1. Introduction

Since Quandt [1,2] and Chow [3] initiated the research on switching regressions, it had been widely studied and applied in psychology, economics, social science and music perception [4–8]. For fuzzy clustering, the fuzzy c-means (FCM) algorithm is the best known and most commonly used method [9–12]. Hathaway and Bezdek [13] first combined switching regressions with FCM, which they referred to as fuzzy c-regressions (FCR). To increase the speed of FCR, Wang et al. [14] combined the concept of Newton’s law of gravity with FCR. Since these FCRs are sensitive to noise and outliers, Leski [15] considered an ε-insensitive loss function to improve the robustness against noise and outliers. However, these FCRs present a problem whose results always depend heavily on initial values, even more seriously than FCM.

The mountain method proposed by Yager and Filev [16,17] can obtain the initial values for a clustering algorithm. In the mountain method, extracting initials is restricted to the grid nodes such that the computation time will increase with the dimensionality of data. In order to implement the mountain method on a high-dimensional data set, Chiu [18] modified the original mountain method by considering the data points instead of the grid nodes. Pal and Chakraborty [19] used the mountain method to extract clusters of different shapes. Yang and Wu [20] modified both mountain function and revised mountain function to reduce its computation time where a method of estimating parameters and cluster number was also considered. Pal et al. [21] used the mountain method to find typical objects of a relational data set. In this paper, we will use the modified mountain method to extract the initial-value problem of FCR.

We first transform the switching regression data set into a parameter space. We then implement the modified mountain method on the transformed data set. From our experiments, we find that the proposed mountain c-regressions (MCR) method actually gives good c-fitted regression models for the switching regression data set. Several experimental results demonstrate its effectiveness and superiority of the proposed method. The rest of the paper is organized as follows. In Section 2, we review switching regressions and the FCR algorithm. We also use an example to demonstrate the influence of initial values on FCR. The original mountain method and the modified mountain method are reviewed in Section 3. In Section 4, we proposed the data transformation technique that can transform the original switching regression data set into the parameter space. The properties of transformation are also discussed. We use the modified mountain method to extract cluster centers from the transformed data set. These cluster centers can correspond to
the switching regression models in the original data set. We call this method the mountain c-regressions. In Section 5, the comparisons of the proposed MCR with other methods are made. The numerical examples and comparisons show the effectiveness and superiority of the proposed method. Conclusions are finally stated in Section 6.

2. Switching regression and fuzzy c-regressions

Regression analysis is a technique for modeling the relation between independent and dependent variables. Usually, a single regression model is used for fitting a data set. However, it is sometimes necessary to have more than one regression model, say $c$ number of regression models, for fitting a data set. This kind of model fitting is called switching regressions (see [1–4]). Assume that $\{ (x_1,y_1), \ldots, (x_n,y_n) \}$ is a set of data where each independent observation $x_j = (x_{j1}, \ldots, x_{jp}) \in \mathbb{R}^p$ has its corresponding dependent observation $y_j \in \mathbb{R}$. The switching regression is employed to find $c$ linear regressions

$$
\hat{y}_{ij} = \beta_{i0} + \beta_{i1}x_{j1} + \cdots + \beta_{ip}x_{jp}, \quad i = 1, \ldots, c
$$

that best fit the data structure.

Hathaway and Bezdek [13] first combined switching regressions with fuzzy c-means and referred to them as fuzzy c-regressions. The FCR algorithm is to minimize the objective function $J_{FCR}$ defined in

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

with $d(y_j, \hat{y}_{ij}) = (y_j - \hat{y}_{ij})^2$. The update equations for the minimization are

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

and

$$
\mathbf{\hat{p}_i} = \begin{bmatrix} \hat{\beta}_{i0} \\ \vdots \\ \hat{\beta}_{ip} \end{bmatrix} = [X^T D_i X]^{-1} X^T D_i Y,
$$

where $X$ denotes the matrix in $\mathbb{R}^{n \times (p+1)}$ having $1, x_1 = (1, x_{11}, \ldots, x_{1p})$ as its $j$th row, $Y$ denotes the vector in $\mathbb{R}^n$ having $y_j$ as its $j$th component, and $D_i$ denotes the diagonal matrix in $\mathbb{R}^{n \times n}$ having $\mu_{ij}$ as its $j$th diagonal element.

To increase the speed of FCR, Wang et al. [14] combined the concept of Newton’s law of gravity with FCR. Since these FCRs are sensitive to noise and outliers, Leski [15] considered an $\epsilon$-insensitive loss function to improve the robustness against noise and outliers. However, the performance of these FCRs still depends heavily on initial values. In order to analyze the effect of initial values on regression equations, we provide the data sets with two switching regression lines as shown in Figs. 1(a) and (b). We then implement the FCR algorithm with random initials in which the results are shown in Figs. 1(c) and (d), respectively. In fact, we repeated 100 trials with random initial values to implement the FCR algorithm for the data sets of Figs. 1(a) and (b). We found that there are only 53 trials that FCR performs well for Fig. 1(a) and there are 82 trials that FCR performs well for Fig. 1(b). In clustering, the mountain method proposed by Yager and Filev [16] is a well-known technique that can offer a set of suitable initial values and can also be used as an approximate clustering algorithm. However, the mountain method cannot be directly implemented for the switching regression data set. In this paper, we provide a transformation technique for solving it. First, however, we give a brief review of the mountain method and the modified mountain method.

3. Mountain and modified mountain methods

3.1. Mountain method

The mountain method was first proposed by Yager and Filev [16]. Suppose, we have $n$ data points denoted by $(x_1, \ldots, x_p)$ in the $p$-dimensional Euclidean space $\mathbb{R}^p$. We take a grid in the data space with $N_i$ denoting the set of all grid nodes. In the mountain method, cluster centers are restricted to the grid nodes $N_i$ in $\mathbb{R}^p$. The mountain function for each $N_i$ is defined by

$$
M_i(N_i) = \sum_{j=1}^n e^{-ad(x_j, N_i)}, \quad i = 1, 2, \ldots,
$$

where $d(x_j, N_i)$ is the distance between the data point $x_j$ and the grid node $N_i$. The mountain function of the nodes is similar to the density function of the data points in the neighborhood. A node with many neighborhood data points will have a large mountain function value. The parameter $a$ decides the neighborhood radius. A large value of $M_i(N_i)$ represents a large potential ability for node $N_i$ being a cluster center or being a suitable initial value. It is reasonable that the first cluster center estimate is the node with a maximum mountain function score. An objective is to find $N_1^*$ (the first cluster center estimate among all grid nodes) with

$$
M_1(N_1^*) = \max_i M_i(N_i).
$$

Since the nodes close to $N_1^*$ will also have large mountain function values, it is necessary to remove the effect of the identified cluster center before obtaining the next cluster center. The mountain function, after eliminating the $(k - 1)$th cluster center $N_{k-1}^*$, is defined by

$$
M_k(N_k) = M_{k-1}(N_k) - M_{k-1}(N_{k-1}^*) e^{-ad(N_{k-1}^*, N_k)}, \quad k = 2, 3, \ldots,
$$

![Fig. 1.](image-url) (a) The data set with two parallel lines; (b) the data set with two cross lines; (c) and (d) the results of FCR with random initial values for the data sets (a) and (b), respectively.)
where
\[ M_{k-1}(N_{k-1}^*) = \max_i \{ M_{k-1}(N_i) \}, \quad k = 2, 3, \ldots . \] (8)

The function of (7) is called the revised mountain function. The parameter \( \gamma \) determines the neighborhood radius that will have measurable reductions in the mountain function. We can obtain new identified cluster centers by iterations using (7) and (8). For more information on mountain methods, see [17–21].

3.2. Modified mountain method

Since the computation time of the mountain method increases with the dimensionality of the data, Chiu [18] modified the original mountain method by considering the data points instead of the grid nodes. Yang and Wu [20] modified both mountain function and revised mountain function to reduce the computation time where a method of estimating the parameters and number of clusters was also considered. Use of the modified mountain function [20] is analogous to kernel density estimation by defining the modified mountain function on each data vector \( x_i \) with
\[
P_1(x_i) = \sum_{j=1}^{n} e^{-\lambda \|x_i - x_j\|^2}, \quad i = 1, \ldots, n, (9)
\]

where
\[
\rho = \left( \frac{\sum_{j=1}^{n} \|x_i - \bar{x}\|^2}{n} \right)^{-1}
\]
with \( \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n} \). (10)

The role of the parameter \( \rho \) is a normalization term and the parameter \( \lambda \) can be obtained by a correlation self-comparison procedure [20] or selected by the users. The first cluster center estimate \( x^*_1 \) is obtained by
\[
P_1(x^*_1) = \max_i \{ P_1(x_i) \}. (11)
\]

Next search is made for the \( k \)th cluster center using the modified revised mountain function with
\[
P_k(x_i) = P_{k-1}(x_i) - P_{k-1}(x_i)e^{-\rho \|x_i - x_{k-1}^*\|^2}, \quad k = 2, 3, \ldots . (12)
\]

where \( x_i \) is the feature vector and \( x_{k-1}^* \) is the \((k-1)\)th cluster center which satisfies
\[
P_{k-1}(x_{k-1}^*) = \max_i \{ P_{k-1}(x_i) \}, \quad k = 2, 3, \ldots . (13)
\]

New cluster centers are obtained by the iterations of (12) and (13). To demonstrate behaviors of the modified revised mountain function (12), we use the data set as shown in Fig. 2(a).

The values of the modified mountain function and the modified revised mountain function after extracting the first, second, third, and fourth clusters are shown in Figs. 2(b)–(f), respectively. Note that the above example shows the efficiency and robustness of the modified mountain method. The locations of the extracted cluster centers are shown in Fig. 2(a) with solid circle points. However, a cluster center in switching regressions presents a regression model. Since the modified mountain functions cannot extract initial regression models for switching regressions, the method cannot be directly applied in switching regression data set. In the next section, we will present a data transformation method, which transforms a switching regression data set into a parameter space. In this transformed data set, the modified mountain method can be employed to extract good switching regression models without any initializations.

4. Data transformation and mountain c-regressions method

Although the modified mountain method can extract good cluster center estimates as shown in Fig. 2, it cannot be employed to extract suitable switching regression models for a switching regression data set. To apply the modified mountain method to switching regression analysis, we propose a data transformation technique that can transform the original switching regression data set into a parameter space and then implement the modified mountain method on the transformed data set to extract cluster center estimates. These c
extracted cluster center estimates will correspond to $c$ switching regression models and we call this method the mountain $c$-regression clustering method. We now introduce the data transformation technique.

4.1. Data transformation

We know that any two points can form a line. For example, any two points in Fig. 1(a) can have a regression equation $y = b_0 + b_1x$ where the parameters of this equation is denoted by $b=(b_0, b_1)$. If we pair all points of the data set, such as that shown in Fig. 1(a), to form the regression lines, we will have a set of parameters $\vec{b}_l = (b_{l0}, b_{l1})$, $l = 1, \ldots, \binom{n}{2}$ where $\binom{n}{2}$ denotes the number of combinations. For data points located around a regression line, the parameters $\vec{b}_l$ for these points should be very close. If we use a scatter plot to present these $\vec{b}_l$, we will see that these $\vec{b}_l$ will form a cluster in the parameter space.

The scatter plots of the parameters $\vec{b}_l$ obtained by transforming the data sets of Figs. 1(a) and (b) are shown in Figs. 3(a) and (b), respectively. Two significant clusters can be found in these transformed data sets. The $x$-coordinate and $y$-coordinate of each point present the intercept $b_{l0}$ and the slope $b_{l1}$, respectively. We now have a set of data points $\vec{b}_l$ where each point represents a line in the original switching regression data set. If we implement the modified mountain method on these $\vec{b}_l$ to extract $c$ cluster centers, these $c$ extracted points will correspond to $c$ switching regression models for the original data set. Moreover, for a quadratic regression $y = b_0 + b_1x + b_2x^2$, we may use any three points in the data set to form a quadratic regression curve. We have a set of parameters $\vec{b}_l = (b_{l0}, b_{l1}, b_{l2})$, $l = 1, \ldots, \binom{n}{3}$ where $\binom{n}{3}$ denotes the number of combinations. In general, for the case $y = b_0 + b_1x_1 + \cdots + b_px_p$ of $p$ predictors, $(p + 1)$ points can form a regression model with $(p + 1)$ parameters $\vec{b}_l = (b_{l0}, b_{l1}, \ldots, b_{lp})$. For the combinations of all $(p + 1)$ points, we will have a set of parameters $\vec{b}_l=(b_{l0}, b_{l1}, \ldots, b_{lp})$ belonging
Fig. 6. (a and d) The data sets with more noisy points for the data sets of Figs. 1(a) and (b); (b and e) the transformed data sets and their extracted cluster centers; (c and f) the corresponding MCR results.

Fig. 7. (a) The four-cluster switching regression data set; (b)–(e) the transformed data sets with the sample sizes $L = 200, 400, 600$ and $800$, respectively; (f)–(i) the MCR results for the data sets of (b)–(e), respectively.

to a $(p+1)$-dimensional transformed data space. Each point in this parameter space presents a regression model with $p$ predictors in the original switching regression data set.

Assume that $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is a data set where each independent observation $x_j = (x_{j1}, \ldots, x_{jp}) \in \mathbb{R}^p$ has its corresponding dependent observation $y_j \in \mathbb{R}$. Each combination of $(p + 1)$ points will uniquely determine one regression equation and the parameters of each regression equation are denoted by

$$\tilde{b}_l = (b_{l0}, b_{l1}, \ldots, b_{lp}), \quad l = 1, \ldots, C_{p+1}.$$  \hfill (14)

Each $\tilde{b}_l$ is a point in the parameter space $\mathbb{R}^p$ which corresponds to a regression model in the original data space. In Fig. 3(a), the slopes of two transformed clusters are both 1 and the intercepts are 0 and 2, respectively. This is coincident to the original data set of Fig. 1(a). In Fig. 3(b), the intercepts of two transformed clusters are both 0 and the slopes are 0 and 1, respectively. This is also coincident to the original data set of Fig. 1(b).

We can employ any clustering algorithms to extract two cluster centers from the transformed data set where the extracted cluster centers will correspond to two switching regression models in the original switching regression data set. The mountain method is one of these clustering algorithms that need not specify initial values. More details of the reasons that we use the modified mountain method
will be illustrated below. We first show some properties of this data transformation method.

4.2. Properties of transformation

We now add some noisy points to the switching regression data sets of Figs. 1(a) and (b) to form two data sets as shown in Figs. 4(a) and 4(b), respectively. The transformed data sets of Figs. 4(a) and (b) in the parameter space are shown in Figs. 4(c) and (d), respectively. Since most switching regression methods, such as FCR, use the concepts of minimizing the sum of squared residuals to find a best regression model, noise and outliers will influence the results. However, we find that the locations of the cluster centers in Figs. 4(c) and (d) with noisy points are almost equivalent to the locations of the cluster centers in Figs. 3(a) and (b) without noisy points. That means these influences from noise and outliers for switching regression data sets will become less on the structure of the transformed data set. If we can develop an appropriate clustering method for finding the optimal $c$ cluster centers from the transformed data set, the extracted switching regression models will be robust to noise and outliers.

Another property is that the separation between points on the transformed data set in the parameter space will become large. For each parameter in the parameter space, its possible value will belong to the interval $[-\infty, \infty]$. That is, $b_i \in [-\infty, \infty]$ for all $i = 1, \ldots, p$. Since we want to extract $c$ cluster centers from the transformed data set using clustering algorithms, this kind of large separation will become a big challenge for most clustering methods in which a clustering algorithm may get bad clustering results on the transformed data set. This is also the reason why we use the modified mountain method to extract $c$ cluster centers from the transformed data set. We will show that the modified mountain method works better than other fuzzy clustering algorithms with numerical examples. We now summarize the proposed mountain $c$-regressions clustering method in the next section.

4.3. Mountain c-regressions method

We have known that the data transformation method can transform the switching regression data sets into the parameter space and the number of parameters in the switching regression model will correspond to the number of dimensions in the parameter space. After producing the transformed data set, we can use the modified mountain clustering method to extract $c$ cluster centers from the transformed data set. These $c$ cluster centers will be equivalent to $c$ switching regression models in the original switching regression data set. The process of the proposed mountain $c$-regressions can be summarized as follows:

**MCR method**

*Step 1:* Transform the original switching regression data set into the parameter space to acquire the transformed data set.

*Step 2:* Apply the modified mountain method to extract $c$ cluster centers of the transformed data set.

*Step 3:* Transform the extracted cluster centers into the original switching regression data set to acquire $c$ switching regression models.

We now give some simple examples. Figs. 5(a) and (b) show the two cluster centers extracted using the modified mountain method applied to the transformed data sets of Figs. 4(c) and (d); the two corresponding extracted switching regression models are shown in Figs. 5(c) and (d), respectively. We also add more noisy points to test the robustness of the transformed data. Figs. 6(a) and (d) show the data sets after adding 20% and 50% noisy points to the original switching regression data set of Fig. 1(a). The transformed data sets and extracted cluster centers are shown in Figs. 6(b) and (e), respectively. The corresponding extracted regression models are shown in Figs. 6(c) and (f), respectively. By comparing Figs. 6(b) and (e) with Fig. 3(a), we find that those noisy points do not change the structure of the transformed data set. Hence, the extracted regression models are not influenced much by those noisy points as shown in Figs. 6(c) and (f). The results obtained after more experiments reveal that the transformed data structure with its extracted regression models is
Fig. 9. (a)–(f) The extracted regression lines for the data sets in Figs. 8(a)–(f); (a) and (d) the results from FCM; (b) and (e) the results from AFCM; (c) and (f) the results from the modified mountain method.

Fig. 10. The possible results obtained by FCR and MFCR with random initial values.

always robust to those noisy points when the size of noises is \(< 20\%\) of the original switching regression data set.

4.4. Computational cost

A major limitation for applying the proposed MCR is that the data set transformed from the original switching regression data set may become too large. For example, a data set with \(p = 1\) (one predictor case) and \(n = 200\) will have \(C_{p+1}^n = C_{200}^2 = 19,900\) transformed data points \(b_l\) in the parameter space. In the case of \(p = 2\) (the two-predictor case), we have \(C_{p+1}^n = C_{200}^3 = 131,300\) transformed data points \(b_l\) in the parameter space. In this case, the number of transformed data sets is much bigger than the number of original data sets. If we apply MCR to the transformed data set, it may be affected by the memory of hardware used or the time consumed by the algorithm. To overcome this drawback, the concept of data sampling from the transformed data set can be used.

Our method for reducing the transformed data size is to set a constant \(L\) and generate randomly the \(L\) combinations of \(p\) data points from the original switching regression data set. We then apply the modified mountain clustering to the data set \(\{b_l, l = 1, \ldots, L\}\). This is equivalent to obtaining a random sample of size \(L\) from the transformed data set \(\{b_l, l = 1, \ldots, C_{p+1}^n\}\). If the population data set

<table>
<thead>
<tr>
<th>Fig. 1</th>
<th>FCR</th>
<th>MFCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td>(b)</td>
<td>18</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1
Number of inappropriate results out of 100 trials estimated by FCR and MFCR with random initial values.
Fig. 11. (a)–(c) The results from FCR, MFCR, and MCR, respectively, for the data set of Fig. 1(b) with an outlier (−20, 3).

Fig. 12. (a)–(c) The results from FCR, MFCR, and MCR, respectively, for two regression models \( y = -x \) and \( y = -1 + 0.5x \) with 50 noisy points.

Fig. 13. The randomly generated data sets.

\[ \{ \vec{b}_l, l = 1, \ldots, C_{n+1} \} \] contains \( c \) compact clusters, then the random sample \( \{ \vec{b}_l, l = 1, \ldots, L \} \) should also contain \( c \) significant clusters. We give an example for demonstration. Fig. 7(a) shows a four-cluster switching regression data set and random samples with \( L = 200, 400, 600 \) and 800 are shown in Figs. 7(b), (c), (d) and (e), respectively. We now apply the MCR to the transformed random samples \( \{ \vec{b}_l, l = 1, \ldots, L \} \). We find that they can still perform well even with a small sample size \( L = 200 \). This sampling technique enhances greatly the usefulness of MCR in real applications. In general, we suggest the choice of \( L = cnp \). The proposed MCR results for the transformed data sets of Figs. 7(b), (c), (d) and (e) are shown in Figs. 7(f), (g), (h) and (i), respectively. According to the results of Fig. 7, we find that the random sampling technique can greatly reduce computational cost for higher-predictors switching regression data set.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Fig. 13(a)</th>
<th>Fig. 13(b)</th>
<th>Fig. 13(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCR</td>
<td>0.19271</td>
<td>0.18322</td>
<td>0.44472</td>
</tr>
<tr>
<td>MCR</td>
<td>0.00425</td>
<td>0.07002</td>
<td>0.48597</td>
</tr>
</tbody>
</table>
The AFCM objective function

The necessary conditions for a minimizer \((\mu, z)\) of \(J_{\text{AFCM}}\) are the following updating equations:

\[
\mu_i = \frac{(1 - \exp(-\beta|x_j - z_i|^2))^{-1/m}}{\sum_{k=1}^{n} (1 - \exp(-\beta|x_j - z_k|^2))^{-1/m}} \tag{18}
\]

and

\[
Z_i = \frac{\sum_{j=1}^{n} \mu_j^m x_j}{\sum_{j=1}^{n} \mu_j^m} \tag{19}
\]

According to previous discussion, we know that the transformed data set may not only become too large, but also be with big separation between points. In our experiments, both FCM and AFCM cannot give good cluster center estimates for the transformed data sets in which their corresponding extracted regression models cannot match the structure of the original switching regression data. For the simple example shown in Figs. 1(a), 3(a) and (b) show the corresponding transformed data sets. For the transformed data sets shown in Figs. 3(a) and (b), the extracted two cluster centers with solid circle points by FCM, AFCM and modified mountain method are shown in Fig. 8. The two corresponding extracted regression lines are shown in Fig. 9. Fig. 8 shows that the extracted cluster centers obtained by FCM and AFCM cannot be located in the densest area of the data set. Since the modified mountain method can be well used to seek the modes of the data set, it is suitable to be applied to large-sample-size and high-dispersion data sets. In our simulations, most \(c\)-means clustering algorithms cannot work well on the transformed data set even when a set of well-designed initial values are adopted. This is one of the reasons why we choose the modified mountain method to extract the cluster centers from the transformed data set. In the next subsection, we will compare the proposed MCR with other switching regression algorithms.

5.2. Comparisons with FCR and modified FCR methods

In this subsection, we will compare MCR with two switching regression algorithms of fuzzy \(c\)-regrressions and modified fuzzy \(c\)-regrressions (MFCR). First, we take a brief introduction to the MFCR algorithm. It combines the concept of FCR proposed by Hathaway and Bezdek [13] and the concept of AFCM proposed by Wu and Yang [22]. We refer to it as modified FCR (MFCR). The membership update equations of MFCR are

\[
\mu_i = \frac{(1 - \exp(-\beta|y_j - \hat{y}_j|^2))^{-1/m}}{\sum_{k=1}^{n} (1 - \exp(-\beta|y_j - \hat{y}_k|^2))^{-1/m}} \tag{20}
\]

where \(d(y_j, \hat{y}_j) = |y_j - \hat{y}_j|^2\) and \(\beta = (\frac{1}{n} \sum_{j=1}^{n} d(y_j, \hat{y}))^{-1}\). The parameter estimation of \(\hat{y}_i\) is equivalent to Eq. (4). For the data sets of Figs. 1(a) and (b), if we give a set of well-designed initial values, both FCR and MFCR can perform well. These have similar results to that

<table>
<thead>
<tr>
<th>Table 3</th>
<th>True parameter values for the data sets shown in Fig. 14.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 14</td>
<td>n-values interval</td>
</tr>
<tr>
<td>(a)</td>
<td>46</td>
</tr>
<tr>
<td>(b)</td>
<td>28</td>
</tr>
<tr>
<td>(c)</td>
<td>30</td>
</tr>
<tr>
<td>(d)</td>
<td>46</td>
</tr>
</tbody>
</table>

Fig. 14. Four different data sets with four cases of two quadratic regression models.
obtained by MCR as shown in Figs. 9(c) and (f). However, if we give a set of random initial values, the results obtained by FCR and MFCR may become unreliable. Fig. 10 shows the possible results of FCR and MFCR for the data sets in Figs. 1(a) and (b). Table 1 lists the number of inappropriate estimations (out of 100 random trials) that each algorithm with their exact indeterminable initial conditions.

We can find that the results of FCR and MFCR algorithms are actually dependent on initial values, especially for the data set in Fig. 1(a).

We now add an outlier with its coordinates (−20, 3) in the data set of Fig. 1(b). The clustering results of FCR, MFCR, and MCR are shown in Figs. 11(a), (b), and (c), respectively. The lines present the estimated switching regression models. The FCR results are affected by this outlying point where the outlier and some data points form a regression model, but the original data points form another regression model. However, under the same initial values and stopping conditions, the result of MFCR is more robust to this outlier than...
that obtained by FCR as shown in Figs. 11(b). The result of MCR is also robust to the outlier where the estimated switching regression models actually match the data structure well.

We know that the number of noisy points is often more than one that experienced in real applications. We will use an example to demonstrate that the MFCR is affected by these noisy points when our proposed MCR is still robust. The example contains two regression models with $y = -x$ and $y = -1 + 0.5x$. Each model is generated using 50 data points. We then add 50 noisy points to these two regression models. The regression models obtained by FCR, MFCR, and MCR are shown in Figs. 12(a), (b), and (c), respectively. The lines present the estimated regression models. The FCR and MFCR are actually affected by noisy points. However, the result of MCR is still robust to these noisy points even though the size of noise is up to 50% of the original data set. Note that, since the FCR tries to minimize the objective function of the sum of squared errors, a single outlier may even affect the clustering results much more than a set of random noisy points. This phenomenon is similar to the traditional regression analysis. Figs. 11(a) and 12(a) have demonstrated this phenomenon.

We also give an example to demonstrate the performance of FCR and MCR in a highly overlapping case. The randomly generated data
sets are shown in Figs. 13(a), (b) and (c). The results of 100 repeated trials are shown in Table 2. The true parameters of the switching regression models are used as the initial values for FCR. The means of squared errors are computed by averaging the sum of the squared errors between the estimated parameter and the true parameter. Both FCR and MCR cannot perform well in these highly overlapping cases where FCR needs initial values, but MCR does not. The better results obtained by FCR as shown in Table 2 are obtained by adopting good initial values. The worse results of FCR occur for some initial values. For the cases in Figs. 13(a) and (b), we find that MCR gives a more precise estimation than that obtained by FCR without any initializations.

5.3. Comparisons for quadratic regression models

Now we use another example to demonstrate the results of two quadratic regression models with the following form:

\[
\begin{align*}
y_1 &= \beta_{10} + \beta_{11} x_1 + \beta_{12} x_1^2, \\
y_2 &= \beta_{20} + \beta_{21} x_1 + \beta_{22} x_1^2.
\end{align*}
\]

(21a) (21b)

The comparisons include five clustering algorithms of FCM, AFCM, FCR, MFCR and our proposed MCR. Four cases of quadratic regression models are shown in Table 3 where the corresponding data sets are shown in Fig. 14. We mention that these cases had been considered in Hathaway and Bezdek [13]. In a quadratic regression model, the total number of transformed data sets is equal to \( C_n^2 \) where \( n \) is the number of original data set. The size of the transformed data set will be larger than that of the original data set where the scatter plot of the transformed data set is a three-dimensional data set. The specified initial values for this series data sets are \((-19, 2, 0)\) and \((-31, 2, 0)\). The clustering results of FCM, AFCM, FCR, MFCR and MCR are shown in the first, second, third, fourth and fifth row of Fig. 15, respectively. The lines in the figures present the estimated regression models obtained by the clustering algorithms. Note that well-designed initial values can improve the performance of FCR and MFCR such as the results obtained by MCR without any initial values. We now add an outlying point and use the true model parameters as the initial values (see Table 3). The results are shown in Fig. 16. Only the proposed MCR can give good estimated regression models. It is not necessary to specify initial values for the proposed MCR method. Moreover, the results of MCR are also robust to noise and outliers.

6. Conclusions and discussion

The c-regressions clustering algorithm, such as FCR, depends heavily on initial values and is not robust to noise and outliers. In this paper, we propose a new robust algorithm that can be robust to initial values, noisy points and outliers. We construct a transformation technique that can transform the original switching regression data set into a parameter space. The concept of transformation on switching regression data set is employed to form a new data space. We then apply the modified mountain clustering to the transformed data set to extract c cluster centers. In this manner, the mountain c-regression clustering method is created. These centers extracted by the MCR in the parameter space correspond well to c switching regression models in the original switching regression data set.

We know that any c-means clustering algorithm, such as fuzzy c-means and alternative fuzzy c-means, can be employed to extract cluster centers from the transformed data set in the parameter space. In the paper, we have compared the performance of the proposed MCR with FCM and AFCM. According to our analysis on properties of the transformed data, most c-means algorithms cannot work well for the transformed data sets. This is the main reason why we use the modified mountain method to extract cluster centers from the transformed data set. More numerical examples are also employed to compare the proposed MCR with other c-regression algorithms, such as fuzzy c-regressions and modified fuzzy c-regressions. As a whole, the MCR method actually presents good results for the analysis on switching regression data sets. Several experiments demonstrate better accuracy and effectiveness of the MCR.

Acknowledgments

The authors are grateful to the anonymous referees for their critical and constructive comments and suggestions for improving the presentation of the research in the paper. This work was supported in part by the National Science Council of Taiwan, under Kuo-Lung Wu’s Grant: NSC-98-2118-M-168-001 and Miin-Shen Yang’s Grant: NSC-97-2118-M-033-002-MY2.

References

About the Author—KUO-LUNG WU received the BS degree in Mathematics in 1997, the MS and PhD degrees in Applied Mathematics in 2000 and 2003, all from the Chung Yuan Christian University, Chungli, Taiwan. Since 2003, he has been an Assistant Professor in the Department of Information Management at Kun Shan University of Technology, Tainan, Taiwan. He is a Member of the Phi Tau Phi Scholastic Honor Society of Taiwan. His research interests include fuzzy theorem, cluster analysis, pattern recognition, and neural networks.

About the Author—MIIN-SHEN YANG received the BS degree in Mathematics from the Chung Yuan Christian University, Chungli, Taiwan, in 1977, the MS degree in Applied Mathematics from the National Chiao-Tung University, Hsinchu, Taiwan, in 1980, and the PhD degree in Statistics from the University of South Carolina, Columbia, USA, in 1989. In 1989, he joined the Faculty of the Department of Mathematics in the Chung Yuan Christian University as an Associate Professor, where, since 1994, he has been a Professor. From 1997 to 1998, he was a Visiting Professor with the Department of Industrial Engineering, University of Washington, Seattle. During 2001–2005, he was the Chairman of the Department of Applied Mathematics in the Chung Yuan Christian University. His current research interests include applications of statistics, fuzzy clustering, neural fuzzy systems, pattern recognition and machine learning. Dr. Yang is an Associate Editor of the IEEE Transactions on Fuzzy Systems, and Editorial Board of the Applied Computational Intelligence and Soft Computing.

About the Author—JUNE-NAN HSIEH received the BS degree in Mathematics in 1997, the MS degrees in Applied Mathematics in 1999, all from the Chung Yuan Christian University, Chungli, Taiwan. Since 2003, he has been a PhD Student in the Department of Applied Mathematics at Chung Yuan Christian University, Chung-Li, Taiwan. His research interests include fuzzy theorem, cluster analysis, and switching regressions.