International Journal of Engineering & Technology, 7 (3.25) (2018) 44-48

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Increasing T-Method Accuracy Through Application of Robust M-Estimatior

N Harudin[1]*, Jamaludin, K R[2] , M Nabil Muhtazaruddin[3] , Ramlie F[4] ; S.H , Ismail[5], Wan Zuki Azman Wan Muhamad[6], NN Jaafar[7]

[1,2,3,4,5,6,7]UTM Razak School in Engineering & Advanced Technology Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia
[1]Department of Mechanical Engineering, Universiti Tenaga Nasional, 43000 Kajang Selangor, Malaysia
**Corresponding author E-mail: nolia.harudin@gmail.com*

Abstract

Mahalanobis Taguchi System is an analytical tool involving classification, clustering as well as prediction techniques. T-Method which is part of it is a multivariate analysis technique designed mainly for prediction and optimization purposes. The good things about T-Method is that prediction is always possible even with limited sample size. In applying T-Method, the analyst is advised to clearly understand the trend and states of the data population since this method is good in dealing with limited sample size data but for higher samples or extremely high samples data it might have more things to ponder. T-Method is not being mentioned robust to the effect of outliers within it, so dealing with high sample data will put the prediction accuracy at risk. By incorporating outliers in overall data analysis, it may contribute to a non-normality state beside the entire classical methods breakdown. Considering the risk towards lower prediction accuracy, it is important to consider the risk of lower accuracy for the individual estimates so that the overall prediction accuracy will be increased. Dealing with that intention, there exist several robust parameters estimates such as M-estimator, that able to give good results even with the data contain or may not contain outliers in it. Generalized inverse regression estimator (GIR) also been used in this research as well as Ordinary Lease Square Method (OLS) as part of comparison study. Embedding these methods into T-Method individual estimates conditionally helps in enhancing the accuracy of the T-Method while analyzing the robustness of T-method itself. However, from the 3 main case studies been used within this analysis, it shows that T-Method contributed to a better and acceptable performance with error percentages range 2.5% ~ 22.8% between all cases compared to other methods. M-estimator is proved to be sensitive with data consist of leverage point in x-axis as well as data with limited sample size. Referring to these 3 case studies only, it can be concluded that robust M-estimator is not feasible to be applied into T-Method as of now. Further enhance analysis is needed to encounter issues such as Airfoil noise case study data which T -method contributed to highest error% prediction. Hence further analysis need to be done for better result review.

Keywords: *T-Method; Robust M-estimator; Prediction,*

1. Introduction

In 1936, Dr. Prasanta Chandra who is a famous Indian statistician has introduced a theory called Mahalanobis distance(MD) which used in representing a mathematical way in defining distance between two objects. Years later, in 2000 the quality guru, Dr. Taguchi who interested in combining the idea of this MD as the normal target region identification with signal to noise ratio and orthogonal array as fusion element in creating a new analysis tool called Mahalanobis Taguchi System. One of the governing tool under MTS was T-Method which specifically developed for designing a prediction model that adapting the signal to noise ratio (SNR) as weightage element inside the model to predict the forecasting value relying on historical data. The theory underlying T-Method consist of unit space concept and zero-proportional linear regression while orthogonal array as part of its optimization which not include in this analysis. The key different of T-method are the application of unit space concept in defining the average normalize data as well as the fusion of signal to noise ratio as weightage element in the overall model formula. Most of the researchers agreed and aware that MTS including T-method is not relying on any probability distribution theory due to its function in only measuring the descriptive statistic (1-3) . As an example, we can refer to the normalization stage of T-Method which involved differences between signal space and unit space data only without standard deviation consideration.

Its simplicity to be understood and its limited sample size consideration are the two main advantages of T-Method. The practitioner able to do a prediction even with very limited sample size. Application of T-Method are still limited but keep increasing I yearly basis. There are a few case study on T-method application conducted within these six years past which dealing several sort prediction analysis such as (4-6). The practitioner of T-Method required to have clear understanding on the selected data behavior since the effect of outliers is not been considered within the population data selected. There is no discussion mentioned so far on the robustness of T-method towards outliers. This grey area need to be exercised and further enhanced so it can lead to the robustness of T-method. The very well-known method in considering outliers effect was robust estimator. There is various enhanced robust estimator existed such as S-estimator and MM-estimator, however this study only focused on classical M-estimator.

Back then, it is a common practiced among the statisticians that any regression analysis should be free from the effect of outliers.

Due to this, the normality assumption was assumed to be normally distributed, randomized, independently and identically disseminated as well equally dependable. Robust regression was attempted to limit the influence of outliers by replacing the square of the residuals in the estimation of β. Regression M-estimation (7) which developed decades back is an alternative robust procedure to improve least-square by minimizing the residual function with respect to the same scale (σ). The letter M indicates that M-estimation is an estimation of the maximum likelihood type. Nowadays, with the aid of several available prevalence of programs like Maple, Mathematica, R programming and many more, the calculation towards M-estimation complexity is just the matter of daily routine manipulation. As stated by Stefanski and Boos, (8) as well, the key advantages of M-estimator is a very large class of asymptotically normal statistics including delta method transformation can be put in general M-estimator framework. Despite the advantages, there are several findings that highlight the limitation of M-estimation in certain cases which are M-estimators are vulnerable to leverage points (9) as well as Huber type M-estimates are robust against outliers in y-direction, but they are not robust against leverage points (outliers in x-direction) (10). These limitations seem to be true for this research since in most cases, M-estimator is not really lowering the effect of outliers in doing the prediction. Generalized regression analysis (GIR) (11) and Ordinary Least Square (OLS) are two methods used in this study as a comparison method to T-Method and M-estimator in improving the estimator or probability coefficient. The main intention of this study is to see the feasibility in applying M-estimator proportional coefficient (β_M) into T-Method proportional coefficient (β_M) to increase the prediction accuracy.

2. Methodology/Materials

2.1. T-Method

The theory behind T-Method formulation, are a practical combination between Mahalanobis Distance theory and the concept of S/N Ratio (SNR). The core elements of SNR (*sensitivity, linearity and variability*) are still the main focus while considering the dynamic environment bring an added value towards the whole concept. By following the concept of reference point proportional equation, the linear regression line generated will pass the zero point (origin) of the graph. In making the prediction, the establishment on unit space and signal space is done for the normalization data. In fact, selection of unit space is one of the most crucial process to be define in the early stage prior any analysis to be conducted. The selection of unit space should be homogenous as possible as well as taken from a highly dense data within the population. Taguchi clearly mentioned in (12) by following these rules, the selection of unit space will always be in the middle position between low and high data of the population selected. Figure 1 explained about this unit space concept in much clear picture. By having this clear concept in minds as well as very knowledgeable on the data trend, one can definitely increase the prediction accuracy as well as strong justification can be made.

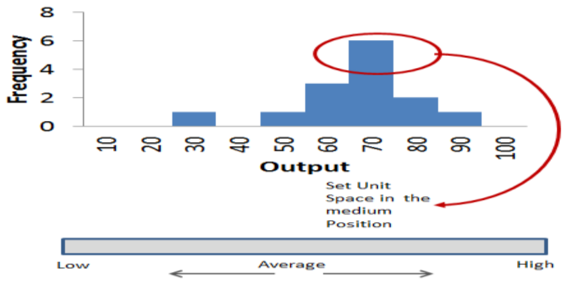

Fig 1: The concept of unit space selection for T-Method

In order to fulfill the prediction model which is the Integrated estimate output value $\hat{M}$ as in equation 2.1, the proportional coefficient (β) and SN ratio (η) need to be computed item by item basis with the use of normalized data (X_{ij}) calculated by equation 2.2. Equation 2.3 until equation 2.9 are the formulation of proportional coefficient (β) and SN ratio (η). If the value of SN ratio η calculated to be negative value, it need to be assumed as zero. It is clearly seen that higher S/N ratio of an item, will definitely contribute to a greater degree of contribution on overall model estimation.

$$M_i = \frac{\eta_1 \times \left(\frac{X_{i1}}{\beta_1}\right) + \eta_2 \times \left(\frac{X_{i2}}{\beta_2}\right) + \cdots . + \eta_j \times \left(\frac{X_{ij}}{\beta_j}\right)}{\eta_1 + \eta_2 + \cdots + \eta_j} \tag{2.1}$$

$$\textit{normalized data } (X) = \text{ signal data} - \text{average of unit space} \tag{2.2}$$

$$\text{Effective Divider, r} = \mathrm{M_1}^2 + \mathrm{M_2}^2 + \cdots + \mathrm{M_i}^2 \tag{2.3}$$

$$\text{Total Variation}, S_T = \mathrm{X_1}^2 + \mathrm{X_2}^2 + \cdots + \mathrm{X_i}^2 \tag{2.4}$$

$$\text{Variation of proportional term}, S_\beta = \frac{(M_1X_{11} + M_2X_{21} + \cdots + M_lX_{l1})^2}{r} \tag{2.5}$$

$$\text{Error variation}, S_e = S_T - S_\beta \tag{2.6}$$

$$\text{Error Variance}, V_e = \frac{S_e}{l-1} \tag{2.7}$$

$$\text{SN ratio}, \eta = \left(\frac{(S_\beta - V_e)}{r\,V_e}\right) \tag{2.8}$$

$$\beta_M = \frac{M_1X_{11} + M_2X_{21} + \cdots + M_lX_{l1}}{r} \tag{2.9}$$

Once the overall model elements been identified, the integrated estimate value (Ŷ) for unknown data can be calculated easily. Average unit space value of the output called (M_0) will need to be added with the Integrated estimate output value $\hat{M}$ will provide the estimate value of ($\hat{Y}_1$) for the unknown data No.1. The procedure repeated accordingly for the remaining unknown data available.

$$\hat{Y}_1 = \hat{M}_1 + M_0 \tag{2.10}$$

2.2. Robust M-Estimator

M-estimation principle is to minimize the residual function ρ.

$$\hat{\beta}_M = \min_\beta \rho\left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) \tag{2.11}$$

As summarized by Susanti, Y ,the following steps were the keys in applying the M-estimator (13).
1. Estimate regression coefficients on the data using OLS.
2. Test assumptions of the regression model
3. Detect the presence of outliers in the data.
4. Calculate estimated parameter $\hat{\beta}_0$with OLS.
5. Calculate residual value $e_i = y_i - \hat{y}_i$.
6. Calculate value $\hat{\sigma}_i$= 1.4826 MAD.
7. Calculate value $u_i = e_i / \hat{\sigma}_i$.
8. Calculate the weighted function value w_i, using c = 4.685 for Tukey’s bisquare weighted function

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & , |u_i| \le 4.685 \\ 0, & , |u_i| > 4.685 \end{cases} \tag{2.12}$$

9. Calculate $\hat{\beta}_M$ using iteratively reweighted least squares (IRLS) with weighted w_i.
10. Repeat steps 5-8 to obtain a convergent value of $\hat{\beta}_M$.

$$\sum_{i=1}^{n} x_{ij} w_i^0 \left(y_i - \sum_j^k x_{ij}\,\beta^0\right) = 0\,, \quad j = 0,1,2,\ldots k \tag{2.13}$$

2.3. M-Estimator into T-Method Application

As explained and discuss in section 2.1 and 2.2 above, the adaption of M-estimator into T-method was done by taking the minimum $\hat{\beta}_M$ value converged through equation 2.12 into equation 2.1 from T-Method which lead to the following equation.

$$M_i = \frac{\eta_1 \times \left(\frac{X_{i1}}{\hat{\beta}_{M1}}\right) + \eta_2 \times \left(\frac{X_{i2}}{\hat{\beta}_{M2}}\right) + \cdots . + \eta_j \times \left(\frac{X_{ij}}{\hat{\beta}_{Mj}}\right)}{\eta_1 + \eta_2 + \cdots + \eta_j} \tag{2.14}$$

2.4. Generalized Inverse Regression Estimator (GIR Estimator) into T-Method

The enhancement of individual estimator accuracy in T-Method initially done by (11) by applying the concept of generalize inverse regression (GIR) into T-Method as function of β in overall T-Method equation as shown in 2.1 and 2.14. The concept helps in reducing the risk on the issues of unstable and infinite mean square error (MSe) of the estimate produced by the classical estimator since it is the ratio of two random variables that follow normal distributions. The fluctuations in the new data in the final model are incorporated into the predictions for each item. The GIR estimator can be used as a linear calibration when new data are obtained. Equation 2.15 represent the meaning of applying the β from GIR-estimator theory into equation 2.1 or 2.14.

$$\hat{\beta}_M = \frac{{\beta_j}^2 + c\hat{V}(\beta_j)}{\beta_j} \tag{2.15}$$

The Value of β_j is obtained using equation 2.9 while $c\hat{V}(\beta_j)$ is following the remaining equation listed below with c is constant 1. Variance of the mean square error (MSE) of the data is calculated using equation 2.16 while the MSE is calculated by equation 2.17. The $\emptyset$ is following the linear equation calculated in equation 2.2.

$$\hat{V}(\beta_j) = \frac{S_j^2}{r} \tag{2.16}$$

$$S_j^2 = \frac{S_{ej}}{\emptyset} \tag{2.17}$$

$$S_{ej}^* = error\ square\ of\ normalize\ data \tag{2.18}$$

$$\emptyset = L \tag{2.19}$$

2.5. Data Collection, Equipment and Tool Setup

Data for the 1st case study which involved 15 variables and 14 sample sizes was taken from a real industrial case conducted at Nuclear Agency Malaysia on power consumption prediction. 2nd case study on yield% prediction used to be one of the reference example data which involved 6 variables with only 7 sample sizes was taken from Teshima (12) book. While for the airfoil-noise prediction case study, the data was taken from the UCI Machine learning data collection (14) which involve huge sample size data. The main intention of this paper is to see how robust these methods react to outliers especially extreme outliers. Due to that data of extreme outliers were added purposely for each case to visualize the impact to all method. Yield% prediction case involve only 1 unknown data while the power consumption case involved 13 unknown data to be analyzed and airfoil-noise are 20 sample data. The actual output from the unknown data is used to calculate the error % differences from output calculated over the prediction formula. Matlab R2015a application software was used to construct the T-Method and robust estimator formulation for comparable study. R programming is applied in order to calculate the robust M-estimator proportional coefficient (β) value before adapting the result into Matlab code.

3. Results and Findings

As been mentioned earlier, 3 differences set of case study were used as comparison in this research. The actual data of the unknown trend is taken for a comparison purpose. Throughout the entire analysis process, it been clearly seen that the behavior of the data was very crucial which may affect the entire analysis. If the analysts are very well trained and well known with the data trend, that might be low risk on making high error prediction but those without so knowledgeable on the data trend might face difficulty in doing accurate prediction. This section will discuss the detail findings across these 3 case studies.

3.1. Results of Four Different Methods Comparison

Table 1 shows that M–estimator represent moderate performance which slightly better than OLS except for Airfoil noise case study. T-method shows more accurate and not much error percentages prediction result in most cases except for Airfoil noise with extreme outliers. Error percentages result analysis in table 1 seems synchronize with R-squared for model estimator shown in figure 2. Extreme outliers were purposely added to see the impact to regression model for each case. Overall R-squared value is not varying much among 4 comparison method except for:

a) GIR in Airfoil noise prediction with extreme outlier case (23.8%)
b) GIR & OLS in power consumption case study (57.8% & 58.5% respectively)

The performance of GIR estimator as well not vary much with T-Method. In fact, it shows better performance for case of Airfoil self-noise. However, dealing with minimum sample data and non-normal data trend still contribute to lower performance to the overall prediction provided by GIR estimator. The disadvantages of M-estimator which is sensitive to leverage point as mentioned in section 1, seems to be true since the data for instance the Airfoil-noise scatter plot in figure 3c are showing various outliers with high leverage point in x as well as y direction for most of the parameter involved. Yield percentages scatter plot in figure 3a shows data is having low residual even with extreme outliers added to it. The outliers seem to provide good prediction towards the overall analysis. While power consumption plot in figure 3b are showing various outliers cases including high leverage point in x and y but not all the parameters are having bad residuals trend. These are the key important things need to be further understand since T-method is performing worse with data that having very bad outliers trend (non-not such as Airfoil self-noise case study. The result as well shows that M-estimator is quite sensitive towards data with lower sample size. Limitation on the need of normal trend data as well as outliers effect in overall T-Method is still the major area to be enhanced since the value of error% among all case study is still higher and should be able to improve further. The selection of these 3 case studies seem to be not enough to conclude that neither M-estimator is not feasible to be applied into T-Method nor concluding that T-Method is already robust to outliers.

Table 1: Summary of prediction error percentages between T-method, OLS, GIR and M-estimator

Case Study	**OLS-estimator**	**M-estimator**	**GIR estimator**	**T-method**
Power cons prediction with extreme outliers	7.27%	7.12%	5.07%	5.00%
yield % prediction	12.22%	12.22%	11.02%	9.67%

with extreme outliers				
Airfoil_noise prediction with extreme outliers	10.40%	10.40%	11.02%	22.81%
Airfoil_noise prediction	7.21%	11.47%	7.08%	7.26%
yield % prediction	12.43%	12.22%	6.80%	2.50%
Power consumption prediction	4.82%	4.72%	2.78%	3.22%

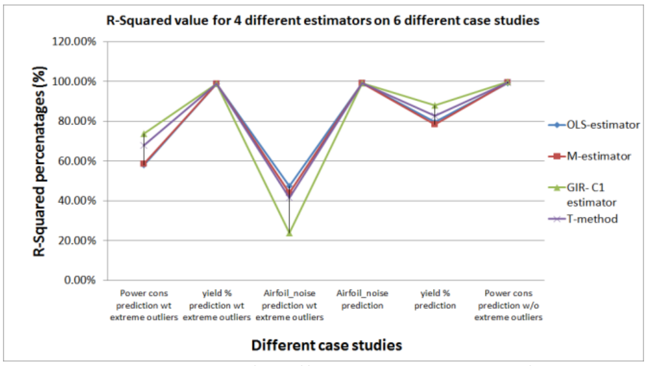

Fig. 2: R-squared value for different cases and method of analysis

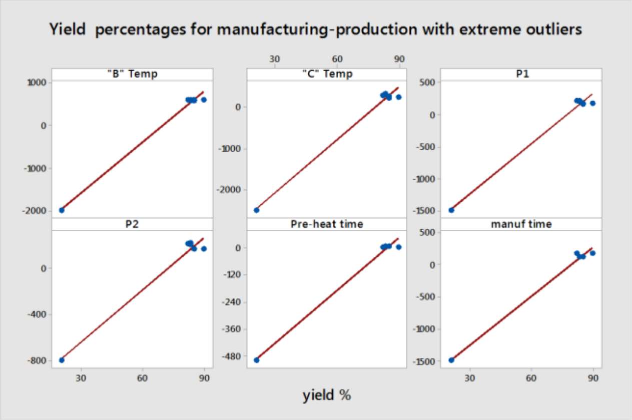

Fig. 3a: Yield percentages with extreme outliers

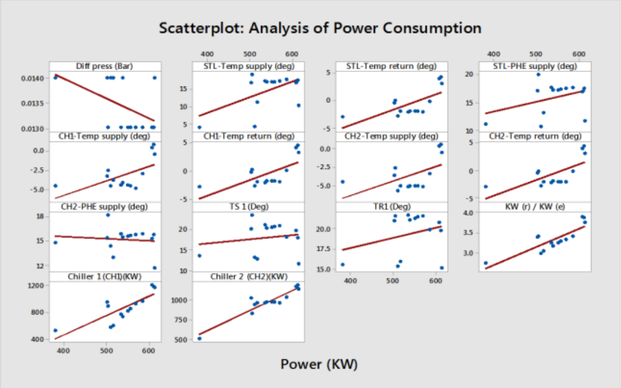

Fig. 3b: Scatter plot for Power consumption

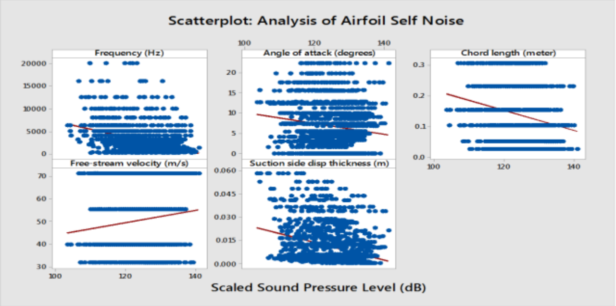

Fig. 3c: Scatter plot for Airfoil self-noise

3.2. Other Potential Robust Estimator Method

The methods selected to be adapted in this study is mainly the early regression theory available for line fitting. Over the years, there are various enhancement been made to the ordinary least square (OLS) method as well as M-estimator which is part of the robust line fitting estimator that enhance the OLS. Robust line fitting is massively used nowadays in outlier rejection. In robust line fitting the goal is to find the regression parameter values of the line model that minimize the residual errors. Many other robust estimators are available for robust line fitting such S-estimator and MM-estimator. They are function to minimize the dispersion of the residuals (15). S estimation is based on residual scale (standard deviation) of M-estimation intentionally to overcome the weakness of median as weightage value in M-estimator. MM-estimator procedure is to estimate the regression parameter using S-estimator which minimize the scale of the residual from M-estimator and then proceed with M-estimator. MM-estimator aims to obtain estimates that have a high breakdown value and more efficient (13). There are various opportunity to be explored in future for the purpose of improving T-method prediction accuracy especially when dealing with outliers data as well as huge data with multiple issues including heteroscedastic, multicollinearity and autocorrelation. MM-estimator, S-estimator or even Robust least square method is a good start for that.

4. Conclusion

Through all the discussion mentioned earlier, it is not sufficient enough to summarize the overall effectiveness of these 4 methods especially on T-Method by just relying on these 3 case studies. Results of this study are mainly proven to be effective only for these 3 case study. A lot more case study need to be used as trial run so that better review can be made across several issues and the robustness of T-method can be represent in more accurate manner. T-Method in this study chosen to be a better approach dealing with:

a) *Data consist outliers with lower residual effect*
b) *Data with limited sample size*
c) *Data with moderate outliers and lower variation of data.*
d) *Data that is normally distributed*

M-estimator is quite sensitive to outliers with high leverage point as well as data with limited sample size. It can be concluded that M-estimator is not feasible for this research study, however this is limited to these 3 case studies analysis. Further enhance analysis is needed to encounter issues such Airfoil noise data which T-method contribute to highest error percentages prediction. Overall, T-Method theory which easily been used and practice as well as easy to understand, can be applied in various area including economic trend prediction, weather forecasting, failure rate prediction, health prediction, product inspection, building & structure monitoring, and many more.

Acknowledgement

Special thanks to Mr Yusnan for his guidance and help during the data collection stage at nuclear agency Malaysia and not to forget the MOHE, Malaysia as well as UNITEN for their financial support during the author research duration period. This research paper publication is mainly supported by Tier 1 Research University Grant No: 15H43(Universiti Teknologi Malaysia).

References

[1] Jugulum R, Taguchi G, Taguchi S, Wilkins JO. Discussion. Technometrics. 2003;45(1):16-21.

[2] Kim SB, Tsui K-L, Sukchotrat T, Chen VC. A comparison study and discussion of the Mahalanobis-Taguchi System. International Journal of Industrial and Systems Engineering. 2009;4(6):631-44.

[3] Woodall WH, Koudelik R, Tsui K-L, Kim SB, Stoumbos ZG, Carvounis CP. A review and analysis of the Mahalanobis—Taguchi system. Technometrics. 2003;45(1):1-15.

[4] Daniels B, Corns S, Cudney E, editors. A comparison of representations for the prediction of ground-level ozone concentration. Evolutionary Computation (CEC), 2012 IEEE Congress on; 2012: IEEE.

[5] Nedeltcheva GN, Ragsdell KM. TECHNICAL QUALITY ANALYSIS OF GLOBAL STOCK MARKETS.

[6] Negishi S, Morimoto Y, Takayama S, Ishigame A. Daily Peak Load Forecasting by Taguchi's T Method. Electrical Engineering in Japan. 2017;201(1):57-65.

[7] Huber PJ. Robust regression: asymptotics, conjectures and Monte Carlo. The Annals of Statistics. 1973;1(5):799-821.

[8] Stefanski LA, Boos DD. The calculus of M-estimation. The American Statistician. 2002;56(1):29-38.

[9] Víšek JÁ, editor Advantages And Disadvantages, Challanes And Threads Of Robust Methods. In Proceedings of the seminar “Analyza dat 2012/II” (Data Analysis 2012/II), 14; 2012.

[10] Arslan O, Edlund O, Ekblom H. Algorithms to compute CM-and S-estimates for regression. Metrika. 2002;55(1-2):37-51.

[11] Kawada H, Nagata Y. An application of a generalized inverse regression estimator to Taguchi's T-Method. Total Quality Science. 2015;1(1):12-21.

[12] Teshima S. Quality Recognition & Prediction: Smarter pattern technology with the Mahalanobis-Taguchi System: Momentum Press; 2012.

[13] Susanti Y, Pratiwi H. M estimation, S estimation, and MM estimation in robust regression. International Journal of Pure and Applied Mathematics. 2014;91(3):349-60.

[14] Lichman M. UCI machine learning repository. Irvine, CA; 2013.

[15] Pitselis G. A review on robust estimators applied to regression credibility. Journal of Computational and Applied Mathematics. 2013;239:231-49.