4

3D FCRM MODELING IN MILES PER GALLON OF CAR

Mohd Saifullah Rusiman
Robiah Adnan
Efendi N. Nasibo

The new fuzzy c-regression modeling (FcRM) are widely used in order to fit switching regression models. Minimization of objective function yields immediate estimates for different c regression models. The functions of model, estimation technique and results are discussed in this paper. A case study in miles per gallon (MPG) of different cars using the FcRM modeling was carried out. The 3D graph for significant independent variables for FcRM clustering is shown in this study. The comparison between multiple linear regression and FcRM modeling were done. The mean square error (MSE) was used to find the better model. It was found that the FcRM modeling with lower MSE to be the better model and has great capability in predicting the dependent variable effectively.

4.1 OVERVIEW

For the past few years, new fuzzy modeling has become popular because of their better explanation of describing complex systems. The modified version of FCM called the fuzzy c-regression model (FcRM) clustering algorithm [1] develops hyper-plane-shaped clusters. The FcRM assumes that the input–output data are drawn from *c* different regression models. This FcRM modeling is also known as a switching regression modeling. The minimization of the objective function in FcRM clustering simultaneously yields a fuzzy *c*-partitioning matrix of the data and the *c* regression models. This new technique has been widely used in engineering, science, medicine, economic and other fields.

4.2 DATA BACKGROUND

A measure of fuel economy in automobiles or miles per gallon (MPG) used similarly in North America and the United Kingdom, although the Imperial gallon used in the UK is about 20% larger than the U.S. gallon. A metric term measures how many miles a vehicle can travel on one gallon of fuel. Most countries other than the US and UK use the SI (aka metric) units liter (0.22 Imperial gallon or 0.264 US liquid gallon) and km (0.621 statute miles). These can be combined to either km/L (efficiency) or L/100km (consumption). MPG figures are 20.095% higher in the UK than in the U.S. for the same real fuel economy.

The research on MPG done by Larson et. al in [2] was about the city and highway MPG prediction models (a pilot study). In this research, in order to comply with federal law, automobile manufacturers must display window stickers on automobiles leaving their factories. This paper presents an inexpensive statistical method of predicting city and highway mpg estimates which could be possible alternative methods. Another research was done by Jun Meng and Xiangyin Liu in [3]. They use the data mining theory to construct a BP (back propagation) neural network model to predict MPG (mile per gallon). The six independent variables are number of cylinders, displacement, horsepower, weight, acceleration and model year.

In this paper, the data were gained fom SPSS Software version 10. The data were collected involving 400 cars. The dependent variable is miles per gallon (MPG), whereas the independent variables are engine displacement in cubic inches, horsepower, weight of cars in lbs., time to accelerate from 0 to 60 mph, model year, country of origin (USA, European and Japanese), number of cylinders and cylinder filter.

The FCRM modeling proposed by Harthway and Bezdek in 1993 develops hyper-plane-shaped clusters. Kim et al. [4] successfully applied FcRM to construct fuzzy models in two phases algorithms which is *c* clusters are firstly identified through the FCRM cluster and a supervised learning algorithm further adjusts the obtained parameters to improve the modeling accuracy. The number of clusters (rules), *c*, is fixed and assigned by the user. In [5], for an unknown system, the appropriate number of clusters (rules) is supposed to be unknown and could be gained by using Equation (4.6).

4.3 METHODOLOGY

Firstly, the analysis of influential and outlier data should be done to the data to discard unimportant data due to human error, machine error or environment error. The analysis used like Pearson standardized residual (outliers *Y* test) in [6] and [7], Leverage (outliers *X* test) and DFBETA (influential test) in [7]. However, there are no conditions needed in FcRM modeling.

A switching regression model is specified by

$$y_i = f_i(x;\theta_i) + \varepsilon_i, \qquad 1 \le i \le c$$

The optimal estimate of θ depends on assumptions made about the distribution of random vectors or ε_i. Generally, the ε_i are assumed to be independently generated from some pdf $p(\varepsilon;\ \eta,\ \sigma)$ such as the Gaussian distribution with mean 0 and unknown standard deviation σ_i,

$$p(\varepsilon;\eta,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\varepsilon-\eta)^2}{2\sigma^2}}$$

Based on the Hathaway and Bezdek algorithm in [1] and [8];

(1) Fix the number of cluster c, $2 \leq c$. Choose the termination tolerance $\delta > 0$. Fix the weight, w, $w > 1$ and initialise $\mathrm{U}^{(0)} \in M_{fc}$ randomly.

(2) Estimate $\theta_1,...,\theta_c$ simultaneously by modifying the fuzzy c-means algorithm (FCM). If the regression functions $f_i(x;\theta_i)$ are linear in the parameters θ_i, the parameters can be obtained as a solution of the weighted least squares:

$$\theta_i = [\mathrm{X}_\mathrm{b}^\mathrm{T}\mathrm{W}_\mathrm{i}\mathrm{X}_\mathrm{b}]^{-1}\mathrm{X}_\mathrm{b}^\mathrm{T}\mathrm{W}_\mathrm{i}\mathrm{Y}$$

(3) Calculate the objective function:

$$E_w[\mathrm{U},\{\theta_i\}] = \sum_{i=1}^{c}\sum_{j=1}^{d} u_{ij}^{w}.\| y_j - f_i(x_j;\theta_i)\|^2$$

(4) Make iterations in order to minimize the objective function. Repeat for $l = 1, 2, \ldots$ until $\| \mathrm{U}^{(l)} - \mathrm{U}^{(l-1)} \| < \delta$. The steps are described as follows:

Step 1 : Calculate model parameters $\theta_i^{(l)}$ to globally minimize (4).

Step 2 : Update U with $E_{ij} = E_{ij}[\theta_i^{(l-1)}]$, to satisfy:

$$u_{ij}^{(l)} = \frac{1}{\sum_{k=1}^{c}\left(\frac{E_{ij}}{E_{kj}}\right)^{\frac{2}{w-1}}} \quad \text{for } E_{ij} > 0$$

and $1 \leq i \leq c$

otherwise $u_{ij} = 0$ if $E_{ij} > 0$, and

$$u_{ij} \in [0, 1] \text{ with } (u_{1j} + \ldots + u_{cj}) = 1$$
$$\text{until } \| \mathrm{U}^{(l)} - \mathrm{U}^{(l-1)} \| < \delta$$

In the FCRM clustering algorithm, the number of clusters, c, is fixed and assigned by the user. In practice, the appropriate number of clusters is usually decided with the aid of the cluster validity criterion like the Bezdek's partition coefficient [9] as follows,

$$V_{PC} = \frac{\sum_{h=1}^{N}\sum_{i=1}^{c}(\mu_{ih})^2}{N} \quad .$$

The optimal number c is chosen when V_{PC} is closest to 1.

Takagi and Sugeno [10,11] has introduced a fuzzy rule-based model and also called as an affine T-S fuzzy model [12], described as follows:

$$R^i : \text{ IF } x_1 \text{ is } A_1^i \text{ and } \cdots \ x_n \text{ is } A_n^i$$
$$\text{THEN} \quad y^i = a_1^i x_1 + \cdots + a_n^i x_n + a_0^i$$

where R^i denotes the ith IF-THEN rule

$i = 1, 2, \ldots, c$ where c is the number of rules

$x_m,\ m = 1,\ldots, n,$ are individual input variables

A_m^i are individual antecedent fuzzy sets

$a_k^i,\ k = 1,\ldots,n$ are consequent parameters

a_0^i denotes a constant

$y^i \in \Re$ is the output of each rule

The output of the fuzzy model $\hat{y}$ is inferred in [13,14] if the singleton fuzzifier, the product fuzzy inference and the centre average defuzzifier are applied:

$$\hat{y} = \frac{\sum_{i=1}^{c} w^i(\mathbf{x})\ y^i}{\sum_{i=1}^{c} w^i(\mathbf{x})}$$

where

$$w^i(\mathbf{x}) = A_1^i(x_1) \times A_2^i(x_2) \times \ldots A_n^i(x_n) = \prod_{m=1}^{n} A_m^i(x_m)$$

denotes the degree of fulfillment of the antecedent, that is, the level of firing of the *i*th rule.

The consequent parameters can be found directly from the FcRM program output. The antecedent fuzzy sets A_m^i are achieved by projecting the membership degrees in the fuzzy partitions matrix U onto the axes of individual antecedent variable x_m and then to approximate it by a normal bell-shaped membership function. Hence, each antecedent fuzzy set A_m^i is calculated from the sampled input data $x_h = [x_{1h},..., x_{nh}]^{\mathrm{T}}$ and the fuzzy partition matrix $\mathrm{U} = [\mu_{ig}]$ as follows [15,16] :

$$A_q^i(z) = \exp\left\{ -\frac{1}{2}\left(\frac{z - \mu_q^i}{\sigma_q^i} \right)^2 \right\},$$

with the mean and standard deviation are respectively given as

$$\mu_q^i = \frac{\sum_{h=1}^{N} \mu_{ih} x_{qh}}{\sum_{h=1}^{N} \mu_{ih}} \quad \text{and} \quad \sigma_q^i = \sqrt{\frac{\sum_{h=1}^{N} \mu_{ih}(x_{qh} - \alpha_q^i)^2}{\sum_{h=1}^{N} \mu_{ih}}}.$$

4.4 NUMERICAL EXAMPLES

Firstly, the analysis of influential and outlier data should be done to the data. The analysis used are Pearson standardized residual (outliers of *Y*), Leverage (outliers of *X*) and DFBETA (influential test). Nine data are discarded due to missing value of MPG. From the analysis of influential and outlier data, it was found that seven set of data should be discarded.

The dependent variable is miles per gallon (MPG), whereas the significant independent variables are endisplacement (x_1), horsepower (x_2), weight (x_3), year (x_4), origin (x_5), cylinderno (x_6), cylinderflt (x_7) and accellerationt (x_8). Figure 4.1 shows the individual scatter plots for MPG versus x_1 to x_8. The plots indicate the negative relationship between MPG versus endisplacement, horsepower, weight and cyclinder number. The positive relationship is shown between MPG versus cylinderflt whereas the scatter plot for MPG versus year, origin and accellerationt indicate no any relationship.

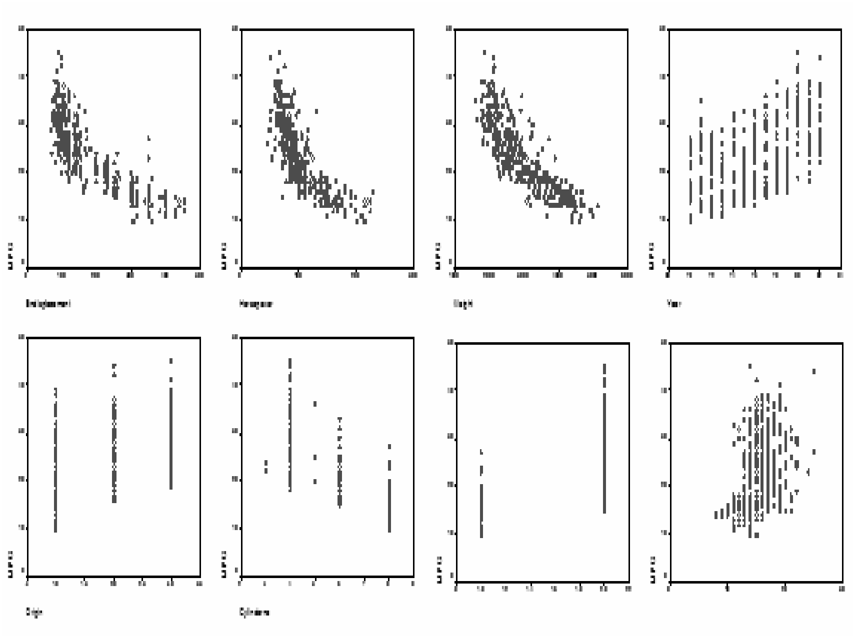

Figure 4.1. Scatter plot for MPG versus independent variables.

The FcRM clustering for the data were analysed by using Matlab software. Table 4.1 shows the optimal value for the number of clusters is two for x_1 to x_8 since the value V_{PC} is close to 1 when $c = 2$.

Table 4.1. The value of V_{PC} for Y versus x_1 to x_8.

Cluster Number (c)	2	3	4	5
Y vs x_1	0.898	0.881	0.885	0.868
Y vs x_2	0.915	0.887	0.877	0.887
Y vs x_3	0.914	0.908	0.902	0.889
Y vs x_4	0.892	0.872	0.889	0.883
Y vs x_5	0.937	0.912	0.897	0.909
Y vs x_6	0.937	0.906	0.906	0.908
Y vs x_7	0.924	0.916	0.906	0.912
Y vs x_8	0.936	0.920	0.912	0.905

Graphs in Figure 4.2 indicate the individual cluster plot for x_1 to x_8.

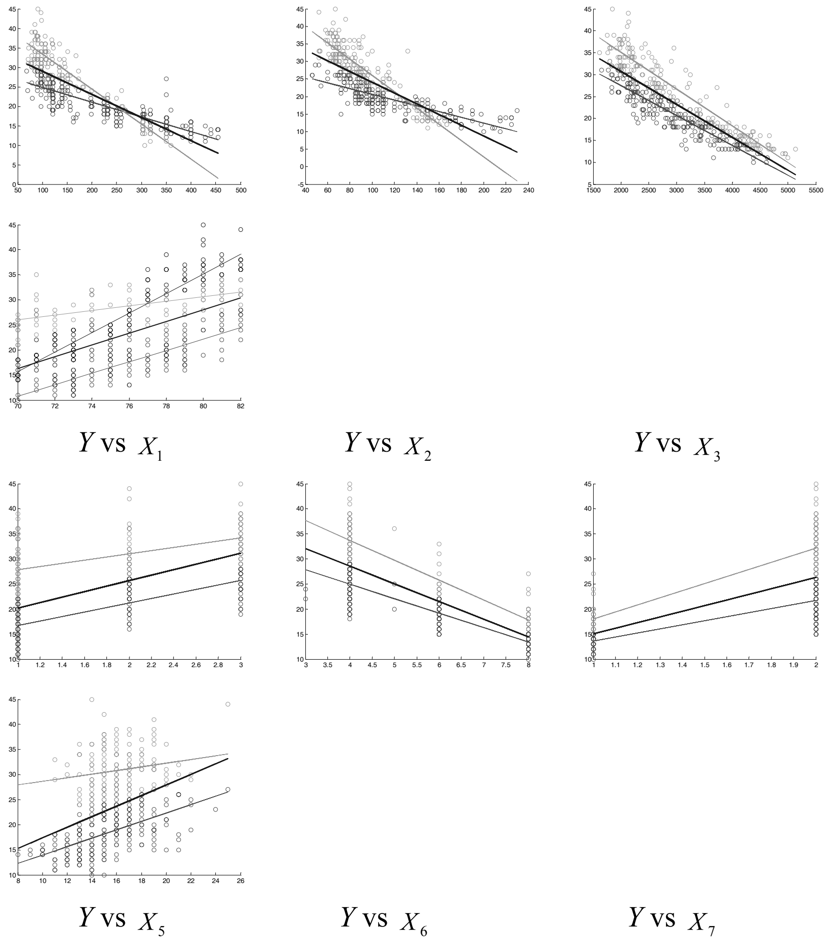

Figure 4.2. Individual cluster plot.

The memberships for antecedent parameters are calculated using formula (4.10). Hence, only two parameters are significant which are origin and cyclinderflt with the smallest MSE. The number of clusters chosen is two since its V_{PC} is the closest to 1 as summarized in Table 4.2.

Table 4.2. The value of V_{PC} for Y versus x_5 and x_7.

Number of clusters, c	*2*	*3*	*4*	*5*
V_{PC}	0.935	0.917	0.908	0.919

A significant affine T-S fuzzy model described as follows:

$$R^1: \text{ IF } x_1 \text{ is } A_5^1 \text{ and } x_2 \text{ is } A_7^1$$
$$\text{THEN } \quad y^1 = 2.578\,x_5 + 6.411\,x_7 + 4.627$$

$$R^2: \text{ IF } x_1 \text{ is } A_5^2 \text{ and } x_2 \text{ is } A_7^2$$
$$\text{THEN } \quad y^2 = 2.359\,x_5 + 11.455\,x_7 + 4.187$$

where the antecedent parameter is described in detail in Table 4.3.

Table 4.3. Details of the antecedent parameter .

R^i	$i = 1$	$i = 2$
A_1^i in x_1 : μ : σ	215.210 109.09	168.169 90.629
A_2^i in x_2 : μ : σ	114.871 41.777	94.938 31.173
A_3^i in x_3 : μ : σ	2939.726 717.119	3047.907 991.107
A_4^i in x_4 : μ : σ	75.753 3.475	76.142 3.812
A_5^i in x_5 : μ : σ	1.399 0.709	1.81 0.864
A_6^i in x_6 : μ : σ	5.628 1.658	5.248 1.746
A_7^i in x_7 : μ : σ	1.701 0.458	1.782 0.413
A_8^i in x_8 : μ : σ	15.268 2.926	16.154 2.216

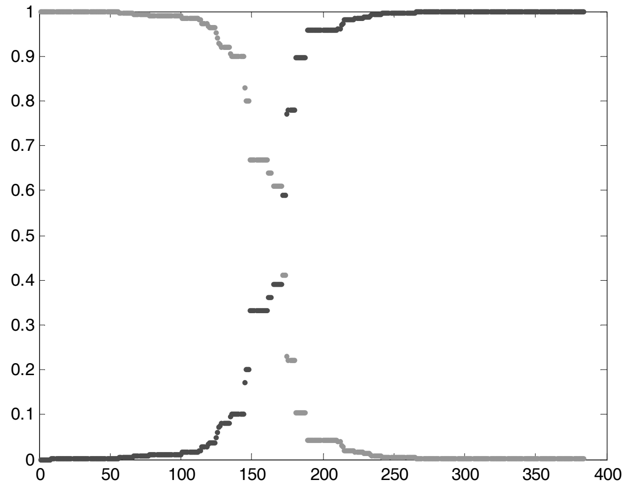

Figure 4.3. Membership function plot for y versus x_5 and x_7.

Figure 4.3 represents the membership function graph for MPG versus x_5 and x_7 with the optimal two clusters. Figure 4.4 shows the three-dimension graph of clustering for MPG versus x_5 and x_7 with two clusters. The middle plane is the multiple linear regression plane for all data. The other two planes are the two clustering planes with two different multiple linear regression planes.

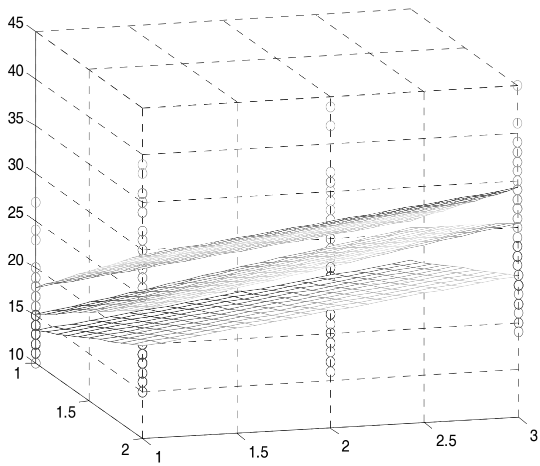

Figure 4.4. Three-dimension FcRM clustering graph for y versus x_5 and x_7.

In finding the better model, the mean square error (MSE) is used as follow:

1. For Multiple Linear Regression Model

$$MSE = \frac{1}{N-p}\sum\left(Y_i - \hat{Y}_i\right)^2$$

2. For FcRM Model

$$MSE = \frac{1}{N}\sum\left(Y_i - \hat{Y}_i\right)^2$$

where Y_i denotes the real data,

$\hat{Y}$ represents the predicted value of Y_i,

N is the number of data, and

p is the number of parameters.

The comparisons between these two model are summarized in Table 4.4.

Table 4.4. The comparisons between 2 models.

Model	*Variable Chosen*	*MSE*
Multiple Linear Regression Model	$X_1, X_2, X_3, X_4,$ X_5, X_6, X_7 (Significant Variable)	8.246
FcRM Model	$X_1, X_2, X_3, X_4,$ X_5, X_6, X_7, X_8 (All Variables)	8.932
	X_5 and X_7 (Significant Variable)	7.848

4.5 CONCLUSION

A new model called FcRM has been proposed in analyzing a continuous data where no assumptions are needed in the analysis. The minimization of objective function yields immediate estimates for different *c* regression models. The comparison modeling between FcRM and multiple linear regression modeling indicate that FcRM modeling appeared to be the better model with the lower MSE. This FcRM modeling could be proposed as one of the best model in analyzing mainly in a complex system. Hence, the value of MPG could be predicted based on the variable of origin and cylinderflt. The best cars that could increase MPG come from Japan, followed by Europe and America. Cars with cylinder filter have better MPG than cars without cylinder filter.

4.6 REFERENCES

[1] Hathaway, R. J., and Bezdek, J. C.: 'Switching regression models and fuzzy clustering', *IEEE Trans. Fuzzy Syst.*, 1993, 1, (3), pp. 195–204.

[2] Larson, E. C., Foster, K. A., Hopfe, M. W. : 'City and highway MPG prediction models (a pilot study)', *Proceedings Decision Sciences Institute 1994 Annual Meeting*, Part vol.2, 1994, pp. 1316-17.

[3] Jun Meng and Xiangyin Liu **:** 'MPG Prediction based on BP Neural Network', *IEEE on Industrial Electronics and Applications*, International Conference, 2006, pp. 1-3.

[4] Kim, E., Park, M., Ji, S., and Park, M.: 'A new approach to fuzzy modeling', *IEEE Trans. Fuzzy Syst.*, 1997, 5, (3), pp. 328–337.

[5] Chuang, C.C., Su, S.F., and Chen, S.S.: 'Robust TSK fuzzy modeling for function approximation with outliers', *IEEE Trans. Fuzzy Syst.*, 2001, 9, (6), pp. 810–821.

[6] Agresti, A. : '*An Introduction to Categorical Data Analysis*' ,John Wiley & Sons, Inc, New York, 1996.

[7] Norusis, M. J. : '*SPSS for Windows Advanced Statistics Release 6.0*' , SPSS Inc, USA, 1993.

[8] Abonyi, J. and Feil, B. : 'Cluster analysis for data mining and system identification', Springer, USA, 2007.

[9] Bezdek, J.C. : 'Cluster validity with fuzzy set ', *J. Cybern.*, 1974, 3, pp. 58–72.

[10] Takagi, T., and Sugeno, M.: 'Fuzzy identification of systems and its applications to modeling and control', *IEEE Trans. Syst., Man Cybernatic*, 1985, 15, pp. 116–132.

[11] Sugeno, M., and Kang, G.T.: 'Structure identification of fuzzy model', *Fuzzy Sets Syst.*, 1988, 28, (1), pp. 15–33.

[12] Babuska, R.: 'Fuzzy modeling for control', Kluwer Academic Publishers, Boston, 1988.

[13] Wang, L.X. : 'Adaptive fuzzy systems and control: design and stability analysis', Prentice-Hall, NJ, 1994.

[14] Wang, L.X.: 'A course in fuzzy systems and control', Prentice-Hall, NJ, 1997.
[15] Hoppner, F., Klawonn, F., Kruse, R., and Runkler, T.: 'Fuzzy cluster analysis, methods for classification, data analysis and image recognition', John Wiley & Sons, NJ, 1999.
[16] Kim, E., Park, M., Ji, S., and Park, M.: 'A new approach to fuzzy modeling', *IEEE Trans. Fuzzy Syst.*, 1997, 5, (3), pp. 328–337.