Outlier Detection in Balanced Replicated Linear Functional Relationship Model
(Pengesanan Data Terpencil dalam Model Hubungan Fungsian Linear Bereplika Seimbang)

AZURAINI MOHD ARIF, YONG ZULINA ZUBAIRI* & ABDUL GHAPOR HUSSIN

ABSTRACT
Identification of outlier in a dataset plays an important role because their existence will affect the parameter estimation. Based on the idea of COVRATIO statistic, we modified the procedure to accommodate for replicated linear functional relationship model (LFRM) in detecting the outlier. In this replicated model, we assumed the observations are equal and balanced in each group. The derivation of covariance matrices using Fisher Information Matrices is also given for balanced replicated LFRM. Subsequently, the cut-off points and the power of performance are obtained via a simulation study. Results from the simulation studies suggested that the proposed procedure works well in detecting outliers for balanced replicated LFRM and we illustrate this with a practical application to a real data set. The implication of the study suggests that with some modification to the procedures in COVRATIO, one could apply such a method to identify outliers when modelling balanced replicated LFRM which has not been explored before.

Keywords: Covariance matrix; covratio; influential observation; linear functional relationship model; outliers

INTRODUCTION
The linear functional relationship model (LFRM) is one of the families in the error-in-variables model (EIVM) besides the linear structural relationship model and ultrastructural relationship model. Many authors have considered LFRM over the years in fitting the parameters (Barnett 1970; Cheng & Van Ness 1994; Kendall & Stuart 1979; Lindley 1947). Furthermore, the LFRM can be extended to unreplicated and replicated functional relationship models, with certain recommendation (Dorff & Gurland 1961). In unreplicated LFRM, the assumption on the ratio of error variances, \( \lambda = \frac{\sigma^2}{\tau^2} \), is needed to estimate the parameters and usually equal to 1. However, in the absence of knowledge on the ratio of error variances, \( \lambda \), the data can be transformed into pseudo replicates which is called as replicated and used maximum likelihood method to estimate all parameters (Hussin et al. 2005).

However, the estimation of the parameter becomes inconsistent when outliers occur in the dataset. Outliers are observations in the dataset which follow unusual patterns and occur because of gross measurement and recording errors (Aggarwal 2013). As mentioned by Hampel et al. (1986), “A routine data set typically contains about 1-10% outliers and even the highest quality data set cannot be guaranteed free of outliers”. Hence, to evaluate their influence on the model, it is important to
locate the outliers. The presence of outliers changes the parameter estimates for example in the linear regression for circular variables when estimating the parameters on a wind direction dataset (Hussin et al. 2013; Rambli et al. 2015) and also on an eye dataset (Alkasadi et al. 2019).

Extensive works on outlier detection have been well established in linear models (Cheng & Van Ness 1994; Ibrahim et al. 2013; Satman et al. 2021; Wong 1989). Applications using COVRATIO statistics can be seen in physiological, epidemiology, medicine and many different disciplines (Alcaraz-Ibáñez et al. 2021; Satari & Khalif 2020; Viechtbauer & Cheung 2010). Several researchers proposed a group deleted version to identify outliers based on the COVRATIO statistic because this procedure is simple, widely used and had been well established in both linear and circular regression modelling (Belsley et al. 1980; Rambli et al. 2016). As for errors-in-variable models, outlier detection using COVRATIO statistics has been developed for the linear model (Ghapor et al. 2014; Mamun et al. 2019) as well as for circular variables (Hussin et al. 2010; Mokhtar et al. 2019). Additionally, the COVRATIO method has been used in detecting the outlier in unreplicated linear functional relationship model (Ghapor et al. 2014). However, works on identifying outliers in replicated linear functional relationship model are somewhat limited; this could largely due to the complexity of the model with little use in modelling real data sets. Nevertheless, this does not deem the model less important; in fact, this is the motivation of this study. Thus, in this article, we propose the COVRATIO statistic in detecting a single outlier in balanced replicated LFRM and investigate on the suitability of the procedure.

This paper is organized as follows: Firstly, we review the replicated linear functional relationship model and derive the covariance matrix of the balanced replicated LFRM. Secondly, we present the COVRATIO statistic in identifying the outlier. Thirdly, the procedures are described by simulation study to determine the cut-off point. Next, the performance of the statistic proposed is investigated. Finally, we illustrate the detection of the outlier using examples of some data.

**REPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL**

In replicated LFRM, suppose \( x_{ij} \) and \( y_{ij} \) are the observed values of the linear variables \( X_i \) and \( Y_i \). For any fixed \( X_i \), we assume that may be replicated observations of \( X_i \) and \( Y_i \) occurring in \( p \) groups and measured with errors \( \delta_{ij} \) and \( \epsilon_{ij} \), respectively. This can be written as

\[
X_{ij} = X_i + \delta_{ij} \quad \text{and} \quad y_{ij} = Y_i + \epsilon_{ij},
\]

where \( \alpha \) is the intercept and \( \beta \) is the slope parameters, respectively. We assume the error terms \( \delta_{ij} \) and \( \epsilon_{ij} \) follow a normal distribution where \( \delta_{ij} \sim N(0, \sigma^2) \) and \( \epsilon_{ij} \sim N(0, \tau^2) \), respectively (Mohd Arif et al. 2020). For balanced replicated LFRM, the log-likelihood function can be expressed as

\[
\log \log L\left( \alpha, \beta, \sigma^2, \tau^2, X_1, \ldots, X_p \right) = \text{constant} - \frac{1}{2} \sum_{i=1}^{p} m \left( \log \log \sigma^2 + \log \log \tau^2 \right) - \frac{1}{2} \sum_{i=1}^{p} \frac{(x_{ij} - \bar{X}_i)^2}{\sigma^2} - \frac{1}{2} \sum_{i=1}^{p} \frac{(y_{ij} - \bar{Y}_i)^2}{\tau^2} \tag{2}
\]

There are \( p + 4 \) parameters to be estimated and can be obtained by the first partial derivative of the log-likelihood function as given in (2) with respect to \( \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2, \hat{\tau}^2 \) and \( \hat{X}_i \) respectively, and equating to zero (Barnett 1970). From Barnett (1970), by setting the equal size in each group, \( m \), the estimates of \( \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2, \hat{\tau}^2 \) and \( \hat{X}_i \) can be solved iteratively by setting initial estimates from the unreplicated linear functional relationship model by assuming \( \lambda = 1 \) or \( \sigma^2 = \tau^2 \) or until all parameters converge. Thus, we can obtain the parameters in the order of \( \hat{X}_i \) followed by \( \hat{\sigma}^2 \) or \( \hat{\tau}^2 \) and lastly \( \hat{\alpha} \) or \( \hat{\beta} \) as given by

\[
\hat{X}_i = \frac{1}{\hat{\lambda}} \left( \frac{m^2 \hat{\alpha}}{\hat{\sigma}^2} + \frac{m \hat{\beta}}{\hat{\tau}^2} \bar{Y}_i \right) \tag{3},
\]

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} (x_{ij} - \hat{X}_i)^2}{\sum_{i=1}^{p} m}, \quad \hat{\tau}^2 = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} (y_{ij} - \hat{\alpha} - \hat{\beta} \hat{X}_i)^2}{\sum_{i=1}^{p} m},
\]

\[
\hat{\alpha} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} x_{ij} (y_{ij} - \hat{\beta} \hat{X}_i)}{\sum_{i=1}^{p} m}, \quad \hat{\beta} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} x_{ij} (y_{ij} - \hat{\alpha})}{\sum_{i=1}^{p} \sum_{j=1}^{m} x_{ij}^2} \tag{4}.
\]
where $\bar{x}_i = \frac{\sum y_{ij}}{m_i}, \bar{y}_i = \sum y_{ik}$ and $\Delta_i = m_i + m_i \beta^2$. Thus, for balanced replicated LFRM, all parameters can be estimated.

Next, we derived the asymptotic covariance of parameters for balanced replicated LFRM using the Fisher Information matrix. By considering the first partial derivative and minus the expected value of the second partial derivative of the log-likelihood function, we obtain the estimated Fisher information matrix, $\mathbf{F}$, for $x_1, \ldots, x_p, \tau^2, \sigma^2, \alpha, \beta$ and $\hat{\theta}$ given by

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

where $B$ is a $p \times p$ diagonal matrix having same elements equal to $\frac{m_i}{\sigma^2} + \frac{m_i \beta^2}{\tau^2}$ while $E$ is a $p \times 2$ matrix given by

$$E = \begin{bmatrix} m_i \beta^2 & m_i \beta \tau^2 \\ \tau^2 & \tau^2 \\ m_i \beta^2 & m_i \beta \tau^2 \end{bmatrix}$$

and $C$ is a $2 \times 2$ matrix given by

$$C = \begin{bmatrix} \frac{n}{2 \alpha^4} & 0 \\ 0 & \frac{2 \alpha^4}{n} \end{bmatrix}$$

and $D$ is a $2 \times 2$ matrix given by $D = \begin{bmatrix} m_p \tau^2 & m \sum_{i=1}^p x_i^2 \\ m \sum_{i=1}^p x_i^2 & m \sum_{i=1}^p x_i^2 \end{bmatrix}$ respectively.

The asymptotic covariance matrix of $\hat{\sigma}^2, \hat{\tau}^2, \hat{\alpha}$, and $\hat{\beta}$ is the bottom right minor of order $4 \times 4$ of the inverse of matrix $\mathbf{F}$ which is our main interest. From the theory of partitioned matrices (Graybill 1961), this is given by,

$$\text{Var} \left[ \begin{bmatrix} \hat{\sigma}^2 \\ \hat{\tau}^2 \\ \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \right] = \begin{bmatrix} C^{-1} & 0 \\ 0 & (D - E^TB^{-1}E)^{-1} \end{bmatrix}$$

It can be shown that for $D - E^TB^{-1}E)^{-1}$ is as follows:

$$(D - E^TB^{-1}E)^{-1} = \frac{m \tau^2 + m \beta^2 \alpha^2}{m^2 \left\{ \sum_{i=1}^p x_i^2 - \left( \sum_{i=1}^p x_i \right)^2 \right\}^2} \left( \sum_{i=1}^p x_i \right)$$

In particular, we have the following results:

$$\text{Var} \left[ \hat{\alpha} \right] = \frac{\left( m \tau^2 + m \beta^2 \alpha^2 \right) \sum_{i=1}^p x_i^2}{m^2 \left\{ \sum_{i=1}^p x_i \right\}^2}$$

The COVRATIO statistic has been used in detecting the outliers in the linear regression matrices (Graybill 1961), this is given by,

$$\text{COVRATIO} = \frac{\text{covariance matrix for a full data set and a reduced data set (Belsley et al. 1980). This is given}}{\text{determinant for the reduced data set by excluding the observation. If the ratio is close}}$$

In particular, we have the following results:

$$\text{Var} \left[ \hat{\beta} \right] = \frac{\left( m \tau^2 + m \beta^2 \alpha^2 \right) \sum_{i=1}^p x_i^2}{m^2 \left\{ \sum_{i=1}^p x_i \right\}^2}$$

COVRATIO STATISTIC FOR BALANCED REPLICATED LFRM

The COVRATIO statistic has been used in detecting the outliers in the linear regression model and also in unreplicated LFRM (Belsley et al. 1980; Ghapor et al. 2014). The idea of COVRATIO statistic is based on the determinant ratio between determinant of the covariance matrix for a full data set and a reduced data set (Belsley et al. 1980). This is given by

$$| \text{COVRATIO}_{(-i)} - 1 | = \frac{| \text{COV} |}{| \text{COV}_{(-i)} |}$$

where $| \text{COV} |$ is the determinant of the covariance matrix for a full data set and $| \text{COV}_{(-i)} |$ is the determinant of the reduced data set by excluding the $i^{th}$ observation. If the ratio is close to one, then the $i^{th}$ observation is consistent with other observations. In this study, the idea is extended to the replicated LFRM in which the cut-off values and formula will be derived.

For balanced replicated LFRM, the ratio of statistic is suggested by a slightly different procedure. The proposed procedure is based on the determinant of the asymptotic covariance of the parameters from (8). Although the $| \text{COV} |$ is the determinant of the covariance matrix for a full data set, the $| \text{COV}^{*}_{(-i)} |$ is obtained by deleting $i^{th}$ observation of every group and this observation is repeated with the mean of every group to make balanced replication of all sample groups. The use of mean substitution may be based on the fact that the mean is a reasonable guess of a value for a randomly selected observation from a normal distribution (Acocck 2005). This is given by

$$| \text{COVRATIO}_{(-i)} - 1 | = \frac{| \text{COV}^{*} |}{| \text{COV}^{*}_{(-i)} |}$$

where $| \text{COV}^{*}_{(-i)} |$ is a determinant of the covariance matrix by the proposed method. Any $| \text{COVRATIO}_{(-i)} - 1 |$ with observation exceeds the cut-off points will be considered as an outlier. As mentioned earlier, for the balanced replicated LFRM, the cut-off points are obtained through the simulation studies by following the idea from (Ghapor et al. 2014; Mamun et al. 2019).
TABLE 1. Values of sample size, group and elements

<table>
<thead>
<tr>
<th>Sample size, n</th>
<th>Group, p</th>
<th>Elements, m</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>60</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>80</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>132</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>180</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>300</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

DETERMINATION OF CUT-OFF POINTS FOR COVRATIO STATISTIC

We carried out a simulation study to obtain the cut-off points of COVRATIO statistic for balanced replicated LFRM. Eight different sample sizes, \( n = 20, 40, 60, 80, 100, 132, 180 \) and 300 are used according to the division of sample size in Table 1.

Furthermore, five values of \( r^2 = 0.2, 0.4, 0.6, 0.8, \) and 1.0, respectively, are used. For each sample of size \( n \) and \( r^2 \), a set of normal random errors are generated from the normal distribution with mean 0 and variance \( r^2 \), respectively. By adopting the steps suggested by Ghapor et al. (2014), the procedure of COVRATIO statistic in Step 6 was slightly modified to accommodate for balanced replicated LFRM. The steps are listed down in detail:

**Step 1** Generate a fixed \( X = 10(\frac{1}{p}) \) of size \( p \), with \( i = 1, 2, \ldots, p \) where \( p \) is the number of groups. Without loss of generality, the intercept, slope and error variance parameters of balanced replicated LFRM are fixed at \( \alpha = 1, \beta = 1 \) and \( \sigma^2 = 1 \), respectively. **Step 2** Generate two random error terms \( \delta_y \) and \( \epsilon_y \) from \( N(0, \sigma^2) \) and \( N(0, r^2) \), respectively. **Step 3** Calculate the observed values of \( x_j \) and \( y_j \) and also the value of \( Y_i \) using (1). **Step 4** Fit the generated data to balanced replicated LFRM and estimate parameters using (3), (4) and (5), respectively. **Step 5** Find the variance-covariance matrix and calculate the \( |COV| \) for all data. **Step 6** Delete the \( i^{th} \) row of every group and replicate with the mean for observation in every group from the generated sample of both \( x_j \) and \( y_j \) where \( i = 1, 2, \ldots, p \) and \( j = 1, 2, \ldots, m \). Repeat steps 4 till steps 6 to obtain \( |COVRATIO_{(i)}| \). **Step 7** Calculate \( (COVRATIO_{(i)} - 1) \) and find the value of \( |COVRATIO_{(i)}| \) for all \( i \). **Step 8** Specify the maximum value of \( |COVRATIO_{(i)}| \).

These steps are repeated for 5000 times for each combination of sample size \( n \) and \( r^2 \). Then, the 5% upper percentiles of the maximum values of \( |COVRATIO_{(i)}| \) is calculated. This upper percentile is used as the cut-off points in identifying the outliers for the balanced replicated linear functional relationship model. Table 2 shows the value of 5% upper percentiles of each value of \( n \) and \( r^2 \), respectively. From this table, the cut-off points show a monotonic decreasing function of sample size \( n \).

TABLE 2. The 5% upper percentile points of \( |COVRATIO_{(i)}| - 1 \) at \( r^2 = 0.2, 0.4, 0.6, 0.8 \) and 1.0

<table>
<thead>
<tr>
<th>Sample size, n</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.1321</td>
<td>1.9974</td>
<td>2.5915</td>
<td>2.8700</td>
<td>3.0111</td>
</tr>
<tr>
<td>40</td>
<td>1.3591</td>
<td>1.2633</td>
<td>1.3131</td>
<td>1.3179</td>
<td>1.3985</td>
</tr>
<tr>
<td>60</td>
<td>0.9804</td>
<td>0.9529</td>
<td>0.9281</td>
<td>0.9078</td>
<td>0.8913</td>
</tr>
<tr>
<td>80</td>
<td>0.9625</td>
<td>0.9158</td>
<td>0.8807</td>
<td>0.8498</td>
<td>0.8227</td>
</tr>
<tr>
<td>100</td>
<td>0.9401</td>
<td>0.8797</td>
<td>0.8322</td>
<td>0.7931</td>
<td>0.7632</td>
</tr>
<tr>
<td>132</td>
<td>0.8967</td>
<td>0.8086</td>
<td>0.7491</td>
<td>0.7049</td>
<td>0.6705</td>
</tr>
<tr>
<td>180</td>
<td>0.8245</td>
<td>0.7139</td>
<td>0.6431</td>
<td>0.5977</td>
<td>0.5613</td>
</tr>
<tr>
<td>300</td>
<td>0.6741</td>
<td>0.5412</td>
<td>0.4719</td>
<td>0.4288</td>
<td>0.3969</td>
</tr>
</tbody>
</table>
The arithmetic mean of the values for the respective $n$ are calculated and the best fit is obtained by using the least squares method as in Figure 1. For 5% significant level, we obtain the equation of the series trend line line $y = 9.6293n^{-0.526}$. This trend line will be used as a cut-off point in detecting the outliers. The next step of the study is to determine the power performance of the proposed method.

**POWER OF PERFORMANCE FOR COVRATIO STATISTIC**

Again, a Monte Carlo simulation method is used to investigate the performance of $|\text{COVRATIO}_{(i)}|^{-1}$ in identifying the outlier in the balanced replicated LFRM. Four different sample sizes of 40, 80, 100, and 180 are considered in this study by using the procedures described earlier to generate the data set. The power of performance $|\text{COVRATIO}_{(i)}|^{-1}$ are tested when correctly detecting the outlier. To determine the performance of the COVRATIO statistic, contamination is randomly applied; for example at observation $c$, where $y_c$ is contaminated as follows:

$$y_c = a + \beta X_c + \varphi_c$$

where $y_c$ and $X_c$ are the value of the $c^{th}$ observation of both variables $y$ and $X$, respectively, after contamination and $\varphi_c$ is error taken from a normal distribution with mean zero and different variances of 6,8,10,12,14, and 16, respectively (Ghapor et al. 2014; Mamun et al. 2019). The generated data are refitted and the maximum of $|\text{COVRATIO}_{(i)}|^{-1}$ statistic is specified. This procedure has correctly identified the outlier in the data set if the values of $|\text{COVRATIO}_{(i)}|^{-1}$ is maximum and exceed the stated cut-off point. The process is repeated 5000 times and the power of performance is then examined by calculating the percentage of the correct detection of the contaminated observation at $c$th position. Figure 2 shows the power of performance of $|\text{COVRATIO}_{(i)}|^{-1}$ statistics for $n = 80$ $(8 \times 10)$ for $\tau^2 = 0.2, 0.4, 0.6, 0.8$, and 1.0. From this plot, it can be concluded that as $\tau^2$ decreases, the power of performance in detecting the correct outlier increases.

**APPLICATION**

We consider two data sets where both have 10 groups and three elements in each group. The first data set is from simulated data generated from replicated LFRM by setting the parameters $a = 0$, $\beta = 1$, $\sigma^2 = 1$ and $\tau^2 = 0.2$. Next, we contaminated the observation randomly. The COVRATIO statistic for each value is calculated and the...
TABLE 3. The COVRATIO statistic for simulated data

<table>
<thead>
<tr>
<th>Index</th>
<th>COVRATIO(_{-i}^{-1})</th>
<th>Index</th>
<th>COVRATIO(_{-i}^{-1})</th>
<th>Index</th>
<th>COVRATIO(_{-i}^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9726</td>
<td>11</td>
<td>0.3911</td>
<td>21</td>
<td>0.4404</td>
</tr>
<tr>
<td>2</td>
<td>0.9760</td>
<td>12</td>
<td>0.4826</td>
<td>22</td>
<td>0.7032</td>
</tr>
<tr>
<td>3</td>
<td>0.9781</td>
<td>13</td>
<td>0.0230</td>
<td>23</td>
<td>0.6234</td>
</tr>
<tr>
<td>4</td>
<td>0.8263</td>
<td>14</td>
<td>0.2483</td>
<td>24</td>
<td>0.6850</td>
</tr>
<tr>
<td>5</td>
<td>0.1331</td>
<td>15</td>
<td>0.2449</td>
<td>25</td>
<td>0.9171</td>
</tr>
<tr>
<td>6</td>
<td>0.7490</td>
<td>16</td>
<td>0.0253</td>
<td>26</td>
<td>0.9071</td>
</tr>
<tr>
<td>7</td>
<td>0.3807</td>
<td>17</td>
<td>0.0068</td>
<td>27</td>
<td>0.9234</td>
</tr>
<tr>
<td>8</td>
<td>7.2855</td>
<td>18</td>
<td>0.0121</td>
<td>28</td>
<td>0.9686</td>
</tr>
<tr>
<td>9</td>
<td>0.4360</td>
<td>19</td>
<td>0.4510</td>
<td>29</td>
<td>0.9709</td>
</tr>
<tr>
<td>10</td>
<td>0.3444</td>
<td>20</td>
<td>0.2116</td>
<td>30</td>
<td>0.9635</td>
</tr>
</tbody>
</table>

FIGURE 2. Power of performance for \(|COVRATIO\(_{-i}^{-1}\)| for \(n = 80\)

FIGURE 3. Power of performance for \(|COVRATIO\(_{-i}^{-1}\)| when \(\tau^2 = 0.2\)
results are given in Table 3. Based on the formulation as given in Table 2, the cut-off point for \( \alpha \) is calculated and the value 1.609 obtained as the cut-off point at 5% significant level. Figure 4 clearly shows that the COVRATIO value for 8th observation is 7.286 which exceeds the cut-off points of 1.609. Hence, the developed test statistic and the cut-off points correctly detect the 8th observation as an outlier in the simulated data.

![Figure 4: The graph of the \(|COVRATIO_{(-i)}|\) statistic for simulated data](image)

Next, we consider another dataset taken from Altman and Bland (1999). We use a subsample of the original data containing 30 observations. The data set measures the systolic blood pressure which simultaneous measurements were made by two experienced observers denoted as J and R. In this case, we have 10 groups (or subjects) and each group have three sets of readings that were made in quick succession. Since there is no outlier in the original data, we insert the outlier randomly into the original data by following Kim (2000) and Imon and Hadi (2008). The COVRATIO statistic for each value is calculated and the results are given in Table 4. The cut-off point for \( n = 30 \) is 1.609 obtained as before at 5% significant level. From Figure 5, we observe that the value of \(|COVRATIO_{(-1)}|\) for the 11th observation is 1.849, which exceeds the cut-off point value of 1.609 at 5% significant level. To conclude, our cut-off point correctly identifies that the 11th observation as an outlier in this data set.

![Table 4: The COVRATIO statistic for real data](image)
CONCLUSION
We extend the idea of COVRATIO statistics for detecting a single outlier because it is a new topic and has not been explored in balanced replicated LFRM. The COVRATIO statistic is used because it is a simple procedure, widely used and easy to implement to conform with balanced replicated LFRM. In adopting the COVRATIO procedure to balanced replicated LFRM, we need to determine the cut-off point and measure the power of performance as well as a slight modification to COVRATIO to accommodate the replicates of the model. Using simulation studies, we identified the cut-off point for this statistic and have shown that this statistic performs well in identifying an outlier. A practical example is illustrated with real data set. As an implication, this study provides empirical evidence that COVRATIO can be utilized for detecting outliers on balanced replicated LFRM; thus providing a better understanding to the approach in outlier detection in the model.

ACKNOWLEDGEMENTS
We are most grateful to University of Malaya (BKS005-2019 and GPF006H-2018), National Defence University of Malaysia and Ministry of Higher Education (MoHE), Malaysia for the financial support. We also wish to thank to referees for their helpful comments and suggestions.

REFERENCES

FIGURE 5. The values of the $|\text{COVRATIO}_{(-i)}^{-1}|$ statistic for real data


Azuraini Mohd Arif Institute of Advanced Studies University of Malaya 50603 Kuala Lumpur, Federal Territory Malaysia

Azuraini Mohd Arif Centre for Foundation Studies National Defence University of Malaysia 57000 Kuala Lumpur, Federal Territory Malaysia

Yong Zulina Zubairi* Centre for Foundation Studies in Science University of Malaya 50603 Kuala Lumpur, Federal Territory Malaysia

Abdul Ghapor Hussin Faculty of Defence Science and Technology National Defence University of Malaysia 57000 Kuala Lumpur, Federal Territory Malaysia

*Corresponding author; email: yzulina@um.edu.my

Received: 2 February 2021
Accepted: 19 June 2021