qquad (20)$$

Substituting (18) and (20) in (17), we get

$$Bias(\tilde{\alpha}) = \lambda\sqrt{V(\hat{\alpha})}\left[\frac{\mathrm{a}}{(2\mathrm{b}+1)^{3/2}}e^{-\frac{b\lambda^2}{2b+1}} - 1\right] \qquad (21)$$

In a similar way, we get

$$Bias(\tilde{\beta}) = \lambda\sqrt{V(\hat{\beta})}\left[\frac{\mathrm{a}}{(2\mathrm{b}+1)^{3/2}}e^{\frac{-b\lambda^2}{2b+1}} - 1\right] \qquad (22)$$

Now, to find the MSE for $\tilde{\alpha}$, we just apply the definition of MSE as follows:

$$MSE(\tilde{\alpha}) = E\left[(\hat{\alpha} - \alpha_0)ae^{\frac{-b(\hat{\alpha}-\alpha_0)^2}{V(\hat{\alpha})}} + (\alpha_0 - \alpha)\right]^2$$
$$= C_1 - 2C_2 + C_3 \qquad (23)$$

where

$$C_1 = a^2E\left[(\hat{\alpha} - \alpha_0)^2 e^{\frac{-2b(\hat{\alpha}-\alpha_0)^2}{V(\hat{\alpha})}}\right] \qquad (24a)$$

$$C_2 = aE\left[(\hat{\alpha} - \alpha_0)(\alpha - \alpha_0)e^{-\frac{b(\hat{\alpha}-\alpha_0)^2}{V(\hat{\alpha})}}\right]$$
$$= a(\alpha - \alpha_0)A_1 \qquad (24b)$$

$$C_3 = E(\alpha - \alpha_0)^2 = \lambda^2V(\hat{\alpha}) \qquad (24c)$$

Now, we find the bias, MSE and RE for the estimators $\tilde{\alpha}$ and $\tilde{\beta}$. For this, we use

$$\hat{\alpha} \sim N(\alpha, V(\hat{\alpha})) \quad , \quad \hat{\beta} \sim N(\beta, V(\hat{\beta})) \qquad (16)$$

Then

$$\begin{aligned} \text{Bias } (\tilde{\alpha}) &= E(\tilde{\alpha} - \alpha) \\ &= E\left[(\hat{\alpha} - \alpha_0) a \exp\left[-b(\hat{\alpha} - \alpha_0)^2 / V(\hat{\alpha})\right] + (\alpha_0 - \alpha)\right] \\ &= a\,A_1 - A_2 \end{aligned} \qquad (17)$$

where

$$A_1 = E\left[(\hat{\alpha} - \alpha_0)\exp\left[-b(\hat{\alpha} - \alpha_0)^2 / V(\hat{\alpha})\right]\right]$$
$$A_2 = E(\alpha - \alpha_0)$$

Now,

$$A_1 = E\left[(\hat{\alpha} - \alpha_0)\exp\left[-b(\hat{\alpha} - \alpha_0)^2 / V(\hat{\alpha})\right]\right]$$

$$= \int_{-\infty}^{\infty} (\hat{\alpha} - \alpha_0) e^{-\frac{b(\hat{\alpha}-\alpha_0)^2}{V(\hat{\alpha})}} \frac{1}{\sqrt{2\pi V(\hat{\alpha})}} e^{-\frac{(\hat{\alpha}-\alpha)^2}{2V(\hat{\alpha})}} d\hat{\alpha}$$

$$= e^{\frac{-2b\alpha_0^2 - \alpha^2}{2V(\hat{\alpha})}} \int_{-\infty}^{\infty} (\hat{\alpha} - \alpha_0) \frac{1}{\sqrt{2\pi V(\hat{\alpha})}} e^{\frac{-\left[(2b+1)\hat{\alpha}^2 - 2(2b\alpha_0 + \alpha)\hat{\alpha}\right]}{2V(\hat{\alpha})}} d\hat{\alpha}$$

After making change of variables, integration by parts, using (14) and simplifying the result, we get

$$A_1 = \frac{\lambda\sqrt{V(\hat{\alpha})}}{(2b+1)^{3/2}} e^{-\frac{b\lambda^2}{2b+1}} \qquad (18)$$

where

$$\lambda = (\alpha - \alpha_0) / \sqrt{V(\hat{\alpha})} \qquad (19)$$

and

The estimations of regression parameters are similar to that given in (3), (4) and (5), and the estimation of σ^2 is

$$\hat{\sigma}^2 = (n-1)\left[\widetilde{S}_{yy} - \hat{\beta}\widetilde{S}_{xy}\right] / (n-k+1) \qquad (10)$$

3. ESTIMATION BY USING A NON–CONSTANT WEIGHTED SHRINKAGE FUNCTION.

In this section, we consider the general weighted shrinkage function $\psi(\hat{\theta})$ as a non constant weighted shrinkage function of the exponential type in the following form

$$\psi(\hat{\theta}) = a\exp\left[-b(\hat{\theta} - \theta_0)^2 / V(\hat{\theta})\right] \qquad (11)$$

where a and b are constants and $0 \le a \le 1$, $b > 0$. We know from (ii) that the shrinkage estimator $\widetilde{\theta}$ of θ is given as a linear combination using θ_0 as an initial value and $\hat{\theta}$ as an estimated value calculated from a small sample using one of the classical methods. Replacing $k(\hat{\theta})$ by $\psi(\hat{\theta})$, then (1) becomes

$$\widetilde{\theta} = (\hat{\theta} - \theta_0)\psi(\hat{\theta}) + \theta_0 \qquad (12)$$

Now, using (11) in (12), we get

$$\widetilde{\theta} = (\hat{\theta} - \theta_0)a\exp\left[-b(\hat{\theta} - \theta_0)^2 / V(\hat{\theta})\right] + \theta_0 \qquad (13)$$

Using (13), then the shrinkage estimator for $\hat{\alpha}$ and $\hat{\beta}$ are

$$\widetilde{\alpha} = (\hat{\alpha} - \alpha_0)a\exp\left[-b(\hat{\alpha} - \alpha_0)^2 / V(\hat{\alpha})\right] + \alpha_0 \qquad (14)$$

and

$$\widetilde{\beta} = (\hat{\beta} - \beta_0)a\exp\left[-b(\hat{\beta} - \beta_0)^2 / V(\hat{\beta})\right] + \beta_0 \qquad (15)$$

The use of complete observations analysis leads to considerable loss in data which then affect the estimators of the parameters in which that some of its components have been omitted. To avoid considerable loss in data, the available observations analysis method is used. The steps of this method are summarized below:

1. Computing the variance–covariance matrix of the explanatory variables.
2. Computing the vector of covariances between the response variable with each of the explanatory variable.

To estimate the means and the variance–covariance matrix, a basic condition must be satisfied which is that the mechanism of missing data has to be of MAR type, and then computing of the covariance between X_j and X_k variables using availability variables is as follows:

$$S_{jk} = \sum_{i=1}^{n_{jk}} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)/(n_{jk} - 1) \qquad (7)$$

where j,k = 1, ... , p; i = 1, ... , n_{jk}, and n_{jk} represents the number of observations where X_j and X_k appear together and

$$\bar{X}_j = \sum_{i=1}^{n_{jk}} X_{ij} / n_{jk} \quad , \quad \bar{X}_k = \sum_{i=1}^{n_{jk}} X_{ik} / n_{jk} \qquad (8)$$

and $\bar{X}_j$ and $\bar{X}_k$ are computed from the same common observations for the variables X_j and X_k. The variances of X_j and X_k are calculated as follows

$$\tilde{S}_{jj} = \sum_{i}^{n_j} (X_{ij} - \tilde{X}_j)^2 \quad , \quad \tilde{S}_{kk} = \sum_{i=1}^{n_k} (X_{ik} - \tilde{X}_k)^2 \qquad (9)$$

where

$$\tilde{X}_j = \sum_{j=1}^{n_j} X_{ij} / n_j \quad , \quad \tilde{X}_k = \sum_{i=1}^{n_k} X_{ik} / n_k \quad ,$$

and n_j is the number of complete observations for the variable X_j and n_k is the number

of complete observations for the variable X_k. Then $\tilde{X}_j$, $\tilde{X}_k$, $\tilde{S}_{jj}$ and $\tilde{S}_{kk}$ are computed using the available observations for each variable in the sample.

some numerical computation for each one of them have been made for different sample sizes.

2. OBSERVATIONS ANALYSIS

2.1. Complete Case Analysis (CC)

Yates (1932) was the first to define and to use the complete observations analysis to treat the incomplete data for the SLRM which is defined as:

$$Y_i = \alpha + \beta X_{ij} + e_i , \quad i = 1, 2, \dots , n_c ; \quad j = 1, 2, \dots , n_x \qquad (2)$$

where n_x represents the number of observations of explanatory variable and n_c represents the number of common observations of each of the explanatory variable and response variable. In order to find the estimation of the SLRM's parameters in the case of incomplete data, we use the ordinary least squares (OLS) criterion and the estimators forms are as follows:

$$\hat{\alpha} = \bar{Y}_c - \hat{\beta}\, \bar{X}_c , \qquad \hat{\beta} = S_{XcYc} / S_{XcXc} \qquad (3)$$

where

$$\bar{X}_c = \sum_{j=1}^{n_c} X_j / n_c , \quad \bar{Y}_c = \sum_{j=1}^{n_c} Y_j / n_c , \quad S_{XcYc} = \sum_{j=1}^{n_c} (X_j - \bar{X}_c)(Y_j - \bar{Y}_c) \qquad (4)$$

The $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators and have MVUE and their variances are respectively

$$V(\hat{\alpha}) = \sigma^2 \left[\frac{1}{n_c} + \frac{\bar{X}_c^2}{S_{XcXc}} \right], \qquad V(\hat{\beta}) = \sigma^2 / S_{XcXc} \qquad (5)$$

2.2. Available Case Analysis (AC)

In (1964), Glesser [4], defined the analysis of available observations method for estimating the SLRM's parameters, which is given by

$$\hat{\beta} = (\mathrm{cov}(X_k, X_j))^{-1} \mathrm{cov}(X_j, Y) \qquad (6)$$

differ for some reasons (see [1], [5], [8], [14] [17], [21]). Some of these are not done on purpose , for example, the failure of recording machines or because of natural disasters, wars or others. And some are done on purpose because of the risk or the highly cost or because of the unavailability of sources. In this case, the observations and data are incomplete (see [2], [3], [14], [18]) .

The incomplete observations data in SLRM's may miss one or more of the explanatory variables for some observations for any reason (see [7], [15], [16]). Thus, the estimators that is gotten from the incomplete data will be inefficient estimators. In this case, the statistical analysis which deal with the incomplete data must be used and accordingly accurate results can be reached.

The mechanism of missing observations differs from one sample observations to another sample observations. The missing data either are done on purpose, i.e, not missing at random (not MAR) or are not done on purpose. There are two types of the second case which are (i) missing at random (MAR), (ii) missing completely at random (MCAR) (see [6], [10], [11], [12], [19]).

In the linear regression model one or more explanatory variables may miss part of its observations with the assumption that the dependent (response)variable has complete observations. And this what is called by (incomplete observation data). In this paper, we study the SLRM when the explanatory variable miss part of its observations. Thus, this work describes two statistical methods for estimating the SLRM's parameters. These methods are (i) the complete observations and (ii) the available observations. The two methods treat the case of incomplete observations data. The second aim of this work is to combine the two methods together through a linear combination, by considering that the available observations method gives the classical estimator ($\hat{\theta}$ say), whereas the complete observations method gives the initial value (θ, say). The linear combination is given as follows:

$$\tilde{\theta} = k(\hat{\theta})\hat{\theta} + (1 - k(\hat{\theta}))\theta \qquad (1)$$

where $k(\hat{\theta})$, $0 \le k(\hat{\theta}) \le 1$ is a weighted function which represents the one stage shrinkage estimator (see [9], [13], [20]). In this paper, $k(\hat{\theta})$ is considered as a varying weighted function. For the proposed shrinkage estimator of (1), the MSE, the bias and the RE are derived along with

مقدار التقلص لانموذج الانحدار الخطي باستخدام دالة وزن متغيرة للبيانات غير تامة المشاهدات

د. سامي داود كبارة *

المستخلص :

في الواقع العملي هناك الكثير من المتغيرات التي تتضمن بيانات غير تامة المشاهدات، ويمكن وصف هذه المتغيرات بنماذج انحدار خطي بسيط وهناك عدة طرائق لمعالجة وتقدير معالم نموذج الانحدار الخطي البسيط في حالة البيانات غير تامة المشاهدات. يهدف هذا البحث الى دراسة طريقتين من طرائق تقدير معالم النموذج الخطي البسيط عندما تكون البيانات غير تامة المشاهدات ومن ثم ربط وتكوين توليفة خطية بين مقدرات هاتين الطريقتين وذلك باستخدام المقدرات المقلصة ذات المرحلة الواحدة واشتقاق متوسط مربعات الخطأ ومقدار التحيز والكفاءة النسبية لمقدر التقلص. وباستخدام المحاكاة تم حساب مقدرات الطرائق جميعا بالاضافة الى متوسط مربعات الخطأ ومقدار التحيز والكفاءة النسبية، وتوصلت الدراسة بأن مقدرات التقلص ذات المرحلة الواحدة هي افضل من بقية مقدرات الطرائق المستخدمة في البحث.

1. INTRODUCTION

Most of the statistical analysis methods have shown that the sample observations data which have been studied are complete observation data. In most of the phenomenons, part of their data are exposed to either missing or not observed. The reasons of missing or not observing the data

---

* استاذ مساعد/ الجامعة المستنصرية /كلية العلوم / قسم الرياضيات

مقبول للنشر بتاريخ 2006/2/15