Novel Approach to Predict Ground-Level Ozone Concentration Using S-estimation and MM-Estimimation

Ahmad Zia Ul-Saufie
Universiti Teknologi Mara, Malysia
ahmadzia101@uitm.edu.my

Dhiya Al-Jumeily
Liverpool John Moores University
United Kingdom
d.aljumeily@ljmu.ac.uk

Abir Hussain
Liverpool John Moores University
United Kingdom
d.aljumeily@ljmu.ac.uk

Muqhlisah Muhamad
Universiti Teknologi Mara, Malysia
muqhlismuhd@gmail.com

Jamila Musafina
Kazan Federal University
Russia
DNMustafina@kpfu.ru

Fawaz Ghali
Liverpool John Moores University
United Kingdom
F.Ghali@ljmu.ac.uk

Thar Baker
Liverpool John Moores University
United Kingdom
T.M.Shamsa@ljmu.ac.uk

***Abstract*— Ground-level ozone concentration is one of the main concerns for air pollution, due to the negative impacts on human health, animals, foliage, climate and the whole ecosystem. The aim of this paper is to reduce the influential outliers by including weightages within robust method to avoid the bias of the model. The influential outliers from x-space (predictors) have been identified using leverage values. Furthermore, Cook's distance and standardized residual have been computed to clarify the influential outliers from both of x-space and y-direction. S-estimation and MM-estimation have been introduced as a new approach for reducing the influential outliers from x-space and both of y-direction and x-space respectively. The comparison between the robust method and the ordinary least square method shows that, the accuracy measures of the robust method have been improved by around 0.94% (D+1), 0.56% (D+2) and 1.85% (D+3) respectively.**

Keywords*— *Ozone prediction model, S-estimation, MM-estimation, Robust Regression, OLS

I. Introduction

Ground-level ozone (O3) has become a pollutant of concern in Malaysia. Even though it is not the primary pollutant, it does produce a large negative impact on human health, living things and crop yield. According to Sanna (2009), the emission of NOx and VOCs, which would be extremely active in heavy traffics on the roads, together with the heat and sunlight will form the smog of O3. Regression analysis is widely used as a method to predict the concentration level of air pollutants. This study used several independent variables to form a new equation to predict the level of O3 concentration. The method was known as multiple linear regression. The outliers is one of the common issues in developing a preidction model using regression models. Even a single point of outlier could lead to incorrect inferences and distort the regression analysis (Sarkar, et al., 2011). The outliers are a huge different observation point from another point of observation. The presence of outliers will affect the accuracy of the prediction model in the forming of regression coefficients (Yahaya, 2014). Real data contain outliers as mentioned in (Ul-Saufie, 2012). According to Osborne and Overbay (2004), outliers occur because of data recording/sampling, the damage of monitoring instrument in data recording and the error in data acquisition/data management. According to Kudryavtzev (2009), OLS is not suitable for non-normal distribution data. As an alternative method, the robust regression method will provide acceptable results when the data contains influential. Robust regression is an iterative procedure that has many steps to identify outliers and minimize the influential outlier including M estimation, MM estimation, S-estimation, Least Absolute Value Estimation (LTS), Least Trimmed Squares estimation (LTS) and others. In this paper, we focus on S-estimation and MM estimation for predicting O3 concentration for three days.

Different variables have been used in this study, which consist of ozone (O3, ppb), carbon monoxide (CO, ppb), nitric oxide (NO: ppb), sulphur dioxide (SO2, ppb), nitrogen dioxide (NO2, ppb) and three meteorological parameters; i.e. ambient temperature (T, oC), wind speed (WS, km/h), relative humidity (RH, %). In this study, the level of O3 concentration for the following period (next day (D+1), next two days (D+2) and next three days (D+3)) are used as the predictors. The selection of variables is summarized in Table 1.

TABLE I. Variables Selection in the Prediction of O3 Concentration Level

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Authors	WS	T	RH	NOx	NO	SO2	NO2	CO	O3, D-n
Agirre-Basurko, Ibarra-Berastegi, and Madariaga (2006)							X		X
Musa, Jemain, and Wan Zin (2013)									X
Jaioun, Saithanu, and Mekparyu (2014)								X	X
Wang, Lu, Wang, and Leung (2003)	X	X		X	X	X	X	X	X
Ghazali, et al. (2010)							X		X
Heo, Kim, and Kim (2004)	X	X	X			X	X	X	X
Delcloo and Backer (2005)		X	X	X					X
Ramli, Ghazali, and Yahaya (2010)	X	X		X					X
Banan, Latif, Juneng, and Firoz Khan (2014)	X	X	X	X	X	X	X	X	X
Schlink, et al. (2006)	X	X			X		X		X
This study	X	X	X	X	X	X	X	X	X

II. Methodology

Pasir Gudang, specifically the coordinate N 1.470750o, E 103.895702o is located at Sekolah Menengah Pasir Gudang 2 of the city of Johor. It is the main location of many industrial activities and the main road for transportation (Ahmat, Yahaya, & Ramli, 2015). Currently, Pasir Gudang has high impact industries with total population of 100,000 (Afzali, et al., 2018). This helped to attract many investments from local and foreign investors (Ahmad, Majid, M. Yusoff, Abdullah, & Othman, 1994). In air pollution monitoring procedures, United States Environmental Protection Agency (EPA) has written some guidelines to measure air pollutants and meteorological variables (Banan, et al., 2013). The secondary hourly data was obtained from Department of Environmental Malaysia (DoE) from 1st January 2002 until 31st December 2014. In this study, the hourly concentrations for each variable selected were transformed into daily 12 hours average concentration, from 7am to 7pm, as and according to (Awang, et al., 2013), the O3 concentration level was suspected to be highly active during this period (Awang, Ramli, Mohammed, & Yahaya, 2013). Mohammed et al. (2013) have confirmed that, the most of areas in Malaysia have a large emission of O3 concentration level during the day time activities i.e. from 7am to 7pm. The most popular methods in the imputation of missing value for air pollution data is what has been suggested by Yahaya, et al., (2005), the mean imputation technique. In this method, the missing values will be replaced by the mean, which is obtained between the above value and the below value, known usually as the Mean Above Below method (MAB) as implemented and recommended by Noor et al. (2008).

Based on the randomized dataset presented by Paarsch and his colleagues (Paarsch and Golyaev, 2016), in the period of Year 2002 and Year 2012, 80% of the dataset were selected as a training dataset and the other 20% of the data were selected to be used for validating the model as suggested by Ul-Saufie (2012). The 2013 and 2014 dataset were used to verify the models obtained. According to Sarkar et al., (2011), the data containing influential outlier will lead to incorrect inferences as welll as to reducing the accuracy. There are several methods to identify the influential outliers such as leverage values, standardized residual (ZRE) and Cook's distance.

III. Model Development for O3 Concentrations level

MLR is an extension model that was developed from a simple linear regression. It has one response variable and several predictors. In this study, 89% (n=3214 observations) were randomly selected as a training model to form the equation for developing the prediction model. Table 2 lists the general equations that were used for this study.

TABLE II. Multiple Linear Regression Model for O3 Concentrations Level

Day of prediction	Model
Next day prediction (D+1)	O3, D+1 = β0 + β1WS + β2T + β3RH + β4NO + β5SO2 + β6NO2 + β7O3 + β8CO
Next two days prediction (D+2)	O3, D+2 = β0 + β1WS + β2T + β3RH + β4NO + β5SO2 + β6NO2 + β7O3 + β8CO
Next three days prediction (D+3)	O3, D+3 = β0 + β1WS + β2T + β3RH + β4NO + β5SO2 + β6NO2 + β7O3 + β8CO

where O3, D+1: Next day prediction of O3 concentration (ppb); O3, D+2: Next two days prediction of O3 concentration (ppb); O3, D+3: Next three days prediction of O3 concentration (ppb); WS: Wind speed (km/h); T: Temperature (0C); RH: Relative humidity (%); NO: Nitric oxides (ppb); SO2: Sulphur dioxides (ppb); NO2: Nitrogen dioxides (ppb); O3: Ozone (ppb); CO: Carbon monoxides (ppb).

Rousseeuw and Yohai (1984) proposed the S-estimation as a method of breakdown value, that is based on estimation scale, which was reflected in the name of it. Another study by Alma, (2011) defined the breakdown point/value as a contamination percentage of the data which can cause regression coefficient (parameter estimate), that will lead to the loss of accuracy. Hubert and Michiel Debruyne (2009) in their study, also defined breakdown point as a result of the smallest fraction from the impurity of the outliers in data observation. The MM-estimation is a method introduced by Yohai (1987), as a combination between two methods, M-estimation and S-estimation. In the other words, this method is a combination between efficient estimation and high breakdown value. The performance indicators are used to measure the performance of the model using accuracy and error measures. In this study, four performance indicators were used, prediction accuracy (PA), Index of Agreement (IA) Root Mean Square Error (RMSE) and normalized absolute error (NAE).

IV. Discussion

Descriptive Statistics analysis is used to describe the characteristics of ozone concentration in Pasir Gudang monitoring station. The results are shown in Table 3. The Pasir Gudang monitoring station is surrounded by major roads and industrial area. The average level of O3 from year 2002 until year 2014 was recorded at 19.906 ppb. The maximum value of O3 was recorded at 57.00 ppb. The reason behind such level was reported and justified by Mahmud et al., (2012). It was reported that the reason behind such high value of O3 could relate to the fact that transboundary haze pollution occurred as a result of the huge biomass burning activities in the west and central Kalimantan region. The distribution of the data was skewed to the right with a standard deviation of (7.759) and a skewness value of 0.572.

TABLE III. Descriptive of O3 Concentration for the Pasir Gudang Region

Pasir Gudang	O_3
Mean	19.906
Median	19.250
Mode	19.906
Standard Deviation	7.759
Variance	60.209
Skewness	0.572
Kurtosis	0.584
Maximum	57.000
Percentiles 95%	33.923
Percentiles 99%	41.154

A. *S-estimation.*

In order to reduce the influential outliers from x-space, the breakdown points for both Bisquare and Yohai were specified at 11.93% and 13.48% respectively when the data approximately normally are distributed at 95% level of efficiency. Table 5 shows the models and the performance indicators of using the Bisquare and Yohai methods for next day prediction (D+1) of O3 concentration level. The results have shown that, Bisquare method has performed better than the Yohai method, subsequently the values of IA and PA are higher than in Yohai method and the value of NAE is smaller.

TABLE IV. Performance Indicators for Next Day (D+1) Prediction of O3 Concentration Level

Method	T/Constant	Models	NAE	RMSE	IA	PA
Bisquare	4.685	$O_{3, D+1} = 21.56716 + 0.280961WS - 0.461792T - 0.040229RH - 0.059840NO - 0.058701SO_2 + 0.197210NO_2 + 0.494698O_3 + 0.001463CO$	0.253355	6.572068	0.588135	0.435088
Yohai	1.0600	$O_{3, D+1} = 21.90695 + 0.294545WS - 0.462654T - 0.044088RH - 0.059333NO - 0.070086SO_2 + 0.204291NO_2 + 0.479965O_3 + 0.001679CO$	0.253678	6.572062	0.582360	0.433808

B. *MM-estimation.*

In MM-estimation approach, the influential outliers were minimized for y-direction through the tuning of constant at 95% level of efficiency. Besides, the outliers from x-space were reduced by 25% from the breakdown value. Table 6 shows the models and the performance indicators for Bisquare and Yohai methods for next day prediction (D+1) of O3 concentration level. The results indicated that, Bisquare method performs better as a model compared to Yohai method, since the values of IA and PA are higher than in Yohai method and the value of NAE and RMSE are smaller than in Yohai method.

TABLE V. Performance Indicators for Next Day (D+1) Prediction of O3 Concentration Level

Method	T/Constant	Models	NAE	RMSE	IA	PA
Bisquare	4.685	$O_{3, D+1} = 21.50252 + 0.283241WS - 0.462100T - 0.039965RH - 0.058993NO - 0.058707SO_2 + 0.197262NO_2 + 0.496285O_3 + 0.001442CO$	0.253360	6.573392	0.588631	0.435059
Yohai	1.0600	$O_{3, D+1} = 21.82301 + 0.293819WS - 0.461138T - 0.043761RH - 0.057916NO - 0.071421SO_2 + 0.203102NO_2 + 0.481871O_3 + 0.001640CO$	0.253711	6.574745	0.582747	0.433520

Selection of the best method. A comparison study to identify the best method of predicting the level of O3 concentration was carried out between the model developed by the ordinary least square estimate and the best preformed model that is selected from each S-estimation and MM-estimation. The results of the comparison study for the next day prediction (D+1) in Pasir Gudang region are presented in Table 7. It is clear from the results presented in Table 8, that S-estimation using Bisquare method preforms better than the other listed models, since the rank sum of the performance indicators is the smallest.

TABLE VI. Selection of the Best Method

Method	T/Constant	Models for Pasir Gudang (D+1)	NAE	RMSE	IA	PA
OLS	-	$O_{3,\ D+1} = 22.425489 + 0.205303WS - 0.432579T - 0.042338RH - 0.079944NO - 0.054896SO_2 + 0.187287NO_2 + 0.460949O_3 + 0.001810CO$	0.253471	6.547462	0.576533	0.435650
S (Bisquare)	4.685	$O_{3,\ D+1} = 21.56716 + 0.280961WS - 0.461792T - 0.040229RH - 0.059840NO - 0.058701SO_2 + 0.197210NO_2 + 0.494698O_3 + 0.001463CO$	0.253355	6.572068	0.588135	0.435088
MM (Bisquare)	4.685	$O_{3,\ D+1} = 21.50252 + 0.283241WS - 0.462100T - 0.039965RH - 0.058993NO - 0.058707SO_2 + 0.197262NO_2 + 0.496285O_3 + 0.001442CO$	0.253360	6.573392	0.588631	0.435059

TABLE VII. PERFORMANCE INDICATORS RANKING OF ALL MODELS FOR NEXT PREDICTION (D+1) IN PASIR GUDANG

Method	Models for Pasir Gudang (D+1)	NAE	RMSE	IA	PA	Sum
OLS	$O_{3,\ D+1} = 22.425489 + 0.205303WS - 0.432579T - 0.042338RH - 0.079944NO - 0.054896SO_2 + 0.187287NO_2 + 0.460949O_3 + 0.001810CO$	3	1	3	1	8
S (Bisquare)	$O_{3,\ D+1} = 21.56716 + 0.280961WS - 0.461792T - 0.040229RH - 0.059840NO - 0.058701SO_2 + 0.197210NO_2 + 0.494698O_3 + 0.001463CO$	1	2	2	2	7
MM (Bisquare)	$O_{3,\ D+1} = 21.50252 + 0.283241WS - 0.462100T - 0.039965RH - 0.058993NO - 0.058707SO_2 + 0.197262NO_2 + 0.496285O_3 + 0.001442CO$	2	3	1	3	9

C. Verification of the proposed Model.

A verifying procedure has been carried out to examine the best model (to be used) obtained for each day of the prediction, for the period of year 2013 and year 2014. The verification procedure involved the comparison of four performance indicators i.e. IA, RMSE, PA and NAE. This research results show that, all the models selected for Pasir Gudang region have a strong possibility to be used in 2013 and 2014 to predict the level of O3 concentration for next day (D+1), next two days (D+2) and next three days (D+3) when the values of performance indicators between models' verification and models' validation are close (Table 9).

TABLE VIII. PERFORMANCE INDICATORS OF THE MODEL VALIDATION AND MODEL VERIFICATION

Method	Evaluation	NAE	RMSE	IA	PA
S (Bisquare): D+1	Validation	0.253355	6.572068	0.588135	0.435088
	Verification	0.279497	7.636684	0.623867	0.528242
S (Bisquare): D+2	Validation	0.273855	6.926628	0.459301	0.306606
	Verification	0.276181	8.277184	0.527250	0.413508
MM (Bisquare): D+3	Validation	0.293860	7.336129	0.344515	0.227065
	Verification	0.289815	8.661464	0.461398	0.344348

V. CONCLUSION

The paper showed that the proposed robust method is better than the ordinary least square method. Previous research used the approach of reducing the weightage for the influential outliers in air pollution data. The average accuracies of the proposed robust method are 0.5116 (D+1), 0.3830 (D+2) and 0.2858 (D+3), which are better than the ordinary least square method in which the average accuracy is 0.5061 (D+1), 0.3806 (D+2) and 0.2811 (D+1). Moreover, the proposed robust regression models showed the improvement from the ordinary least square method, when the average accuracy for the D+1 model, D+2 model, and D+3 model has been increased by 0.94%, 0.56%, and 1.85% respectively. Therefore, these proposed models could be implemented by different interested parties including, the public health agencies, government, citizen and the other authorities to take proactive action to avoid the negative impact of O3 concentration.

ACKNOWLEDGMENT

A special appreciation to the Department of Environmental Malaysia (DoE) for providing the air quality dataset to support this research, with a special thanks to Universiti Teknologi Mara, Malaysia and the Ministry of Higher Education Malaysia for funding this study under the grant number (600-IRMI/FRGS 5/3 (289/2019)).

REFERENCES

[1] Abd Hamid, N.Z., Md Noorani, M. S., and Adenan, N. H. (2017). Chaotic analysis and short-term prediction of ozone pollution in Malaysian urban area, Journal of Physics: Conference Series, Volume 890.

[2] Afzali, A., Rashid, M., Afzali, M., and Younesi, V. (2017). Prediction of air pollutants concentrations from multiple sources using AERMOD coupled with WRF prognostic model J. Clean. Prod., 166 (2017), pp. 1216-1225.

[3] Ahmad, R., Majid, Z., M. Yusoff, M., Abdullah, M., & Othman, A. (1994). Air Quality of Pasir Gudang Industrial Estate. FNEHR Conference.

[4] Ahmat, H., Yahaya, A., & Ramli, N. (2015). Particulate Matter Analysis for Three Industrialized Areas Using Extreme Value. Sains Malaysiana, 44 (2): 175-185.

[5] Alma, O. G. (2011). Comparison of Robust Regression Methods in Linear Regression. Int. J. Contemp. Math. Sciences, 6, 9, 409-421.

[6] ASMA. (2013). Alam Sekitar Malaysia Sdn. Bhd. http://www.environmalaysia.com.

[7] Awang, N. R., Ramli, N., Mohammed, N., & Yahaya, A. (2013). Times Series Evaluation of Ozone Concentrations in Malaysia Based on Location of Monitoring Stations. International Journal of Engineering and Technology, 3.

[8] Banan, N., Latif, M. L., Juneng, L., & Ahamad, F. (2013). Characteristics of Surface Ozone Concentrations at Stations with Different Backgrounds in the Malaysian Peninsula. Aerosol and Air Quality Research, 13: 1090-1106.

[9] Banja, M., Papanastasiou, D., Poupkou, A., & Melas, D. (2012). Development of a Short-Term Ozone Prediction Tool in Tirana Area Based on Meteorological Variables. Atmospheric Pollution Research, 3: 32-38.

[10] Berry, W. D., & Feldman, S. (1985). Multiple Regression in Practice. Quantitative Application in Social Sciences. Newbury Park, : Sage University Paper.

[11] Blatna, D. (2005). Outliers in Regression. University of Economic Prague.

[12] Bohrnstedt, G., & Knoke, D. (1982). Norusis's SPSS 11 Chapter 22 on "Analyzing Residual:" Hamilton's Chapter on "Robust Regression". In Pindyck, & Rubinfeld, Statistics for Social Data Analyis.

[13] Chatterjee, S., & Hadi, A. (2006). Regression Analysis by Example (Fourth ed.). New Jersey: John Wiley.

[14] Chen, C. (2002). Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Sugi 27, Statistics and Data Analysis.

[15] Cohen, A., Bromberg, S., Buechley, R., Heiderscheit, L., & And Shy, C. (1972). Asthma and Air Pollution from a Coal-Fuelled Power Plant. Public Health, 62, 1181-1188.

[16] (2008). Commission for Environmental Cooperation (CEC). The North American Mosaic: An Overview of Key Environmental Issues. Canada.

[17] DoE. (2014). Department of Environment, Malaysia. Malaysia Environmental Quality Report 2013. Kuala Lumpur: Department of Environment, Ministry of Natural Resources and Environment, Malaysia.

[18] EPA. (2012). Ground Level Ozone Make it Harder to Breathe, United State Environmental Protection Agency. Online [Accessed 23 April 2015] Available from World Wide: http://www.epa.gov./.

[19] Field, A. (2005). Discovering Statistics Using SPSS (Second ed.). London: Sage.

[20] Gervasi, O. (2008). Computational Science and Its Applications, Italy. Springer.

[21] Ghazali, N. A., Ramli, N., Yahaya, A., MD Yusof, N. F., Sansuddin, N., & Al Madhoun , W. (2010). Transformation of Nitrogen Dioxide into Ozone and Prediction of Ozone Concentrations Using Multiple Linear Regression. Environ Monit Assess, 165: 475-489.

[22] Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo, Ann. Stat.

[23] Hubert, M., & Michiel Debruyne. (2009). Breakdown Value. Inc. WIREs Comp Stat, 1, 296-302.

[24] Jammalamadaka, S. R., & Lund, U. (2003). The Effect of Wind Direction on Ozone Levels - A Case Study.

[25] Jones, & Bartlett. (2006). Chemistry of the Troposphere.

[26] Klym, A. H. (2007). Detecting Outliers and Influential Data Points in Receiver Operating Characteristic (ROC) Analysis.

[27] Mahmud, M., & Mohd Shafie, S. (2012). Perbandingan Kualiti Udara Antara Tahun 2002 dan 2003 di Kawasan Padi Sewaktu Pembakaran Jerami di Mergong, kedah. Malaysian Journal of Society and Space, 8(3), 33-42.

[28] Mohammed, N. I., Ramli , N., & Yahaya, A. (2013). Ozone Phytotoxicity Evaluation and prediction of Crops Production in Tropical Regions. Atmospheric Environment, 343-349.

[29] Osborne, J. W., & Overbay, A. (2004). The Power of Outliers (and why researcher should always check for them). Practical Assessment, Research & Evaluation.

[30] Paarsch, H. J., & Golyaev, K. (2016). A Gentle Introduction to Effective Computing in Quantitative Research. United States.

[31] Pandey, & Chasta. (2006). Evalutation on Air Pollution Phytotoxicity in Urban Environment of Udaipur, Rajasthan. Plant Achieves, 6(1).

[32] Render, Stair, & Hanna. (2013). Quantitative Analysis for Managemet.

[33] Robenson, S. M., & Steyn, D. (1990). Evaluation and Comparison of Statistical Forecast Models for Daily Maximum Ozone Concentrations. Atmospheric Environment, 24, 303-312.

[34] Rousseeuw, P. J., & Yohai, V. (1984). Least Median of Squares Regression. Journal of the American Statistical Association, 79, 871-880.

[35] Salibian, M., & Yohai, V. (2006). A Fast Algorithm for S-Regression Estimates. Journal of Computational and Graphical Statistics, 15, 414-427.

[36] Sanna, E. (2009). Air Pollution and Health . New York: Health and the Environment.

[37] Sarkar, S. K., Midi, H., & Rana, R. (2011). Detection of Outliers and Influential Observations in Binary Logistic Regression: An Empirical Study. Applied Sciences, 11 (1): 26-35.

[38] Souza, A. d., Aristones, F., Pavao, H., & Fernandes, W. (2014). Development of a Short-Term Ozone Prediction Tool in Campo Grande-MS-Brazil Area Based on Meteorological Variables. Open Journal of Air Pollution.

[39] Stevens, D. L. (2002). Sampling Design and Statistical Analysis Methods for the Integrated Biological and Physical Monitoring of Oregon Streams. Corvalis, Oregon.

[40] Susanti, Y., Pratiwi, H., & Sulistijowati H., S. (2013). Optimasi Model Regresi RobustUntuk Memprediksi Produksi Kedelai di Indonesia. ISBN: 978-979-16353-9-4.

[41] Susanti, Y., Pratiwi, H., Sulistijowati H, S., & Twenty, L. (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. International Journal of Pure and Applied Mathematics.

[42] Ul-Saufie, A. Z. (2012). Future Daily Particulate Matter Concentrations Prediction Using Regression Artificial Neural Network and Hybrid Models in Malaysia. Pulau Pinang : Universiti Sains Malaysia.

[43] Wahab, S. A., Bouhamra, W., Sowerby, B., & Crittenden, B. (1996). Predicting Ozone Levels: A Statistical Model for Predicting Ozone Levels in the Shuaiba Industrial Area, Kuwait. Environmental Science and Pollution Research, 3 (4), 195-204.

[44] Wang, Y.-G., Xu Lin, Min Zhu, & Zhidong Bai. (2007). Robust Estimal Using the Huber Function With a Data-Dependent Tuning Constant. Journal of Computational and Graphical Statistics, 16, 1-14.

[45] Yahaya, A. S. (2014). Applied Regression Models Using SPSS.

[46] Yohai, V. J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression, Annals of Statistics.

[47] Yuliana. (2014). Penerapan Model Regresi Linear Robust Dengan Estimasi M Pada Data Nilai Kalkulus II Mahasiswa Universitas Widya Dharma Klaten.