MAXIMUM LIKELIHOOD ESTIMATION OF REPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL

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

AZURAINI MOHD ARIF

1 Institute of Advanced Studies, [3] Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia,

[2] Centre for Foundation Studies, [4] Faculty of Defence Science and Technology, National Defence University of Malaysia, 57000 Kuala Lumpur, Malaysia

4 - 5 AUGUST 2021

CONTENTS

Introduction

- Errors-in-variables model :
  - functional relationship model (LFRM),
  - structural relationship model and
  - ultrastructural relationship model.
- Measurement error can occur in many disciplines such as in econometrics, environmental sciences, engineering, manufacturing, and many others (Buonaccorsi, 2010; Doganaksoy and Van Meer, 2015; Hu and Wansbeek, 2017).
- Linear functional relationship model
  - the variable X is fixed or deterministic
  - can be divided into unreplicated and replicated LFRM with certain recommendations

Objectives

- Discuss the parameter estimates as well as the asymptotic covariance in replicated linear functional relationship model (LFRM).

- Emphasize on a balanced replicated LFRM
  - Review the maximum likelihood estimation for the balanced replicated model
  - Derive the asymptotic closed-form of the variance-covariance matrix
  - Investigate the precision of the estimated parameters and its variance-covariance matrix by simulation study

Literature Review

Findings	Authors
the unidentifiability problem ;the assumption on the ratio of error variance in unreplicated LFRM	Lindley, 1947; Kendall and Stuart, 1979
parameter estimation in the errors-in-variables model	Lindley, 1947; Villegas, 1961; Kendall and Stuart, 1979; Fuller, 1987; Buonaccorsi, 2010
maximum likelihood estimation method in estimating the parameters in both linear and circular models	Barnett, 1970; Chan and Mak, 1979; Hussin et al., 2005; Mokhtar et al., 2017
derived the asymptotic variance-covariance matrix in the errors-in-variables model	Hussin, 2005; Hussin et al., 2010; Mamun et al., 2013

Motivation: The balanced replicated LFRM can be used to overcome the inconsistencies i.e the unidentifiability problem and also the assumption on the ratio of error variance in unreplicated LFRM.

Methodology

- ❑ Replicated Linear Functional Relationship Model
- ❑ Maximum Likelihood Estimation of the Model
- ❑ Variance-covariance Matrix of the Model
- ❑ Simulation Study

Replicated Linear Functional Relationship Model

- A linear relationship between X_i and Y_i are given by

$$Y_i = \alpha + \beta X_i$$

$\alpha \Rightarrow$ intercept parameter

$\beta \Rightarrow$ slope parameter

- x_{ij} and y_{ij} are subject to random errors δ_{ij} and ε_{ij}

$$x_{ij} = X_i + \delta_{ij} \quad \Rightarrow \quad \delta_{ij} \sim N(0, \sigma^2)$$

$$y_{ij} = Y_i + \varepsilon_{ik} \quad \Rightarrow \quad \varepsilon_{ik} \sim N(0, \tau^2)$$

for $\mathrm{i} = 1,2,\cdots,p$, $j = 1,2,\cdots,m$ and $k = 1,2,\cdots,m$

- The replicated model is balanced and equal.

Maximum Likelihood Estimation of the Model

- Maximum Likelihood Estimation method which involves an iterative technique.
- The log likelihood function can be expressed as

$$\log L\left(\alpha,\beta,\sigma^2,\tau^2,X_1,\dots,X_p\right) = constant - \frac{1}{2}\sum_{i=1}^{p} m(\log\sigma^2 + \log\tau^2) - \frac{1}{2}\left\{\sum_{i=1}^{p}\sum_{j=1}^{m}\frac{\left(x_{ij}-X_i\right)^2}{\sigma^2} + \sum_{i=1}^{p}\sum_{j=1}^{m}\frac{\left(y_{ij}-\alpha-\beta X_i\right)^2}{\tau^2}\right\}$$

Maximum Likelihood Estimation of the Model

- There are $(p+4)$ parameters to be estimated and can be obtained by differentiating the log-likelihood function with respect to $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2, \hat{\tau}^2$ and $\hat{X}_i$:

$$\hat{X}_i = \frac{1}{\hat{\Delta}}\left\{\frac{m\bar{x}_{i.}}{\hat{\sigma}^2} + \frac{m\hat{\beta}}{\hat{\tau}^2}(\bar{y}_{i.} - \hat{\alpha})\right\}, \quad \hat{\sigma}^2 = \frac{\sum\sum\left(x_{ij}-\hat{X}_i\right)^2}{\sum m}, \quad \hat{\tau}^2 = \frac{\sum\sum\left(y_{ij}-\hat{\alpha}-\hat{\beta}\hat{X}_i\right)^2}{\sum m},$$

$$\hat{\alpha} = \frac{\sum m\left(\bar{y}_{i.}-\hat{\beta}\hat{X}_i\right)}{\sum m}, \quad \hat{\beta} = \frac{\sum m\hat{X}_i(\bar{y}_{i.}-\hat{\alpha})}{\sum m\hat{X}_i^2}$$

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

- The parameters can be solved iteratively by using unreplicated LFRM

Variance-covariance Matrix of the Model

- By inverting the estimated Fisher information matrix for balanced replicated LFRM.
- The second derivative for the log-likelihood function is obtained followed by their negatives expected values.
- The estimated Fisher information matrix, *F*, for $\hat{X}_1, \cdots, \hat{X}_p, \hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$:

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

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.

Variance-covariance Matrix of the Model

- From the theory of partitioned matrices, (Graybill, 1961), this is given by,

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

- The asymptotic covariance matrix for $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$ and $\hat{\beta}$ is given by

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

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 is 0.

Simulation Study

- To evaluate the performance of the parameters.
- Fixed the true value of $\alpha = 0$ and 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.
- The values of x_{ij} and y_{ij} are divided into p-subgroups with m elements such that $p \times m = n$.
- The parameters of interest can be solved iteratively.
- The variance-covariance matrix of the parameters can be obtained.
- The steps are repeated for 5000 simulations.
- The performance of the estimated parameters is measured by estimated bias, mean square error, and standard deviation.

Simulation Results and Discussion

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	(0.8,1)				(1,0.8)			
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
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

- The estimated bias for estimated parameters, $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ and $\hat{\tau}^2$ become smaller and are 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 and the standard deviation also show similar trends
- The estimated values of parameters are consistent.

Simulation Results and Discussion

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	(0.8,1)				(1,1)			
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
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

- 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 unbiased and consistent.

Conclusion

- The estimated parameters, $\hat{\sigma}^2$, $\hat{\tau}^2$, $\hat{\alpha}$ and $\hat{\beta}$ can be obtained iteratively using the maximum likelihood estimation method.
- The variance-covariance matrix can be obtained using the Fisher information matrix and partitioned matrix.
- The results from simulation study suggest that the estimated parameters unbiased and consistent.

THANK YOU