Analysis and Assessment of Boxplot Characters for Extreme Data

Babangida Ibrahim Babura$^{1,2,a}$, Mohd Bakri Adam$^2$, Abdul Rahim Abdul Samad$^3$, Anwar Fitrianto$^4$ and Bashir Yusif$^1$

$^1$Department of Mathematics, Federal University Dutse, Nigeria,
$^2$Institute for Mathematical Research, Universiti Putra Malaysia,
$^3$Department of Economics, Universiti Putra Malaysia,
$^4$Department of Statistics, Bogor Agricultural University, Indonesia.

E-mail: “bibabura@gmail.com

Abstract. The robust procedure used in constructing boxplot makes it to remain a vital tool for the display of distributional summaries, with no or less deviation from the empirical model characters which the data possess. In this paper, we investigate the embedded characters of the extreme dataset as richly displayed by a boxplot. We discuss and assess boxplot characters such as; the display of asymmetry, the outliers cut off using the outside rate per sample for three different types of boxplot implementations. The performance of the three boxplot fence implementations on extreme data was further assessed by introducing a new measure called fence sensitivity ratio. The fence sensitivity ratio is an attempt to propose an alternative to the conventional routine of data contamination in assessing the boxplot outlier rules. The findings in this paper highlighted the significance of boxplot as an exploratory data analysis tool in diagnosing some extreme data modelling tools and stress on the weakness of the existing boxplot methods and recommend useful suggestion in addressing such weaknesses for further investigation.

1. Introduction

Boxplot techniques of exploratory data analysis (EDA) has employed a resistant rule for identifying possible outliers in a single batch of univariate dataset. The lower and upper fourth or first and third quartiles, $Q_L$ and $Q_U$ are approximate estimation of the two quartiles of a dataset with $Q_M$ as the second quartile or median of the dataset. The resistance rule of a boxplot labels an observation from the data set as “outside” if the observation fall below $Q_L - 1.5IQR$, or above $Q_U + 1.5IQR$, where $IQR = Q_U - Q_L$ according to the Tukey’s classical boxplot rule [2, 3]. The Tukey’s classical boxplot rule fail in marking a correct outliers for skewed dataset as highlighted in [3]. To address the problem of incorrect marking of outliers a substantial amount of literature were achieved.

A critical examination for advantages of different boxplot implementations is given in [5], which is based on the choices of five number summaries with description of their consequences. Some of the suggestions given by [5] are in selection of the fence constant other than 1.5. The use of semi-interquartile ranges to reflect asymmetric whiskers was proposed in [6] with adjustment of the resistance rule towards skewed dataset. Another approach for the adjustment of Tukey’s resistance rule is proposed by [7], in which the two quartiles $Q_L$ and $Q_U$ were replaced with the...
median ($Q_M$). An enhancement to the idea in [7] was given in [9, 10] in which the resistance rule is proposed according to a pre-specified outside rate. The outside rate is the probability that an observation is wrongly marked as an outlier by a resistance rule. The limitations of [9, 10] is attached to the estimated fence constants which are sample size dependent and require some properties from uncontaminated distribution in which their estimation is difficult to achieve. A more general implementation of boxplot rule that requires an adjustment of the position of fences to account for skewness using medcouple (MC), a robust measure of skewness was proposed in [11].

The performance of a boxplot method depends significantly on the resistance rule utilized in its implementation and the dataset characteristics. This research work attempts to assess three variants of boxplot resistance rules namely; the classical method [2], the asymmetric method [7] and the adjusted method [11] on extreme dataset. We refer to an extreme dataset as a records of events that are the most extreme than any value that have already been observed within a particular uniform block of time such as annual maximum events. Such observations are assumed to be low extreme as minima or high extreme as maxima. The current development in global warming which signifies a considerable interest in environmental research and financial crisis a consequence from so much volatility in the financial sector, are some of the events that give rise to a universal interest in modelling and forecasting of extreme events.

This research considers boxplot as an EDA tool that enhance the quality of extreme dataset before embarking on any empirical and confirmatory analysis. As we reviewed earlier, the existence of different boxplot outlier rules has posed a challenge in selecting an appropriate method to implement for extreme dataset. Upon this development, we implement a simulation study that extend the work of [4]. The study in [4] examines the boxplot resistance rules for dataset that follows the Gaussian and chi-squared distributions set up. This paper presents an investigation of some boxplot characteristics such as: visualizing asymmetry, the outside rate and a newly introduced measure of fence sensitivity ratio for extreme dataset according a simulation from the GEV distribution. The research was concluded by suggesting the best boxplot method from among; the Tukey’s classical method [2], the Kimber’s asymmetric method [6] and the Hubert’s adjusted method [11] in diagnostic visualization of extreme dataset.

2. Methods

2.1. Boxplot Construction and Methods

Let $X = \{x(1), x(2), \ldots, x(n)\}$ be an ordered sample of size $n$. Then any boxplot construction require the three sample quartiles $Q_L$, $Q_M$ and $Q_U$ from $X$, which describe the two boxplot hinges to be a versions of the first and third quartile, i.e., close to quantiles of sample $X$ at $p = \{0.25, 0.75\}$ such that the first and third quartiles of sample $X$ are respectively given by $Q_L = x(k)$ and $Q_U = x(n-k+1)$ where $k = \left\lfloor \frac{(n+3)}{2} \right\rfloor$ with $\left\lfloor \cdot \right\rfloor$ as the greatest integer function. The second quartile $Q_M$ is referred to be the median of sample $X$. However, any boxplot implementation require a robust center spread estimate usually with interquartile range given by $IQR = Q_U - Q_L$ or semi interquartile ranges given by $SIQR_L = Q_M - Q_L$ and $SIQR_U = Q_U - Q_M$. The boxplot fence estimate (resistance rule) is the tool that determine a particular boxplot method of implementation. In this paper we consider three boxplot methods namely; the classical boxplot method [2], the asymmetric boxplot method [6] and the adjusted boxplot method [11]. The classical, asymmetric and adjusted boxplot methods are constructed in a similar way except in the fence definition.

Generally the lower fence estimate $F_L$ and upper fence estimate $F_U$ can be expressed as:

$$F_L = Q_L - k_1C_1$$
$$F_U = Q_U - k_2C_2$$
Figure 1. A labelled boxplot display of a single batch of dataset

where $C_1$ and $C_2$ are the robust centre spread, usually $C_1 = C_2 = IQR$ for classical [2] and adjusted [11] boxplot methods while $C_1 = SIQR_L$ and $C_2 = SIQR_U$ for asymmetric [6] boxplot method. The spread coefficient is given as a constant $k_1 = k_2 = 1.5$ for classical [2] and $k_1 = k_2 = 3$ for asymmetric [6] boxplot methods while in adjusted [11] method the coefficient is determine according to the medcouple [12] $MC$ a skewness coefficient in the dataset $X$ given by $\{k_1 = 1.5e^{-3MC}, k_2 = 1.5e^{3MC} : MC \geq 0\}$ or $\{k_1 = 1.5e^{-3MC}, k_2 = 1.5e^{3MC} : MC < 0\}$. Then the fence markup is the interval given by

$$F = [\min\{x \in X : x \geq F_L\}, \max\{x \in X : x \leq F_U\}]$$

and any observation outside $F$ is referred as the outlying observation (outlier).

Then the boxplot is constructed by first drawing a rectangular box to account for central 50% data points of sorted dataset, in which the hight of box equals the value of centre spread usually the $IQR$ or $SIQR_L + SIQR_U$. A line representing the median $Q_M$ was drawn to divides the box into two parts. Whisker lines are drawn from upper and lower hinges (lower and upper quartiles) to the upper and lower fence markup position respectively. Finally the outlying observations are drawn as points below and above the fence markup $F$. Figure 1 is a boxplot construction for a univariate dataset with labelling of all the important boxplot component.

2.2. Extreme Data in Block Maximum Method

The asymptotic models of extreme events focus usually on statistical behavior of $X = \max\{X_1, \cdots, X_n\}$, where $X_1, \cdots, X_n$ is a sequence independent random variables, with common distribution function $F$. In practical form, the $X_i$ usually refers to representation of values measured on a process based on regular time scale such as hourly measurement making $X$ representing the maximum of the process upon $n$ time units of observations. In this regard, the theory of distribution of $X$ were generally derived for all values of $X_i, i = 1, \ldots, n$. 

Figure 2. Typical boxplot display for three batches of samples from GEV distribution family with same location ($\mu = 50$) and scale ($\sigma = 5$) but different shapes parameters $\xi = \{0.5, 0, -0.5\}$ corresponding to respectively Fréchet, Gumbel and Weibull distributions.

The theory on modelling the extreme event was consolidated by Jenkinson in 1955 [1] by combining the three existing limiting distributions of maxima into one with the distribution function of the form

$$G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

(1)
defined over $\{ x : 1 + \xi \left( \frac{x - \mu}{\sigma} \right) > 0 \}$ such that the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The expression in Equation (1) becomes the Generalized Extreme Value (GEV) family of distributions [8]. The model compose of three parameters: the location parameter $\mu$, the scale parameter $\sigma$ and the shape parameter $\xi$. When $\xi > 0$ and $\xi < 0$ the model respectively correspond to Fréchet and Weibull classes of extreme value distribution. Furthermore, the subset of the GEV family with $\xi = 0$ is regarded as the limit of Equation (1) as $\xi \to 0$, leading to the Gumbel family with re-expression of the distribution function as

$$G(x) = \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\}, \quad -\infty < x < \infty.$$

Hoaglin [4] study characteristics of boxplot outlier labelling rules according to simulation from Gaussian and Chi squared distributions setup and make some recommendations. This paper extend the work of Hoaglin [4] to observe the boxplot resistance rules characteristics of extreme data according a simulation from the GEV distribution.
3. Assessment of Boxplot Characters

3.1. Boxplot Display of Symmetry by Extreme Data

Figure 2 is a boxplot display of batches of 100 samples each for the three family of GEV distribution. A strong display of characters to distinguish each of the family type were illustrated. The first batch in the right is a boxplot display of the right skewed Fréchet distribution. The long tail can be viewed from the spread of the upper whisker together with marked outlier points. The middle batch is also a boxplot display of near symmetric Gumbel distribution. The display show that the Gumbel distribution has little longer right tail by comparing the spread of the upper versus the lower whisker lines. But the symmetry were mostly seen around the median, from the lower upto the upper hinges. The Weibull type represented as the left batch boxplot display which show a sign of left skewness from the spread of the lower whisker all together with outlier points as compared to the spread of the upper whisker.

3.2. Fence Rule Performance with Outside rate for Extreme Data

Figure 2 shows all the three boxplots indicate some number of marked points above upper or lower whisker which are referred to as the potential outliers. This point are usually data points that if removed may significantly distort the accuracy of some parameter estimates of the empirical models about the dataset. We investigates through simulation study if the three points that if removed may significantly distort the accuracy of some parameter estimates of or lower whisker which are referred to as the potential outliers. This point are usually data

One important factor to investigate in this regard is the outside rate per sample. For an extreme sample $X_n$ of size $n$ that is assumed to follow a three parameter GEV distribution (i.e. $X \sim GEV(\mu, \sigma, \xi)$). Let $B(\xi, n)$ be the probability that the sample $X_n$ contains no outlier, then $1 - B(\xi, n)$ is the probability that the sample contains at least one outlying observation and is referred to as the outside rate per sample $[3]$.

So, let observe the behaviour of the outside rate from different sample sizes $n \in \{10, 20, 50, 100, 250, 500\}$ as generated from the GEV distribution to represent extreme samples. We also fixed the location parameter $\mu = 0$ and scale parameter ($\sigma = 1$) then varied the shape parameter to assumes values $\xi \in \{-0.5, -0.25, 0, 0.15, 0.30, 0.50\}$ so that when $\xi \in \{-0.50, -0.25\}$ we represent the Weibull family, $\xi = 0$ representing the Gumbel family and $\xi \in \{0.15, 0.30, 0.50\}$ to be representing the Fréchet family.

Table 1 illustrates that by varying values of $n$ and $\xi$, we obtain different values of the outside rate per sample $1 - B(\xi, n)$ according to classical, asymmetric and adjusted boxplot methods. While comparing two or more boxplot resistance rule, we consider a method which demonstrate a better performance than the other by considering a method with the lowest value of the outside rate. So, in the case of $GEV(0, 1, -0.5)$ the classical boxplot method did better than the asymmetric and adjusted for samples with $n = 10$ and $n = 20$ whereas the asymmetric method do better well when $n = 50$ and the adjusted method outperform both classical and asymmetric methods for large sample sizes of 100, 250 and 500. Also the classical method outperform both asymmetric and adjusted methods in visualizing the Weibull scenario with $GEV(0, 1, -0.25)$ in all the sample sizes with outside rate values between 0.1819 to 0.7234. In the Gumbel case $GEV(0, 1, 0)$ a better performance for classical method ($n = 10$ and 20), for asymmetric method when the sample size $n = 50$ and the adjusted method for a large sample sizes of 100, 250 and 500. The performance of the three boxplot methods worsen for the Fréchet scenarios, $GEV(0, 1, 0.15)$, $GEV(0, 1, 0.15)$ and $GEV(0, 1, 0.15)$ with the outside rate exploding upto 0.99 when the sample sizes $n \geq 100$. The adjusted boxplot method demonstrate a relatively better performance almost all the Fréchet scenarios except for lower sample sizes $n = 10$ and $n = 20$ in which we have some instances which the asymmetric method do better.

The overall performance of the outside rate $1 - B(\xi, n)$ in Table 1 is not fairly in sync with the ideal values described by the Gaussian samples of between 0.198 for small sample of 10 to
Table 1. Comparison of simulation result of some outside rate per sample \((1 - B(\xi, n))\) from GEV distribution samples, between classical, asymmetric and adjusted boxplot methods.

| \(n\)  | \(GEV(0, 1, -0.5)\) | \(GEV(0, 1, -0.25)\) | \(GEV(0, 1, 0)\) |
|-------|-----------------|-----------------|-----------------|
|       | classical | asymmetric | adjusted | classical | asymmetric | adjusted | classical | asymmetric | adjusted |
| 10    | 0.2063     | 0.2914     | 0.3924     | 0.1819     | 0.3035     | 0.4084     | 0.2911     | 0.3591     | 0.4529     |
| 20    | 0.2533     | 0.3045     | 0.4747     | 0.1838     | 0.2980     | 0.4990     | 0.4336     | 0.4451     | 0.5892     |
| 50    | 0.4337     | 0.3990     | 0.5036     | 0.2556     | 0.3552     | 0.5912     | 0.7314     | 0.6604     | 0.7255     |
| 100   | 0.6415     | 0.5482     | 0.5349     | 0.3249     | 0.4211     | 0.6410     | 0.9241     | 0.8538     | 0.8358     |
| 250   | 0.9006     | 0.8090     | 0.5950     | 0.5235     | 0.5822     | 0.7577     | 0.9983     | 0.9862     | 0.9663     |
| 500   | 0.9897     | 0.9552     | 0.6981     | 0.7234     | 0.7555     | 0.8682     | 0.9998     | 0.9999     | 0.9973     |

0.523 for large sample of upto 100 when visualized according to the classical boxplot fence rule. This signifies that, even though the adjusted method demonstrate an overall good performance than the other two methods, a more scientific adjustment of boxplot fence rule need to be done.

3.3. Assessment of Fence Rule with Fence Sensitivity Ratio

We consider as insufficient to discuss the outside rate per sample of any particular boxplot method alone in fence rule assessment. Therefore, we consider another important measure in order to assess how sensitive a particular fence rule is in detecting the outliers. For many instances the fence rule can over/under estimate the actual fence position. Example of such scenario is the case of an extreme sample which are mostly right skewed. The upper fence markup usually under estimated the fence position leading to a higher outside rate, while the scenario is the case of an extreme sample which are mostly right skewed. The upper fence instances the fence rule can over/under estimate the actual fence position. Example of such order to assess how sensitive a particular fence rule is in detecting the outliers. For many methods alone in fence rule assessment. Therefore, we consider another important measure in

Definition 3.1. : Let \(X(n) = \{x_{(1)}, \ldots, x_{(n)}\}\) be an ordered sample of size \(n\) from a univariate distribution \(F\) with \(Q_L\) and \(Q_U\) as the lower and upper quartile of \(X(n)\) respectively. Then the boxplot lower fence sensitivity ratio is given by \(LF_R = \frac{k_1C_i}{Q_L - x_{(1)}}\) and the boxplot upper fence sensitivity ratio is analogously given by \(UF_R = \frac{k_2C_i}{Q_U - x_{(n)}}\), where \(k_i, i = 1, 2\) and \(C_i, i = 1, 2\) are determine from a particular boxplot implementation as described in Section 2.1 with \((Q_L - x_{(1)}) > 0, (x_{(n)} - Q_U) > 0\).

By observing the behaviour of the two ratios as both signifies a very important property about boxplot in the following way:

- Both \(LF_R\) and \(UF_R\) are non negative measures i.e. \(LF_R \geq 0\) and \(UF_R \geq 0\) for all univariate samples.
- The values of \(LF_R\) and \(UF_R\) signifies a good fence estimate if they clustered closely around 1.0 for re-sampling over \(F\). That is, the measures indicate the extend of overestimation/underestimation if the fence estimate falls above/below 1.0 respectively.
- Both \(LF_R\) and \(UF_R\) are location scale invariant.
We assess the sensitivity of the three boxplot fences with the fence sensitivity ratios $LF_R$ and $UF_R$ in a similar way we assess the outside rate per sample for extreme sample from GEV distribution. This is done by varying both the sample size $n \in \{10, 20, 50, 100, 250, 500\}$ and the GEV shape parameter $\xi \in \{-0.5, -0.25, 0, 0.15, 0.30, 0.50\}$. Now consider $LF_R$ and $UF_R$ as the median lower and upper fence sensitivity ratios obtainable by re-sampling of samples of size $n$ from $GEV(0, 1, \xi)$ as tabulated in Table 2.

### Table 2. Simulation result of the two fence sensitivity ratios per sample from GEV distribution samples, over selected values of $\xi$ and $n \ (GEV(\mu, \sigma, \xi))$

| $n$ | classical | asymmetric | adjusted | classical | asymmetric | adjusted | classical | asymmetric | adjusted |
|-----|-----------|------------|----------|-----------|------------|----------|-----------|------------|----------|
|     | $LF_R$    | $UF_R$     | $LF_R$   | $UF_R$    | $LF_R$    | $UF_R$   | $LF_R$   | $UF_R$     | $LF_R$   |
| 10  | 2.0291    | 3.5688     | 2.5039   | 3.0845    | 3.4704    | 3.8011   | 1.2422   | 1.2792     | 1.0666   |
| 20  | 1.4365    | 2.7702     | 1.5280   | 2.5352    | 1.9153    | 1.8171   | 1.9189   | 1.7478     | 1.8148   |
| 50  | 1.0632    | 2.3088     | 1.1466   | 2.1072    | 1.4365    | 1.5385   | 1.4761   | 1.3566     | 1.4331   |
| 100 | 0.9044    | 2.1308     | 0.9607   | 1.9583    | 1.2422    | 1.4089   | 1.2792   | 1.1858     | 1.2428   |
| 250 | 0.7658    | 1.9918     | 0.8189   | 1.8365    | 1.0464    | 1.3117   | 1.1096   | 1.0497     | 1.0872   |
| 500 | 0.6918    | 1.9324     | 0.7395   | 1.7801    | 0.9424    | 1.2832   | 1.0219   | 0.9806     | 0.9982   |
|     | $LF_R$    | $UF_R$     | $LF_R$   | $UF_R$    | $LF_R$    | $UF_R$   | $LF_R$   | $UF_R$     | $LF_R$   |
| 10  | 2.3752    | 3.1176     | 2.4160   | 2.1308    | 0.2287    | 0.3422   | 1.1467   | 1.5054     | 1.1146   |
| 20  | 1.8717    | 1.3215     | 0.6160   | 2.0389    | 0.4152    | 0.7783   | 0.5554   | 0.2890     | 0.7783   |
| 50  | 1.2428    | 1.4331     | 2.3409   | 3.0344    | 1.9517    | 1.4761   | 1.0439   | 1.0497     | 1.0872   |
| 100 | 1.2428    | 1.4331     | 2.3409   | 3.0344    | 1.9517    | 1.4761   | 1.0439   | 1.0497     | 1.0872   |
| 250 | 1.2428    | 1.4331     | 2.3409   | 3.0344    | 1.9517    | 1.4761   | 1.0439   | 1.0497     | 1.0872   |
| 500 | 1.2428    | 1.4331     | 2.3409   | 3.0344    | 1.9517    | 1.4761   | 1.0439   | 1.0497     | 1.0872   |
|     | $LF_R$    | $UF_R$     | $LF_R$   | $UF_R$    | $LF_R$    | $UF_R$   | $LF_R$   | $UF_R$     | $LF_R$   |
| 10  | 3.4738    | 2.3377     | 0.8449   | 1.4089    | 1.9393    | 1.0632   | 1.4285   | 2.0761     | 0.9806   |
| 20  | 2.3088    | 3.7839     | 1.7478   | 0.1878    | 3.5688    | 1.2072   | 1.2215   | 2.1270     | 1.1217   |
| 50  | 1.0632    | 2.3088     | 1.1466   | 2.1072    | 1.4365    | 1.5385   | 1.4761   | 1.3566     | 1.4331   |
| 100 | 0.9044    | 2.1308     | 0.9607   | 1.9583    | 1.2422    | 1.4089   | 1.2792   | 1.1858     | 1.2428   |
| 250 | 0.7658    | 1.9918     | 0.8189   | 1.8365    | 1.0464    | 1.3117   | 1.1096   | 1.0497     | 1.0872   |
| 500 | 0.6918    | 1.9324     | 0.7395   | 1.7801    | 0.9424    | 1.2832   | 1.0219   | 0.9806     | 0.9982   |

Table 2 shows the result of simulation study on fence rule using the fence sensitivity ratios. The median of the estimates of $LF_R$ and $UF_R$ obtained from re-sampling of fixed choices of sample size $n$ and GEV distribution shape parameter $\xi$ was recorded. The values recorded when the sample size became 10 and 20 shows an indication of over estimation except some few instances when $\xi = 0.3, 0.5$. The table however shows the effect of change in the sample size $n$ and shape parameter $\xi$ in both three fence rule implementation. But in terms of the overall performance of the three boxplot implementations, the adjusted method is the overall winner as the estimation has responded better especially towards the change in $\xi$. The worse case scenarios are observed all against the classical fence rule, especially overestimation of the lower fence when $LF_R(10, 0.5)$ and underestimation of the upper fence when $UF_R(500, 0.5)$.

The value at $LF_R(10, 0.5)$ of the classical method indicates that, more than 50% of the estimated fence in the simulation were underestimated by not less-than 6 times the actual fence position. Also the value at $UF_R(500, 0.5)$ for the classical method shows that, for more than 50% of the estimated fence in the simulation underestimate the fence position by a fraction of 0.06 or less.
4. Discussion and Conclusion

The performance of some of the most important features of boxplot in relation to extreme dataset was critically explored. Initially we observe and analyse symmetry and tail behaviour of an extreme samples according to the three GEV family of distributions as displayed by the boxplot. The exploratory analysis indicates that the three GEV distribution family are identifiable based on the tail and asymmetric display of the dataset. However, we assess the performance of boxplot fence mark-up base on the Tukey’s classical boxplot method [2], Kimber’s asymmetric boxplot method [6] and Hubert’s adjusted boxplot method [11] according to a measure called outside rate per sample. The outside rate per sample is a probability that atleast one regular observation will be mark as outlier by the fence rule. We evaluate the outside rate per sample by varying both the sample size and the GEV shape parameter. The result of this indicates a values above the benchmark requirement of the outside rate for all the three boxplot methods.

We introduce a new measures for fence assessment called the lower fence sensitivity ratios \( LFR \) and upper fence sensitivity ratio \( UFR \). The measures were similarly utilize to assess the fence mark-up according the three boxplot methods. We simulates the estimates of \( LFR \) and \( UFR \) by varying sample sizes and shape parameter of the GEV distribution.

The result of the two performance studies indicates that, the adjusted boxplot method is the overall winner especially for its better response to variation in GEV distribution shape parameter. The better performance of the adjusted method can be attributed due to the incorporation of the skewness property of the dataset in the fence estimation. We recommend at the present that the Hubert’s adjusted boxplot method [11] as the most appropriate in visualising extreme dataset especially when the data size is above 20. Also by considering that the performance of all the three method fail to achieve the benchmark requirement [4], there arise a need for a fence adjustment according to extreme data especially a fence estimate that account for the behaviour of shape parameter and sample size in GEV model. The introduced fence sensitivity ratios \( LFR, UFR \) is a potential alternative to the usage of data contamination in making a robust assessment of a boxplot resistance rules.

5. Acknowledgment

This research is conducted with support from Universiti Putra Malaysia Research Grant (No 02-01-15-1741FR & GPB/2017/9531800) and Federal University Dutse, Nigeria TETFUND Fellowship Grant 2013. The authors also acknowledge the constructive criticism by the unanimous reviewers of the draft paper.

References

[1] Jenkinson, Arthur F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 1955; 81(348) 158–171.
[2] Tukey, JW 1977 Exploratory data analysis. *Addison-Wesley: Reading, Mass.*
[3] Hoaglin DC, Mosteller F, Tukey JW 1983 Understanding robust and exploratory data analysis. Vol.3 *Wiley New York.*
[4] Hoaglin, David C and Iglewicz, Boris and Tukey, John W. Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association.* Taylor & Francis Group 1986; 83(396) 991–999
[5] Frigge M, Hoaglin DC, Iglewicz B. Some implementations of the boxplot. *The American Statistician.* 1989; 43(1):50–54.
[6] Kimber, A. Exploratory data analysis for possibly censored data from skewed distributions. *Applied Statistics* 1990; 21–30.
[7] Carling K. Resistant outlier rules and the non-Gaussian case. *Computational Statistics & Data Analysis.* 2000; 33(3):249–258.
[8] Coles, S 2001 An introduction to statistical modeling of extreme values. *Lecture Notes in Control and Information Sciences* London: Springer. Retrieved from https://books.google.com.my/books?id=2nuqUEaKqFEC
[9] Schwertman NC, Owens MA, Adnan R. A simple more general boxplot method for identifying outliers. *Computational statistics & data analysis.* 2004; 47(1):165–174.
[10] Schwertman NC, de Silva R. Identifying outliers with sequential fences. *Computational statistics & data analysis*. 2007;51(8):3800–3810.

[11] Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*. 2008;52(12):5186–5201.

[12] Brys G, Hubert M, Struyf A. A robust measure of skewness. *Journal of Computational and Graphical Statistics*. 2004;13(4).