International Journal of Business, Human and Social Sciences
World Academy of Science, Engineering and Technology
ISSN: 2517-9411
Vol:4 2010-11-27
Vol:4, No:11, 2010

Robust Regression and its Application in Financial Data Analysis

Mansoor Momeni, Mahmoud Dehghan Nayeri, Ali Faal Ghayoumi, Hoda Ghorbani

Abstract**—**This research is aimed to describe the application of robust regression and its advantages over the least square regression method in analyzing financial data. To do this, relationship between earning per share, book value of equity per share and share price as price model and earning per share, annual change of earning per share and return of stock as return model is discussed using both robust and least square regressions, and finally the outcomes are compared. Comparing the results from the robust regression and the least square regression shows that the former can provide the possibility of a better and more realistic analysis owing to eliminating or reducing the contribution of outliers and influential data. Therefore, robust regression is recommended for getting more precise results in financial data analysis.

Keywords**—**Financial data analysis, Influential data, Outliers, Robust regression.

I. Introduction

THE assumption of normality is a crucial basis for most statistical methods of data analysis. However, various papers have shown this is true only with 10%-15% of data. This may be attributed to unnormal distribution of errors or the effect of outliers in observations [14]. Outliers are the observations not fitted to the pattern developed for the majority of data [19]-[3]. There are two attitudes in statistical modeling in dealing with outliers. The first takes into account the outliers and the second eliminates outliers. Many scholars believe using robust estimation is necessary where outliers are not eliminated from the statistical analysis [14].

A statistical method used in analyzing financial data is the regression analysis which often employs the least square regression (OLS) as its main means. However, as is obvious much deviation is experienced in financial data because of changes in financial policies and commercial cycles that inevitably gives rise to outlier observations in overall data [18]. Since the least square regression is vulnerable to such outlier observations that will ultimately affect results from this technique [7]-[2], and this will end in wrong conclusion and misleading users. The vulnerability of OLS regression to outliers may result from the failure to meet substantial assumptions required for this model.

An essential assumption in regression analysis is the constancy of the error variance which is called homoskedasticity. Outliers, even when the sample is large enough, can lead to the accumulation of error variance and give rise to hetoskedasticity [23].

In most cases, especially when data are acquired in a continuous period of time, as is true with financial data, the correlation of data will be probable [10]. The correlation will make errors interdependent in the regression model and reject the assumption of independent errors. The rejection of the assumption will lead to the inflation of the R square (R2) and erroneous significance of the model [23]. This situation indispensably necessitates using regression models [10].

Hence, in analyzing financial data it is necessary to use a regression model that is not vulnerable to outliers and prevent bias of outcomes. The robust regression is a good substitution for the least square regression concerning these data. This study seeks to introduce applications of the robust regression in financial researches in order to encourage researchers to use this technique and ultimately improve the quality of statistical analyses in financial researches.

II. Literature Review

In this section, we firstly discuss outliers and their types and identification. Subsequent to grasping the concept of outliers, we introduce a number of robust regression models that can attenuate the role of outliers. To formulate a robust regression model we should not restrict out observation to some isolated sporadic cases but we must identify outliers and influential data in order to reduce or eliminate their effects on the model [4].

A. Outliers

Outliers are the observations concomitant of high error residual [11]. The error volume is equal to the difference between the observed quantity and the predicted quantity for ith observation. This can be derived from:

$$e_i = y_i - \hat{y}_i \quad (1)$$

Mansour Momeni, Associate professor of management, Management Faculty, University of Tehran, Iran (e-mail: mmmomeni@ut.ac.ir).
Mahmoud Dehghan Nayeri, PhD Candidate of Operation Research, Management Faculty, University of Tehran, Iran (Corresponding Author, Phone; 00989122880021, e-mail: mdnayeri@ut.ac.ir).
Ali Faal Ghayoumi, Master in Accounting , Management Faculty, University of Tehran, Iran (e-mail:alifaal_grg@yahoo.com)
Hoda Ghorbani, Master in Bussiness Management , Management Faculty, University of Tehran, Iran (e-mail:hodaa_ghorbani@yahoo.com)

The summation of errors is zero in a regression model but the variance of errors can be different. This difference reduces the significance of comparing models. To overcome inequality of the error variance, their standardized values are used [2]-[16]. Standardized errors usually have a normal distribution with zero mean value and standard deviation of 1. The points with standardized error more than 2 or 3 and the standard deviation beyond the mean value (zero) are regarded as outliers [16].

Drawing diagrams is another technique to identify outliers. In this method, a distribution diagram or estimated deviation diagram is prepared, and the points away from the concentration of points or the regression line are outliers [1]. Diagram 1 depicts a number of data, as well as an outlier. It should be noted where we have a great deal of data we will face more constraints in using the schematic presentation.

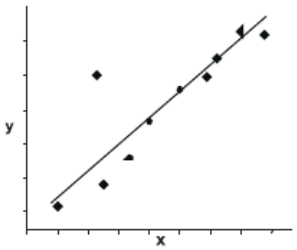

Fig. 1 Identifying an outlier by means of a diagram

Based on this introduction, if the deviation of an error is too large it is designated as an outlier. Therefore, the value of the dependent variable is a quantity used to identify outliers. Another type of outliers that are called influential data are investigated in relation to independent variables, that is, if a data much different from the average of data as concerned the independent variable it is an influential data [16].

B. Influential Observations

Sometimes, one or more data have remarkable effect on estimated parameters of a regression model. These are known as influential data [5]. In other words, influential data are the data whose removal from the model will give rise to crucial alteration in the model. Though eliminating every observation introduces a change in the regression model but when there are notable changes (including change in the slope or intercept) that observation will be influential [16]. Diagram 2 shows an influential data beside the regression line. The regression line in this diagram has a negative slope, and if the influential observation is removed the line's slope will become positive and the intercept will reduce. Evidently, this observation has a greater role in determining the estimated regression.

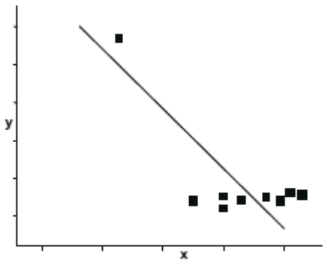

Fig.2 The effect of an influential data on the regression equation

The leverage value can be employed to find out influential observations. The leverage value is the difference in the magnitude of independent variables from their mean value. The leverage value for the ith observation is calculated using this formula:

$$P_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2} \quad (2)$$

Where n is the number of observations, xi the ith observation, mean value of observations and Pi the leverage value of ith observation.

Usually, the points with the leverage value two times the average of leverage values are known as the points with high leverage value (influential) [12]. Of course, some researchers admit the points with a leverage value above 0.5 as the influential data [4]. Statistical tests have been developed for the identification of influential observations. Almost all of these tests use the leverage value as a main tool. One of the most famous tests is the Cook's distance measure [8]. The Cook's distance measure uses both the leverage value and the error magnitude to measure the extent of influence of the data. The equation 3 shows how this test is calculated:

$$D_i = \frac{(e_i)^2}{\sum(k-1)s_e^2}\left[\frac{p_i}{(1-p_i)^2}\right] \quad (3)$$

In this equation, K is the number of independent variables, and Se the standard error of estimation. To know further about this technique, the reader is advised to see the Chaterji and Hadi paper [5].

C. Robust Regression

Robust regression is the regression that tries to minimize or eliminate the effect of outliers or influential data in order to provide a more reliable estimation based on the majority of data. In other words, robust regression is an attempt to find real results out of most data [16]. Hence, various types of robust regression fall in the category of robust estimation methods that function through eliminating or moderating the effect of outliers [14]. It was noted earlier that the normality of errors is one of the primary assumptions in regression models but there is always some divergence from this assumption. In such cases, robust regression can be a substitute for the least square method which is less susceptible to the divergence [23]. The OLS regression is not immune to outliers owing to

its objective-oriented nature of the OLS. This fact is illustrated in the equation 4 that shows the minimum of the summation of errors.

$$Min \quad \sum e_i^2 = \sum \left(y_i - \theta_1 x_{i1} - \cdots - \theta_m x_{im}\right)^2 \tag{4}$$

Where e_i stand for error, yi for the dependent variable and xi the independent variable and $\{\theta_j , j=1,\ldots,m\}$ are the parameters estimated by the OLS model [14].

The model's susceptibility to the error square is obvious; hence outliers significantly contribute to the formation of parameters.

Edge Verst (1887) pioneered the development of the robust regression. He pointed out that outliers, owing to becoming square, crucially impress the OLS. Therefore, he presented the least absolute deviation model (equation 5) [18].

$$Min \quad \sum |e_i| \tag{5}$$

This technique is called the L1 regression model while the model in relation 4 is designated as the L2 regression. Unfortunately, L1 is highly susceptible to the second-type outliers, that are, the data with bad leverage value. The data with bad leverage value is the one that besides having a high leverage effect (influential) is itself an outlier. In other words, it is both far from independent variables and contaminated with a large degree of errors [14].

Hodges (1967) introduced the concept of breakdown point in order to help assessing the robust regression's stability toward outliers. Rosio (1987) defined breakdown point as the lowest ratio of outliers that can impair the regression model. The higher the breakdown point in a regression model more satisfactorily the model functions. For L1 and L2, the breakdown point is equal to 1/n. In other words, as the result of the existence of an outlier in a set of n data, it can render invalid the model by errors [14].

There are various types of robust regression that with different functions try to provide a stable model. Choosing which robust regression is suitable for a case depends on the nature of data and the discretion of the persons using the regression [14].

Next to LAD, least trimmed square (LTS) is another type of robust regression. This regression, introduced by Rosio for the first time in 1984, is a technique to eliminate possible outliers [22]. Coefficients in the equation of the least trimmed square regression are estimated similar to the ordinary regression (lease square). In other words, the coefficients in these two types of regression are estimated in a way to minimize the summation of the second power of errors. However, in least trimmed square, unlike the ordinary regression, not all data is used in estimating the regression model but the data accompanied with high errors are eliminated [23].

The equation 6 explains how data are selected and which mechanism is used in least trimmed square to estimate the equation [21].

$$Min \sum_{i=1}^{q} (y_i - \hat{y}_i)^2 \quad , \quad q = (k+n+1)/2 \tag{6}$$

Where n is the number of observations, y_i the value of i[th] observation, $\hat{y}_i$ the anticipated value of ith observation, q is the number of the data used in least trimmed regression equation, and k the number of parameters including the intercept.

To find the data required to estimate the least trimmed square regression, the second power of each observation is calculated. Then, these values are arranged in an ascending order. Finally, the observations with least error square are selected and this procedure continues up to the qth observation.

It should be noted that some statistical software programs (like S-PLUS), if the data used in estimating the least trimmed square regression is less than 90% of the data only the 90% of the data are used in estimating the equation. Some researches doubt the reliability of such regressions [15] though this regression has improved owing to the introduction of another initiative called the rapid least trimmed square that was set forth by Rosio and Vandersen (1998) [21].

Another version of robust regression is the iteratively reweighted least square that was presented by Jatergi and Machler in 1997 [6]. In iteratively reweighted least square regression, the points with high leverage value and high error are almost prevented from contributing to the outcome in a lesser extent in order to reduce their effects on results of the regression analysis. In this regression, the weight of ith observation is calculated from this equation:

$$W_i^{\,j} = \frac{(1 - P_{ii})^2}{Max\left(\left|r_i^{\,j-1}\right|, med_i\left|r_i^{\,j-1}\right|\right)} \tag{7}$$

Where $W_i^{\,j}$ denotes the weight of i[th] observation, P_{ii} the leverage value of the ith observation, $r_i^{\,j-1}$ the error with the i[th] observation, and $med_i\left|r_i^{\,j-1}\right|$ the average of the absolute value of errors.

As the above equation shows the point with higher leverage value and error are given a lower weight and ultimately reduces their contribution to the result of the analysis. It should be noted that estimating coefficients in the weighted regression seeks to minimize the summation of the square of weighted errors not minimizing the summation of the square of errors.

D. Comparing forecast models

After introducing a number of robust regression methods, now we try to review the forecasting ability of these models concerning an example of financial data as compared to the OLS in order to approve the reliability of the outcome of robust models. To draw a comparison, we will use accuracy indexes, including the root mean square error (RMSE), mean

absolute error (MAD), mean absolute percentage error (MAPE), and the success rate (SR) that counts the number of right forecast of the sign of the true value by the model [18]. If $\hat{y}_t$ and y_t are taken as the true value and the anticipated value in t period and proportionally calculate the forecast from i+l to i+n period then the equations for calculating abovementioned indexes will be as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=i+1}^{i+n}\left(y_t - \hat{y}_t\right)^2} \quad (8)$$

$$MAD = \frac{1}{n}\sum_{t=i+1}^{i+n}\left|y_t - \hat{y}_t\right| \quad (9)$$

$$SR = \frac{1}{n}\sum_{t=i+1}^{i+n} I\left(y_t \hat{y}_t > 0\right) \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{t=i+1}^{i+n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (11)$$

This research is developmental and experimental as far as the objective is concerned and falls in the category of casual correlation researches. It also tries to show the strength and capability of robust regression in analyzing financial data. Ultimately, it tries to prevent probable bias that may be produced by the least square regression in analyzing financial data. In this connection, the input of the research is the data collected from 135 companies in the Tehran stock exchange in the period 1999-2006. These data have been gathered cumulatively and include 1080 company-year.

Financial researches often discuss the relation between accounting data and the stock return or stock price. This research too considers this relation through the Olsen and return models. The Olsen model relates the stock price to dividend and the book value of the stock. On the other hand, the return model links the stock return to the dividend and its annual change. So, it is tried to develop regression models of simple least square and two robust regression models, at least the least trimmed square and the iteratively reweighted least square, and finally draw a comparison between results. The two models referred to be of eliminative and outlier's effect reduction, respectively.

To compare the outcome of regression and explain the performance of the robust regression model, in addition to presenting the modified coefficient, F test ad t test, two forecast indexes introduced in section 2-4, have been used. Statistical analyses have been carried out using SPSS and S-PLUS software programs.

In this research, developing regression models and collecting data have been based on two return model and Olsen model. The return model illustrates the relation between the stock return and accounting profit, and includes dividend per stock and fluctuations in the dividend as independent variables. This model which was presented by Iston and Harris in 1991 [9] can be described as follows:

$$RET_{jt} = a_0 + a_1 E_{jt} / P_{jt-1} + a_2 (E_{jt} - E_{jt-1}) / P_{jt-1} + e_{jt} \quad (12)$$

In this model, RET_{jt} is the annual yield of the j company , E_{jt} denotes dividend, $E_{jt} - E_{jt-1}$ changes in dividend per stock and P_{jt-1} is the last year's stock prices .

Another model, which is known as the Olsen model (price model), was introduced by Olsen for the first time in 1995 [17]. This model explains the relation between the stock price and two independent variables – dividend per stock and the book value of stocks – and can be shown in the equation 13.

$$MV_{jt} = a_0 + a_1 BV_{jt} + a_2 E_{jt} + e_{jt} \quad (13)$$

In this model, MV_{jt} is the market value of stocks from the j company at the end of the month of presenting financial statements, BV_{jt} stands for the book value of each stock of the j company in the year t, and E_{jt} is the accounting profit reported for each stock of the company j in the year t.

All variables of the research, excluding the book value variable, have been calculated for each company. The stock return variable has been calculated from the difference between the price of a stock at the end of the month of presenting financial statements of the company in the last year and the price of a stock at the end of the month of presenting financial statements in the current year plus yields (including the dividend, reward) proportional to the price of a stock price at the end of the month of presenting financial statements in the last year. In the next section, we will discuss the result of employing these various types of regression in estimating related models and will compare their performance in order to ultimately choose the most favorable regression technique.

III. Findings

Results of regression analysis according to the Olsen model are displayed in table 1. As can be observed, the adjusted determination coefficient in the least square regression is equal to 0.474 that shows independent variables of the model (dividend arising from each stock and the book value of the stocks) explains for 47% of the changes in the dependent variable (stock price). Furthermore, the t test shows the insignificance of the book value variable at the lever of 5%. In other words, there is no relation between the stock price and the stock book value in this type of regression analysis.

Results of the IRLS regression are somewhat different. According to this regression and regarding the adjusted determination coefficient, more than 70% of the changes in stock prices are explained by the dividend and the book value of each stock. Additionally, both independent variable (dividend per stock and the book value) are significant at the lever of 5% error. Therefore, using the iteratively reweighted least square regression analysis proves the relation between the variables and the stock price. The LTS model allows explaining more than 60% in the dependent variable. This

model is implemented through the S-Plus software program and is not fitted for running t test. It should be noted that all the three models are significant from the standpoint of the F- test. Therefore, the fitness of the models is approved.

TABLE I
RESULTS OF THE REGRESSION ANALYSIS FOR THE OLSEN MODEL

Model	R^2	β		T- Test			
		BV	EPS	BV		EPS	
				t	P-value	t	P-value
OLS	0.47	0.19	5.56	0.89	0.37	28.54	0.00
LTS	0.62	0.42	3.56	---	---	---	---
IRLS	0.72	0.23	5.09	2.31	0.02	46.59	0.00

The difference between results of these three regressions can be explained in the light of reducing the effect of outliers and influential data. These data damages the efficiency of the model and the lever of significance of one of independent variables in the lease square regression. Based on this, if the researchers relies on the least square regression to draw result he will go wrong and the analysis will end in false findings. Consequently, it can be concluded that robust models will provide more satisfactory results and we will discuss this later. Table 2 shows the result of regression analysis for the return model. The adjusted coefficient in the least square regression is about 0.14. In other words, only 14% of changes in the dependent variable (stock return) can be explained by independent variables (dividend for a stock and the book value of stocks).

In MTS model, independent variables are able to explain the stock return somewhat better. Of course, in the IRLS this improvement is remarkable and is about 24%. A comparison of regression analyses in the return model shows that reducing the impact of outliers and influential data improves the efficiency of the model. It must also be noted that using robust regression does not always lead to enhancement of results of a regression analysis but it helps making the results more realistic. Anyway, all the three models are significant with respect to the F test.

TABLE II
RESULTS OF THE REGRESSION ANALYSIS FOR THE RETURN MODEL

Model	R^2	β		T- Test			
		EPS/P	EPS/PΔ	EPS/P		EPS/PΔ	
				t	P-value	T	P-value
OLS	0.139	110.99	32.06	8.06	0.00	5.23	0.0
LTS	0.142	115.25	19.08	---	---	---	---
IRLS	0.244	111.65	32.24	11.28	0.00	6.86	0.0

We have used the same criteria of assessment described in the previous section in order to scrutinize the performance of the developed models. Calculation of the related indexes for abovementioned regression models – Olsen model and the return model, respectively – are depicted in tables 3 and 4. In evaluating the models, the lower RMSE, MAD and MAPE indexes the more the SR index the more desirable will be the model considered.

TABLE III
COMPARISON OF OLS, IRLS AND LTS REGRESSION IN OLSEN MODEL

Regression	RMSE	MAD	SR	MAPE
OLS	8008.56	4075.77	92.2%	67.7%
LTS	8656.61	3760.44	93.1%	50%
IRLS	8087.73	3802.18	92.2%	56.8%

As is seen, as MAD, MAPE and SR indexes are concerned robust models have functioned better than simple models. On the other hand, the result is inversed when RMSE index is used as a criterion for comparison. It can be concluded that robust model's performance is less satisfactory when this index is measured. However, in order to prevent the accumulation of outlier-simulated error in RMSE it is recommended to use the MAD index. Some researchers approve using this initiative [14].

TABLE IV
COMPARISON OF OLS, IRLS AND LTS REGRESSIONS IN RETURN MODEL

Regression	RMSE	MAD	SR	MAPE
OLS	82.49	52.92	65.2%	21.81%
LTS	84.98	50.03	65.5%	14.17%
IRLS	83.30	50.51	65.3%	16.75%

As it is observed, the value of SR and LTS indexes for OLS regression is approximately 92% and 93%, respectively. The most favorable value is 100% and shows all predictions are of the same sign with real values. From the value 92% it is inferred that in 8% of cases the sign of the predicted value is contrary to the sign real values. In other words, the model is extremely weak in 8% of cases. The MAD index for OLS, LTS and IRLS regressions are 4076, 3760 and 3802, respectively that reveals the fact that robust models are more suitable. And ultimately, the MAPE index is smaller with robust models and approves former deductions. The abovementioned indexes produce results similar to the Olsen model.

IV. CONCLUSION

Regression analysis is an important statistical tool in financial researches. Regression analysis must be free from bias and bent if realistic results are to be obtained. The least square regression, which is used in many financial researches, is strongly impressed by outliers and influential data, and in case of the existence of such items the conclusion will be distorted to a large extent. In order to prevent such faults and insufficiencies in results of a research we can identify and control outliers and influential data. Unfortunately, this is a time-consuming process and naturally reduces the number of statistical population and as a constraint hinders generalizing results to statistically smaller populations.

However, there are solutions to overcome this problem; robust regression is one of these solutions. Robust regression

is aimed at achieving stable and reliable results versus outliers and influential data. Findings show that robust regression, through reducing or eliminating this type of data, can give rise to more logical results and consequently to more precise results. Additionally, this research indicates that MAD, MAPE and SR indexes are good criteria to evaluate the model's function concerning outliers. Furthermore, it is recommended using robust regression in regression analysis of data in financial researches. Using this technique can thwart the effect of misleading results and improve the quality of statistical analyses in financial researches.

ACKNOWLEDGMENT

The authors would like to thank to Iranian National Elite Foundation for its encouragement and financial support.

REFERENCES

[1] Azar, A. and Momeni,M., Business statistics, 3th Edition, Samt Publications,Tehran, 2008.
[2] Anderson, D. R. and Sweeney, D. J, Statistics for Business and Economics, 7th edition, Williams, T.A., South western college, 1998.
[3] Bamett, V. and Lewis. T, Outliers in Statistical data, Third ed., Wiley, Chichester, 1993.
[4] Chatterjee, S. and Hadi A.S., Regression analysis by example, 4th edition, Wiley, New Jersey, 2006.
[5] Chatterjee, S. and Hadi A.S., "Influential Observations, High Leverage Points, and Outliers in Linear Regression", Statical Science, Vol. 1, No. 3, 1986, pp.379-416.
[6] Chatterjee, S. and Mächler, M., "Robust regression: A weighted least squares approach", Communications in Statistics — Theory and Methods, No.26, 1997, pp.1381–1394.
[7] Chen, C., "Robust Regression and Outlier Detection with the ROBUSTREG Procedure", presented at SUGI, No.27, 2002, pp.265.
[8] Cook, R. D., "Detection of influential observations in linear regression", Technometrics, No. 19,1977, pp.15-18.
[9] Easton, P., & Harris. T., "Earnings as an Explanatory Variable for Returns", Journal of Accounting Research, 1991, No.29, pp.19–36.
[10] Field, C. and Zhou, J., "Confidence intervals based on robust regression", Journal of Statistical Planning and Inference, No.115, 2003, pp.425 – 439.
[11] Gujarati, D.N., Basic Econometrics, Third edition, Mc Graw-Hill international edition, 1995.
[12] Hoaglin,D.C. and Welsch,R.E., "The hat matrix in regressin and ANOVA", The American Statistician, No.32, 1978, pp.17-22.
[13] Hodges J.L., Proc. Fifth Berkeley Symp. Math. Stat. Probab.,No.1, 1967, pp.163-168. http: // www. 2sas . com / proceedings /sugi27/pdf.
[14] Liang,Y.Z. and Kvalheim O.M., "Robust methods for multivariate analysis - a tutorial review", Chemometrics and Intelligent Laboratory Systems, No.32, 1996, pp.1-10.
[15] Martin, R. D. "Robust Statistics with the S-Plus Robust Library and Financial Applications", Vol. 1 and 2, Insightful Corp Presentation, New York, NY, 2002, No.17-18.
[16] Neter, J. and Kunter, M. H., Nachtsheim, C. J. and Wasserman. W., Applied Linear Regression Models, Third edition, Mc Graw-Hill, USA, 1996.
[17] Ohlson, J., "Earnings, Book Values, and Dividends in Equity Valuation", Contemporary Accounting Research, (Spring), 1995, pp.661–687.
[18] Preminger, A. and Franck, R., "Forecasting exchange rates: A robust regression approach", International Journal of Forecasting, No.23, 2007, pp.71– 84.
[19] Rousseeuw P. J. and Zomeren B. C., "A comparison of some quick algorithms for robust regression", Computational Statistics & Data Analysis, Vol.14, 1992, pp.107-116.
[20] Rousseeuw, P.J. and Leroy, A.M., Robust Regression and Outlier Detection, Wiley, New York, 1987.
[21] Rousseeuw, P.J. and Van Driessen, K., "Computing LTS Regression for Large Data Sets", Technical Report, University of Antwerp, 1998.
[22] Rousseeuw, P.J., "Least Median of Squares Regression", Journal of the American Statistical Association, No.79, 1984, pp. 871-880.
[23] Yaffee, R. A., "Robust Regression Analysis: Some Popular Statistical Package Options", Statistics, Social Science, and Mapping Group,Academic Computing Ser, 2002. www.nyu.edu/its/socsci/docs/RobustReg2.pdf.