International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 14, No. 1 (2006) 87–100

FUZZY PRINCIPAL COMPONENT REGRESSION (FPCR) FOR FUZZY INPUT AND OUTPUT DATA

JIH-JENG HUANG*, GWO-HSHIUNG TZENG† and CHORNG-SHYONG ONG‡,

*Department of Information Management, National Taiwan University,
No.1, Sec. 4, Roosevelt Road, Taipei, Taiwan
d91725010@ntu.edu.tw
† Institute of Management of Technology, National Chiao Tung University,
Hsinchu 30050, Taiwan
ghtzeng@cc.nctu.edu.tw
‡ Department of Information Management, National Taiwan University,
No.1, Sec. 4, Roosevelt Road, Taipei, Taiwan
ongcs@im.ntu.edu.tw

Received 27 May 2004
Revised 1 April 2005

Although fuzzy regression is widely employed to solve many problems in practice, what seems to be lacking is the problem of multicollinearity. In this paper, the fuzzy centers principal component analysis is proposed to first derive the fuzzy principal component scores. Then the fuzzy principal component regression (FPCR) is formed to overcome the problem of multicollinearity in the fuzzy regression model. In addition, a numerical example is used to demonstrate the proposed method and compare with other methods. On the basis of the results, we can conclude that the proposed method can provide a correct fuzzy regression model and avoid the problem of multicollinearity.

Keywords: Fuzzy regression; fuzzy centers principal component analysis; fuzzy principal component scores; multicollinearity; fuzzy principal component regression (FPCR).

1. Introduction

Since the fuzzy linear regression was originally proposed by Tanaka et al. in 1982[1,2], it has been applied to various problems to consider the situation of uncertainty and vagueness in practice. However, the problem of multicollinearity seems to have been ignored. The problem of multicollinearity will result in incorrect fuzzy regression coefficients e.g. a positive effect becomes a negative effect. The purpose here is to propose a new kind of fuzzy regression which can overcome the problem of multicollinearity and is suitable for both fuzzy input and output variables.

Although the partial index of confidence was proposed to try to cope with the problem of multicollinearity in fuzzy regression, it is clear that the index is more suitable for measuring the impact of a new input variable rather than the impact of multicollinearity. As we know, several methods have been proposed to deal with the problem of multicollinearity in traditional statistics and one of them is principal component regression. In this paper, fuzzy principal component regression (FPCR) is derived to overcome the problem of multicollinearity. First, the fuzzy input variables are transformed to the fuzzy principal component scores and then the shape-preserving operations are used to form FPCR. Since the fuzzy principal component scores are uncorrelated with each other, the problem of multicollinearity can be avoided.

In addition, a numerical example is used to demonstrate the procedures and the advantages of the proposed method. We first employ the fuzzy correlation analysis to display the problem of multicollinearity in fuzzy regression and then two kinds of the fuzzy principal component analysis (FPCA) are employed to derive the fuzzy principal component scores. Finally, the fuzzy principal component scores are used to form FPCR using the shape preserving operations. On the basis of the numerical results, we can conclude that the proposed method can provide the rational fuzzy regression coefficients and avoid the problem of multicollinearity.

The rest of this paper is organized as follows. The problems of multicollinearity in fuzzy regression are presented in Section 2. Fuzzy principal component analysis is proposed in Section 3 and fuzzy principal component regression is presented in Section 4. A numerical example is implemented in Section 5 to demonstrate the effectiveness of the proposed method. Section 6 contains discussions of the implementation, and the final section presents conclusions.

2. Problem of Multicollinearity in Fuzzy Regression

The form of the fuzzy linear regression in Tanaka's model can be expressed as

$$y = \tilde{A}_0 + \tilde{A}_1 x_1 + \cdots + \tilde{A}_n x_n = \tilde{\boldsymbol{A}}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]'$ is a non-fuzzy input vector, and $\tilde{A}_i$ is a symmetrical fuzzy number denoted as $(a_i, c_i)_L$. The form of the membership function of Eq. (1) can be obtained for $\boldsymbol{x} \neq \boldsymbol{0}$ as

$$\mu_Y(y) = L((y - \boldsymbol{a}'\boldsymbol{x}) / \boldsymbol{c}' \,|\, \boldsymbol{x} \,|) \tag{2}$$

for $\boldsymbol{x} = \boldsymbol{0}$ and $y = 0$, $\mu_Y(y) = 1$, and for $\boldsymbol{x} = \boldsymbol{0}$ and $y \neq 0$, $\mu_Y(y) = 0$. The h-level set of y denoted as $[y]_h$ can be obtained as the following setting:

$$L((y - \boldsymbol{a}'\boldsymbol{x}) / \boldsymbol{c}' \,|\, \boldsymbol{x} \,|) = h. \tag{3}$$

Then, $[y]_h$ can be obtained as

$$[y]_h = [(\boldsymbol{a}'\boldsymbol{x} - \boldsymbol{c}' \,|\, \boldsymbol{x} \,||\, L^{-1}(h) \,|), (\boldsymbol{a}'\boldsymbol{x} + \boldsymbol{c}' \,|\, \boldsymbol{x} \,||\, L^{-1}(h) \,|)]. \tag{4}$$

Using the conditions above, the fuzzy data $(y_j, \boldsymbol{x}_j, h_j), j = 1, \ldots, m$ can be formed a fuzzy regression model by solving the following mathematical programming model:

$$\begin{aligned} \min_{\boldsymbol{a},\boldsymbol{c}} \quad & J = \sum_{j=1,\ldots,m} h_j \boldsymbol{c}' \,|\, \boldsymbol{x}_j \,| \\ s.t. \quad & y_j \geq \boldsymbol{a}'\boldsymbol{x}_j - \boldsymbol{c}' \,|\, \boldsymbol{x}_j \,||\, L^{-1}(h_j) \,|, \\ & y_j \leq \boldsymbol{a}'\boldsymbol{x}_j + \boldsymbol{c}' \,|\, \boldsymbol{x}_j \,||\, L^{-1}(h_j) \,|, \quad j = 1, \ldots, m \\ & \boldsymbol{c} \geq \boldsymbol{0}. \end{aligned} \tag{5}$$

Although the fuzzy regression model can cope well with the fuzzy functional relationships between the input and the output variables in many real-world problems, it has been criticized for the problem of estimation[3], robustness[4,5], fuzzy input variables[6,7], and multicollinearity[8]. In order to overcome these shortcomings, various approaches have been proposed to extend or modify the original method, such as fuzzy least square regression[9-11], fuzzy regression for both fuzzy/crisp input and output variables[12,13], robust fuzzy regression[14,15], and quadratic fuzzy regression[16,17].

In this paper, we focus on the problem of multicollinearity. The cause of multicollinearity is the high correlation among input variables. As we know, the problem of multicollinearity may result in the incorrect fuzzy regression coefficients and the irrational interpretation. In order to deal with the problem of multicollinearity in fuzzy linear regression, Wang and Tsaur[18] proposed the partial index of confidence (*IC*) to select the optimal input subset. Their method can be described as follows. Let Tanaka's fuzzy linear regression with two input variables can be described as $\tilde{y} = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2$. The coefficient of the partial *IC* between $\tilde{y}$ and x_1 given x_2 ($IC_{1|2}$) can be expressed as

$$IC_{1|2} = \frac{SSE(x_2) - SSE(x_1, x_2)}{SSE(x_2)} = \frac{SSE(x_1 \mid x_2)}{SSE(x_2)}. \tag{6}$$

Next, the forward selection algorithms are employed to select an entering input variable according to the biggest partial *IC*.

From Eq. (6), it can be seen that the partial *IC* index is similar to the coefficient of partial determination in traditional statistics. However, as we know, the coefficient of partial determination is used to measure the impact of a new input variable rather than the effect of multicollinearity. In statistics, a better index of measuring the problem of multicollinearity is variance inflation factors (*VIF*). The *VIF* index measures the degree of multicollinearity among input variables and the formulation can be expressed as

$$VIF_j = \frac{1}{1 - R_j^2} \tag{7}$$

where R_j^2 denotes the coefficient of determination when x_j is regressed on the other input variables. If x_j is exactly multicollinear with the other input variables, then $R_j^2 = 1$ and $VIF_j = \infty$. On the other hand, if x_j is completely uncorrelated with the other input variables, then $R_j^2 = 0$ and $VIF_j = 1$.

Many methods have been proposed to guarantee $VIF_j = 1$ in statistics to deal with the problem of multicollinearity and one of these methods is principal component analysis (PCA). Next, in order to consider the fuzzy number in PCA, the autoassociative neural network and the fuzzy centers PCA methods are proposed in Section 3.

3. Fuzzy Principal Component Analysis

In order to extend PCA to consider the situation of interval or fuzzy numbers, several algorithms such as linear programming (LP) method[19], vertices PCA (V-PCA)[20], midpoint range PCA (MR-PCA)[21], and symbolic data analysis (SDA) approach[22] have been proposed. Here, we propose two methods called autoassociative neural network[23] and fuzzy centers PCA to derive the fuzzy principal component scores.

3.1. *Autoassociative neural network method*

The autoassociative neural network method was proposed by[23] to obtain fuzzy principal components based on the viewpoint of machine learning. Let a fuzzy input vector $\tilde{x} = [\tilde{x}_1, \cdots, \tilde{x}_p]$, the *j*th fuzzy principal component and the *k*th fuzzy output can be defined as

$$\tilde{z}_j = \sum_{i=1}^{p} e_{ji}\tilde{x}_i = (z_j^l, z_j^c, z_j^r) \quad \forall j, j = 1, \ldots, q \tag{8}$$

and

$$\tilde{w}_k = \sum_{j=1}^{q} e_{kj}\tilde{z}_j = (w_k^l, w_k^c, w_k^r) \quad \forall k, k = 1, \ldots, p \tag{9}$$

where e_{ji} denotes the $p \times q$ matrix of input-to-hidden weights, e_{kj} is the $q \times p$ matrix of hidden-to-output weights, (z_j^l, z_j^c, z_j^r) is the left value, the center value, and the right value in the *j*th fuzzy principal component, respectively, and (w_k^l, w_k^c, w_k^r) is the left value, the center value, and the right value in the *k*th fuzzy output, respectively. Next, let $\tilde{z}$ be the fuzzy hidden vector and $\tilde{w}$ be the fuzzy output vector. The autoassociative neural network can be considered to minimize the following error function to obtain the fuzzy principal components:

$$e(\tilde{x}, \tilde{w}) = \| \tilde{x} - \tilde{w} \|^2 = \sum_{k=1}^{p} (\tilde{x}_k - \tilde{w}_k)^2 \tag{10}$$

where

$$\sum_{k=1}^{p} (\tilde{x}_k - \tilde{w}_k)^2 = \sum_{k=1}^{p} [(x_k^l - w_k^l)^2 + (x_k^c - w_k^c)^2 + (x_k^r - w_k^r)^2]. \tag{11}$$

The framework of the autoassociative neural network can be depicted as shown in Fig. 1.

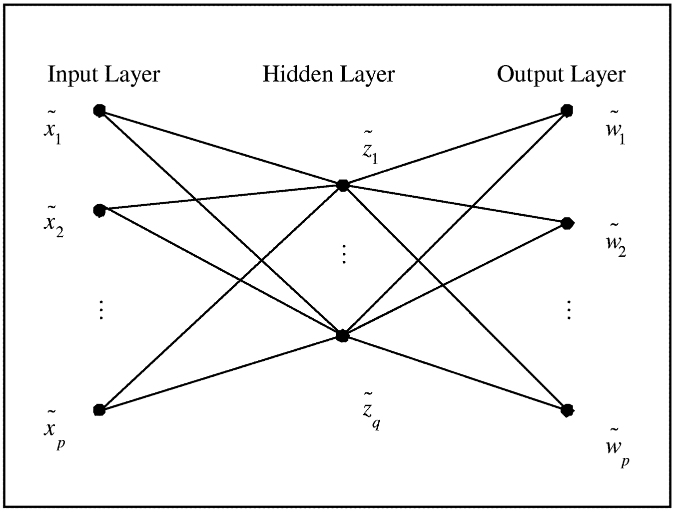

Fig. 1. The structure of the autoassociative neural network.

In addition, we can depict Fig. 2 to describe the concept of fuzzy PCA using the autoassociative neural network. Let a fuzzy input vector $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \tilde{x}_2)'$ with triangular fuzzy number. The hidden unit projects $\tilde{\boldsymbol{x}}$ onto the line L spanned by $\boldsymbol{e}$ so that we can maximize the distance among z_1^l, z_1^c, and z_1^r.

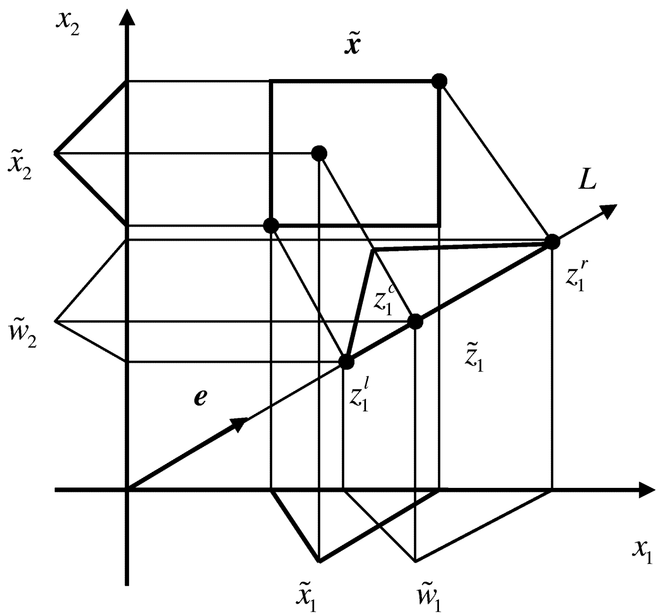

Fig. 2. Geometric views in the autoassociative neural network.

When the eigenvector is obtained by using the autoassociative neural network, the *j*th fuzzy principal component can be expressed as

$$\tilde{z}_j = e_{j1}\tilde{x}_1 + e_{j2}\tilde{x}_2 + \cdots + e_{jp}\tilde{x}_p \,. \tag{12}$$

Next we provide another way to derive the fuzzy principal components using fuzzy centers PCA.

3.2. *Fuzzy centers PCA*

Assume a fuzzy matrix $\tilde{\boldsymbol{X}} = [\boldsymbol{X}^l, \boldsymbol{X}^c, \boldsymbol{X}^r]$, where the center matrix of $\tilde{\boldsymbol{X}}$ can be expressed as

$$\boldsymbol{X}^c = \begin{pmatrix} x_{11}^c & \cdots & x_{1p}^c \\ \vdots & \ddots & \vdots \\ x_{n1}^c & \cdots & x_{np}^c \end{pmatrix} \tag{13}$$

and the left and the right spread matrices of $\tilde{\boldsymbol{X}}$ can be, respectively, expressed as

$$\boldsymbol{X}^a = \begin{pmatrix} x_{11}^a & \cdots & x_{1p}^a \\ \vdots & \ddots & \vdots \\ x_{n1}^a & \cdots & x_{np}^a \end{pmatrix} \tag{14}$$

and

$$\boldsymbol{X}^b = \begin{pmatrix} x_{11}^b & \cdots & x_{1p}^b \\ \vdots & \ddots & \vdots \\ x_{n1}^b & \cdots & x_{np}^b \end{pmatrix} \tag{15}$$

where $x_{ij}^a = x_{ij}^c - x_{ij}^l$ and $x_{ij}^b = x_{ij}^r - x_{ij}^c$, $i = 1, \ldots, n; j = 1, \ldots, p$.

Using the conventional PCA method, we can obtain the center and the spread eigenvalues and eigenvectors. Then the *i*th fuzzy eigenvalue and eigenvector can be obtained as

$$\tilde{\lambda}_i = [\lambda_i^c - \lambda_i^a, \lambda_i^c, \lambda_i^c + \lambda_i^b] \tag{16}$$

and

$$\tilde{\boldsymbol{e}}_i = [\boldsymbol{e}_i^c - \boldsymbol{e}_i^a, \boldsymbol{e}_i^c, \boldsymbol{e}_i^c + \boldsymbol{e}_i^b] \tag{17}$$

Then, the *j*th fuzzy principal component can be obtained as

$$\tilde{z}_j = \tilde{e}_{j1}\tilde{x}_1 + \tilde{e}_{j2}\tilde{x}_2 + \cdots + \tilde{e}_{jp}\tilde{x}_p \tag{18}$$

where

$$\tilde{e}_{ji}\tilde{x}_i = [\min\{e^l_{ji}x^l_i, e^l_{ji}x^r_i, e^r_{ji}x^l_i, e^r_{ji}x^r_i\}, x^c y^c, \max\{e^l_{ji}x^l_i, e^l_{ji}x^r_i, e^r_{ji}x^l_i, e^r_{ji}x^r_i\}]. \tag{19}$$

In other to show that the fuzzy centers PCA method can ensure the fuzzy principal components are uncorrelated with each other, we define the fuzzy *VIF* index to measure whether or not a new fuzzy input variable is related to the others. The fuzzy *VIF* formulation can be expressed as

$$\widetilde{VIF}_j = \frac{\tilde{1}}{\tilde{1} - \tilde{R}^2_j} \tag{20}$$

where $\tilde{1}$ denotes the fuzzy number $(1,1,1)$, and $\tilde{R}^2_j$ denotes the fuzzy coefficient of determination when a fuzzy variable $\tilde{x}_j$ is regressed on the other fuzzy input variables. Similar to traditional statistics, if $\tilde{x}_j$ is exactly multicollinear with the other fuzzy input variables, then $\tilde{R}^2_j = \Phi(\tilde{1})$ and $\widetilde{VIF}_j = \Phi(\tilde{\infty})$ where $\Phi(a)$ denotes the concept that the left and the right values are close to the center value a. On the other hand, if $\tilde{x}_j$ is completely uncorrelated with the other fuzzy input variables, then $\tilde{R}^2_j = \Phi(\tilde{0})$ and $\widetilde{VIF}_j = \Phi(\tilde{1})$. Now, we can show that these fuzzy principal components are uncorrelated so that we can ensure $\widetilde{VIF}_j = \Phi(\tilde{1})$.

Proof:

$$\begin{aligned} \text{Since} \quad Cov(\tilde{z}'_i, \tilde{z}'_j) &= Cov(\tilde{e}'_i\tilde{x}, \tilde{e}'_j\tilde{x}) \\ &= \tilde{e}'_i\tilde{\Sigma}\tilde{e}_j = \tilde{e}'_i(\tilde{\lambda}_j\tilde{e}_j) \\ &= \tilde{\lambda}_j\tilde{e}'_i\tilde{e}_j = \Phi(\tilde{0}) \end{aligned}$$

$$\text{and } \tilde{R}_j = \frac{Cov(\tilde{z}'_i, \tilde{z}'_j)}{\sqrt{Var(\tilde{z}'_i)}\sqrt{Var(\tilde{z}'_j)}} = \Phi(\tilde{0})$$

$$\text{then } \widetilde{VIF}_j = \frac{\tilde{1}}{\tilde{1} - \tilde{R}^2_j} = \frac{\tilde{1}}{\tilde{1} - \Phi(\tilde{0})} = \Phi(\tilde{1}).$$

On the basis of the proof above, it can be seen that the proposed method can provide the fuzzy eigenvectors to ensure these fuzzy principal components are uncorrelated with each other. This characteristic can avoid the problem of multicollinearity. Next, FPCR is proposed to overcome the problem of multicollinearity in fuzzy regression.

4. Fuzzy Principal Component Regression

In Tanaka's fuzzy linear regression, the input variables are limited to crisp values, which are not suitable in this paper. In order to extend Tanaka's method to consider fuzzy input variables, several methods have been proposed[12,13]. In this paper, we adopt the approach of Hong et al.[24,25] which uses shape-preserving operations to cope with fuzzy input variables in FPCR. Note that other methods are also available to execute FPCR.
Based on Zadeh's extension principle[26] with a triangular norm T, the fuzzy number arithmetic operations can be described as follows:

$$(\tilde{M} \oplus \tilde{N})(z) = \sup_{x+y=z} T(\tilde{M}(x), \tilde{N}(y)) \qquad \text{(Fuzzy number addition)}$$

$$(\tilde{M} \otimes \tilde{N})(z) = \sup_{x \cdot y=z} T(\tilde{M}(x), \tilde{N}(y)) \qquad \text{(Fuzzy number multiplication)}$$

Then, the FPCR can be derived as the following equation:

$$\tilde{\boldsymbol{y}} = (\tilde{A}_0 \oplus \tilde{A}_1 \otimes \tilde{z}_1 \oplus \cdots \oplus \tilde{A}_p \otimes \tilde{z}_p) = \tilde{\boldsymbol{A}} \otimes \tilde{\boldsymbol{Z}}, \tag{21}$$

where $\tilde{y}_i = (y_i, e_i)_L$ is fuzzy output variables with center y_i and radius e_i, $\tilde{z}_{ij} = (z_{ij}, \gamma_{ij})_L$, $j = 1, \ldots, p$ is fuzzy input variables with center z_{ij} and radius γ_{ij} and $\tilde{A}_j = (a_j, c_j)_L$, $j = 0, \ldots, p$ is fuzzy regression parameters with center a_j and radius c_j.

Let the input-output relationships be $(\tilde{z}_i, \tilde{y}_i)$, $i = 1, \ldots, n$ and a threshold h; the following equation holds

$$\mu_{\tilde{y}_i}^{-1}(h) \subset \mu_{\tilde{y}_i^*}^{-1}(h) \quad i = 1, \ldots, n. \tag{22}$$

From the relationship above, it can be seen that the index of fuzziness of the possibilistic linear model can be expressed as

$$J = \sum_{i=1}^{n} \max_{1 \le j \le p}(|a_j| \gamma_{ij}, |z_{ij}| c_j). \tag{23}$$

On the basis of concepts above, FPCR can be obtained by solving the following mathematical programming problem:

$$\min \quad J = \sum_{i=1}^{n} \max_{1 \le j \le p}(|a_j| \gamma_{ij}, |z_{ij}| c_j) \tag{24}$$

$$\text{s.t.} \quad \left| y_i - \sum_{j=1}^{p} a_j z_{ij} \right| \le \; |L^{-1}(h)| \max_{1 \le j \le p}(|a_j| \gamma_{ij}, |z_{ij}| c_j) - |L^{-1}(h)| e_i,$$

$$c_j \ge 0,$$

$$\text{for all} \quad i = 1, \ldots, n; \quad j = 1, \ldots, p.$$

In the next section, a numerical example is used to demonstrate the procedures of the proposed method. In addition, ridge fuzzy regression[27,28] are also employed to compare with the proposed method.

5. Numerical Example

The fuzzy data set contains the relationships among the advertised price of a newspaper ($\tilde{x}_1$), the coupon ($\tilde{x}_2$) and the product sales quantity per month ($\tilde{y}$) in 12 branches as shown in Table 1. Note that the symbol *d* denotes the spread in each fuzzy variable so that we can consider the fuzzy data as the memberships with *L-L* function.

Table 1. Fuzzy raw data of the numerical example.

Branch	Sales(y)	d	Newspaper(x_1)	d	Coupons(x_2)	d
1	104	3	47	3	22	3
2	70	3	46	3	20	3
3	90	3	37	2	13	2
4	56	3	24	1	2	1
5	84	3	43	3	17	2
6	120	3	54	4	29	3
7	62	3	35	2	7	1
8	76	3	39	2	14	2
9	66	3	31	2	6	1
10	96	3	49	3	26	3
11	70	3	45	3	19	2
12	114	3	51	4	24	3

In order to understand the relationships among the advertised price, the coupon and the product sales quantity, fuzzy regression is used. Let $L^{-1}(h)$ =0.8. From Eq. (24), we can obtain the fuzzy regression equation as

$$\tilde{y} = (55.1181, 0) + (0, 0.5450)\tilde{x}_1 + (1.6269, 1.0148)\tilde{x}_2 \,.$$

Next, we can calculate the fuzzy correlation coefficient to show the irrationality in the fuzzy regression model above. Note that the formulation of the fuzzy correlation coefficient can refer to the Appendix.

$$\tilde{R}_{x_1, y} = \frac{Cov(\tilde{x}_1, \tilde{y})}{\sqrt{var(\tilde{x}_1)}\sqrt{var(\tilde{y})}} = [0.4973, 0.9581]$$

$$\tilde{R}_{x_2, y} = \frac{Cov(\tilde{x}_2, \tilde{y})}{\sqrt{var(\tilde{x}_2)}\sqrt{var(\tilde{y})}} = [0.5589, 0.9677]$$

$$\tilde{R}_{x_1, x_2} = \frac{Cov(\tilde{x}_1, \tilde{x}_2)}{\sqrt{var(\tilde{x}_1)}\sqrt{var(\tilde{x}_2)}} = [0.7426, 1.0000]$$

It can be seen that since the fuzzy correlation coefficients between $\tilde{y}$ and $\tilde{x}_1$ is positive, it is irrational that the fuzzy regression coefficient of $\tilde{x}_1$ has the zero center value and the negative left value. That is, the high correlation coefficients between $\tilde{x}_1$ and $\tilde{x}_2$ ([0.7426,1.0000]) results in the problem of multicollinearity.

In order to overcome the problem of multicollinearity, the autoassociative neural network method is first used to derive the fuzzy principal component scores as shown in Table 2.

Table 2. Fuzzy principal component scores derived by the autoassociative neural network.

Sales(y)	d	Prin1(z_1)	d	Prin2(z_2)	d
104	3	47.347	3.049	21.238	3.0337
70	3	46.312	3.049	19.261	3.0337
90	3	37.197	2.0327	12.418	2.0225
56	3	24.015	1.0163	1.6495	1.0112
84	3	43.262	3.032	16.316	2.037
120	3	54.464	4.0483	28.109	3.0482
62	3	35.093	2.0157	6.4719	1.0257
76	3	39.213	2.0327	13.386	2.0225
66	3	31.079	2.0157	5.5331	1.0257
96	3	49.415	3.049	25.193	3.0337
70	3	45.296	3.032	18.279	2.037
114	3	51.378	4.0483	23.173	3.0482

From the fuzzy principal component scores above, we can derive the fuzzy regression equation as follows:

$$\tilde{y} = (55.6768, 0) + (0, 0.5413)\tilde{z}_1 + (1.6597, 1.0482)\tilde{z}_2 .$$

Compared with the original fuzzy regression equation, the fuzzy regression equation above still has the irrational situation. The reason is that no matter how the original fuzzy variables are transformed into the fuzzy principal components, the effects of the advertised price of a newspaper or the coupons to the product sales quantity are never negative.

Next we propose fuzzy centers PCA method to derive the fuzzy principal component scores and the results of the fuzzy principal component scores can be presented as shown in Table 3.

Table 3. Fuzzy principal component score using the fuzzy centers PCA.

Sales(y)	d	Prin1(z_1)	d	Prin2(z_2)	d
104	3	7.538	0.811	0.285	0.384
70	3	5.433	0.811	−0.469	0.384
90	3	−5.909	0.131	0.694	0.331
56	3	−22.907	0.131	1.730	0.278
84	3	1.191	0.131	−0.563	0.278
120	3	17.435	1.544	0.505	0.278
62	3	−11.502	0.131	−2.260	0.278
76	3	−3.773	0.131	0.034	0.331
66	3	−15.084	0.131	−0.217	0.278
96	3	11.748	0.811	1.794	0.384
70	3	4.019	0.131	−0.500	0.278
114	3	11.811	1.544	−1.034	0.278

By solving the mathematical programming, the fuzzy regression equation can be obtained as

$$\tilde{y} = (79.8011, 24.8932) + (1.9236, 1.4278)\tilde{z}_1 + (5.8366, 0)\tilde{z}_2$$

From the equation above, it can be seen that the fuzzy regression coefficients of $\tilde{z}_1$ and $\tilde{z}_2$ are both positive. That is, the advertised price of the newspaper and the coupon always provide the positive effects to the product sales quantity. The fuzzy regression model is rational and conforms to our intuition.

Next, we propose the ridge fuzzy regression model[27,28] to compare with the proposed method. In order to determine the appropriate λ value in ridge fuzzy regression, the following formulation is employed:

$$\lambda = p\hat{\sigma}^2 / \| A \|^2 \tag{25}$$

where p denotes the number of the input variables, $\hat{\sigma}^2$ is the least-squares estimates of variance, and A denotes the regression coefficient vector. Note that we use the center value to derive the appropriate λ value in this paper. Using Eq. (25), we can determine the appropriate λ =2.8. Then, the ridge fuzzy regression can be derived as

$$\tilde{y} = (37.0592, 0.3659) + (0.6567, 0)\tilde{x}_1 + (1.2026, 0)\tilde{x}_2 \,.$$

From the results of the equation above, we can conclude that the ridge fuzzy regression model can also provide the rational and correct results. However, it should be highlight that the fuzzy regression coefficients vary with the different λ value. Only the appropriate λ value can determine the correct ridge fuzzy regression model.

6. Discussions

Since the fuzzy correlation coefficients between $\tilde{x}_1$, $\tilde{y}$ and $\tilde{x}_2$, $\tilde{y}$ are positive, it is rational to assume that the fuzzy regression coefficient of $\tilde{x}_1$ and $\tilde{x}_2$ are all positive. However, the fuzzy regression model shows that the fuzzy coefficients of $\tilde{x}_1$ has the negative left value. It can be seen that the irrational results are caused by the high correlation coefficients between $\tilde{x}_1$ and $\tilde{x}_2$.

In this paper, we proposed the autoassociative neural network and the fuzzy centers PCA methods to derive the fuzzy principal component score for dealing with the contravention above. The fuzzy principal components are considered as the new fuzzy input variables in fuzzy regression to ensure the uncorrelated relationship between $\tilde{z}_1$ and $\tilde{z}_2$. However, from the result of the autoassociative neural network method, the contravention still happened. The problem of the autoassociative neural network is that it cannot provide the uncorrelated fuzzy principal components. We can calculate the fuzzy correlation coefficient of the two fuzzy principal components which are derived using the autoassociative neural network as follows:

$$\tilde{R}_{z_1,z_2} = [0.7378, 1.000]. \qquad \text{(autoassociative neural network)}$$

Unlike the autoassociative neural network, the fuzzy centers PCA method can actually overcome the problem of multicollinearity in fuzzy regression. We can calculate the fuzzy correlation coefficients of the two fuzzy principal components which are derived using fuzzy centers PCA to show the uncorrelated situation as follows:

$$\tilde{R}_{z_1,z_2} = [-0.2475, 0.2499]. \qquad \text{(fuzzy centers PCA)}$$

In addition, from the results of FPCA, it can be seen that the fuzzy coefficients of $\tilde{z}_1$ and $\tilde{z}_2$ are all positive. It shows that both $\tilde{x}_1$ and $\tilde{x}_2$ can provide the positive effects to the product sales quantity. Therefore, the results of FPCA are rational and intuitive. From the results of ridge fuzzy regression, we can conclude that the ridge fuzzy regression model can also overcome the problem of multicollinearity and can be another alternative. However, it should be highlight that the ridge fuzzy regression may spend more time to determine the appropriate λ value.

7. Conclusions

Although fuzzy regression has been widely employed to cope with many practical problems, the basic problem of multicollinearity has received little attention. In this paper, two kinds of the fuzzy PCA techniques including the autoassociative neural network and the fuzzy centers PCA methods are used to transform the original fuzzy variables into the fuzzy principal components. Then these fuzzy principal components are considered as the new fuzzy input variables to ensure the uncorrelated relationship among input variables. On the basis of the numerical results, we can conclude that fuzzy centers PCA can provide the more accurate fuzzy principal component scores than the autoassociative neural network method and can overcome the problem of multicollinearity using FPCA. In addition, the ridge fuzzy regression can be another alternative to deal with the problem of multicollinearity.

Appendix

In this paper, the formulation of fuzzy correlation analysis can be described as follows. Let the fuzzy correlation coefficient can be expressed using a interval number $\tilde{R}_{x,y} = [R^-_{x,y}, R^+_{x,y}]$. The left and the right correlation coefficient can be described as

$$R^-_{x,y} = \min\{R_{x,y} \mid x_i \in \tilde{x}_i, y_i \in \tilde{y}_i\} \tag{A.1}$$

and

$$R^+_{x,y} = \max\{R_{x,y} \mid x_i \in \tilde{x}_i, y_i \in \tilde{y}_i\}. \tag{A.2}$$

By solving the following mathematical programming, the fuzzy correlation coefficient between x and y can be obtained.

$$\min/\max \quad R_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{A.3}$$

$$\begin{aligned} s.t. \quad & x_i^l \le x_i \le x_i^r, \\ & y_i^l \le y_i \le y_i^r, \\ & \overline{x} = \sum_{i=1}^{n} x_i \Big/ n, \\ & \overline{y} = \sum_{i=1}^{n} y_i \Big/ n \end{aligned}$$

where x_i^l and x_i^r denote the left and the right values of the ith sample in variable x, and y_i^l and y_i^r are the left and the right values of the ith sample in variable y.

References

1. H. Tanaka, S. Uejima, and K. Asai, "Linear Regression Analysis with Fuzzy Model", *IEEE Transaction on System, Man and Cybernetics* **12** (1982), 903-907.
2. H. Tanaka, "Fuzzy Data Analysis by Possibilistic Linear System", *Fuzzy Sets and Systems* **24** (1987), 363-375.
3. C. Kao and C. L. Chyu, "Least-Squares Estimates in Fuzzy Regression Analysis", *European Journal of Operational Research* **148** (2003), 426-435.
4. Y. S. Chen, "Outlier Detection and Interval Modification in Fuzzy Regression", *Fuzzy Sets and Systems* **119** (2001), 259-272.
5. G. Peters, "Fuzzy Linear Regression with Fuzzy Intervals", *Fuzzy Sets and Systems* **63** (1994), 15-55.
6. D. T. Redden and W. H. Woodall, "Further Examination of Fuzzy Linear Regression", *Fuzzy Sets and Systems* **79** (1996), 203-211.
7. M. Sakawa and H. Yano, "Multiobjective Fuzzy Linear Regression Analysis for Fuzzy Input-Output Data", *Fuzzy Sets and Systems* **47** (1992), 173-181.
8. K. J. Kim, H. Modkowitz, and M. Koksalan, "Fuzzy Versus Statistical Linear Regression", *European Journal of Operational Research* **92** (1996), 417-434.
9. P. Diamond and R. Körner, "Extended Fuzzy Linear Models and Least Squares Estimates", *Computers and Mathematics with Applications* **33** (1997), 15-32.
10. P. T. Chang and E. S. Lee, "A Generalized Fuzzy Weighted Least-Squares Regression", *Fuzzy Sets and Systems* **82** (1996), 289-298.
11. M. S. Yang and H. H. Liu, "Fuzzy Least-Squares Algorithms for Interactive Fuzzy Linear Regression Models", *Fuzzy Sets and Systems* **135** (2003), 305-316.

12. P. D'Urso, "Linear Regression Analysis for Fuzzy/Crisp Input and Fuzzy/Crisp Output Data", *Computational Statistics & Data Analysis* **42** (2003), 47-72.
13. H. C. Wu, "Linear Regression Analysis for Fuzzy Input and Output Data Using the Extension Principal", *Computers and Mathematics with Applications* **45** (2003), 1849-1859.
14. J. Watada and Y. Yabuuchi, "Fuzzy Robust Regression Analysis, Proceeding of 3rd IEEE International Conforance on Fuzzy Systems", Florida, (1994), 1370-1376.
15. J. Watada and Y. Yabuuchi, "Fuzzy Robust Regression Analysis based on A Hyperelliptic Function", Proceedings of the International Joint Conference of the 4th IEEE International Conference on Fuzzy Systems and the 2nd International Fuzzy Engineering Symposium, Yokohama, (1995), 1841-1848.
16. Y. S. Chen, "Fuzzy Ranking and Quadratic Fuzzy Regression", *Computers and Mathematics with Applications* **38** (1999), 265-279.
17. H. Tanaka and H. Lee, "Interval Regression Analysis by Quadratic Programming Approach", *IEEE Transactions on Fuzzy Systems* **5** (1998), 473-481.
18. H. F. Wang, and R. C. Tsaur, "Insight of a Fuzzy Regression Model", *Fuzzy Sets and Systems* **112** (2000), 355-369.
19. Y. Yabuuchi, J. Watada, and Y. Nakamori, "Fuzzy Principal Component Analysis for Fuzzy Data", The 6th IEEE International Conference on Fuzzy Systems, Barcelona, (1997), 83-92.
20. P. Cazes, A. Chouakria, E. Diday, and Y. Schektman, "Extension De L'analyse En Composantes Principales a des Donnees de Type Intervalle", *Revue de Statsitique Appliquee* **14** (1997), 5-24.
21. F. Palumbo and C. N. Lauro, "A PCA for Interval Valued Data Based on Midpoints and Radii, in: New developments in Psychometrics", H. Yanai, A. Okada, K. Shigemasu, Y. Kano, and J. Meulman (eds.), Psychometric Society, Tokyo: Springer-Verlag, 2003.
22. H. H. Bock and H. Mannila, "Random Projection in Dimensionality Reduction: Applications to Image and Text Data", Proceeding of 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2000), 245-250.
23. T. Denoeux and M. Masson, "Principal Component Analysis of Fuzzy Data Using Autoassociative Neural Networks", *IEEE Transactions on Fuzzy Systems* **12** (2004), 336-349.
24. D. H. Hong, S. Lee, and H. Y. Do, "Fuzzy Linear Regression Analysis for Fuzzy Input-Output Data Using Shape-Preserving Operations", *Fuzzy Sets and Systems* **122** (2001), 513-526.
25. D. H. Hong, S. Lee, and H. Y. Do, "Fuzzy Least-Squares Linear Regression Analysis Using Shape Preserving Operations", *Information Sciences* **138** (2001), 185-193.
26. L. A. Zadeh, "Fuzzy Sets", *Information Control* **8** (1965), 338-353.
27. D. H. Hong and C. Hwang, "Ridge Regression Procedures for Fuzzy Models Using Triangular Fuzzy Numbers", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **12** (2004), 145-159.
28. D. H. Hong, C. Hwang, and C. Ahn, "Ridge Estimation for Regression Models with Crisp Inputs and Gaussian Fuzzy Output", *Fuzzy Sets and Systems* **142** (2004), 307-319.