Tikrit Journal of Administrative and Economic Sciences, Vol. 21, No. 70, Part (2): 355-373
Doi: www.doi.org/10.25130/tjaes.21.70.2.19

Tikrit Journal of Administrative and Economics Sciences
مجلة تكريت للعلوم الإدارية والاقتصادية
EISSN: 3006-9149
PISSN: 1813-1719

Improving the Performance of the Andrews Weighted Function in a Robust Multiple Linear Regression Model

Diyar lezghin ramdan*[A], Sarbast Saeed Ismael [B], Sarween Asaad Othman[C], Taha Hussein Ali [C]

[A] College of Health Sciences/ Duhok University
[B] College of Administration and Economics/ Duhok University
[C] College of Administration and Economics/ Salahaddin University

Keywords:
Multiple Regression model, Outliers, Robust estimation, Andrews's function, and Tune parameter.

ARTICLE INFO

Article history:

Received	22 Nov. 2024
Accepted	22 Dec. 2024
Available online	30 Jun. 2025

*Corresponding author:
Diyar lezghin ramdan
College of Health Sciences/Duhok University

Abstract: Outliers affect the accuracy of estimating multiple linear regression model parameters and lead to estimated parameters that are inaccurate and far from their true values; thus, robust estimator methods such as the Andrews weighted function must be used to obtain more accurate parameters and robustness versus outliers. The proposed method involves choosing the optimal tuning parameter value that produces the minimum mean square error of the parameters and treating outliers. Simulation and real data were used to compare the efficiency of the models estimated based on the classical robust method for Andrews weighted function that uses the default tune parameter and the proposed algorithm through the MATLAB program dedicated to this purpose. The research results revealed the efficiency of the proposed algorithm in estimating optimal tuning parameters for the Andrews weighted function and treating outliers and the accuracy of estimating the model parameters.

Tikrit Journal of Administrative and Economic Sciences, Vol. 21, No. 70, Part (2): 355-373
Doi: www.doi.org/10.25130/tjaes.21.70.2.19

تحسين أداء دالة أندروز الموزونة في نموذج الانحدار الخطي المتعدد القوي

طه حسين علي	سروين أسعد عثمان	سربست سعيد إسماعيل	ديار لزگين رمضان
كلية الإدارة والاقتصاد	كلية الإدارة والاقتصاد	كلية الإدارة والاقتصاد	كلية العلوم الصحية
جامعة صلاح الدين	جامعة صلاح الدين	جامعة دهوك	جامعة دهوك

المستخلص

تؤثر القيم المتطرفة على دقة تقدير معلمات نموذج الانحدار الخطي المتعدد وتؤدي إلى معلمات مقدرة غير دقيقة وبعيدة عن قيمها الحقيقية؛ وبالتالي، يجب استخدام طرق تقدير قوية مثل دالة أندروز المرجحة للحصول على معلمات أكثر دقة ومتانة مقابل القيم المتطرفة. تتضمن الطريقة المقترحة اختيار قيمة معلمة الضبط المثلى التي تنتج الحد الأدنى لخطأ مربع متوسط المعلمات ومعالجة القيم المتطرفة. تم استخدام المحاكاة والبيانات الحقيقية لمقارنة كفاءة النماذج المقدرة بناءً على الطريقة القوية الكلاسيكية لدالة أندروز المرجحة التي تستخدم معلمة الضبط الافتراضية والخوارزمية المقترحة من خلال برنامج MATLAB المخصص لهذا الغرض. كشفت نتائج البحث عن كفاءة الخوارزمية المقترحة في تقدير معلمات الضبط المثلى لدالة أندروز المرجحة ومعالجة القيم المتطرفة ودقة تقدير معلمات النموذج.

الكلمات المفتاحية: نموذج الانحدار المتعدد، القيم المتطرفة، التقدير القوي، دالة أندروز، ومعلمة الضبط.

1. Introduction

To estimate the relationships between independent variables and the dependent variable in statistical data analysis, regression models in general and multiple linear regression, in particular, are essential tools, as we can predict the values of the dependent variable whether they are values that do not exist within the period studied or future values. These models can be used in many fields such as economics, medicine, engineering, marketing, and others (Omer et al., 2024: 112-113).

Sometimes some values differ significantly from the rest of the data and are called outliers. These values are either harmful or necessary. Both cases must be handled properly, as they affect the statistical analysis and their excitation may lead to incorrect conclusions. A single outlier can extract parameters far from the model's true parameters. There are many ways to deal with these values, including deletion. Still, sometimes these values are important in the data and cannot be deleted and must be handled with extreme caution to recognize their importance and not mislead the results (Ali et al. 2023: 9).

Some statistical methods have been developed to deal with outliers to reduce their impact on the results scientifically called robust estimators, one of these methods is called Andrews' weighted function, which is a robust estimator for dealing with data containing outliers. This method gives weights to the variables. This method has proven successful in reducing the impact of outliers on the model estimates, ensuring reliable and non-misleading results. Andrews' weighted method gives less weight to outliers. The further away the value is from the data mean, the less weight it is given to reduce its influence on parameter estimates (Maronna et al., 2019: 357-366).

Much previous research has addressed the tuning parameter and focused on strategies that give us a robust tuning parameter against outliers. Many of them concluded that these methods reduce the influence of outliers on the model's conclusions and improve the accuracy of the estimates. Some other research has been completed by testing and examining the statistical performance of these variables to expand our understanding of how to enhance the stability of statistical models (Gladwell, 2008: 25-29).

This research aims to find a new approach that improves the effectiveness of the Andrews function by determining the value of the optimal tuning parameter for it, by finding parameters that are closest to the truth parameters by reducing the mean square error between the model parameters without outliers and the estimated parameters after adding outliers, all of this to increase the stability and improve the accuracy of the estimates in the presence of outliers. Through this approach, researchers seek to extract results for more reliable and accurate statistical models.

2. **Linear Regression Model:** Many statistical techniques are used to examine relationships between variables, among which regression analysis is the most important technique for this purpose. (Legendre, 1805) and (Gauss, 1809) used the least squares method to provide the earliest version of linear regression. This method was initially used to calculate the orbits of the celestial planets until Gauss developed the theory further in 1821 when he proposed a theory called the Gauss-Markov theory, which is a fundamental principle in estimating linear models. (Ali & Awaz, 2017: 38).

A statistical model is a simplified representation of a situation or process that may produce testable hypotheses but does not represent reality. Linear regression requires linear regression parameters to determine the relationship between the independent variables and the dependent or response variable. (Omar et al., 2020: 58-67).

Due to its straightforward structure, direct computation and clear interpretation, the Multiple Linear Regression (MLR) model is one of the most widely used estimation models in research. The MLR model indicates that there is a linear connection between the independent variables $(x_1, x_2, \dots, x_n)$ and the dependent variable (Y). The regression coefficients are denoted by $\beta_0, \beta_1, \dots,$ and β_p Where β_0 is a parameter that determines the point of intersection of the regression line with the Y-axis and the $(\beta_1, \dots,$ and $\beta_p)$ are slope parameters. The error term ε in classical regression is assumed to have a normal distribution with constant variance $\mathrm{Var}(\varepsilon) = \sigma^2$ and E(ε) = 0 (Xin & Xiao, 2009, p. 3; Chagas et al., p. 5, 2016; Zhang et al. 2017, p. 6; Xie et al. 2021, p. 3; Ali et al. 2023, p. 3-6).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \qquad (1)$$

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \cdots + \widehat{\beta_p} x_p \qquad (2)$$

3. **Ordinary Least Squares Estimators**: The ordinary least squares (OLS) method works to minimize the sum of the squared differences of the errors or between the predicted values in the model and the actual values. Therefore, the OLS method is one of the most widely used methods when estimating the parameters of a linear regression model for data analysis. It is simple and effective for predicting linear relationships between variables (Ali et al., 2023: 7). There are three basic steps in OLS: First: Determine the model that represents the relationship between the dependent variable and the independent variables. Second: Estimate the parameters that give us the least summation of the squared differences between the actual and predicted data. Third: Test the model that was extracted using a set of statistics such as (R^2) or statistical tests such as hypothesis tests through some statistical tests such as the t-test. The least squares method also depends on a set of basic assumptions including homogeneity of variance, independence of errors, and normal distribution of errors. When these assumptions are met, ordinary least squares provide objective and efficient parameter estimates. In many fields such as econometrics, statistics, and social sciences, the ordinary least square is an essential tool in predictive modelling and data analysis. (Buylov, 2022: 20-22):
4. **Outliers**: Outliers can stimulate creativity and provide important insights, even though they are sometimes seen as abnormal. In his book Outliers: A Success Story, Malcolm Gladwell explains these outliers, their cumulative

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

advantages, opportunities, and how to uncover the mechanisms of their success (Gladwell, 2008: 25-29). Outliers can reveal new market trends and uncommon disease patterns in fields such as business and epidemiology. They are also useful for detecting bank card theft. However, sometimes, robust strategies are needed to control outliers because they tend to mislead the results and success of studies. Strategies such as robust regression and statistical tests resistant to outliers are used to prevent them from significantly impacting the conclusions (Ali et al., 2023: 5-7).

5. **Some Robust Estimators**: Classical estimation techniques such as OLS produce skewed parameters and misleading and inaccurate conclusions when there are outliers in the data. They are sensitive to these values and cannot deal with these values. Therefore, statistical techniques called robust estimators were created. They were specifically designed to extract accurate parameters and estimates in the presence of outliers. These techniques give weights to the data to reduce the effect of these values or extract their effect when estimating the parameters to obtain accurate conclusions. Among these methods, there is a method called the weighted least squares (WLS) method, where this method gives small weights to outliers to reduce their effect on the conclusions and estimation of parameters (Hampel et al. 1986). There is also robust logistic regression. This technique modifies classical logistic regression in a way that makes it less sensitive to outliers. When we reduce the influence of outliers, we derive estimates that more accurately and reliably represent data that contain outliers (Pregibon, 1982). Huber estimation is also a powerful strategy for outliers. In this method, to minimize the impact of these values, unconventional loss functions are used. Compared to the OLS method, this method is considered better in the case of outliers, as it punishes large deviations less severely. (Huber, 1981). Many studies have used these strategies and others to prove the superiority of these techniques in the event of the presence of outliers in the data, such as: (Ali et al. 2024); (Kareem et al. 2019); (Ali and Awaz, 2017).
6. **Weighted Andrews Function**: Andrews function is a robust technique against outliers and is one of the reliable statistical techniques. This method gives weights to data according to their distance from the data center using the weighting method. Outliers are assigned smaller weights to reduce their effect. One of the main components of the Andrews weighted function is the adjustment factor or adjustment constant or scale factor. This factor

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

determines the sensitivity of the Andrews function to outliers. The analyst can adjust this value to reduce or increase the sensitivity of the function to outliers. According to (Marona et al. 2006), a lower value of the adjustment coefficient results in less aggressive weighting of outliers, while a larger value reduces the sensitivity of the function to deviations from the central tendency (Huper and Ronchetti, 2009: 225). When using classical techniques such as ordinary least squares (OLS), which are very susceptible to outliers and provide inaccurate and time-consuming estimates, this function is particularly useful (Huber, 1992: 86-95). Andrews Weighted Function Formula is:

$$w(r) = \frac{\sin(r)}{r} \quad \text{for } |r| < \pi \text{ and } w(r) = 0 \quad \text{otherwise} \qquad (3)$$

Where, r represents the residuals scaled by a tuning constant k, which is crucial for the robustness of the estimator (Huber, 1992: 86-95).

Tuning parameters: The power and effectiveness of the estimator are mostly determined by the fine-tuning parameter k. For evenly distributed data without outliers, it manages the balance between efficiency and resilience against outliers. The fine-tuning constant for the Andrews function is about 1.339 by default. This value can be changed to change the weights given to outliers; smaller constants provide more robust estimates against outliers by assigning less weight to larger residuals, (Maronna et al. 2006: 12). Decreasing the fine-tuning constant increases, the negative weight given to large residuals, which improves robustness but may reduce efficiency. The required degree of flexibility must be taken into account when choosing the fine-tuning constant in real applications. On the other hand, if the fine-tuning constant is increased, efficiency increases but robustness decreases. In robust regression, the residuals are formulated as follows using the fine-tuning parameter (Wang et al. 2005, p. 1-5):

$$r = \frac{\text{resid}}{k.s\sqrt{1-h}} \qquad (4)$$

Where h represents the effect (or leverage) values from the least squares fit, k: is the adjustment constant, resid: are the residuals from the previous iteration (residuals or errors) and s is an estimate of the standard deviation of the error component, calculated using the formula: MAD / 0.6745 where MAD represents the absolute median of the deviations (Wang et al. 2005: 1-5).

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

7. **Mean Square Error of parameters**: The mean square error (MSE) is a popular and widely used criterion when evaluating the effectiveness of estimators in regression analysis. Therefore, in this context, the (MSE) statistic is used. This statistic includes variance and bias to improve quality and expresses the accuracy of the estimator through them. The following is the definition of the mean square error of the estimator (McLean et al. 2012: 6106-6109):

$$\mathrm{MSE}(\hat{\beta}) = \mathrm{E}(\hat{\beta} - \beta)^2 \qquad (5)$$

This measure is the difference between the expected value and the true value of the estimator or provides a balance between the estimator's variance around its expected value and its true value and the extent of its bias or variance. Robust estimates are used when dealing with outliers frequently in linear regression models to reduce and control their influence. Compared to traditional techniques such as (OLS) in the presence of extreme values, robust estimates often produce a reduced (MSE) because they reduce the influence of extreme values on the parameter estimates. Determining the correct tuning parameter is a challenge for all researchers, especially when using the Andrews weighted function. Choosing the correct tuning parameter will reduce (MSE) and is the goal of all researchers trying to obtain results with the least error and closest to the truth so that the predictions and estimates are accurate in all areas. Robust estimators and their mean square error in the presence of extreme values have been studied extensively in studies such as Huber, (1981) and Hampel et al. (1986). More recent studies, such as the work by Maruna et al. (2019), highlight the importance of the regularization parameter in improving the mean squared error while exploring developments in robust estimation methodologies (McLean et al., 2012: 6106-6109).

8. **Proposed Method**: The proposed method involves selecting the optimal tuning parameter value for the weighted Andrews function that handles outliers and produces the minimum mean square error of the coefficients in the linear regression model in the following steps:

Step 1: Set Tune = 0.1, 0.11, 0.12 …, 100. To estimate the model parameters, use the weighted Andrews function for robust estimators for all Tune values.

Step 2: Determines the optimal tune parameter value that gives the minimum mean square error of the model parameters.

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

Step 3: Use a weighted Andrews function depending on the value of the optimal tuning parameter in estimating the parameters of the robust linear regression model.

The proposed algorithm gives the largest value for the tuning parameter if there are no outliers in the data, which provides estimates of the parameters of the linear regression model completely like the ordinary least square estimators.

9. Simulation Study: To prove the superiority of the proposed method, we will compare it with classical methods such as OLS and Andrews Classic. The researchers will generate data using MATLAB randomly for simple and multiple linear regression, then add extreme values. Let us make the observation (12 and 25) for simple linear regression. We will give these two values positive numbers to show in the drawing how to extract extreme values for simple linear regression using the (OLS) method, and for multiples we will give a negative value with a positive value to prove the superiority of the proposed method in most cases of extreme values in the data relative to the sample sizes. We will use distinct sample sizes (30, 50 and 100), and repeat the sample (1000) times. Based on the differences between the real parameter values and the parameter values of the methods used, we will discover the best method or the method that is less affected by extreme values. MATLAB is designed for this purpose for simple and multiple linear regression in the appendices and we can modify it like changing the numbers or values of the extreme values or their location or sample sizes as changing their location can change the results but every time as an average we will get close results and we will start with simple linear regression and figure 1 is a regression line for the three methods with two extreme values as we mentioned earlier for the simple linear regression model as shown below.

Figure 1 illustrates how outliers affected the ordinary least squares approach by showing the regression line indicating the direction of the outliers. On the other hand, outliers did not affect the classical and proposed Andrew's function. For the three methods, the result of simple linear regression is shown in Table 1.

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

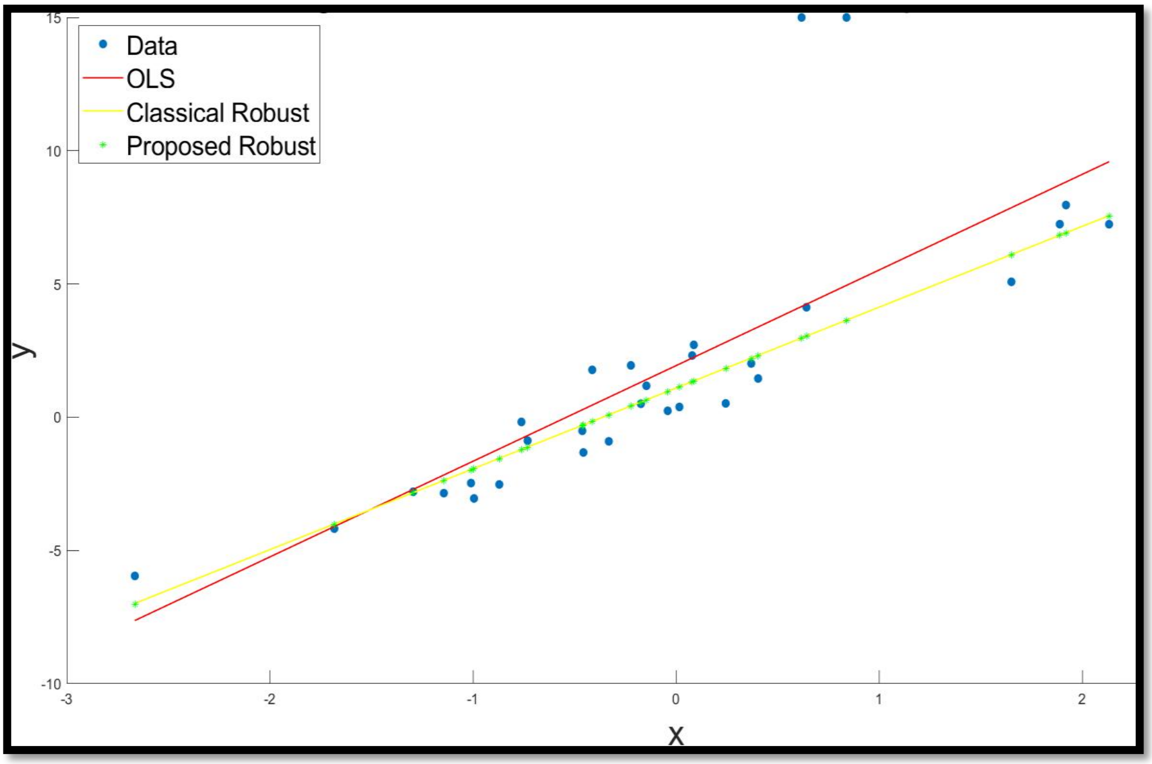

Figure (1): Regression Lines for OLS, Classical Robust, and Proposed Robust

Table (1): Simple Linear Regression

method	Sample size	Error of parameters	Tuning parameter	T	F
OLS	100	0.1459		1	99
Classic robust		0.0210	1.3390	31	69
Proposed robust		0.0185	4.62	68	32
OLS	50	0.5318		0	100
Classic robust		0.0422	1.3390	27	73
Proposed robust		0.0371	1.01	73	27
OLS	30	0.6581		0	100
Classic robust		0.0840	1.3390	36	64
Proposed robust		0.0800	1.01	64	36

T represents the method that is better than other methods and F is worse. The data presented in Table 1 indicate that the proposed robust method outperforms the classical robust methods and ordinary least squares (OLS) when there are outliers in the data for sample size (100, 50, and 30), as measured by the mean β error where the proposed robust method value was (0.0185, 0.0371, and 0.0800) for the tuning parameters (4.62, 1.01, and

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

1.01) respectively and the classical robust method (0.0210, 0.0422, and 0.0840) respectively for the fixed tuning parameter (1.3390) and OLS (0.1459, 0.5318, and 0.6581) respectively. Specifically, the proposed robust method showed superior performance in (68, 73, and 64) out of 100 experiments, respectively, while the classical robust method was more effective in (31, 27, and 36, respectively) experiments, and the ordinary least squares method was in (1, 0, and 0, respectively) experiments due to its sensitivity to outliers.

Table 2 shows the results of multiple linear regression for three different sample sizes. The results in multiple linear support the proposed robust methods same as simple linear because in all sample sizes, the proposed robust method has less error of parameters for sample size (100,50 and 30) which are (0.0455, 0.0872 and 0.1731) respectively and higher T numbers (70, 69 and 67) respectively which represents the method is better. The results in both simple linear regression and multiple regression for all sample sizes used were in favor of the proposed robust method, but it should be noted that the value of the tuning parameter varies in each experiment, unlike the value of the classical robust method, which is constant.

Table (2): Multiple Linear Regression

Method	Sample Size	Error of parameters	Tuning parameter	T	F
OLS	100	0.2326		5	95
Classic robust		0.0488	1.3390	25	75
Proposed robust		0.0455	1.01	70	30
OLS	50	0.7215		4	96
Classic robust		0.0989	1.3390	27	73
Proposed robust		0.0872	3.99	69	31
OLS	30	2.2359		1	99
Classic robust		0.1810	1.3390	32	68
Proposed robust		0.1731	1.33	67	33

10. Real Data: The actual data were taken from the website of the Central Statistical Organization in Iraq, the General Authority for Statistics and Geographic Information Systems, where the data were about the average living conditions in all governorates of Iraq for the year 2017-2018, and the variables taken are the average monthly per capita expenditure as a dependent variable and the variables of food and clothing with shoes as

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

independent variables, and the researchers extracted the parameters using all three methods, the OLS method, the classical Andrew method, and the proposed Andrew method, before taking the results we have to make sure that there is no outliers in data to compare the results after add outliers to the same data.

Figure (2): shows that there are no outliers in data between the range (2.5 and -2.5), with start taking out the parameters.

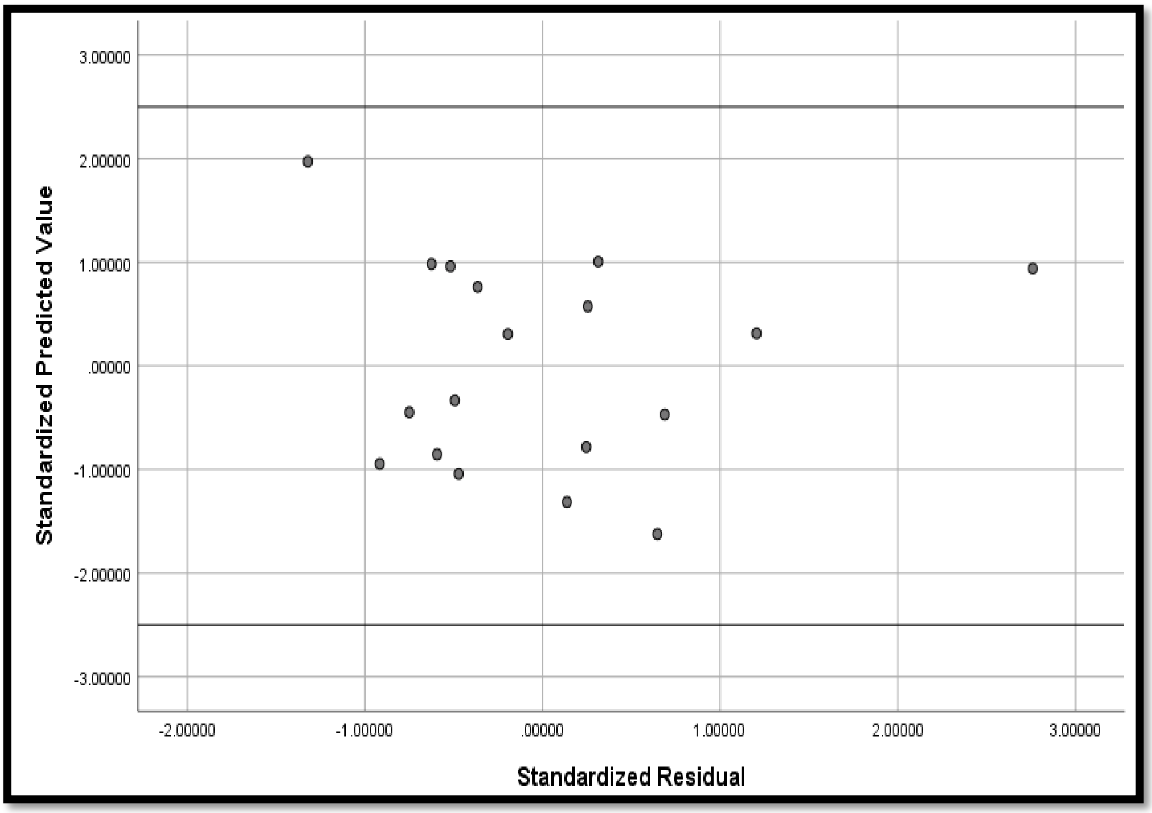

Figure (2): Standardized predicted values with standardized residual

Table 3 shows the parameter values for this data using the three methods. Since the least squares method parameters make the errors as small as possible under normal conditions, we will take them as a basis for this stage and the stage after adding the outliers to the data to know the extent of the impact of these outliers on the three methods. However, it must be noted that the parameters for the classical Andrews method are far from the parameter values of the OLS methods, and it has a parameter error (42.7335), but the parameter values of the proposed method are very close to the parameter values of OLS methods with parameter error (0.0016). It uses the moving tuning parameter according to the data and not a fixed value. Figures (3 and 4) show the results for data before adding outliers.

Tikrit Journal of Administrative and Economic Sciences, Vol. 21, No. 70, Part (2): 355-373
Doi: www.doi.org/10.25130/tjaes.21.70.2.19

Table (3): Real data

Method	Error of parameters	Tuning parameter	Parameters
OLS	-----------	--------------	-21.8197
			2.6469
			1.5570
Classic robust	42.7335	1.3390	-15.2947
			2.3990
			1.8608
Proposed robust	0.0016	11.329	-21.78
			2.6451
			1.5595

Figure (3) shows the true values with the estimated values as a line for each method. Since there are no outliers, all methods are convergent.

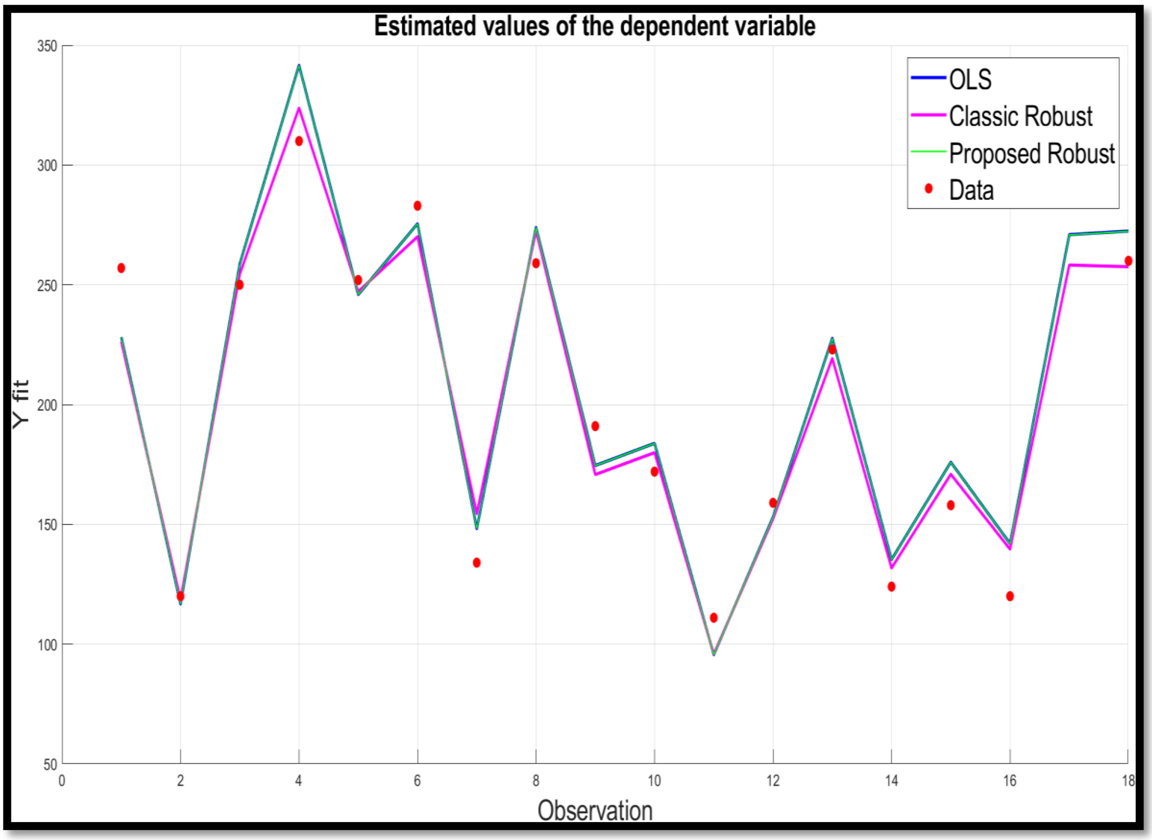

Figure (3): Predicted line of methods with data

Figure (4) shows the residuals for each method when estimating each value, where all methods were close together.

Tikrit Journal of Administrative and Economic Sciences, Vol. 21, No. 70, Part (2): 355-373
Doi: www.doi.org/10.25130/tjaes.21.70.2.19

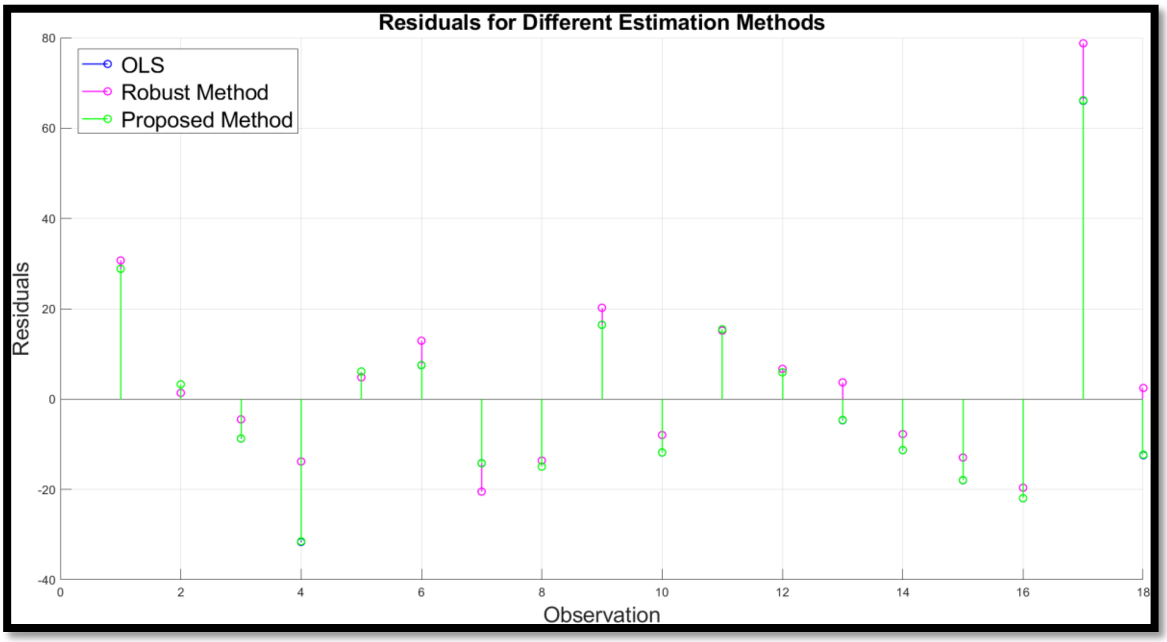

Figure (4): Residuals of methods

After extracting the real data results, we will add two outliers to the data, let be the second and sixth observations, y (2) = 500 and y (6) = 650, After adding these outliers, we will extract results for the data to see the extent of the impact of each method on the outliers and which one has the least impact on these values by comparing the differences between the parameter values for each method with the parameter values of the OLS method before adding the outliers.

Table (4): Real data after adding outliers

Method	Error of parameters	Tuning parameter	Parameters (without outliers)	Parameters (with outliers)
OLS	2961.3	-------	-21.8197	32.5739
			2.6469	1.9833
			1.5570	3.0424
Classic robust	33.8334	1.3390	-15.2947	-16.0076
			2.3990	2.4708
			1.8608	1.6921
Proposed robust	0.7813	3.539	-21.78	-20.9365
			2.6451	2.6260
			1.5595	1.5387

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

Table 4 shows that the parameters of the proposed method are better than those of the classical Andrew method based on the errors between the estimated parameters in the event of outlier values in the data, where the error value for the proposed method reached (0.7813) for optimal tuning parameter (3.539), and the parameter values were (β_0=-20.9365, β_1=2.6260, β_2=1.5387). For the classical Andrew method, the error value was (33.8334) and the parameters (β_0=-16.0076, β_1=2.4708, β_2=1.6921) which are less influential than the OLS method because the error values in it were equal to (2961.3) and the parameters (β_0=462.3474, β_1=27.9663, β_2=-22.4318) which are very far from the parameter values without the outliers . Hence, it is the most influential with the outliers. The value of the tuning parameter in the proposed method is not fixed like the value of the tuning parameter in the classical Andrew method. The proposed method changes its value according to the number of outliers and their location. This was explained in the simulation data. Whenever we change the sample size or location of outliers, the value of the tuning parameter changes in our proposed method. Figures 5 and 6 show the effect of the outliers on each method in charts.

Figure (5) shows that the estimated value line of the OLS method has been affected by the outliers and has been pulled in the direction of these two outliers, while the classical robust method has less influence and the proposed robust method is better than the two methods, as it is less influenced by the outliers than OLS and classic robust method.

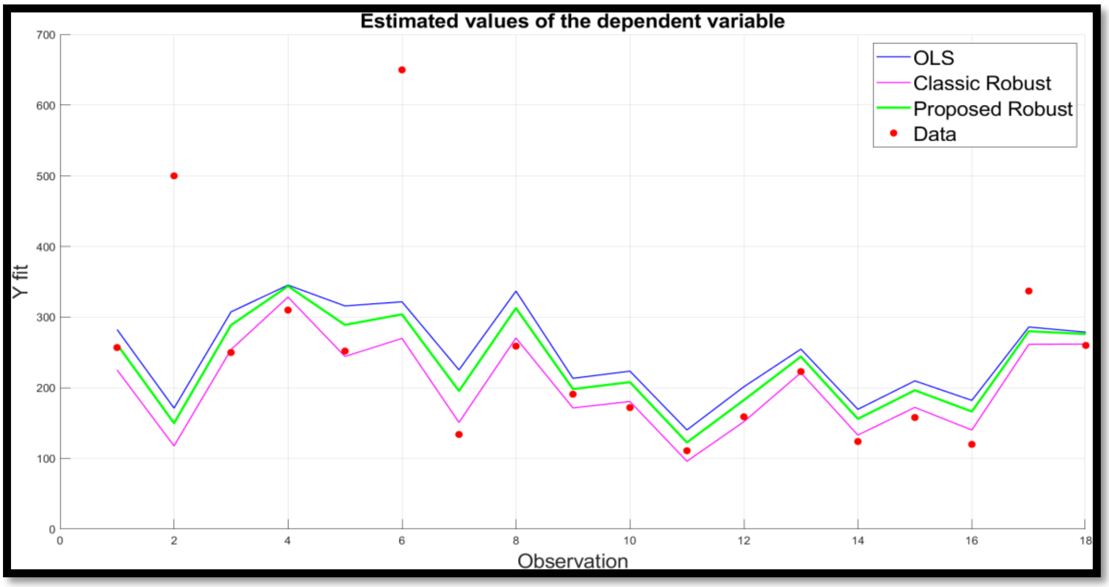

Figure (5): Line of estimate data after adding outliers

Doi: www.doi.org/10.25130/tjaes.21.70.2.19

Figure (6) shows that the OLS method was affected by the outliers in a way that made the differences for all values larger, while the classical method did not give the appropriate weight to the outliers. However, the proposed method gave the best weight to these values, making the errors as small as possible.

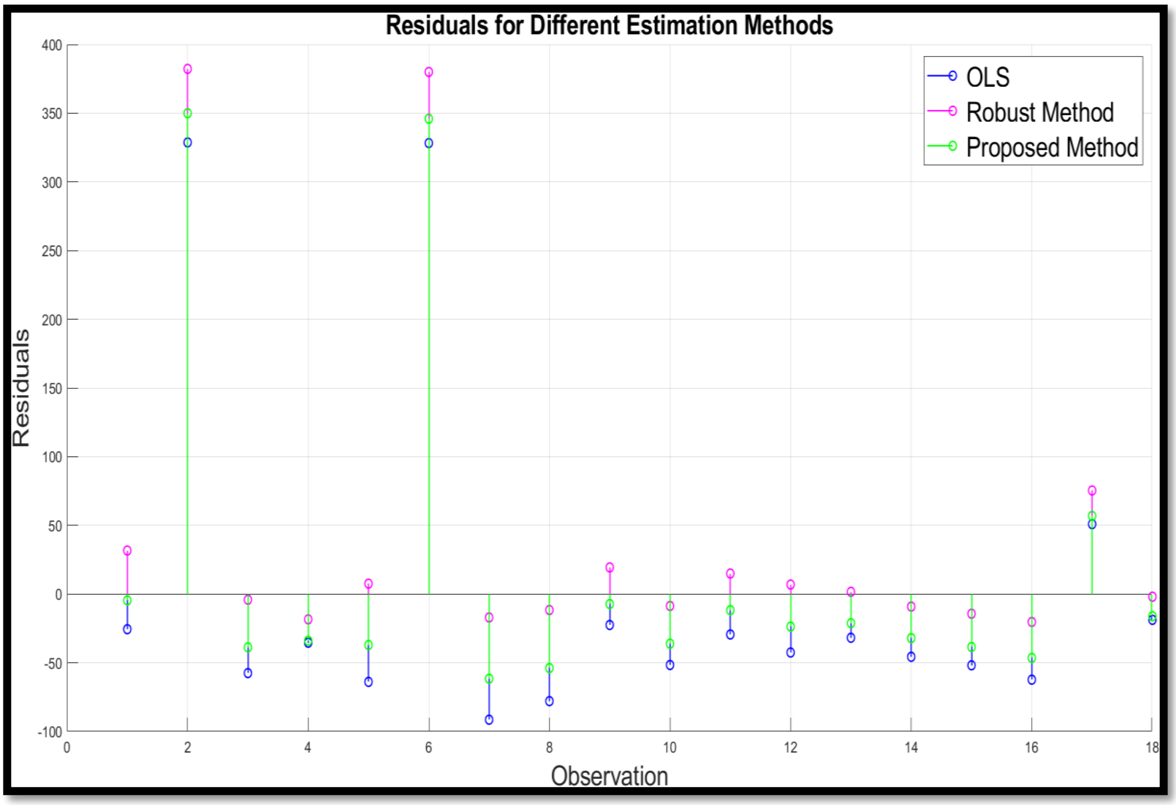

Figure (6): Residuals after adding outliers

Conclusion: This research has proven that the proposed robust Andrew method is less sensitive to extreme values that it extracts better parameters that are closer to the true parameter values if outliers are found in the data, and that the value of the tuning parameter in the classical robust Andrew method is not the best in all cases, as it uses a fixed tuning parameter value; although it is better than the OLS method, this value must change according to the number and location of outliers in the data. The proposed method uses an algorithm that always changes the value of the tuning parameter until it obtains the tuning parameter, which makes the errors of the parameters as low as possible. Based on the simulated data, the proposed method showed its superiority in all sample sizes (30, 50, 100) and repeating experiments (1000 times), which enhances the credibility of its results and its effectiveness in dealing with outliers, and also in real data in this research.

Recommendations: This study recommended some points, including:

1. Use the proposed robust Andrew function instead of the classic Andrew function in case of outliers in the data.
2. Connecting Andrew's function to quantile regression to extract the best outlier-proof equation.
3. Developing the proposed method so that its superiority over the classical method and OLS is close to 100%.

References

1. Ali, T. H., Sedeeq, B. S., Saleh, D. M., & Rahim, A. G. (2024). Robust multivariate quality control charts for enhanced variability monitoring. Quality and Reliability Engineering International, 40(3), 1369-1381. *https://doi.org/10.1002/qre.3472*.
2. Ali, T. H., & Awaz Shahab, M. (2017). Uses of Waveshrink in Detection and Treatment of Outlier Values in Linear Regression Analysis and Comparison with Some Robust Methods. Journal of Humanity Sciences, 21, 38-61.
3. Ali, T., Albarwari, N. H. S., & Ramadhan, D. L. (2023). Using the hybrid proposed method for Quantile Regression and Multivariate Wavelet in estimating the linear model parameters. Iraqi Journal of Statistical Sciences, 20(1), 9-24.
4. Ali, T., Al-Saffar, A., & Ismael, S. S. (2023). Using Bayes weights to estimate parameters of a Gamma Regression model. Iraqi Journal of Statistical Sciences, 20(1), 43-54.
5. Ali, T. H., Mahmood, A. P. D. S. H., & Wahdi, A. P. D. A. S. (2022). Using Proposed Hybrid method for neural networks and wavelet to estimate time series model. Tikrit Journal of Administrative and Economic Sciences, 18(57 part 3).
6. Buylov, V. (2022). Practical application of econometric modelling to market approach method using Tesla Motors, Inc. as an example.
7. da Silva Chagas, C., de Carvalho Junior, W., Bhering, S. B., & Calderano Filho, B. (2016). Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena, 139, 232-240.
8. Gladwell, M. (2008). Outliers: The story of success. Little, Brown.
9. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons.
10. Huber, P. J. (1992). Robust estimation of a location parameter. In Breakthroughs in statistics: Methodology and distribution (pp. 492-518). New York, NY: Springer New York.
11. Huber, P. J. (1981). Robust Statistics. John Wiley & Sons.
12. Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics, John Wiley & Sons, Inc. Publication.
13. Kareem, N.S., Ali, T. H., and Mohammad, A. S. (2019) "Construction robust simple linear regression profile Monitoring" Journal of Kirkuk University for Administrative and Economic Sciences, 9.1.: 242-257.
14. Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). Robust Statistics: Theory and Methods. John Wiley & Sons.

15. Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). Robust Statistics: Theory and Methods (with R). John Wiley & Sons.

16. McLean, K. A., Wu, S., & McAuley, K. B. (2012). Mean-squared-error methods for selecting optimal parameter subsets for estimation. Industrial & Engineering Chemistry Research, 51(17), 6105-6115.

17. Omar, C., & Ali, T. (2020). Using Bayes weights to remedy the heterogeneity problem of random error variance in linear models. Iraqi Journal of Statistical Sciences, 17(2), 58-67.

18. Omer, A. W., Sedeeq, B. S., & Ali, T. H. (2024). A proposed hybrid method for Multivariate Linear Regression Model and Multivariate Wavelets (Simulation study). Polytechnic Journal of Humanities and Social Sciences, 5(1), 112-124.

19. Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. Biometrics, 38(2), 485-498.

20. Wang, Y. G., & Lin, X. (2005). Effects of variance-function misspecification in analysis of longitudinal data. Biometrics, 61(2), 413-421.

21. Xie, X., Wu, T., Zhu, M., Jiang, G., Xu, Y., Wang, X., & Pu, L. (2021). Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. Ecological Indicators, 120, 106925.

22. Yan, X., & Su, X. (2009). Linear regression analysis: theory and computing. world scientific.

23. Zhang, H., Wu, P., Yin, A., Yang, X., Zhang, M., & Gao, C. (2017). Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. Science of the Total Environment, 592, 704-713.

Program-1

```
clc; clear all
n=100; beta=[1 3]';p=1;T=0;F=0; rng ('default') % For reproducibility
for j=1:100; clear k
x1 = randn(n,1); X=[ones(n,1) x1 ]; x=[x1]; y = X*beta + randn(n,1);y(12)=
15;y(25)=15;
% OLS
b = regress(y,X); mse(j) = sum((beta-b).^2);
% Classical robust
bc = robustfit(x,y,'andrews',1.339); mserc = sum((beta-bc).^2); msrc(j)
=mserc;
% robust (bisquare)
k(1) = 1.339;rmse1(1)=100;
for i= 2:100;    [BR,stats1] = robustfit(x,y,'andrews',k(i-1)); rmse1(i) =
sum((BR-beta).^2);
  e(i-1)=abs(rmse1(i)- rmse1(i-1));   k(i)=k(i-1)+0.01; end
[t N] = min(e); R(j)=k(N); br = robustfit(x,y,'andrews',k(N));
mserp = sum((beta-br).^2); msr(j) = mserp;
if msr(j) < msrc(j); T= T+1; else; F=F+1; end; end
msec = mean(mse), mmserp = mean(msr), Optimal = mean(R),
mmserc=mean(msrc)
```

Program-2

```
clc; clear all
n=30;beta=[1 1.5 2 -2.5]';p=3;T=0;F=0; rng ('default') % For reproducibility
for j=1:100; clear k
x1 = randn(n,1);x2 = randn(n,1);x3= randn(n,1);
X=[ones(n,1) x1 x2 x3]; x=[x1 x2 x3]; y = X*beta + randn(n,1);y(12)= -
15;y(25)=15;
% OLS
b = regress(y,X); mse(j) = sum((beta-b).^2);
% Classical robust
bc = robustfit(x,y,'andrews',1.339); mserc = sum((beta-bc).^2); msrc(j)
=mserc;
% robust (bisquare)
k(1) = 1.339;rmse1(1)=100;
```

```
for i= 2:100; [BR,stats1] = robustfit(x,y,'andrews',k(i-1)); rmse1(i) =
sum((BR-beta).^2);
  e(i-1)=abs(rmse1(i)- rmse1(i-1));   k(i)=k(i-1)+0.01; end
[t N] = min(e); R(j)=k(N); br = robustfit(x,y,'andrews',k(N));
mserp = sum((beta-br).^2); msr(j) = mserp;
if msr(j) < msrc(j); T= T+1; else; F=F+1; end; end
msec = mean(mse), mmserp = mean(msr), Optimal = mean(R),
mmserc=mean(msrc)
```