Applied Mathematics and Computational Intelligence
Volume 10, No.1, Dec 2021 [301-308]

Maximum Likelihood Estimation of Replicated Linear Functional Relationship Model

Azuraini Mohd Arif[1&2], Yong Zulina Zubairi[3*] and Abdul Ghapor Hussin[4]

[1]Institute of Advanced Studies, University of Malaya, 50603 Kuala Lumpur.
[2]Centre for Defence Foundation Studies, National Defence University of Malaysia, 57000 Kuala Lumpur.
[3]Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur.
[4]Faculty of Defence Science and Technology, National Defence University of Malaysia, 57000 Kuala Lumpur.

* Corresponding author: yzulina@um.edu.my

Received: 25 October 2021; Accepted: 8 November 2021; Available online: 20 December 2021

ABSTRACT

This paper discusses the parameter estimates as well as the asymptotic covariance in replicated linear functional relationship model (LFRM). The model is assumed to be balanced and equal in each group. The maximum likelihood estimation is used to estimate four parameters in this model namely the intercept, the slope, and two error variances. Although the closed-form of the estimates is not available, it is shown that the closed-form for the asymptotic covariance matrix of the model using the Fisher Information matrix can be obtained. Using a simulation study, we showed that the estimated values of the parameters are unbiased and consistent suggesting the proposed model's superiority.

Keywords: errors-in-variable model, parameter estimation, replicated, variance-covariance matrix

1 INTRODUCTION

The functional relationship model is a family of the errors-in-variables model including the structural relationship model and the ultrastructural relationship model. Errors-in-variable models is the extension of the regression model where both variables X and Y are continuous linear and measured with errors. Measurement error can occur in many disciplines such as in econometrics, environmental sciences, engineering, manufacturing, and many others [1-3]. For example, instrument problems could occur in the industrial field due to the variation in the measuring process.

In the linear functional relationship model (LFRM), the variable X is fixed or deterministic and it can be further extended into two models namely the unreplicated and replicated linear functional relationship model. The replicated LFRM has been used to overcome the inconsistencies i.e the unidentifiability problem and also the assumption of error variance in unreplicated LFRM [4]. Extensive works on parameter estimation in the errors-in-variables model had been well explored and discussed [1], [4-7]. Several researchers have proposed a maximum likelihood estimation method in estimating the parameters in both linear and circular models [8–11]. A number of studies have discussed the asymptotic variance-covariance matrix in the errors-in-variables model [10], [12-13].

In this paper, we emphasize a balanced replicated LFRM. This model has equal observation in each group which will be discussed in the next section. Next, the maximum likelihood estimation for the balanced replicated model is given although the estimates cannot be obtained algebraically. Nevertheless, the estimates can be solved iteratively. Then, we derive the asymptotic algebraically or closed form of the variance-covariance matrix followed by a simulation study to investigate the precision of the estimated parameters and its variance-covariance matrix.

2 REPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL

A linear relationship between X_i and Y_i is given by

$$Y_i = \alpha + \beta X_i \tag{1}$$

where α is the intercept parameter and β is the slope parameter. For any fixed X_i, we assume that the observations x_{ij} and y_{ij} have been measured with errors δ_{ij} and ε_{ik} respectively. This can be written as:

$$x_{ij} = X_i + \delta_{ij} \text{ and } y_{ij} = Y_i + \varepsilon_{ij} \tag{2}$$

for $i = 1,2,\cdots,p$ and $j = 1,2,\cdots,m$. The errors terms δ_{ij} and ε_{ij} follow a normal distribution with mean zero and variance σ^2 and τ^2 respectively. Given a particular pair (X_i, Y_i), there may be replicated observations of X_i and Y_i occurring in p groups and each group has a sample size of m. In this case, the replicated model is balanced and equal.

3 MAXIMUM LIKELIHOOD ESTIMATION OF THE MODEL

In balanced replicated LFRM, the parameters to be estimated are $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}, \hat{\beta}$, and $\hat{X}_i$ although our interests are $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$. The estimation of parameters can be obtained by the maximum likelihood estimation method. By using the maximum likelihood estimation method, one can easily get the asymptotic variance-covariance matrix of the estimators. For balanced replicated LFRM, the log-likelihood function can be expressed as

$$\log L\left(\alpha,\beta,\sigma^2,\tau^2,X_1,\dots,X_p;x_{ij},y_{ij}\right) = constant\ -\frac{n}{2}(\log\sigma^2 + \log\tau^2)$$
$$-\frac{1}{2}\left\{\sum\sum\frac{\left(x_{ij}-X_i\right)^2}{\sigma^2} + \sum\sum\frac{\left(y_{ij}-\alpha-\beta X_i\right)^2}{\tau^2}\right\} \tag{3}$$

The function $\log L$ is differentiated with respect to parameters $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2, \hat{\tau}^2$ and $\hat{X}_i$.

$$\frac{\partial logL}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{p}\sum_{j=1}^{m}\left(x_{ij}-X_i\right)^2}{2\sigma^4} \tag{4}$$

$$\frac{\partial logL}{\partial \tau^2} = -\frac{n}{2\tau^2} + \frac{\sum_{i=1}^{p}\sum_{j=1}^{m}\left(y_{ij} - \alpha - \beta X_i\right)^2}{2\tau^4} \tag{5}$$

$$\frac{\partial logL}{\partial \alpha} = \frac{1}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m}\left(y_{ij} - \alpha - \beta X_i\right) \tag{6}$$

$$\frac{\partial logL}{\partial \beta} = \frac{X_i}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m}\left(y_{ij} - \alpha - \beta X_i\right) \tag{7}$$

$$\frac{\partial logL}{\partial X_i} = \frac{1}{\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{m}\left(x_{ij} - X_i\right) + \frac{\beta}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m}\left(y_{ij} - \alpha - \beta X_i\right) \tag{8}$$

By setting (4) until (8) to zero and simplifying, we obtain the estimate of $\hat{\sigma}^2$, $\hat{\tau}^2$, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{X}_i$ as follows:

$$\hat{\sigma}^2 = \frac{\sum\sum\left(x_{ij} - \hat{X}_i\right)^2}{\sum m} \tag{9}$$

$$\hat{\tau}^2 = \frac{\sum\sum\left(y_{ij} - \hat{\alpha} - \hat{\beta}\hat{X}_i\right)^2}{\sum m} \tag{10}$$

$$\hat{\alpha} = \frac{\sum m\left(\bar{y}_{i.} - \hat{\beta}\hat{X}_i\right)}{\sum m} \tag{11}$$

$$\hat{\beta} = \frac{\sum m\hat{X}_i(\bar{y}_{i.} - \hat{\alpha})}{\sum m\hat{X}_i^2} \tag{12}$$

$$\hat{X}_i = \frac{1}{\hat{\Delta}}\left\{\frac{m\bar{x}_{i.}}{\hat{\sigma}^2} + \frac{m\hat{\beta}}{\hat{\tau}^2}(\bar{y}_{i.} - \hat{\alpha})\right\} \tag{13}$$

where $\bar{x}_{i.}$, $\bar{y}_{i.}$ are sample means for each group and $\hat{\Delta} = \frac{m}{\hat{\sigma}^2} + \frac{m\hat{\beta}^2}{\hat{\tau}^2}$.

The estimates of (9) until (12) are dependent on $\hat{X}_i$ (13) which suggests that there is no closed-form available. Thus, the estimates may be obtained iteratively and starting values for the iteration can be chosen using parameters from unreplicated LFRM by assuming $\lambda = 1$ until all parameters converge[14]. This iteration procedure will continue until all parameters converge.

4 VARIANCE-COVARIANCE MATRIX OF THE MODEL

The asymptotic properties of $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$ can be found by inverting the estimated Fisher information matrix for balanced replicated LFRM. Next, the second derivative for the log-likelihood function is obtained followed by their negatives expected values. This is given by:

$$\frac{\partial^2 logL}{\partial(\sigma^2)^2} = \frac{\sum_{i=1}^{p} m}{2\sigma^4} - \frac{\sum_{i=1}^{p}\sum_{j=1}^{m}\left(x_{ij} - X_i\right)^2}{\sigma^6}. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial(\sigma^2)^2}\right] = \frac{mp}{2\sigma^4}$$

$$\frac{\partial^2 logL}{\partial(\tau^2)^2} = \frac{\sum_{i=1}^{p} m}{2\tau^4} - \frac{\sum_{i=1}^{p}\sum_{j=1}^{m}\left(y_{ij} - \alpha - \beta X_i\right)^2}{2\tau^6}. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial(\tau^2)^2}\right] = \frac{mp}{2\tau^4}$$

$$\frac{\partial^2 logL}{\partial\alpha^2} = -\frac{1}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m}(1). \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial\alpha^2}\right] = \frac{mp}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial\beta^2} = -\frac{1}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m} X_i^2. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial\beta^2}\right] = \frac{m\sum_{i=1}^{p} X_i^2}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial\alpha\partial\beta} = -\frac{1}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m} X_i. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial\alpha\partial\beta}\right] = \frac{m\sum_{i=1}^{p} X_i}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial X_i^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{m}(1) - \frac{\beta^2}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{m}(1). \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i^2}\right] = \frac{m}{\sigma^2} + \frac{m\beta^2}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial X_i\partial\alpha} = -\frac{\sum_{i=1}^{p}\sum_{j=1}^{m}\beta}{\tau^2}. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i\partial\alpha}\right] = \frac{\sum_{i=1}^{p}\sum_{j=1}^{m} m_i\beta}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial X_i\partial\beta} = -\frac{\sum_{i=1}^{p}\sum_{j=1}^{m} m_i X_i\beta}{\tau^2}. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i\partial\beta}\right] = \frac{\sum_{i=1}^{p}\sum_{j=1}^{m} m_i X_i\beta}{\tau^2}$$

$$\frac{\partial^2 logL}{\partial X_i X_j} = 0. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i X_j}\right] = 0 \text{ for } i \neq j.$$

$$\frac{\partial^2 logL}{\partial X_i\partial\sigma^2} = 0. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i\partial\sigma^2}\right] = 0 \quad \text{and}$$

$$\frac{\partial^2 logL}{\partial X_i\partial\tau^2} = 0. \text{ Hence, } E\left[-\frac{\partial^2 logL}{\partial X_i\partial\tau^2}\right] = 0$$

Next, the estimated Fisher information matrix, F, for $\hat{X}_1, \cdots, \hat{X}_p, \hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$ is given by

$$F = \begin{bmatrix} B & 0 & E \\ 0 & C & 0 \\ E^T & 0 & D \end{bmatrix}$$

where B, C and D are a square matrix with sizes p, 2 and 2 respectively and E is a $p \times 2$ matrix.

The asymptotic covariance matrix of our interest, $\hat{\sigma}^2, \hat{\tau}^2$, $\hat{\alpha}$ and $\hat{\beta}$ is the bottom right minor of order $4 \times p$ of the inverse of matrix F. From the theory of partitioned matrices, [15], this is given by,

$$\widehat{Var}\begin{bmatrix}\hat{\sigma}^2\\ \hat{\tau}^2\\ \hat{\alpha}\\ \hat{\beta}\end{bmatrix} = \begin{bmatrix} C^{-1} & 0 \\ 0 & (D - E^T B^{-1} E)^{-1}\end{bmatrix}$$

After lengthy algebraic manipulation, the asymptotic covariance matrix for $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$ is given by

$$M = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} \tag{14}$$

where $a_{11} = 2\sigma^4/n$, $a_{22} = 2\tau^4/n$, $a_{33} = Q\sum_{i=1}^{p} X_i^2$, $a_{34} = -Q\sum_{i=1}^{p} X_i$, $a_{44} = Qp$ and $a_{43} = -Q\sum_{i=1}^{p} X_i$ and $Q = \frac{m\tau^2 + m\beta^2\sigma^2}{m^2\left\{p\sum_{i=1}^{p} X_i^2 - \left(\sum_{i=1}^{p} X_i\right)^2\right\}}$. Other elements in matrix M are 0.

5 SIMULATION STUDY

A simulation study has been carried out using the R software to evaluate the performance of the parameters of the balanced replicated LFRM. Two factors for performance to consider namely the value of errors variances and the sample sizes. Without loss of generality, we fixed the true value of $\alpha = 0$ and choose different true values of β, σ^2 and τ^2. The sample size, n are 50,100 and 180 with p-subgroups of 5, 10, and 12 respectively. These sample sizes are chosen in such a way that represents the small and large datasets. The details of the algorithm can be described as follows:

Step 1: Generate two random error terms δ_{ij} and ε_{ij} from $N(0, \sigma^2)$ and $N(0, \tau^2)$ respectively with $i = 1,2, \ldots, p$ where p is the number of group and $j = 1,2, \ldots, m$ where m is the number of elements in each subgroup.

Step 2: Generate the data using (1) and (2) in which data of x_{ij} and y_{ij} variables are of equal sample size. The values of x_{ij} and y_{ij} are divided into p-subgroups with m elements such that $p \times m = n$.

Step 3: Solve the parameters of interest iteratively using (9) until (13), The possible initial estimates for $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ and $\hat{\tau}^2$ can be obtained from unreplicated LFRM with the assumption $\lambda = 1$ to start the iteration process.

Step 4: Calculate the variance-covariance matrix of the parameters using (14).

Step 5: Repeat the steps for 5000 simulations.

The performance of the estimated parameters is measured by estimated bias, mean square error, and standard deviation. The estimated bias and the mean square are given by:

EB= $|\hat{b} - b|$ and MSE= $\frac{1}{s}\sum(\hat{b}_j - b)^2$

with b be a generic term for the parameters and s is the number of simulation. The standard deviation for parameters is calculated from the diagonal element of the asymptotic variance-covariance matrix.

Table 1 presents a result from the simulation study for $\alpha = 0$ and $\beta = 1$ with different sets of σ^2 and τ^2. It can be seen from the data in Table 1, when the value $\sigma^2 < \tau^2$, the estimated bias for estimated parameters, $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ and $\hat{\tau}^2$ are consistently small and approximately close to 0 when the sample size is increased from 50 to 180. This shows the unbiasedness of the parameters. The mean square error of each parameter estimate tends to decrease with the increase in sample sizes. Furthermore, the standard deviation is generally small for all parameter estimates. Similar trends can also be observed when $\sigma^2 > \tau^2$, such as these values tend to decrease with the increase in sample sizes which show that the estimated values of parameters are unbiased and consistent.

Table 1. Results for $\alpha = 0$ and $\beta = 1$ with different sets of (σ^2, τ^2) with n is the sample size

Statistics	**Sample size, n**	$\sigma^2 = 0.8$ **and** $\tau^2 = 1$				$\sigma^2 = 1$ **and** $\tau^2 = 0.8$			
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
Estimated Bias	50	0.0117	0.0024	0.0550	0.0659	0.0137	0.0027	0.0618	0.0582
	100	0.0088	0.0014	0.0499	0.0529	0.0089	0.0015	0.0527	0.0500
	180	0.0010	0.0004	0.0315	0.0351	0.0004	0.0004	0.0329	0.0332
Mean Square Error	50	0.1995	0.0045	0.0279	0.0422	0.1997	0.0045	0.0430	0.0275
	100	0.0881	0.0023	0.0153	0.0227	0.0877	0.0023	0.0229	0.0151
	180	0.0466	0.0012	0.0081	0.0123	0.0464	0.0012	0.0122	0.0081
Standard Deviation	50	0.4299	0.0648	0.1490	0.1868	0.4306	0.0649	0.1876	0.1484
	100	0.2812	0.0453	0.1061	0.1339	0.2814	0.0453	0.1340	0.1061
	180	0.2090	0.0341	0.0810	0.1017	0.2091	0.0341	0.1019	0.0808

Table 2 shows result from the simulation study for $\alpha = 0$ and $\beta = 1.2$ with different sets of σ^2 and τ^2. These results are similar to Table 1 i.e when the sample size is increased, the value of the estimated bias, the mean square error and the standard deviation also decreased. These results clearly show that the estimated values of parameters are not so bias and consistent.

Table 2. Results for $\alpha = 0$ and $\beta = 1.2$ with different sets of (σ^2, τ^2) with n is the sample size

Statistics	Sample size, n	$\sigma^2 = 0.8$ and $\tau^2 = 1$				$\sigma^2 = 1$ and $\tau^2 = 1$			
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
Estimated Bias	50	0.0146	0.0029	0.0504	0.0715	0.0177	0.0034	0.0597	0.0749
	100	0.0104	0.0017	0.0436	0.0608	0.0115	0.0019	0.0498	0.0655
	180	0.0018	0.0006	0.0272	0.0403	0.0017	0.0006	0.0308	0.0435
Mean Square Error	50	0.2383	0.0054	0.0276	0.0428	0.2707	0.0061	0.0428	0.0432
	100	0.1049	0.0027	0.0148	0.0234	0.1188	0.0031	0.0227	0.0239
	180	0.0556	0.0015	0.0078	0.0126	0.0629	0.0017	0.0121	0.0129
Standard Deviation	50	0.4704	0.0709	0.1499	0.1857	0.5017	0.0756	0.1881	0.1850
	100	0.3076	0.0496	0.1070	0.1328	0.3278	0.0528	0.1344	0.1322
	180	0.2286	0.0373	0.0815	0.1012	0.2436	0.0397	0.1022	0.1008

6 CONCLUSION

For balanced replicated LFRM, the estimated parameters namely the error variances, $\hat{\sigma}^2$ and $\hat{\tau}^2$, the intercept, $\hat{\alpha}$ and the slope, $\hat{\beta}$ can be obtained iteratively using the maximum likelihood estimation method. Although the closed-form of parameter estimation is not available, the variance-covariance matrix can be obtained using the Fisher information matrix and partitioned matrix. Taken together, the results from simulation study suggests the maximum likelihood estimation method performs well in estimating the parameters of the balanced replicated linear functional relationship model by showing that the estimated parameters are unbiased and consistent. Further research could be focusing on the unbalanced and unequal replicated linear functional relationship model.

ACKNOWLEDGEMENT

We are most grateful to the University of Malaya and the National Defence University of Malaysia for supporting this work.

REFERENCES

[1] J. P. Buonaccorsi, *Measurement Error Models, Methods and Applications*. New York: Chapman and Hall, 2010.

[2] N. Doganaksoy and H. Van Meer, "An Application of the Linear Errors-in-Variables Model in Semiconductor Device Performance Assessment," *Qual. Eng.*, vol. 27, no. 4, pp. 500–511, 2015.

[3] Y. Hu and T. Wansbeek, "Measurement error models: Editors' introduction," *J. Econom.*, vol. 200, no. 2, pp. 151–153, 2017.

[4] M. G. Kendall and A. Stuart, *The Advanced theory of statistics. vol. 2: Inference and relationship*, 4th ed. London, 1979.

[5] D. V. Lindley, "Regression lines and the linear functional relationship," *Suppl. to J. R. Stat. Soc.*, vol. 9, no. 2, pp. 218–244, 1947.

[6] C. Villegas, "Maximum Likelihood Estimation Of A Linear Functional Relationship," *Ann. Math. Stat.*, vol. 32, pp. 1048–1062, 1961.

[7] W. A. Fuller, *Measurement Error Models*. John Wiley & Sons, 1987.

[8] V. D. Barnett, "Fitting straight lines-the linear functional relationship with replicated observations," *Appl. Stat.*, vol. 19, no. 2, pp. 135–144, 1970.

[9] L. K. Chan and T. K. Mak, "Maximum likelihood estimation of a linear structural relationship with replication," *J.R. Stat. Soc. Ser. B*, vol. 41, no. 2, pp. 263–268, 1979.

[10] A. G. Hussin, "Approximating Fisher's Information for the Replicated Linear Circular Functional Relationship Model," *Bull. Malaysian Math. Sci. Soc.*, vol. 28, no. 2, pp. 131–139, 2005.

[11] N. A. Mokhtar, Y. Z. Zubairi, A. G. Hussin, and R. M. Yunus, "On parameter estimation of a replicated linear functional relationship model for circular variables," *Matematika*, vol. 33, no. 2, pp. 159–163, 2017.

[12] A. G. Hussin, A. Abuzaid, F. Zulkifili, and I. Mohamed, "Asymptotic covariance and detection of influential observations in a linear functional relationship model for circular data with application to the measurements of wind directions," *ScienceAsia*, vol. 36, no. 3, pp. 249–253, Sep. 2010.

[13] A. S. M. Al Mamun, A. G. Hussin, Y. Z. Zubairi, and A. H. M. R. Imon, "Maximum likelihood estimation of linear structural relationship model parameters assuming the slope is known," *Scienceasia*, vol. 39, no. 5, pp. 561–565, 2013.

[14] A. Mohd Arif, Y. Z. Zubairi, and A. G. Hussin, "Parameter estimation in replicated linear functional relationship model in the presence of outliers," *Malaysian J. Fundam. Appl. Sci.*, vol. 16, no. 2, pp. 158–160, 2020.

[15] F. A. Graybill, *An Introduction to Linear Statistical Models*. New York: McGraw-Hill, 1961.