Mathematics and Statistics 9(4): 566-573, 2021
DOI: 10.13189/ms.2021.090415 http://www.hrpub.org

# Robust Estimation for Proportional Odds Model through Monte Carlo Simulation

**Faiz Zulkifli[1,2], Zulkifley Mohamed[2,*], Nor Afzalina Azmee[2], Rozaimah Zainal Abidin[1]**

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia
[2]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia

*Received January 28, 2021; Revised March 18, 2021; Accepted May 30, 2021*

***Cite This Paper in the following Citation Styles***

***(a):*** *[1] Faiz Zulkifli, Zulkifley Mohamed, Nor Afzalina Azmee, Rozaimah Zainal Abidin , "Robust Estimation for Proportional Odds Model through Monte Carlo Simulation," Mathematics and Statistics, Vol. 9, No. 4, pp. 566 - 573, 2021. DOI: 10.13189/ms.2021.090415.*

***(b):*** *Faiz Zulkifli, Zulkifley Mohamed, Nor Afzalina Azmee, Rozaimah Zainal Abidin (2021). Robust Estimation for Proportional Odds Model through Monte Carlo Simulation. Mathematics and Statistics, 9(4), 566 - 573. DOI: 10.13189/ms.2021.090415.*



**Abstract** Ordinal regression is used to model the ordinal response variable as functions of several explanatory variables. The most commonly used model for ordinal regression is the proportional odds model (POM). The classical technique for estimating the unknown parameters of this model is the maximum likelihood (ML) estimator. However, this method is not suitable for solving problems with extreme observations. A robust regression method is needed to handle the problem of extreme points in the data. This study proposes Huber M-estimator as a robust method to estimate the parameters of the POM with a logistic link function and polytomous explanatory variables. This study assesses ML estimator performance and the robust method proposed through an extensive Monte Carlo simulation study conducted using statistical software, R. Measurement for comparisons are bias, RMSE, and Lipsitzs' goodness of fit test. Various sample sizes, percentages of contamination, and residual standard deviations are considered in the simulation study. Preliminary results show that Huber estimates provide the best results for parameter estimation and overall model fitting. Huber's estimator has reached a 50% breakdown point for data containing extreme points that are quite far from most points. In addition, the presence of extreme points that have only a distance of two times far from most points has no major impact on ML estimates. This means that the estimates for ML and Huber may yield the same results if the model's residual values are between -2 and 2. This situation may also occur for data with a percentage of contamination below 5%.

**Keywords** Proportional Odds Model, M-Estimation, Ordinal Response Model, Robust Estimator

## 1. Introduction

The Fourth Industrial Revolution needs a more detailed and in-depth data analysis for the development of the latest technologies and applications to improve the daily lives of people. The collected data can be categorised as quantitative or qualitative data. In recent years, many researchers have shifted to qualitative data to produce more extensive research. It is important to know the type of measurement scale for qualitative data which usually involves nominal and ordinal scales. Ordinal data usage is more popular among researchers and is widely used in various fields such as scientific research, education, sociological psychology, and economics (Zulkifli, Mohamed, Azmee, and Abidin [1]). Interest in this type of data increased as the development of the item instruments becomes easier and convenient. In addition, the cost of data collection is cheaper with the existence of an online system. Generally, category data appear when the item is in the form of opinion, judgement, and rating which are stated as

the ordered categories.

Literature has substantially contributed robustness to the discrete and continuous data as produced by Hampel, Ronchetti, Rousseeuw, and Stahel [2]; Huber [3]; Huber and Ronchetti [4]; and Maronna, Martin, and Yohai [5]. However, the robustness of the ordinal model has somewhat neglected its improvements. Sometimes, a situation may occur where respondents willfully or accidentally chose the wrong category. This situation caused a major error in the model of contaminated distribution, therefore, changing the estimator's nature and modeling (Iannario, Monti, Piccolo, and Ronchetti [6]). Through extensive reading, only several studies have suggested a robust approach to the ordinal regression model.

A recent study by Iannario et al. [6] which uses a robust M-estimator introduced by Hampel et al. [2] as an alternative to maximum likelihood estimates of ordinal response models. They have tested the estimator's performance through extensive numerical experiments into five difference models with a mix of dichotomous, polytomous, and continuous explanatory variables.

Therefore, this study only focuses on the estimation of the proportional odds model (POM) with a logistic link function and polytomous explanatory variables in overcoming the presence of extreme data. This model is selected as it is widely used by many researchers who find the results easier to explain compared to other ordinal models (Zulkifli, Mohamed, Azmee, and Abidin [7]). This study proposed robust M-estimator using the weighting method introduced by Huber. A comparative analysis will be made against ML estimator and the proposed M-estimator via Monte Carlo simulation. The performance of both estimators will be observed through its model fitting on simulation data for different sample sizes, for different percentages of contamination, and for different distances of extreme points. All estimates parameters are tested using bias and RMSE, while the overall performance of the model is measured using the Lipsitz's goodness of fit test.

## 2. Literature Review

Among the weaknesses identified in the study of Iannario et al. [6] are random data generated in a simulation study have been coded for polytomous categorized covariates to several dichotomous variables. This approach is only suitable for using against normally categorized covariates rather than ordinals. This means that the proposed estimator is not tested against random data containing contamination data on polytomous categorized covariates. In addition, the total sample size of 200 and the percentage of contamination data of 10% are assumed to be insufficient to measure the estimator's robustness of the increased sample size and efficiency in managing samples containing a larger percentage of contamination data.

In addition, the coding of polytomous data to the dichotomous has an impact on model fitting. Francis [8] has seen the impact of coding techniques on statistical model and multicollinearity. Among the advantages of coding over the original data is that it can reduce the calculation time which makes it more effective. Meanwhile, the results of the analysis of the actual data found that the coding method can reduce the multicollinearity effect on the regression model. However, the dummy encoding method has drastically reduced the goodness of fit test. The value of the determination coefficient for the original data of 98.9% was reduced to only 15.4% due to dummy encoding. Similarly, the F-value decreased from 692,184 (original data) to 1.369 (dummy data).

Iannario, Clara, and Piccolo [9] examined the robustness of the Cub model, which is a discrete mix of uniform random variables and binomial shifts. Croux, Haesbroeck, and Ruwet [10] proposed a robust estimator by adding a weighting step to the maximum likelihood estimator for an ordinal response model with a logistic link function. Moustaki and Victoria-Feser [11] developed a robust alternative to LISREL and a robust estimator for generalised linear models on latent variable models. Finally, Albert and Chib [12], and Albert and Chib [13] calculated the residual Bayesian for both binary and polytomous response data in detecting outliers for categorical and ordinal data. Unfortunately, none of the studies has used M-estimator to estimate ordinal model with polytomous explanatory variables or ordered categories.

## 3. Methodology

### 3.1. Proportional Odds Model

The proportional odds model was originally proposed by Walker and Duncan [14] and was known as an odds ratio model in 1980 as in McCullagh [15]. Assuming variable $Y$ as multinomial categorical response and variable $X$ as m-dimensional vector for covariates with polytomous ordered categories, the latent regression model is translated into the equation:

$$P\left(Y_i \le j\right) = Y_i^* = \sum_{k=1}^{p} x_{ik}^t \beta_k + \varepsilon_i, i = 1,2,...,n; \quad j = 1,2,...,m; \alpha_{j-1} < Y_i^* < \alpha_j. \tag{1}$$

The probability mass function of $Y_i$ is given by:

$$P\left(Y_i = j \middle| x_{ij}\right) = P\left(\alpha_{j-1} < Y_i^* < \alpha_j\right)$$
$$= F\left(\alpha_j - \sum_{k=1}^{p} x_{ik}'\beta_k\right) - F\left(\alpha_{j-1} - \sum_{k=1}^{p} x_{ik}'\beta_k\right), \quad (2)$$
$$-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_m = +\infty$$

where: $F(.)$ is a distribution function for a random variable error, $\varepsilon_i$,

$\varepsilon_i$ has follow a cumulative logistic distribution,

and $\alpha_j - \sum_{k=1}^{p} x_{ik}'\beta_k$ is a function of log odds.

Based on equation (2), the vector of the regression coefficient does not depend on *i*. This means that the relationship between $x_j$ and $y_i$ is independent of *i*. McCullagh [15] presents it as a proportional odds assumption of equality in logarithmic proportions that crosses *k* point deductions.

The proportional odds model (POM) is the most widely used ordinal regression model as it provides an easy-to-understand estimate. In addition, it can be used against variables in the continuous form that has been made discrete during the data collection process.

### 3.2. Maximum Likelihood (ML) Estimator

The parameters in equation (2) are easily estimated using an ordinary least square (OLS) method compared to maximum likelihood (ML) estimator. This is because the logit function is a complex function because it involves the calculation of cells and causes the probability function to not necessarily obtain the approximate form. However, the OLS method is not able to estimate the parameters of the ordinal regression model that contains covariates in the form of ordered categories and continuous. A more suitable method for estimating model parameters with polytomous ordered categories of covariates is maximum likelihood (ML) estimator.

The log-likelihood function is:

$$L\left(y_i, x_i; \alpha_j, \beta_k\right)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{r} I\left(y_i = j\right) \log P\left(Y_i = j \middle| x_i\right)$$
$$= I\left(y_i = j\right) \times \sum_{i=1}^{n}\sum_{j=1}^{r} \log\left[\begin{array}{l} F\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) \\ -F\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right) \end{array}\right], \quad (3)$$

where $I\left(y_i = j\right) = \begin{cases} 1, y_i = j \\ 0, \text{otherwise} \end{cases}$.

The first derivative of the log-likelihood can be written as:

$$\frac{\partial L\left(y_i, x_i; \alpha_j, \beta_k\right)}{\partial \beta_k}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{r} I\left(y_i = j\right) \frac{1}{P\left(Y = j \middle| x_i\right)} \frac{\partial P\left(Y = j \middle| x_i\right)}{\partial \beta_k}$$
$$I\left(y_i = j\right) \times$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{r} \frac{\left[f\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) - f\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right)\right] x_{ik}}{F\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) - F\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right)} \quad (4)$$

where,

$$\frac{\partial P\left(Y = y_i \middle| x_i\right)}{\partial \beta_k}$$
$$= -\left[f\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) - f\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right)\right] x_{ik}.$$

Franses and Paap [16] had expressed the generalized residuals:

$$e_{ij} = \frac{f\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) - f\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right)}{F\left(\alpha_j - \sum_{k=1}^{p} \beta_k x_{ik}\right) - F\left(\alpha_{j-1} - \sum_{k=1}^{p} \beta_k x_{ik}\right)} \quad (5)$$

Equation (4) can be written as:

$$\frac{\partial L\left(y_i, x_i; \alpha_j, \beta_k\right)}{\partial \beta_k} = -\sum_{i=1}^{n}\sum_{j=1}^{r} I\left(y_i = j\right) e_{ij} x_{ik} \quad (6)$$

By equating with zero, equation (6) becomes

$$-\sum_{i=1}^{n}\sum_{j=1}^{r} I\left(y_i = j\right) e_{ij} x_{ik} = 0 \quad (7)$$

Equation (7) has the same structure as the ML equation of the Gaussian in linear regression model, $\sum_{i=1}^{n} \Psi_i x_i = 0$ that can be solved using Newton-Raphson iterations.

The generalized residuals in equation (5) may be affected by extreme data in the response *Y* that has limited responses, {1,2,...,*r*}. Categories inconsistencies may occur when dissatisfied selection by respondents or collection errors. The resulting residual gives an indelible impression in equation (7).

### 3.3. Robustness

An estimator can be tested for its robustness through the size of the breakdown point (Croux et al. [10]). Breakdown point measures the reliability of estimated procedures by determining the minimum breakdown of the observations before providing unreasonable estimates. The higher the breakdown point, the more robust the estimates are in overcoming extreme data.

The finite sample breakdown for the estimator is the smallest fraction $\alpha$ of data points that is if $n \times \alpha$ approaches infinity, causing the estimator to be unavailable. The formula for the breakdown point is:

$$\eta\left(\hat{\beta}_m; y_1, y_2, ..., y_m\right) = \frac{m^*}{n} \times 100 \tag{8}$$

where: $\hat{\beta}_m$ is the estimator for *m* data points all of which are extreme data, $m^* = \max\left\{m \geq 0 : \hat{\beta}_m < \infty\right\}$ and $n$ is sample size.

Generally, the breakdown point should not exceed 50%. Based on Rousseeuw and Leroy [18], it is impossible to distinguish between the original distribution and the contaminating distribution when more than 50% of the points are contaminated.

### 3.4. Robust Estimation

This research proposes the use of a robust method to overcome the problems faced by the ML estimator for POM. Among the popular robust estimates of the regression model are the least absolute deviations, the least median of squares, the M-estimator, and the MM-estimator Abu-Shawiesh, Riaz, and Khaliq [19], Nugroho, Wardhani, Fernandes, and Solimun [20].

The M-estimator model contains all steps as in the derivative model to the maximum likelihood model and will be used in this study. The idea of this estimator exists by introducing weights into equation (7) such as:

$$\sum_{i=1}^{n} \left\{ \begin{matrix} w\left(e_{ij}\right) s\left(y_i, x_i; \alpha_j, \beta_k\right) \\ -E\left[ w\left(e_{ij}\right) s\left(y_i, x_i; \alpha_j, \beta_k\right) \right] \end{matrix} \right\} = 0, \tag{9}$$

where, $s\left(y_i, x_i; \alpha_j, \beta_k\right) = -\sum_{j=1}^{r} I\left(y_i = j\right) e_{ij} x_{ik}$ is an individual score function that obtained in equation (7).

The M-estimator residuals function must meet three conditions which are not negative value, not a descending function, and a symmetry (Iannario et al. [6]). Given that the M-estimator does not change to scale, it is necessary to modify the residuals of M-estimator to solve the problem of minimisation.

The M-estimator has a breakdown point near to 0.5. From various objectives functions, Huber is one of the most popular functions (Jiang, Wang, Fu, and Wang, [21]). This weight has also been shown to be more efficient than ML estimator for ordinal regression model as per the study conducted by Iannario et al. [6]. Based on its popularity and performance, this method is used for robust estimation. The weight function of Huber is given by:

$$w_H\left(e_{ij}\right) = \begin{cases} 1 \ , \ \sum_{j=1}^{r} I\left(y_i = j\right)\left|e_{ij}\right| \leq c \\ \dfrac{c}{\sum_{j=1}^{r} I\left(y_i = j\right)\left|e_{ij}\right|}, \\ \sum_{j=1}^{r} I\left(y_i = j\right)\left|e_{ij}\right| > c \end{cases} \tag{10}$$

where c is fixed with the best option for linear regression equivalent to 1.345σ 345σ (Susanti, Pratiwi, H., and Liana, [22]), where σ is the standard deviation of the residuals. The standard deviation of the residuals should be estimated using a robust dispersion measure of standard deviation. Usually, the estimator used is $\hat{\sigma} = \frac{AMR}{0.6745}$, where AMR is the median absolute residuals.

### 3.5. Monte Carlo Simulation Procedure

Monte Carlo simulation studies can be used to assess the accuracy of existing statistical models under bad conditions. Monte Carlo simulation is utilised to evaluate the robustness of model estimators on simulation data considering different percentages of contamination for various residuals standard deviations for sample sizes varies. The log odds model used in this study is $Y_i^* = \alpha_j - \sum_{k=1}^{5} \beta_k X_{ik} + \varepsilon_i$, where $i = 1,2,...,n$ and $j = 1,2,...,12$.

All the variables used have ordinal categories scale.

The simulation steps are as follows:

Step 1

Sample size (n) options are 100, 250, and 500. Divide the sample into two groups which consist of good and contamination points. The number of extreme data point is obtained by $n \times p$ where $p$ is the percentage of contamination with level 5%, 10%, 20%, 40%, and 50%. Meanwhile, the rest of the sample will become good points.

Step 2

Generate random data for five covariates, *X*, four of which have an ordinal discrete scale with range values between 0 and 12. While another covariate has a continuous scale normally distributed.

Step 3

Generate the standard deviation of errors for good points,

$e$ and extreme data points, $e_{cont.}$.

Step 4

Both errors are follow a cumulative logistic distribution with the same value of mean and respective values of standard deviation, *e~cumulative logistic(μ=0,σ=1)* and $e_{cont.}$ *~cumulative logistic(μ=0,σ=k)* where *k*=2,3, and 4.

Step 5

Generate all true parameters of slope, $\beta$ from Uniform (0,1).

Step 6

Generate log odds values using equation of the chosen ordinal model. The intercept parameters $\alpha_j$ are fixed and increasing.

Step 7

Calculate the cumulative probability for each category of *Y* using cumulative logistic distribution, $P(Y \le y) = \frac{\exp(\log odds)}{1+\exp(\log odds)}$.

Step 8

From the cumulative probability values in Step 7, variable Y is generated by taking a value between 0 and 12.

Step 9

Fit the model using ML estimator, then find the standardised residuals of model, $e_{ij}$.

Step 10

Fit the model again, but this time, add Huber weighting into the estimates as given in Equation (10).

Step 11

Find the p-values of Wald test for each of the estimates $\hat{\beta}$. Basically, all the p-values must be less than 0.05 (at 5% significance level). Only the fitting models that are significant on each of their variable are accepted in this simulation.

Step 12

The process is replicated 200 times and continued by calculating the value of bias and root mean square error (RMSE) of $\hat{\beta}$ for ML estimator and M-estimator proposed. The formulas for both measurements are given by: $Bias = \frac{\sum_{i=1}^{200}(\hat{\beta}_i - \beta)}{200}$ and $RMSE = \sqrt{\frac{\sum_{i=1}^{200}(\hat{\beta}_i - \beta)^2}{200}}$. The estimator that gives the smallest value of bias and RMSE is considered the best estimator in estimating the parameters $\beta$ (Iannario et al. [6]).

Step 13

Finally, the overall goodness of fit test will be based on the value of Lipsitz statistic. The best model fitting will provide the smallest value of the average test statistic.

From the simulation steps above, steps 1 to 8 will produce simulation data that can be used to test model estimator. The simulation data consider the percentage of contamination and distance of extreme points from the most points. The model estimator is measured using bias, RMSE, and Lipsitz statistic. The bias and RMSE methods will show the most accurate estimates of model parameters by the model estimator. Lipsitz statistic determines if there is a significant difference between the observed frequency and the expected frequency in categories. The results of the study are based on an extensive Monte Carlo simulation using the statistical software, *R*. Measurement results will be further discussed in the analysis section.

# 4. Results and Discussion

Monte Carlo simulation was used to test the estimation of ML estimator and the proposed M-estimator. In general, robust estimates can facilitate model fitting of a variety of data and provide reasonable results. The use of simulation data allows the model to be tested with data designed in various forms. The random data generated should meet the requirements of ordinal regression and have a combination of three constant values that vary. The constant values are percentage of contamination, the standard deviation of the residuals, and sample size. The percentage of selected contamination was 5%, 10%, 20%, 40%, and 50%. The standard deviation of residual for the contamination data are logistically distributed with standard deviation of *2s*, *3s*, and *4s* where *s* is the standard deviation of residual for the most points which are not considered as extreme points. The sample sizes used were 100, 250, and 500 and were replicated for 200 times. The parameter estimates were measured using bias and RMSE. Meanwhile, the overall goodness of fit was measured using Lipsitz statistic. Five parameters were used in the model such as Beta1, Beta2, Beta3, Beta4, and Beta5. The results of the simulation are shown in the diagram.

Based on Figure 1, the ML estimator can refine the model fitting to the data that contain extreme points with a 2s standard deviation of residuals. Overall, the parameter estimates from the combination between the percent of contamination and the sample size produce a consistent pattern. The ML estimator bias shows consistent pattern for different sample sizes. The RMSE ML estimator approaches the zero-horizontal line when the sample size increases. Additionally, the results of the ML estimates on data containing extreme points with 3s standard deviation of the residuals can only complement 5% and 10% of the percentage of contamination. This shows the breakdown point for the ML estimator for data containing extreme

points 3 times far compared to the most points is 10%. However, the estimator of ML produces a 50% breakdown point for the data that contain 2 times the distance of an extreme point.

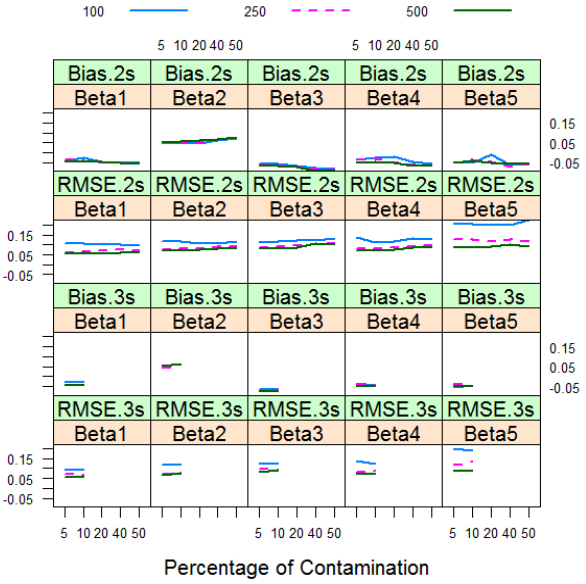


**Figure 1.** Bias and RMSE of ML for different standard deviations of residuals

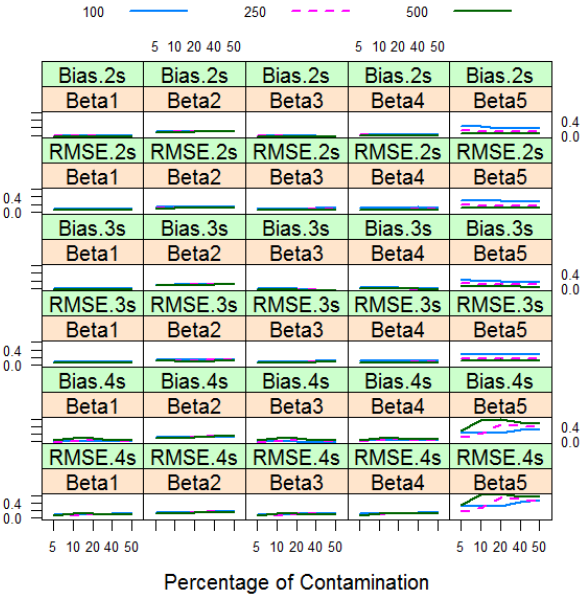


**Figure 2.** Bias and RMSE of Huber for different standard deviations of residuals

The result of parameterisation of the Huber method to simulation data is shown in Figure 2. As ML estimates, the bias and RMSE for Huber show a consistent pattern where the bias has the same pattern according to the contamination level. Meanwhile, the RMSE approaches the zero-horizontal line as the sample size increases, but only for *2s* and *3s* standard deviation of residuals. For data containing extreme points that have a 4-fold distance, the RMSE pattern is not directly related to the sample size. Huber estimator can overcome data problems with 50% of the extreme points, although the distance of the points is very far from most points. This finding corresponds to the results of Huber's robust estimation on the other regression model that provides a 50% breakdown point (Dasiou and Moyssiadis [23]).

Previous analysis only examines the estimation pattern for both estimators separately based on the combination of sample size, percentage of contamination, and standard deviation of residuals. In determining whether the proposed M-estimator is more accurate to the true parameter, the plot of bias and RMSE between ML and Huber in the same diagram are given in Figure 3. The plot is in accordance with the sample size and all the results are taken from the model fitting using *2s* standard deviation of the residuals of extreme points. This is because only this standard deviation provides a complete fitting result for both estimators for various combinations of sample size and percentage of contamination.

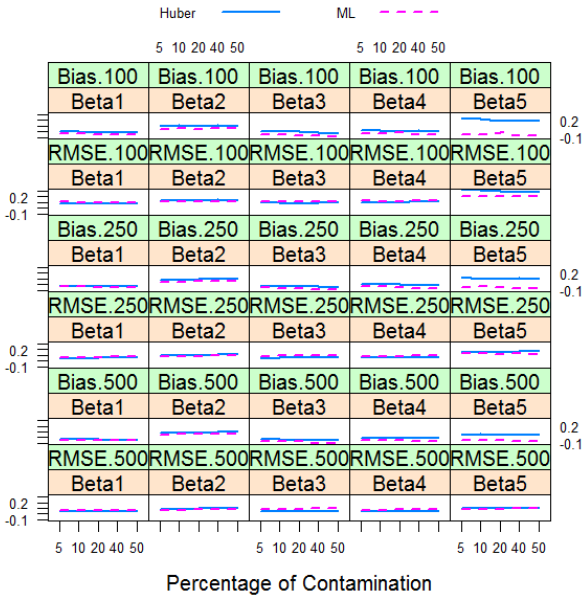


**Figure 3.** Bias and RMSE of ML and Huber for various sample sizes

Based on the diagram, Huber bias produces lower line patterns than ML bias for most parameters estimates except for Beta4 and Beta5. The bias pattern for both estimators is consistent with the percentage of contamination even though the value increased which is in contrast to the RMSE pattern where the Huber estimator has lower lines for Beta1, Beta3, and Beta4 of parameter estimates. Both measurements do not have a direct pattern on the percentage of contamination or have a slightly constant pattern. The comparative analysis concluded that the Huber estimator provides the nearest estimates values to the true parameters whereby three of the five parameters of the measurement test are in favour of Huber through simulation results.

The final part of the analysis for this study aims to determine the best overall model by using the Lipsitz

statistic. The use of p-value for testing is not required as all the model fitted provide significant results to the test. Hence, comparisons based on test statistic are more accurate in determining the best overall model. The estimator method that has the lowest statistic value is chosen as the best model. Lipsitz statistic for both estimation methods are shown in Figure 4 for different sample sizes and various percentages of contamination.

Test statistic for the proposed M-estimator were identified to give the smallest value consistently compared to ML estimates. The results remain despite the combination of sample size, the percentages for contamination, and the standard deviation of residuals are different. It reinforces the conclusion that Huber's estimator can improve the existing estimator in terms of parameter estimation and overall model fit. In other words, the estimator of the Huber is more efficient compared with the ML estimator. However, the value of test statistic for Huber estimators becomes large and less consistent when the distance of the extreme points goes further than most points. The statistics value also increased when the sample size increased. Further research needs to be carried out to investigate the causes of the occurrence. Attention to the increase in replicated numbers and the greater standard deviation of residual values is recommended for future studies. However, there is a constraint since high-end computers with faster processing are needed to ensure results are obtained consistently and accurately.

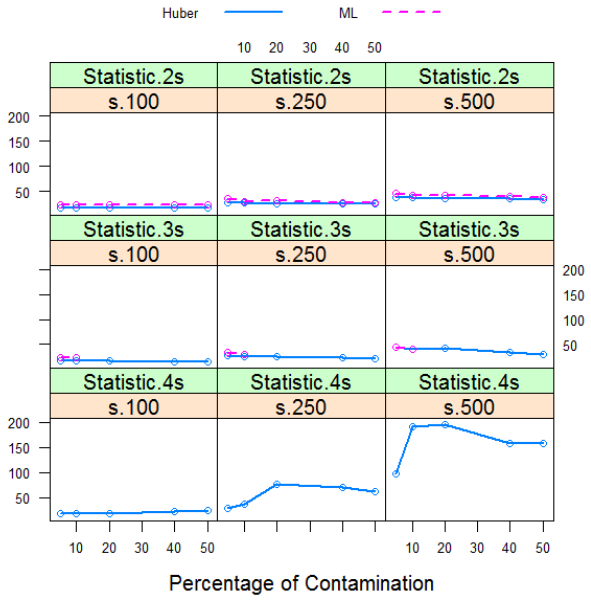


**Figure 4.** Lipsitz Statistic of ML and Huber Estimators

## 5. Conclusions

The focus of this study is to examine the performance of an existing estimate which is ML in producing model fitting for cumulative odds logit model through simulation data that contain extreme data. At the same time, this study has proposed alternative methods in estimating parameters and model fitting by using robust methods. The robust method of the M-estimator was selected based on the recommendation of previous researchers due to easy-to-apply reason and the same problem can be solved with other regression models. Through extensive reading, Huber's robust M-estimator was never tested for its effectiveness on the ordinal cumulative model with polytomous ordered categories of explanatory variables. It is a huge gap for statistical research and should be given attention in improving the existing model.

This study has successfully tested the accuracy of estimating methods of simulation data with a different combination of sample size, percentage of contamination, and standard deviation of residuals. The results of the Monte Carlo simulation of both estimators show Huber estimator delivered the best results for the parameter estimation and overall model fitting. Huber estimator has achieved a breakdown point of 50% for data that contain extreme points that are quite far from most points. In addition, the existence of extreme points that have only a distance of two times far from the most points does not have a major impact on ML estimation. This means that the estimation for ML and Huber is likely to yield the same result if the residual values of the model are at a range between -2 and 2. This situation may also occur for data that have a contamination percentage of below 5%.

In ensuring that the results of the analysis are accurate and consistent for different types of data, the researchers suggest for simulation method to be conducted on data that have an increased sample size, further extreme points distances, and a large number of replicates. For that, simulation results can only be generated with the help of high-end computers and the application of machine learning methods that can improve the speed and accuracy of the test.

## Acknowledgements

Researchers would also like to express gratitude towards UiTM and UPSI for giving them the opportunity to conduct this study.

## REFERENCES

[1] F. Zulkifli, Z. Mohamed, N. A. Azmee, R. Z. Abidin. Predicting Students' Final Exam Grades Using Cumulative Odds Model, Int. J. Psychosoc. Rehabil., Vol. 24, No. 8, 8711–8727, 2020.

[2] F. R. Hampel, E. Ronchetti, P. J. Rousseeuw, W. A. Stahel. Robust Statistics: the Approach Based on Influence Functions, Wiley, 1986.

[3] P. J. Huber. Robust Statistics, J. Wiley & Sons, New York, 1981.

[4] P. J. Huber, E. Ronchetti. Robust Statistics, 2nd ed., J. Wiley & Sons, New York, 2009.

[5] R. A. Maronna, R. D. Martin, V. Yohai. Robust Statistics: Theory and Methods, J. Wiley & Sons, New York, 2006.

[6] M. Iannario, A. C. Monti, D. Piccolo, E. Ronchetti. Robust Inference for Ordinal Response Models, Vol. 11, 3407–3445, 2017.

[7] F. Zulkifli, Z. Mohamed, N. A. Azmee, R. Z. Abidin. Integrating Mirt in Ordinal Regression Modeling to Predict End-of-semester Exam Grades, Int. J. Psychosoc. Rehabil., Vol. 24, No. 8, 8693–8710, 2020.

[8] E. Francis. Effects of Some Coding Techniques on Multicollinearity and Model Statistics, Math. Theory Model., Vol. 8, No. 4, 156–167, 2018.

[9] M. Iannario, A. Clara, D. Piccolo. Robustness Issues for CUB Models, TEST, 2016.

[10] C. Croux, G. Haesbroeck, C. Ruwet. Robust Estimation for Ordinal Regression, J. Stat. Plan. Inference, Vol. 143, No. 9, 1486–1499, 2013.

[11] I. Moustaki, M.P. Victoria-Feser. Bounded-Influence Robust Estimation in Generalized Linear Latent Variable Models, J. Am. Stat. Assoc., Vol. 101, No. 474, 644–653, 2006.

[12] J. H. Albert, S. Chib. Bayesian Analysis of Binary and Polychotomous Response Data, J. Am. Stat. Assoc., Vol. 88, No. 422, 669–679, 1993.

[13] J. Albert, S. Chib. Bayesian Residual Analysis for Binary Response Regression Models, Biometrika, Vol. 82, No. 4, 747, 1995.

[14] S. H. Walker, D. B. Duncan. Estimation of the Probability of an Event as a Function of Several Independent Variables, Biometrika, Vol. 54, 167–179, 1967.

[15] P. McCullagh. Regression Models for Ordinal Data, J. R. Stat. Soc. Ser. B, Vol. 42, 109–142, 1980.

[16] P. H. Franses, R. Paap. Quantitative Models in Marketing Research, Cambridge University Press, 2010.

[17] S. M. Shaharudin, N. Ahmad, N. H. Zainuddin, N. S. Mohamed. Identification of Rainfall Patterns on Hydrological Simulation Using Robust Principal Component Analysis, Indones. J. Electr. Eng. Comput. Sci., Vol. 11, No. 3, 1162–1167, 2018.

[18] P. J. Rousseeuw, A. M. Leroy. Robust Regression and Outlier Detection. Hoboken, John Wiley & Sons, Inc., NJ, USA, 1987.

[19] M. O. A. Abu-Shawiesh, M. Riaz, Q. U. A. Khaliq. MTSD-TCC: A Robust Alternative to Tukey's Control Chart (TCC) Based on the Modified Trimmed Standard Deviation (MTSD), *Mathematics and Statistics*, Vol. 8, No. 3, 262–277, 2020. DOI: 10.13189/ms.2020.080304

[20] W. H. Nugroho, N. W. S. Wardhani, A. A. R. Fernandes, Solimun. Robust Regression Analysis Study for Data With Outliers at Some Significance Levels, *Mathematics and Statistic*s, Vol. 8, No. 4, 373–381, 2020. DOI: 10.13189/ms.2020.080401

[21] Y. Jiang, Y. G. Wang, L. Fu, X. Wang. Robust Estimation Using Modified Huber's Functions With New Tails, Technometrics, Vol. 61, No. 1, 111–122, 2019.

[22] Y. Susanti, H. Pratiwi, S. S. H., T. Liana. M Estimation, S Estimation, and MM Estimation in Robust Regression, Int. J. Pure Appl. Math., Vol. 91, No. 3, 349–360, 2014.

[23] D. Dasiou, C. Moyssiadis. The 50% Breakdown Point in Simultaneous M-estimation of Location and Scale, Stat. Pap., Vol. 42, 243–252, 2001.