DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 44

11-1-2016

# JMASM41: An Alternative Method for Multiple Linear Model Regression Modeling, a Technical Combining of Robust, Bootstrap and Fuzzy Approach (SAS)

Wan Muhamad Amir W Ahmad
*Universiti Sains Malaysia*, wmamir@usm.my

Mohamad Arif Awang Nawi
*Universiti Malaysia Terengganu*, anz_wsh@yahoo.com

Nor Azlida Aleng
*Universiti Malaysia Terengganu*, azlida_aleng@umt.edu.my

Mohamad Shafiq
*Universiti Sains Malaysia, Kelantan, Malaysia*, shafiqmat786@gmail.com



Recommended Citation

W Ahmad, Wan Muhamad Amir; Awang Nawi, Mohamad Arif; Aleng, Nor Azlida; and Shafiq, Mohamad (2016) "JMASM41: An Alternative Method for Multiple Linear Model Regression Modeling, a Technical Combining of Robust, Bootstrap and Fuzzy Approach (SAS)," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 2 , Article 44.
DOI: 10.22237/jmasm/1478004120
Available at: http://digitalcommons.wayne.edu/jmasm/vol15/iss2/44

# JMASM41: An Alternative Method for Multiple Linear Model Regression Modeling, a Technical Combining of Robust, Bootstrap and Fuzzy Approach (SAS)

**Erratum**
This paper was originally published in JMASM Algorithms & Code without its enumeration, JMASM41.

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, pp-pp.
doi: 10.22237/jmasm/1478004120


ISSN 1538 − 9472

# An Alternative Method for Multiple Linear Model Regression Modeling, a Technical Combining of Robust, Bootstrap and Fuzzy Approach

**Wan Muhamad Amir W Ahmad**
University of Science, Malaysia
Kelantan, Malaysia

**Mohamad Arif Awang Nawi**
University Malaysia Terengganu
Kuala Terengganu, Malaysia

**Nor Azlida Aleng**
University Malaysia Terengganu
Kuala Terengganu, Malaysia

**Mohamad Shafiq**
University of Science, Malaysia
Kelantan, Malaysia

Research on modeling is becoming popular nowadays, there are several of analyses used in research for modeling and one of them is known as applied multiple linear regressions (MLR). To obtain a bootstrap, robust and fuzzy multiple linear regressions, an experienced researchers should be aware the correct method of statistical analysis in order to get a better improved result. The main idea of bootstrapping is to approximate the entire sampling distribution of some estimator. To achieve this is by resampling from our original sample. In this paper, we emphasized on combining and modeling using bootstrapping, robust and fuzzy regression methodology. An algorithm for combining method is given by SAS language. We also provided some technical example of application of method discussed by using SAS computer software. The visualizing output of the analysis is discussed in detail.

*Keywords:* Multiple linear regression, robust regression, bootstrap method

## Introduction

Multiple linear regression (MLR) is an extension of simple linear regression. The random error term is added to make the model probabilistic rather than deterministic. The value of the coefficient $\beta_i$ determines the contribution of the independent variables $x_i$, and $\beta_0$ is the $y$-intercept (Ngo & La Puente, 2012; Amir,

*Dr. Wan Muhamad Amir W Ahmad is an Associate Professor. Email him at: wmamir@usm.my. Mohamad Arif Bin Awang Nawi is a postgraduate student in the Faculty Informatics and Computing. Email him at: anz_wsh@yahoo.com.*

Shafiq, Rahim, Liza, & Aleng, 2016). A fuzzy regression model corresponding to equation (1) can be stated as:

$$y = A_0 + A_1 x_1 + A_2 x_2 + \ldots + A_k x_k \tag{1}$$

Explanation variables $x_i$'s are assumed to be precise. However, response variable $Y$ is not crisp; it is fuzzy in nature. That means the parameters are also fuzzy in nature. Hence, the objective is to estimate these parameters.

Assume $A_i$'s are assumes symmetric fuzzy numbers which can be presented by interval. For example, $A_i$ can be expressed as a fuzzy set given by $A_i = < a_{1c}, a_{1w} >$ where $a_{ic}$ is center and $a_{iw}$ is radius or has associated vagueness. The fuzzy set reflects the confidence in the regression coefficients around $a_{ic}$ in terms of symmetric triangular memberships function. Application of this method should be given more attention when the underlying phenomenon is fuzzy which means that the response variable is fuzzy. Thus, the relationship is also considered to be fuzzy.

$A_i = < a_{1c}, a_{1w} >$ can be written as $A_i = [\, a_{1L}, a_{1R} \,]$ with $a_{1L} = a_{1c} - a_{1w}$ and $a_{1R} = a_{1c} - a_{1w}$ (Kacprzyk & Fedrizzi, 1992). In fuzzy regression methodology, parameters are estimated by minimizing total vagueness in the model.

$$y_j = A_0 + A_1 x_{1j} + A_2 x_{2j} + \ldots + A_k x_{kj} \tag{2}$$

Using $A_i = < a_{1c}, a_{1w} >$ write

$$y_j = < a_{0c}, a_{0w} > + < a_{1c}, a_{1w} > x_{1j} + \ldots + < a_{nc}, a_{nw} > x_{nj} = < a_{jc}, a_{jw} > \tag{3}$$

Thus,

$$y_{jc} = a_{0c} + a_{1c} x_{1j} + \cdots + a_{nc} x_{nj} \tag{4}$$

$$y_{jw} = a_{0w} + a_{1w} \left| x_{1j} \right| + \cdots + a_{nw} \left| x_{nj} \right| \tag{5}$$

As $y_{jw}$ represent radius and so cannot be negative, therefore on the right-hand side of equation $y_{jw} = a_{0w} + a_{1w} \mid x_{1j} \mid + \ldots + a_{nw} \mid x_{nj} \mid$, absolute values of $x_{ij}$ are taken. Suppose there m data point, each comprising $a\ (n + 1)$ - *row* vector. Then parameters $A_i$ are estimated by minimizing the quantity, which is total vagueness of the model-data set combination, subject to the constraint that each data point must fall within estimated value of response variable. This can be visualized as the following linear programming problem.

$$\text{Minimized } \sum_{j=1}^{m}\left(a_{0w} + a_{1w}\left|x_{1j}\right| + \cdots + a_{nw}\left|x_{nj}\right|\right)$$

Subject to

$$\left\{\left(a_{0c} + \sum_{i=1}^{n} a_{ic}x_{ij}\right) - \left(a_{0w} + \sum_{i=1}^{n} a_{iw}x_{ij}\right)\right\} \leq Y_j$$

$$\left\{\left(a_{0c} + \sum_{i=1}^{n} a_{ic}x_{ij}\right) + \left(a_{0w} + \sum_{i=1}^{n} a_{iw}x_{ij}\right)\right\} \geq Y_j$$

and $a_{iw} \geq 0$. Simplex procedure is generally employed in order to solve the linear programming problem.

## Calculation for linear Regression using SAS

```
/* First do Multiple linear regression */
procreg data=temp1;
model y=x1  x2;
run;
```

## Approach the MM-Estimation Procedure for Robust Regression

```
/* Then do robust regression, in this case, MM-estimation */
ods graphics on;
procrobustreg  method  =  MM  fwls  data=biostatistics  plot=fitplot(nolimits)
plots=all;
model y = x1  x2 / diagnostics itprint;
output out=resids out=robout r=residual weight=weight outlier=outlier sr=stdres;
run;
ods graphics off;
```

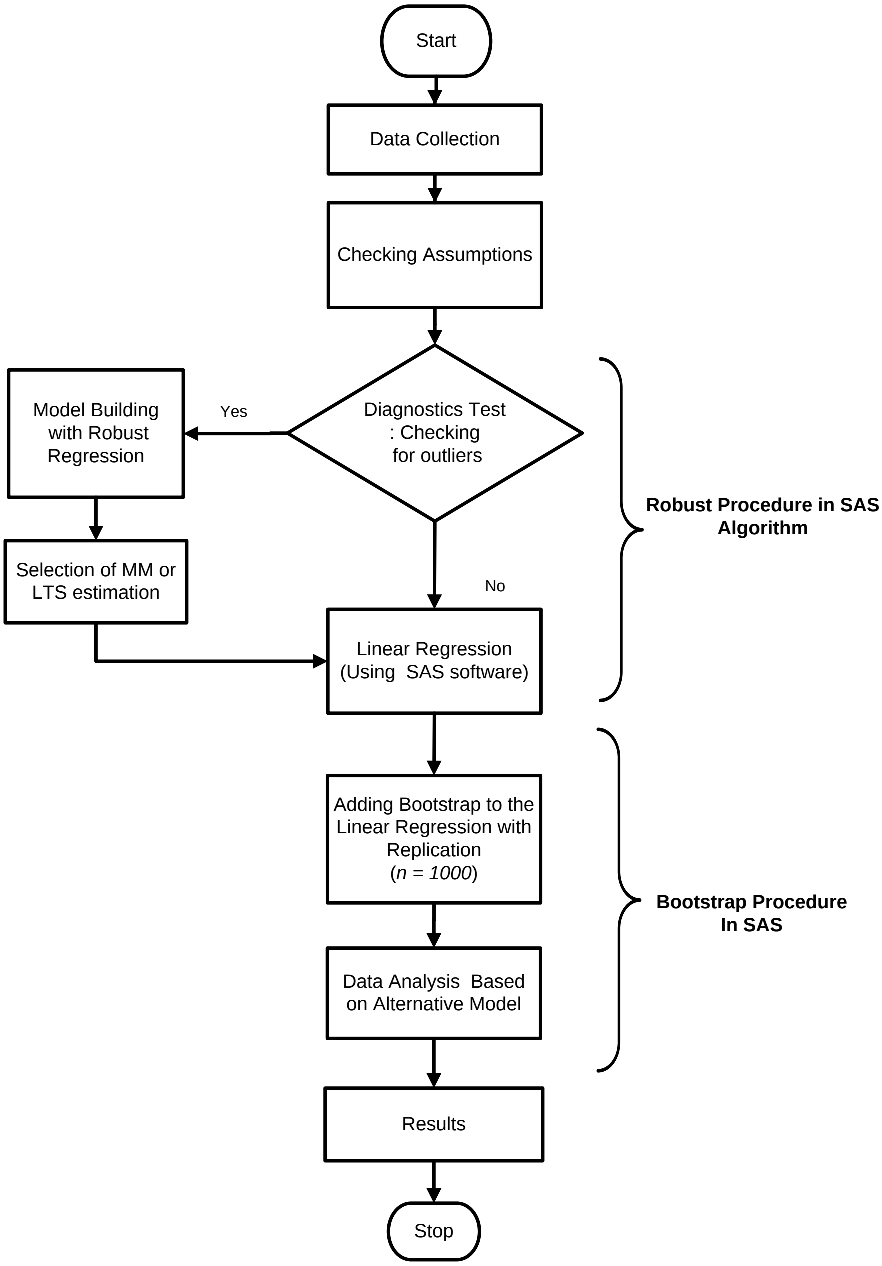


**Figure 1.** Flow Chart of Robust, Bootstrap and Fuzzy Regression

## Procedure for Bootstrap with Case Resampling *n* = 1000

```
/* And finally, use a bootstrap with case resampling */
ods listing close;
procsurveyselect data=temp1 out=boot1 method=urs
samprate=1 outhits rep=1000;
run;
```

## Procedure for bootstrap into fuzzy regression Model

```
/*Combination of Bootstrap Technique with Fuzzy Regression*/
ods listing close;
procoptmodel;
set j= 1..8;
numberFish{j}, weight{j}, height{j};
read data boot1 into [_n_]  Fishweight height;

/*Print Fishweight height*/
printFishweight height;
number n init 8;  /*Total of Observations*/

/* Decision Variables bounded or not bounded*/
/*Theses three variables are bounded*/
var aw{1..3}>=0;

/*These three variables are not bounded*/
var ac{1..3};

/* Objective Function*/
min z1= aw[1]*n + sum{i in j} weight[i]*aw[2]+sum{i in j} height[i]*aw[3];

/*Linear Constraints*/
con c{i in 1..n}:
ac[1]+weight[i]*ac[2]+height[i]*ac[3]-aw[1]-weight[i]*aw[2]-height[i]*aw[3]   <=
Fish[i];
con c1{i in 1..n}:
ac[1]+ weight[i]*ac[2]+ height[i]*ac[3]+aw[1]+ weight[i]*aw[2]+ height[i]*aw[3]
>= Fish[i];
```

```
expand;/* This provides all equations */
solve;
print ac aw;
quit;
ods rtf close;
```

## An Illustration of a Biostatistics Case

### *A Case Study of Aquaculture*

**Table 1.** Description of the Variables

| Variables | Code | Description |
|---|---|---|
| Fish | Y | Number of Fish Caught |
| Weight | X1 | Weight in (g) |
| Height | X2 | Height in (cm) |

*(Talib, Jaafar, & Sirwar, 2007)

### *Full Algorithm for Alternative Multiple Linear Regression Modelling*

```
Title 'Alternative Linear programming with combining robust and bootstrap';
data Biostatistics;
input Fish weigh height;
datalines;
97.32          110.41          103.74
174.52         111.08          104.80
214.56         114.98          105.71
178.44         114.16          105.27
199.48         112.99          105.45
189.92         115.20          105.34
170.48         113.24          105.11
207.16         117.19          105.66
;
run;

ods rtf file='result_ex1.rtf' ;

/*The next step is performing the procedure of modeling linear
```

```
regression model */
procreg data = biostatistics;
modelFish =weigh height;
run;

/* Then do robust regression, in this case, MM-estimation */
ods graphics on;
procrobustreg  method  =  MM  fwls  data=  biostatistics  plot=fitplot(nolimits)
plots=all;
modelFish =weigh height/ diagnostics itprint;
output out=resids out=robout r=residual weight=weight outlier=outlier sr=stdres;
run;
ods graphics off;

/* And finally use a bootstrap with case resampling */
ods listing close;
procsurveyselect data = biostatistics out = boot1 method = urs
samprate =1 outhits rep=1000;
run;
/*Combination of Bootstrap Technique with Fuzzy Regression*/
ods listing close;
procoptmodel;
set j= 1..8;
numberFish{j}, weigh{j}, height{j};
read data boot1 into [_n_]  Fish weigh height;

 /*Print Fish weight height*/
printFish weigh height;

/*Total of Observations*/
number n init 8;

/*Theses three variables are bounded*/
var aw{1..3}>=0;
/*These three variables are not bounded*/
var ac{1..3};

/* Objective Function*/
```

```
min z1= aw[1]*n + sum{i in j} weigh[i]*aw[2]+sum{i in j} height[i]*aw[3];

/*Linear Constraints*/
con c{i in 1..n}:
ac[1]+ weigh[i]*ac[2]+height[i]*ac[3]-aw[1]-weigh[i]*aw[2]-
height[i]*aw[3] <= Fish[i];

con c1{i in 1..n}:
 ac[1]+ weigh[i]*ac[2]+ height[i]*ac[3]+aw[1]+ weigh[i]*aw[2]+
height[i]*aw[3] >= Fish[i];

expand; /* This provides all equations */
solve;
print ac aw;
quit;
ods rtf close;
```

## Results

A higher R-squared value indicated how well the data fit the model and indicates a better model.

**Table 2.** Goodness-of-fit

| Goodness-of-Fit | |
|---|---|
| Statistic | Value |
| R-Square | 0.8199 |
| AICR | 5.5323 |
| BICR | 9.4456 |
| Deviance | 234.4750 |

Method of Multiple linear regression (MLR), we obtained the result as shown in Table 3

Table 4 shows the results by using bootstrapping method for fuzzy regression with $n = 1000$. The aim of bootstrapping procedure is to approximate the entire sampling distribution of some estimator by resampling (simple random sampling with replacement) from the original data (Yaffee, 2002). Table 4 summarizes the findings of the calculated parameter.

**Table 3.** Parameter Estimates for Final Weighted Least Squares Fit

| Parameter Estimates for Final Weighted Least Squares Fit | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
| Intercept | 1 | -6334.91 | 608.3789 | -7527.31 | -5142.51 | 108.43 | <.0001 |
| x1 | 1 | -3.0164 | 2.1608 | -7.2516 | 1.2188 | 1.95 | 0.1627 |
| x2 | 1 | 65.2183 | 7.5704 | 50.3807 | 80.0559 | 74.22 | <.0001 |
| Scale | 0 | 7.1356 | | | | | |

## Method of Fuzzy Regression (FR) (OPTMODEL)

**Table 4.** Value of ac and aw

| | ac | aw |
|---|---|---|
| 1 | -5764.1545 | 0.000000 |
| 2 | -3.0958 | 0.000000 |
| 3 | 59.8722 | 0.075811 |

While using bootstrap procedure, different output for the ac and aw will be obtained:

ac1= -5764.1545
ac2= -3.0958
ac3= 59.8722
aw1= 0
aw2=0
aw3=0.075811.

The next step is to compare the performance of multiple linear regression and fuzzy regression.

### *The Fitted Model for Multiple Linear Regressions*

$$Y = -6334.91 - 3.0164\, weight + 62.21\, height \quad (6)$$

Standard Error (608.3789) (2.1608) (7.5704)

The upper limits of prediction interval are computed by coefficient plus standard error

$$Y = (-6334.91 + 608.3789) + (-3.0164 + 2.1608)\,weight + (65.21 + 7.5704)\,height$$
$$Y = (-5726.53) + (-0.86)\,weight + (72.78)\,height$$

The lower limits of prediction interval are computed by coefficient minus standard error

$$Y = (-6334.91 - 608.3789) + (-3.0164 - 2.1608)\,weight + (65.21 - 7.5704)\,height$$
$$Y = (-6943.29) + (-5.1772)\,weight + (57.6396)\,height$$

### ***The Fitted Model for Fuzzy bootstrap Regression Is***

$$Y = [-5764.1545, 0] + [-3.0958, 0]\,weight + [59.8722, 0.075811]\,height \qquad (7)$$

The upper limits of prediction interval are computed by coefficient plus standard error

$$Y = [-5764.1545 + 0] + [-3.0958 + 0]\,weight + [59.8722 + 0.075811]\,height$$
$$Y = [-5764.15] + [-3.10]\,weight + [60.00]\,height$$

The lower limits of prediction interval are computed by coefficient minus standard error

$$Y = [-5764.1545 - 0] + [-3.0958 - 0]\,weight + [59.8722 - 0.075811]\,height$$
$$Y = [-5764.15] + [-3.10]\,weight + [59.80]\,height$$

The width of prediction intervals in respect of multiple linear regression model and fuzzy regression model corresponding to each set of observed explanatory variables is computed manually.

**Table 5.** Average Width for Former Multiple Linear Regression model and Fuzzy Bootstrap Regression Model

| Multiple Linear Regression model | | | Fuzzy Bootstrap Regression Model | | |
|---|---|---|---|---|---|
| Lower limit | Upper limit | Width | Lower limit | Upper limit | Width |
| -1535.37 | 1728.71 | 3264.09 | 97.23 | 117.98 | 20.75 |
| -1477.74 | 1800.92 | 3278.66 | 154.95 | 179.50 | 24.55 |
| -1445.48 | 1868.16 | 3313.64 | 200.87 | 222.01 | 21.14 |
| -1466.60 | 1836.84 | 3303.44 | 177.10 | 1988.15 | 21.05 |
| -1450.17 | 1850.95 | 3301.12 | 191.49 | 212.58 | 21.09 |
| -1467.93 | 1841.04 | 3308.99 | 178.06 | 199.13 | 21.07 |
| -1471.06 | 1825.99 | 3297.05 | 170.38 | 191.41 | 21.02 |
| -1459.81 | 1862.62 | 3322.43 | 191.03 | 212.16 | 21.13 |
| | Average | 3298.68 | | Average | 21.48 |

From Table 5, average width for former multiple regression was found to be 3298.68 while using fuzzy regression, the average width is 21.48 this indicate that the superiority of fuzzy regression methodology. From this analysis, the most efficient method to obtained relationship between response and explanatory variable is to apply fuzzy regression method compared to linear regression method.

## Conclusion

It was explained how to combine an algorithm between robust, fuzzy regression and the bootstrap method. A small sample size (8 observations only) was used

(a) to apply a bootstrap method in order to achieve an adequate of sample size.
(b) to compare the efficiency between original method and with the bootstrap method.
(c) to give a better understanding on how the algorithm works

According to biostatistics history, all the independent variables that we used in this case were significant to the number of fish caught. Without using bootstrapping, the result shows that two out of eight were significant. It is surprising that, using bootstrapping method (with $n = 1000$) the entire significant variable are included in the model as the finding from the biostatistics record. This algorithm provides us with the improved understanding of the modified method and underlying of relative contributions. For further study, it is possible to approach response surface methodology for every each of significant variables in single algorithm.

## References

Amir, W. M., Shafiq, M., Rahim, H. A., Liza, P., & Aleng, A. (2016). Algorithm for combining robust and bootstrap in multiple linear model regression (SAS). *Journal of Modern Applied Statistical Methods, 15*(1), 884-892. Retrieved from http://digitalcommons.wayne.edu/jmasm/vol15/iss1/44

Kacprzyk, J., & Fedrizzi, M. (1992). *Fuzzy regression analysis*. Warsaw: Omnitech Press.

Ngo, T. H. D., & La Puente C. A. (2012). The steps to follow in a multiple regression analysis. In *Proceedings of the SAS Global Forum 2012 Conference, Orlando, Florida, April 22–25*. Retrieved from http://support.sas.com/resources/papers/proceedings12/333-2012.pdf

Talib, B. A., Jaafar, A. H., & Siwar, C. (2007). Penangkapan ikan oleh nelayan Negeri Sembilan: Satu kajian empirikal. *International Journal of Management Studies*, *14*(1), 219-230. Retrieved from http://myais.fsktm.um.edu.my/10620/

Yaffee, R. A. (2002). *Robust regression analysis: Some popular statistical package options*. ITS Statistics, Social Science, and Mapping Group, New York State University.