Alternative Fuzzy Switching Regression

Kuo-Lung Wu, Miin-Shen Yang, and June-Nan Hsieh

***Abstract*—This paper present an alternative method for fuzzy switching regression analysis. The traditional fuzzy c-regression (FCR) is a method by embedding the fuzzy c-means (FCM) into switching regression. By defining an alternative residual measurement, we modified the fuzzy c-regression (FCR) and proposed an alternative fuzzy c-regression (AFCR). The proposed method is more robust to noise and outlier than the EM and FCR algorithms. Numerical examples show the robustness and superiority of our proposed method.**

***Index Terms*—Switching regression, EM algorithm, fuzzy c-regression, alternative fuzzy c-regression, robustness.**

I. Introduction

Regression analysis is used to model the function relation between the independent and dependent variables. Usually, a single regression model is used for fitting a data set. However, the data set may contain more than one regression model, say c regression models. This kind of model fitting is called switching regressions. Quandt [1,2] and Chow [3] initiated the studies of switching regressions. Subsequently, Quandt [4], Quandt and Ramsey [5] and De Veaux [6] considered the mixture of regressions approach to estimating switching regressions that is widely studied and applied in psychology, economics, social science and music perception [7-10].

Hathaway and Bezdek [11] first combined switching regressions with FCM and referred to them as fuzzy c-regressions (FCR). To increase the speed of FCR, Wang et al. [12] combined the concept of Newton's law of gravity with FCR. However, these FCRs are sensitive to noise and outliers. To improve the robustness against noise and outliers, Leski [13] considered an ε-insensitive loss function that was used in the statistical learning theory (or support vector machine) (see Vapnik [14]). Leski [13] extended the FCR objective function to c simultaneous quadratic programming (QP) problems subject to some bound constraints and one linear equality constraint where the incremental learning method proposed by Cauwenberghs and Poggio [15] was used for solving the QP problem. However, the incremental learning method for solving the QP problem is much more complex than FCR. Yang et al. [16] proposed a robust fuzzy c-regressions method by implementing an alpha-cut technique. However, there does not an objective function for this method.

Manuscript received November 28, 2008. This work was supported in part by the National Science Council of Taiwan, under Kuo-Lung Wu's Grant: NSC-97-2118-M-168-001.

K.L. Wu is with the Department of Information Management, Kun Shan University, Yunk-Kang, Tainan 71023, Taiwan. (phone: +886 (6) 2050-149; e-mail: klwu@mail.ksu.edu.tw).

M.S. Yang and J.N. Hsieh are with the Department of Applied Mathematics, Chung Yuan Christian University, Chung Li 32023, Taiwan. (e-mail: msyanu@math.cycu.edu.tw).

In this paper, we will propose an alternative robust fuzzy switching regression method (AFCR) by replacing the residual measurement in the FCR objective function. In Section II, we review the EM and FCR algorithms. The objective function and the iterative procedure are discussed in Section III. Section IV presents some numerical examples and conclusions are illustrated in Section V.

II. The EM Algorithm and Fuzzy c-Regressions

A. EM Algorithm

In estimating the parameters of a mixture regression model, the EM algorithm is an effective method for approximating maximum likelihood estimates. Let $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ with each independent observation x_j and corresponding dependent observation y_j be a given data set. Let a linear regression model be

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j,\ j = 1, \cdots, n\ , \tag{1}$$

where ε_j are iid from the normal densit $\mathrm{N}(0, \sigma^2)$. We consider a mixture of c numbers of linear regression model as

$$y_j = \beta_{0i} + \beta_{1i} x_j + \varepsilon_{ij},\ \varepsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2) \tag{2}$$

with probability $\alpha_i \in (0,1)$ and $\sum_{i=1}^{c} \alpha_i = 1$, $i = 1, \cdots, c$. That is

$$Y_j \sim f(y_j; \alpha, \theta) = \sum_{i=1}^{c} \alpha_i f_i(y_j; \theta_i)\ , \tag{3}$$

where $Y_1, \cdots, Y_n$ are independent and

$$f_i(y_j; \theta_i) = \mathrm{N}(\beta_{0i} + \beta_{1i} x_j, \sigma_i^2)\ . \tag{4}$$

Consider $Y = \{y_1, \cdots, y_n\}$ to be an incomplete data set and the cluster memberships $\mu_1, \cdots, \mu_c$ to be missing where $\mu_{ij} = \mu_i(y_j) = 1$ if y_j belong to cluster i. In accordance with the mixture density $f(y; \alpha, \theta)$, the log likelihood for the complete data is given by

$$L_c(\alpha, \theta) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij} \ln(\alpha_i f_i(y_j; \theta_i)) \tag{5}$$

where

$$f_i(y_j; \theta_i) = \mathrm{N}(\beta_{0i} + \beta_{1i} x_j, \sigma_i^2)\ . \tag{6}$$

The EM algorithm is applied to the mixture distributions by treating μ as missing data. The algorithm is easy to program in two steps, expectation (E) and maximization (M) [17]. According to the initial value of $r = (\alpha, \theta)$, say $r^{(0)}$, the E step requires the calculation of $\mathrm{E}(L_c(\alpha, \theta) | r^{(0)})$, that is the expectation of log likelihood $L_c(\alpha, \theta)$ of the complete data, conditioning on the observed data and the initial value $r^{(0)}$. The M step is to choose the value of r, say $r^{(0)}$, that maximizes $\mathrm{E}(L_c(\alpha, \theta) | r^{(0)})$ after the E step. Then the EM algorithm for this mixture of linear regression model is the iteration according to the following conditions (see De Veaux [6]):

$$\alpha_i = \frac{1}{n}\sum_{j=1}^{n}\mu_{ij}\,, \tag{7}$$

$$\beta_{0i} = \frac{\sum_{j=1}^{n}\mu_{ij}(y_j - \beta_{1i}x_j)}{\sum_{j=1}^{n}\mu_{ij}}\,, \tag{8}$$

$$\beta_{1i} = \frac{\sum_{j=1}^{n}\mu_{ij}(y_j - \beta_{0i})x_j}{\sum_{j=1}^{n}\mu_{ij}x_j^2}\,, \tag{9}$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{n}\mu_{ij}(y_j - \beta_{0i} - \beta_{1i}x_j)^2}{\sum_{j=1}^{n}\mu_{ij}}\,, \tag{10}$$

$$\mu_{ij} = \frac{\alpha_i f_i(y_j;\theta_i)}{\sum_{k=1}^{c}\alpha_k f_k(y_j;\theta_k)}\,, \tag{11}$$

where $f_k(y_j;\theta_k) = \mathrm{N}(\beta_{ok} + \beta_{1k}x_j, \sigma_k^2)$, $i = 1,\cdots,c$ and $j = 1,\cdots,n$. Note that, EM algorithm for switching regression can not be written as a matrix form such as the traditional regression analysis. The above update equations are available only in a simple case with switching regression models (1). For more complex models, the update equations need to be modified. However, the objective function of fuzzy c-regressions algorithm can be written as a matrix form and the update equations for minimizing the objective function are similar to the traditional weighted least square estimator. We now give a brief review of the fuzzy c-regressions algorithm.

B. Fuzzy C-Regressions

In unsupervised learning clustering literatures, the fuzzy c-means (FCM) [18-22] algorithm is the best-known fuzzy clustering method. FCM is an iterative algorithm using the necessary conditions for minimizing the objective function J_{FCM} with

$$J_{FCM}(\mu, a) = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d(x_j, a_i) \tag{12}$$

where the weighting exponent $m > 1$ is a fuzziness index, $\mu = \{\mu_1,\cdots,\mu_c\}$ with $\mu_{ij} = \mu_i(x_j)$ is a fuzzy c-partition, $a = \{a_1,\cdots,a_c\}$ over the p-dimensional real space R^p is the set of c cluster centers and $d(x_j, a_i)$ is a dissimilarity measure. The necessary conditions for a minimizer (μ, a) of J_{FCM} are the following update equations:

$$\mu_{ij} = \frac{d(x_j, a_i)^{-1/(m-1)}}{\sum_{k=1}^{c} d(x_j, a_k)^{-1/(m-1)}} \tag{13}$$

and

$$a_i = \frac{\sum_{j=1}^{n}\mu_{ij}^m x_j}{\sum_{j=1}^{n}\mu_{ij}^m}\,,\ \ i = 1,\cdots,c\,,\ \ j = 1,\cdots,n \tag{14}$$

where the Euclidean distance $d(x_j, a_i) = \|x_j - a_i\|^2$ is used. Note that other types of dissimilarity $d(x_j, a_i)$ may be used to improve the usage and effectiveness of FCM which will be discussed in next section.

The combination of switching regressions with FCM is referred to as fuzzy c-regression (FCR) by Hathaway and Bezdek [11]. Suppose that, we have a set of data $\{(x_1, y_1),\cdots,(x_n, y_n)\}$ with each independent observation $\vec{x}_j = (x_{j1},\cdots,x_{jp}) \in \Re^p$ and corresponding dependent observation $y_j \in \Re$. The objective of switching regressions is to find c linear regressions

$$\hat{y}_{j,i} = \beta_{1i}x_{j1} + \cdots + \beta_{pi}x_{jp} = \vec{x}_j\vec{\beta}_i\,,\ i = 1,\cdots,c \tag{15}$$

which will fit best for the data structure and $\vec{\beta}_i = (\beta_{1i},\cdots,\beta_{pi})'$. The objective of the FCR is to minimize the objective function J_{FCR} with

$$J_{FCR} = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m d(y_j, \hat{y}_{j,i}) \tag{16}$$

where $d(y_j, \hat{y}_{j,i}) = (y_j - \vec{x}_j\vec{\beta}_i)^2$. Suppose that X denotes the matrix in $\Re^{n\times p}$ having $x_j = (x_{j1},\cdots,x_{jp})$ as its jth row; Y denotes the vector in $\Re^n$ having y_j as its jth component; and D_i denotes the diagonal matrix in $\Re^{n\times n}$ having μ_{ij} as its jth diagonal element. Than we have the following weighted least squares

$$\vec{\beta}_i = (\beta_{1i},\cdots,\beta_{pi})' = [X^t D_i^m X]^{-1} X^t D_i^m Y \tag{17}$$

which minimizes

$$J_{FCR} = \sum_{i=1}^{c} S_i' S_i \tag{18}$$

where $S_i = (D_i)^{\frac{m}{2}}(Y - X\vec{\beta}_i)$.

The update equations for the membership function is

$$\mu_{ij} = \frac{d(y_j, \hat{y}_{j,i})^{-1/(m-1)}}{\sum_{k=1}^{c} d(y_j, \hat{y}_{j,k})^{-1/(m-1)}}\,. \tag{19}$$

Note that, the clustering results of both EM and FCR will be influenced by the noise and outlier. By replacing the residual measurement in the FCR objective function with an alternative term, we propose an alternative fuzzy c-regression method (AFCR). The proposed method is robust to noise and outlier and is quite simple such as the original FCR algorithm.

III. The Alternative Fuzzy C-Regressions

A. The Alternative Residual Measurement

Some alternative distance measurements are used to replace the Euclidean distance measurement to extend the traditional fuzzy c-means algorithm. The first extension to FCM was proposed by Gustafson and Kessel (GK) [23]. They considered the effect of different cluster shapes except for spherical shape by replacing the Euclidean distance in the FCM objective function with the Mahalanobis distance

$$d(x_j, a_i) = \|x_j - a_i\|_{A_i}^2 = (x_j - a_i)^T A_i (x_j - a_i) \tag{20}$$

where A_i is a positive definite $p\times p$ matrix with its determinate $\det(A_i) = \rho_i$ being a fixed constant. This extension became an important extended type of FCM. Krishnapuram and Kim [24] discussed more about the GK algorithm with a new variation. Gath and Geva (GG) [25] also considered the FCM objective function with the Mahalanobis distance that is exactly the same as the GK extension.

However, Gath and Geva [25] did not directly minimize the extended FCM objective function. They used the Bayesian posterior probability via the EM update equations for the normal mixture distributions. Although the GG algorithm is an effective and useful clustering algorithm, it is an ad hoc clustering procedure because there is no obvious objective function to be followed. The GG algorithm can be though as an extension of the EM for the normal mixture with a weighting exponent *m*.

In order to make the traditional k-means and fuzzy c-means robust to noise and outliers, Wu and Yang [26] proposed an alternative distance measurement with

$$d(x_j, a_i) = 1 - \exp\{-\eta \left\| x_j - a_i \right\|^2\}. \tag{21}$$

They then created the alternative k-means and alternative fuzzy c-means that can robust to noise and outliers. They also adopted the above measurement in competitive learn network to propose an alternative learn vector quantization [27]. We now use this alternative distance measurement to define a new residual measurement in the FCR objective function with

$$d(y_j, \hat{y}_{j,i}) = 1 - \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\}. \tag{22}$$

B. Alternative Fuzzy C-Regressions

Based on the concept of Wu and Yang [26], the new metric $1 - \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\}$ is used to replace the residual measurement in the FCR objective function. Thus, an alternative fuzzy c-regression (AFCR) clustering objective function is proposed as follows:

$$J_{AFCR} = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} \{1 - \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\}\}. \tag{23}$$

The parameter η is defined as

$$\eta = \left(\frac{1}{n} \sum_{j=1}^{n} (y_j - \bar{y})^2 \right)^{-1} \tag{24}$$

which is corresponding to the average of total sum of square. Thus, the necessary conditions for the minimization of J_{AFCR} are as follows:

$$\beta_{ki} = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\} (y_j - \vec{x}_j \vec{\beta}_i + \beta_{ki} x_{jk}) x_{jk}}{\sum_{j=1}^{n} \mu_{ij}^{m} \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\} x_{jk}^2} \tag{25}$$

and

$$\mu_{ij} = \frac{d(y_j, \hat{y}_{j,i})^{-1/(m-1)}}{\sum_{k=1}^{c} d(y_j, \hat{y}_{j,k})^{-1/(m-1)}} \tag{26}$$

where $d(y_j, \hat{y}_{j,i}) = 1 - \exp\{-\eta (y_j - \vec{x}_j \vec{\beta}_i)^2\}$, $i = 1, \cdots, c$, $j = 1, \cdots, n$ and $k = 1, \cdots, p$. According to above necessary conditions, we have the following exact AFCR algorithm:

Exact AFCR algorithm

Initializing $\vec{\beta}_i^{(l)}$.

Repeat

Compute $\mu_{ij}^{(l+1)}$ with (26).

Repeat

Use (25) to approximate $\vec{\beta}_i$ by the fixed-point iterative method.

Until Converge.

Get $\vec{\beta}_i^{(l+1)}$ with the result of fixed-point iterative method.

Until Change in $\vec{\beta}_i$ is very small.

Note that, β_{ki} in (25) cannot be solved directly and we need to use the fixed-point iterative method to approximate it. This is not efficiently. We then use a one-step method to approximate it in the following AFCR algorithm:

AFCR algorithm

Let $f(\vec{\beta}_i)$ be the right term of (25) and initialize $\vec{\beta}_i^{(l)}$.

Repeat

Compute $\mu_{ij}^{(l+1)}$ with (26).

Compute $\vec{\beta}_i^{(l+1)}$ with $\vec{\beta}_i^{(l+1)} = f(\vec{\beta}_i^{(l)})$.

Until Change in $\vec{\beta}_i$ is very small

Obviously, if both algorithms converge, then their solutions shall satisfy the necessary condition (25). We recommend using the AFCR algorithm to speed up the process and we also use it in following numerical examples.

IV. Examples

We consider a c=2 simple switching regression model with two parallel lines. Figure 1 present the random generated dataset and add an outlier with its coordinate (10,0). The estimated regression models of EM, FCR and AFCR are shown in Figs. 1(a), 1(b) and 1(c), respectively. The results of EM and FCR are affected by this outlier point as shown in Figs. 1(a) and 1(b), respectively. Under the same initial values and stopping conditions, the result of AFCR is robust to the outlier as shown in Fig. 1(c). We also consider a regression model with two crossed lines and two outlier points. The coordinate of the outliers are (10,10) and (15,5). The data set

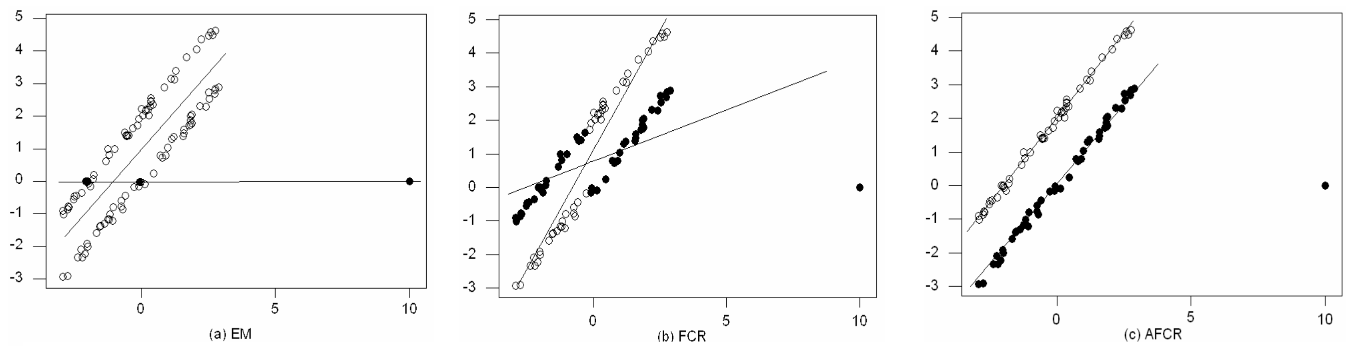


Fig. 1 Clustering results of EM, FCR and AFCR for the two parallel lines data set with one outlier point.

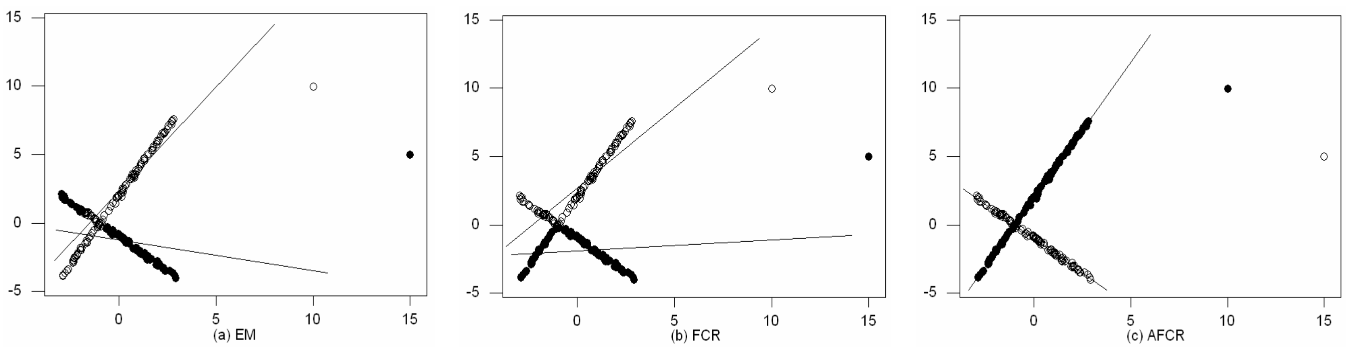
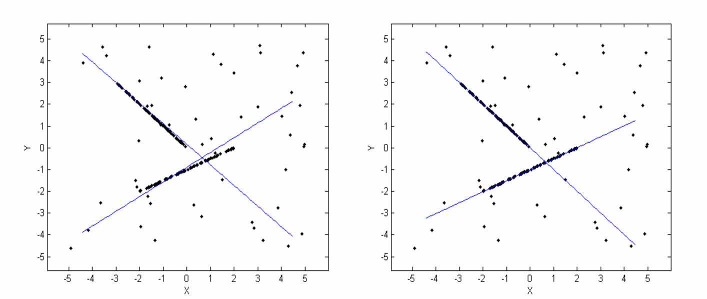


Fig. 2 Clustering results of EM, FCR and AFCR for the two crossed lines data set with two outlier points.

and clustering results are shown in Fig. 2. Similar to above example, AFCR also presents the robust properties in this example as shown in Fig. 2(c). Since the results of EM algorithm are heavily influenced by the noise and outliers, we only compare the performance of FCR and AFCR in the following artificial examples.

We know that the number of outliers is often more than one and the outliers scatter more dispersible in general data sets. We take an example to demonstrate this result. The example illustrates the data set with 2 regression lines. Each line was generated with 50 data points and we scatter 50 noise points around the regression models. The scatter plot of the data set is shown in Fig. 3. The estimated regression models obtained by FCR and AFCR are shown in Figs. 3(a) and 3(b). AFCR also gives precision parameter estimation in this noisy data.



(a) (b)

Fig. 3. The clustering results. (a) FCR. (b) AFCR

Above examples only consider the case of p=2 (i.e. regression lines). We now give an example of two regression curves with

$$y_1 = \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \varepsilon_1 = 21 - 2x + 0.0625x^2 + \varepsilon_1$$

$$y_2 = \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_1 = -5 + 2x - 0.0625x^2 + \varepsilon_2$$

where ε_1 and ε_2 follow the uniform distribution U[-1,1] over the closed interval [-1,1]. This data set is shown in Fig. 4. We add one outlying point into the data set and then estimate the switching regression coefficients using FCR and AFCR. Four kinds of outlier coordinates are adopted in turn and the results are shown in Table 1. We see that, when the outlier location is increasingly distant from the data set (i.e., the distance from outlier 1 to outlier 4), the parameter estimates obtained by FCR are increasingly imprecise. However, parameter estimates obtained by AFCR are not influenced by the outlier. Note that in our simulations, the noise and outlier locations are considered in y-coordinates (i.e. not high leverage points). If a high leverage outlier is added (i.e., an outlier location in x-coordinates), we find that both FCR and AFCR will be influenced. This phenomenon is similar to the traditional regression analysis, where the fitted lines are almost completely determined by this high leverage outlying point.

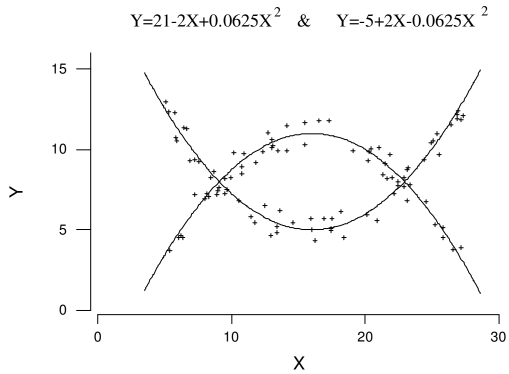


Fig. 5. Randomly generated switching regression data set with two regression curves.

Table 1. Switching regression coefficients obtained by FCR and AFCR.

model values	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}
	21.0000	-2.0000	0.0625	-5.0000	2.0000	-0.0625

FCR

outlier	x	y	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}
1	16	0	21.1215	-2.0291	0.0631	-5.3651	2.0287	-0.0630
2	16	-10	21.5038	-2.0995	0.0652	-5.0333	1.9736	-0.0614
3	16	-20	21.8274	-2.1590	0.0670	-4.6506	1.9097	-0.0594
4	16	-30	22.1695	-2.2221	0.0689	-4.2575	1.8438	-0.0574

AFCR

outlier	x	y	β_{10}	β_{11}	β_{12}	β_{20}	β_{21}	β_{22}
1	16	0	21.0110	-2.0009	0.0623	-5.1271	2.0003	-0.0623
2	16	-10	20.9952	-2.0012	0.0624	-5.1102	1.9995	-0.0624
3	16	-20	21.4607	-1.9243	0.0633	-5.1086	2.0071	-0.0617
4	16	-30	20.9920	-1.9890	0.0620	-5.2242	2.0084	-0.0623

V. Conclusions and Discussions

In this paper, we use an alternative residual measurement to modify the traditional fuzzy c-regressions (FCR) objective function and then propose an alternative fuzzy c-regressions (AFCR) algorithm. The iterative procedure to optimize the AFCR objective function is summarized in Section III. The clustering results of both EM and FCR will be influenced by

ISBN: 978-988-17012-2-0

the noise and outlier. By replacing the residual measurement in the FCR objective function with an alternative term, the proposed method is robust to noise and outlier and is quite simple such as the original FCR algorithm.

Both FCR and AFCR have the constraint that $\sum_{j=1}^{n} \mu_{ij} = 1$. The membership values for the noise and outliers belong to each cluster will equal to 1/c. This is the mean reasonable that FCR will cause trouble in a noisy environment. However, this problem will be solved in AFCR by using the alternative residual measurement. Let $S_{ij} = \exp\{-\eta(y_j - \vec{x}_j\vec{\beta}_i)^2\}$, we find that the AFCR update equation (25) has an extra weighted item S_{ij}. Since S_{ij} decreases monotonically in $(y_j - \vec{x}_j\vec{\beta}_i)^2$, the update equation (25) will reasonable assign a suitable weight to each data point including the noise and outliers. Thus, using (25) will make AFCR more robust to noise and outliers than traditional FCR.

REFERENCES

[1] R.E. Quandt, "The estimation of the parameters of a linear regression system obeying two separate regimes," Journal of the American Statistical Association, vol. 53, pp. 873-880, 1958.
[2] R.E. Quandt, "Tests of the hypothesis that a linear regression system obeys two separate regimes," Journal of the American Statistical Association, vol. 55, pp.324-330, 1960.
[3] G. Chow, "Tests of the equality between two sets of coefficients in two linear regressions," Econometrica, vol. 28, pp. 561-605, 1960.
[4] R.E. Quandt, "A new approach to estimating switching regressions," Journal of the American Statistical Association, vol. 67, pp. 306-310, 1972.
[5] R.E. Quandt and R.B. Ramsey, "Estimating mixtures of normal distributions and switching regressions," Journal of the American Statistical Association, vol. 73, pp. 730-752, 1978.
[6] R.D. de Veaux, "Mixtures of linear regressions," Computational Statistical and Data Analysis, vol. 8, pp. 227-245, 1989.
[7] L.F. Lee, "Simulation estimation of dynamic switching regression and dynamic disequilibrium models - Some Monte Carlo results," Journal of Econometrics, vol. 78, pp. 179-204, 1997.
[8] H.A. Freeman, S.K. Ehui and M. Jabbar, "Credit constraints and smallholder dairy production in the East African highlands: Application of a switching regression model," Agricultural Economics, vol. 19, pp. 33-44, 1998.
[9] G. Peters, "A linear forecasting model and its application to economic data," Journal of Forecasting, vol. 20, pp. 315-328, 2001.
[10] G. Alperovich and J. Deutsch, "An application of a switching regimes regression to the study of urban structure," Papers in Regional Science, vol. 81, pp. 83-98, 2002.
[11] R.J. Hathaway and J.C. Bezdek, "Switching regression models and fuzzy clustering," IEEE Trans. Fuzzy Systems, vol. 1, pp. 195-204, 1993.
[12] S. Wang, H. Jiang and H. Lu, "A new integrated clustering algorithm GFC and switching regressions," International Journal of Pattern Recognition and Artificial Intelligence, vol. 16, pp. 433-446, 2002.
[13] J. Leski, "ε-Insensitive Fuzzy c-Regression Models: Introduction to ε-Insensitive Fuzzy Modeling," IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 34, pp. 4-15, 2004.
[14] V. Vapnik, The Nature of Statistical Learning Theory, New York: Springer, 1995.
[15] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in Proc. Adv. Neu. Inf. Proc. Sys. Cambridge, MA: MIT Press, vol. 13, 2001.
[16] M.S. Yang and K.L. Wu, "Alpha-cut implemented fuzzy clustering algorithms and switching regressions", IEEE Trans. Systems, Man and Cybernetics, Part B, vol. 38, pp. 588-603, 2008.
[17] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," Journal of the Royal Statistical Society B, vol. 39, pp. 1-38, 1977.
[18] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press, 1981.
[19] M.S. Yang, "A survey of fuzzy clustering," Mathematical and Computer Modeling, vol. 18, pp. 1-16, 1993.
[20] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition-part I and II", IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 29, pp. 778-801, 1999.
[21] F. Hoppner, F. Klawonn, R. Kruse and T. Runkler, Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition. New York: Wiley, 1999.
[22] J. Yu and M.S. Yang, "Optimality test for generalized FCM and its application to parameter selection," IEEE Trans. Fuzzy Systems, vol. 13, pp. 164-176, 2005.
[23] D.E. Gustafson and W.C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," Proc. IEEE CDC, San Diego, CA, pp.761-766, 1979.
[24] R. Krishnapuram and J. Kim, "A note on the Gustafson-Kessel and adaptive fuzzy clustering algorithms," IEEE Trans. Fuzzy Systems, vol. 7, pp. 453-461, 1999.
[25] I. Gath and A.B. Geva, "Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Analysis and Machine Intelligence," vol. 11, pp. 73-781, 1989.
[26] K.L. Wu and M.S. Yang, "Alternative c-means clustering algorithms," Pattern Recognition, vol. 35, pp. 2267-2278, 2002.
[27] K.L. Wu and M.S. Yang, "Alternative learning vector quantization," Pattern Recognition, vol. 39, pp. 351-362, 2006.