The Korean Communications in Statistics **Vol. 14** No. 1, 2007, pp. 183–193

Separate Fuzzy Regression with Fuzzy Input and Output*

Seunghoe Choi[1)]

Abstract

This paper shows that a response function for the center of fuzzy output may not be the same as that for the spread in a fuzzy linear regression model and then suggests a separate fuzzy regression model makes a distinction between response functions of the center and the spread of fuzzy output. Also we use a least squares method to estimate the separate fuzzy regression model and compare an accuracy of proposed model with another fuzzy regression model developed by Diamond (1988) and Kao and Chyu (2003).

Keywords: Separate fuzzy regression model; least squares method; accuracy.

1. 서론

사회과학이나 자연과학에서 발생하는 현상을 지배하는 종속변수와 독립변수 사이의 관계를 설명하고, 두 변수에 대한 자료를 이용하여 제안된 모형을 추정하고 분석하는 방법을 회귀분석(regression analysis)이라 한다. 회귀분석에서는 독립변수에 영향을 받는 종속변수를 분석하기 위하여 회귀모형(regression model)

$$y = f(x, \theta) + \epsilon$$

을 이용한다. 여기서 x를 독립변수, y를 종속변수, θ를 회귀모수, ϵ를 통계적 오차라 하고, $f(x,\theta)$를 반응함수라 한다. 회귀분석에서는 정확하게 측정할 수 있는 종속변수와 독립변수 사이의 통계적인 관계를 분석한다. 따라서 종속변수나 독립변수를 정확하게 실수(real number)로 표현할 수 없거나 혹은 두 변수가 불확실성을 포함하고 있을 경우에는 새로운 회귀모형을 연구하여야 한다.

Zadeh (1965)는 애매하고 모호한 문장이나 정보들을 처리하기 위하여 퍼지이론을 제안하고 연구하였으며, Tanaka 등 (1980, 1982)은 퍼지이론을 회귀모형에 도입하여 퍼지회귀모형(fuzzy regression model)

$$Y_i = A_0 + A_1 x_i \tag{1.1}$$

---

* This research has been supported by 2006 Hankuk Aviation Faculty Research Grant.

1) Associate Professor, Department of General Studies, Hankuk Aviation University, Koyang, Kyungkido, 412-791, Korea.
E-mail : shchoi@hau.ac.kr

을 처음으로 소개하였다. 여기서 독립변수 x_i는 실수이고 회귀계수 $A_i = (a_i, l_{a_i}, r_{a_i})_{LR}$와 종속변수 $Y_i = (y_i, l_{y_i}, r_{y_i})_{LR}$는 LR-퍼지수이다.

LR-퍼지수 $A = (a, l_a, r_a)_{LR}$의 소속함수(membership function)는

$$\mu_A(x) = \begin{cases} L_A\left(\frac{a-x}{l_a}\right) & \text{for} \quad x \le a \\ R_A\left(\frac{x-a}{r_a}\right) & \text{for} \quad x \ge a \end{cases}$$

이고, 단조증가(감소)함수 $L_A(R_A)$는

$$L_A(0) = R_A(0) = 1 \quad \text{와} \quad L_A(1) = R_A(1) = 0$$

을 만족한다. 이때 a을 LR-퍼지수 A의 중심이라 하고, l_a와 r_a을 각각 LR-퍼지수 A의 왼쪽과 오른쪽의 폭이라 한다. 만약 $L_A(x) = R_A(x) = 1 - x$이면 LR-퍼지수 A를 삼각퍼지수라고 부르고 $(a, l_a, r_a)_T$와 같이 표시한다.

퍼지회귀모형 (1,1)에서는 독립변수 x_i의 값이 증가하면 종속변수 Y_i의 왼쪽과 오른쪽 폭이 증가하여 퍼지회귀모형에 대한 불확실성이 증가한다. 이와 같은 문제를 해결하기 위하여 Chang과 Ayyub (2001)은 회귀계수가 실수이고 독립변수가 LR-퍼지수 $X_i = (x_i, l_{x_i}, r_{x_i})_{LR}$인 다음과 같은 퍼지회귀모형을 소개하였다.

$$Y_i = a_0 + a_1 X_i. \tag{1.2}$$

퍼지회귀모형 (1.1)과 (1.2)를 추정하기 위하여 여러 가지 방법이 제시되었다. Kao와 Chyu (2003), Savic과 Pedryzc (1991), 그리고 Tanaka 등(1989)은 추정된 퍼지수의 폭을 최소화하는 수치해석적인 방법을 발표하였다. 또한 Diamond와 Korner (1997), Chang과 Ayyub (2001), 그리고 Kim과 Bishu (1998)는 관찰된 퍼지수와 예측된 퍼지수의 차를 최소화하는 통계적인 방법을 제안하였다.

퍼지회귀모형 (1.2)의 회귀계수는 퍼지수가 아닌 실수이므로 독립변수와 종속변수에 대한 소속함수가 동일한 형태이다. 그러나 Kao와 Chyu (2003)가 언급한 것처럼 퍼지회귀모형에서는 종속변수와 독립변수의 소속함수 형태가 일치하지 않는 경우가 존재한다. 또한 퍼지회귀모형 (1.2)는 퍼지수인 종속변수의 폭에 대한 반응함수가 중심에 대한 반응함수와 일치하는 경우에만 적용할 수 있다. 그러나 종속변수의 중심에 대한 회귀모형이 폭에 대한 회귀모형과 반드시 일치하지 않는다(Chen (2001)을 참조). LR-퍼지수인 종속변수의 중심과 폭에 대한 반응함수가 일치하도록 퍼지회귀모형을 설정하는 것은 추정된 퍼지회귀모형의 정확성을 제약할 수 있으므로 새로운 퍼지회귀모형을 연구할 필요가 있다.

본 논문에서는 종속변수와 독립변수가 LR-퍼지수이고, 회귀계수가 실수인 퍼지회귀모형의 중심과 폭에 대한 반응함수가 일치하지 않음을 확인하고 종속변수의 중심과 폭에 대한 회귀모형을 구분한 독립퍼지회귀모형을 소개한다. 최소제곱법을 이용하여 독립퍼지회귀모형을 추정하고 예제를 이용하여 본 논문에서 제시된 독립퍼지회귀모형의 정확성을 Daimond (1988)와 Kao와 Chyu (2003)가 추정한 퍼지회귀모형과 비교한다.

2. 독립퍼지회귀모형

본 절에서는 퍼지수의 중심과 폭을 이용한 독립퍼지회귀모형을 소개하고, 최소제곱법을 이용하여 제안된 독립퍼지회귀모형을 추정하는 방법을 소개한다.

퍼지수 A의 α-수준집합은 소속함수의 값이 α보다 큰 정의구역의 값으로 정의된다. LR-퍼지수 Y_i의 α-수준집합 $Y_i(\alpha) \equiv \{x|\mu_{Y_i}(x) \geq \alpha\}$는 중심이 y_i이고, 왼쪽과 오른쪽 폭이 각각 $l_{y_i} L_{Y_i}^{-1}(\alpha)$와 $r_{y_i} R_{Y_i}^{-1}(\alpha)$인 폐구간 $[y_i - l_{y_i} L_{Y_i}^{-1}(\alpha), y_i + r_{y_i} R_{Y_i}^{-1}(\alpha)]$이므로

$$Y_i(\alpha) \cong (y_i, l_{y_i} L_{Y_i}^{-1}(x), r_{y_i} R_{Y_i}^{-1}(x))$$

와 같이 표현할 수 있다. 또한 LR-퍼지수 Y_i의 1-수준집합과 0-수준집합은 각각

$$Y_i(1) \cong (y_i, 0, 0) \quad \text{와} \quad Y_i(0) \cong (y_i, l_{y_i}, r_{y_i})$$

이다. 따라서 퍼지수의 α-수준집합을 이용하면 퍼지회귀모형 (1.2)는

$$\left(y_i, l_{y_i} L_{Y_i}^{-1}(\alpha), r_{y_i} R_{Y_i}^{-1}(\alpha)\right) = \left(a_0 + a_1 x_i, a_1 l_{x_i} L_{X_i}^{-1}(\alpha), a_1 r_{x_i} R_{X_i}^{-1}(\alpha)\right)$$

와 같이 표현할 수 있다. 이것은 추정된 퍼지회귀모형의 α-수준집합에 대한 중심과 왼쪽 끝점, 그리고 오른쪽 끝점의 기울기가 일치한다는 것을 의미한다. 그러나 퍼지수의 α-수준집합을 구성하는 중심과 왼쪽 끝점, 그리고 오른쪽 끝점에 대한 회귀계수가 반드시 일치하는 것은 아니다. 다음 예제는 퍼지회귀모형에 대한 α-수준집합의 중심과 양 끝점에 대한 추세가 반드시 일치하지 않음을 보여준다.

예제 2.1 Freund (1992)는 시험성적(Y_i)과 시험공부시간(X_i)에 대한 관계를 조사하였다. 표 2.1은 Freund (1992)가 제시한 자료를 LR-퍼지수와 삼각퍼지수로 변형한 자료이다. LR-퍼지수인 Y_i의 소속함수는

$$L_{Y_i}(x) = 1 - x^2$$

와

$$R_{Y_i}(x) = (1 + x)^{\frac{1}{2}}$$

이다.

표 2.1에 제시된 중심(x_i, y_i) 에 대한 회귀분석의 결과는

$$y_i = 21.69 + 3.47x_i$$

이고 오른쪽(왼쪽) 폭에 대한 결과는

$$l_{y_i} = 1.32 + 0.67 l_{x_i} (r_{y_i} = 1.57 + 0.67 r_{x_i})$$

이다. 이것은 종속변수 Y_i의 중심과 폭에 대한 기울기가 반드시 일치하지 않는 것을 보여준다. 종속변수에 대한 중심과 폭의 기울기가 일치하지 않는 경우 퍼지회귀모형 (1.2)을 사용하여 두 변수 사이의 관계를 설명하는 것은 퍼지회귀모형의 정확성을 감소시킬 수 있다. 따라서 이와 같은 경우에는 새로운 퍼지회귀모형을 연구할 필요가 있다.

표 2.1: 퍼지화된 Freund 의 자료

$X_i = (x_i, l_{x_i}, r_{x_i})_T$	$Y_i = (y_i, l_{y_i}, r_{y_i})_{LR}$
$(4, 2.1, 1.8)_T$	$(31, 3.03, 1.9)_{LR}$
$(9, 1.1, 2.8)_T$	$(58, 2.43, 1.5)_{LR}$
$(10, 2.9, 1)_T$	$(65, 1.27, 1.8)_{LR}$
$(14, 2.8, 2.2)_T$	$(73, 3.94, 2.1)_{LR}$
$(4, 2.8, 1.2)_T$	$(37, 3.64, 1.9)_{LR}$
$(7, 1.1, 2.3)_T$	$(44, 1.03, 3.45)_{LR}$
$(12, 2.7, 1.7)_T$	$(60, 3.31, 1.95)_{LR}$
$(22, 2.9, 2.5)_T$	$(91, 3.37, 0.25)_{LR}$
$(1, 2, 1.7)_T$	$(21, 4, 0.25)_{LR}$
$(17, 2, 2.2)_T$	$(84, 2.2, 5.3)_{LR}$

종속변수와 독립변수가 퍼지수인 퍼지회귀모형을 추정하기 위하여 본 논문에서는 퍼지회귀모형을 일반화한

$$Y_i = F(X_i, \theta) + E_i \tag{2.1}$$

와 같은 독립퍼지회귀모형(separate fuzzy regression model)을 생각한다. 여기서 $F(X_i, \theta) = (f(X_i, C), g(X_i, L), h(X_i, R))_{LR}$이고 $E_i \equiv (e_i, l_{e_i}, r_{e_i})_{LR}$는 퍼지오차이다. 또한 함수$f$와 g, 그리고 h는 실가함수이고 $C = (c_0, c_1), L = (l_0, l_1)$과 $R = (r_0, r_1)$에서 c_i, l_i 그리고 $r_i (i = 0, 1)$는 각각 실수이다.

함수 f와 g, 그리고 h가 모두 선형이고 $l_0 = r_0 = 0$와 $c_1 = l_1 = r_1 = a_1$을 만족하면 독립퍼지회귀모형 (2.1)은

$$Y_i = a_0 + a_1 X_i + E_i$$

와 같이 변형되고, 또한 $l_0 \neq 0$이거나 혹은 $r_0 \neq 0$이고 $c_1 = l_1 = r_1 = a_1$이면 독립퍼지회귀모형 (2.1)은

$$Y_i = (c_0, l_0, r_0)_{LR} + a_1 X_i + E_i$$

와 일치한다. 즉, 퍼지회귀모형 (1.2)는 독립퍼지회귀모형 (2.1)의 특별한 경우이다. 본 논문에서는 함수 f와 g 그리고 h가 모두 선형인 경우만 생각한다.

독립퍼지회귀모형 (2.1)의 회귀계수를 추정하는 방법에 대하여 알아보자. 추정된 퍼지수 $\widehat{Y}_i$의 소속함수는 관찰된 퍼지수 Y_i의 소속함수와 동일한 형태이므로 추정된 퍼지수 $\widehat{Y}_i$의 중심과 폭을 추정하면 된다. 두 퍼지수 Y_i와 $F(X_i, \theta)$의 0–수준집합은 각각의 중심과 폭으로 구성되어 있다. 따라서 두 집합 $Y_i(0)$와 $F_0(X_i, \theta)$의 차와 동일한 퍼지오차 E_i의 0–수준집합 $E_i(0)$의 크기를 최소화하는 방법으로 독립퍼지회귀모형을 추정할 수 있다. 이를 위해 Kim과 Bishu (1988)가 제시한 최소제곱법과 퍼지수의 중심과 폭을 분리하여 추정한 Savic과 Pedryzc (1991)의 2단계 방법을 이용하자.

1단계. 독립퍼지회귀모형의 중심에 대한 최소제곱추정량 $(\widehat{c}_0, \widehat{c}_1)$는

$$\sum_{i=1}^{n}(y_i - c_0 - c_1 x_i)^2$$

을 최소화하는 방법으로 구하고, 최소제곱법을 이용하여 추정한 회귀선을 $\widehat{y}_i = \widehat{c}_0 + \widehat{c}_1 x_i$라 하자.

2단계. 독립퍼지회귀모형의 오른쪽과 왼쪽 폭에 대한 최소제곱추정량 $(\widehat{l}_0, \widehat{l}_1, \widehat{r}_0, \widehat{r}_1)$는

$$\sum_{i=1}^{n}\left\{(l_{y_i} - l_0 - l_1 l_{x_i})^2 + (r_{y_i} - r_0 - r_1 r_{x_i})^2\right\}$$

을 최소화하는 방법으로 구하고, 주어진 퍼지회귀모형의 양 폭에 대한 최소제곱추정치는

$$l_{\widehat{y}_i} = -\min\{0, k_i\} + \widehat{l}_0 + \widehat{l}_1 l_{x_i} \quad \text{와} \quad r_{\widehat{y}_i} = \max\{0, k_i\} + \widehat{r}_0 + \widehat{r}_1 r_{x_i}$$

와 같이 정의하자. 여기서 $k_i = y_i - \widehat{y}_i$이다.

표본 $\{(l_{y_i}, l_{x_i}) | i = 1, \ldots, n\}$와 $\{(r_{y_i}, r_{x_i}) | i = 1, \ldots, n\}$을 이용하여 추정한 회귀모형은 1단계에서 추정된 회귀선 $\widehat{y}_i = \widehat{c}_0 + \widehat{c}_1 x_i$에 영향을 받지 않으므로 추정된 퍼지회귀모형의 폭은 추정된 중심과의 거리와 독립적이다. 또한 관찰된 퍼지수중 중심이 비정상적인 자료가 존재하면 추정된 퍼지수 $\widehat{Y}_i$의 오른쪽(왼쪽) 끝점이 추정된 퍼지수의 중심보다 왼쪽(오른쪽)에 위치할 수 있다. 이와 같은 문제를 해결하고 추정된 독립퍼지회귀모형의 정확성을 높이기 위하여 퍼지수의 왼쪽(오른쪽) 폭 $l_{\widehat{y}_i}(r_{\widehat{y}_i})$을 $\min\{0, k_i\}(\max\{0, k_i\})$을 이용하여 추정하였다.

추정된 퍼지수 $\widehat{Y}_i$에 대한 중심과 오른쪽 폭, 그리고 왼쪽 폭에 대한 기울기가 일치하여 $\widehat{c}_1 = \widehat{l}_1 = \widehat{r}_1 = \widehat{s}_1$을 만족하면 추정된 퍼지회귀모형은

$$\widehat{Y}_i = \widehat{S}_0 + \widehat{s}_1(x_i, l_{x_i}, r_{x_i})_{LR} = \widehat{S}_0 + \widehat{s}_1 X_i$$

이다. 여기서 $\widehat{S}_0 = (\widehat{a}_0 - \min\{0, k_i\} + \widehat{c}_0, \max\{0, k_i\} + \widehat{d}_0)$이다.

다음 예제는 표 2.1에서 제시된 퍼지자료에 대한 퍼지회귀모형을 추정한다.

예제 2.2 삼각퍼지수인 독립변수와 LR-퍼지수인 종속변수 사이의 독립퍼지회귀모형을 추정하자. 비록 종속변수와 독립변수의 소속함수가 동일한 형태는 아니지만 본 논문에서 제시된 독립퍼지회귀모형을 이용하여 두 변수 사이의 퍼지관계를 추정할 수 있다.

표 2.1에 주어진 자료의 중심에 대한 최소제곱추정량 $(\widehat{c}_0, \widehat{c}_1)$은 다음과 같다.

$$\widehat{y}_i = 21.69 + 3.47 x_i$$

표 2.2: Freund의 자료에 대한 추정량

$X_i = (x_i, l_{x_i}, r_{x_i})_T$	$\widehat{Y}_i = (\widehat{y}_i, l_{\widehat{y}_i}, r_{\widehat{y}_i})_{LR}$
$(4, 2.1, 1.8)_T$	$(35.57, 4.59, 2)_{LR}$
$(9, 1.1, 2.8)_T$	$(52.92, 1.03, 9.02)_{LR}$
$(10, 2.9, 1)_T$	$(56.39, 1.63, 13.29)_{LR}$
$(14, 2.8, 2.2)_T$	$(70.27, 1.6, 5.74)_{LR}$
$(4, 2.8, 1.2)_T$	$(35.57, 1.6, 3.76)_{LR}$
$(7, 1.1, 2.3)_T$	$(45.98, 2.43, 2.13)_{LR}$
$(12, 2.7, 1.7)_T$	$(63.33, 3.92, 1.98)_{LR}$
$(22, 2.9, 2.5)_T$	$(98.03, 6.6, 2.18)_{LR}$
$(1, 2, 1.7)_T$	$(25.16, 4.27, 1.98)_{LR}$
$(17, 2, 2.2)_T$	$(80.68, 1.33, 6.53)_{LR}$

중심에 대한 회귀식을 이용하여 추정된 퍼지수의 양쪽 폭에 대한 결과는 다음과 같다.

$$\begin{aligned} l_{\widehat{y}_i} &= -\min\{0, k_i\} + 1.32 + 0.67 l_{x_i}, \\ r_{\widehat{y}_i} &= l \max\{0, k_i\} + 1.57 + 0.24 r_{x_i}. \end{aligned}$$

여기서 $k_i = y_i - 21.69 - 3.47x_i$이다.

표 2.2는 독립퍼지회귀모형을 이용하여 추정된 삼각퍼지수 X_i와 LR–퍼지수 $\widehat{Y}_i$를 보여준다.

3. 퍼지회귀모형의 정확성

본 절에서는 2절에서 소개한 독립퍼지회귀모형에 대한 최소제곱추정량를 이용하여 추정된 퍼지회귀모형의 정확성을 조사하기 위하여 Kim과 Bishu (1998)가 제시한 척도를 사용하고, 최소제곱법과 퍼지회귀모형 (1.2)을 이용한 Daimond (1988)와 Kao와 Chyu (2003)의 결과와 본 논문에서 추정된 독립퍼지회귀모형에 대한 정확성을 비교한다.

추정된 퍼지회귀모형의 정확성을 비교하기 위하여 Kim과 Bishu (1998)는 소속함수에 대한 적분을 사용하였다. Kim과 Bishu (1998)는 관찰된 퍼지수 Y_i와 추정된 퍼지수 $\widehat{Y}_i$의 차를 두 소속함수의 차에 대한 적분으로 다음과 같이 정의하였다.

$$D(Y_i, \widehat{Y}_i) = \int_{-\infty}^{\infty} |\mu_{Y_i}(x) - \mu_{\widehat{Y}_i}(x)| dx.$$

위 식으로부터 추정된 퍼지회귀모형에 대한 정확성을 조사하기 위하여 척도

$$M(Y, \widehat{Y}) = \sum_{i=1}^{n} m(Y_i, \widehat{Y}_i) \tag{3.1}$$

를 정의할 수 있다. 여기서

$$m\left(Y_i, \widehat{Y}_i\right) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} |\mu_{Y_i}(x) - \mu_{\widehat{Y}_i}(x)| dx & \text{for} \quad \displaystyle\int_{-\infty}^{\infty} \mu_{Y_i}(x) dx < 1 \\ \dfrac{\displaystyle\int_{-\infty}^{\infty} |\mu_{Y_i}(x) - \mu_{\widehat{Y}_i}(x)| dx}{\displaystyle\int_{-\infty}^{\infty} \mu_{Y_i}(x) dx} & \text{for} \quad \displaystyle\int_{-\infty}^{\infty} \mu_{Y_i}(x) dx > 1 \end{cases}$$

이다. 척도 $M(Y, \widehat{Y})$는 관찰된 퍼지수와 추정된 퍼지수의 소속함수와 두 퍼지수 사이의 거리에 영향을 받는다. 척도 (3.1)의 값이 0에 가까울수록 추정된 퍼지회귀모형의 정확성은 높다고 할 수 있다. 다음 예제는 추정된 퍼지회귀모형의 정확성을 척도 (3.1)를 사용하여 설명한다.

예제 3.1 Diamond (1988)는 학생성적(Y_i)과 가계수입(X_i)에 대한 관계를 조사하기 위하여 표 3.1에 제시된 삼각퍼지수를 이용하여 퍼지회귀모형을 추정하였다. 표 3.1에 주어진 자료에 대한 퍼지회귀모형

$$Y_i^1 = a + bX_i \quad \text{와} \quad Y_i^2 = A + bX_i$$

을 추정하기 위하여 Diamond (1988)는 두 퍼지수 $Y_i^j (j = 1, 2)$와 $a + bX_i (A + bX_i)$사이의 거리 제곱을 최소화하는 방법을 사용하였다. Daimond (1988)가 추정한 학생성적과 가계수입에 대한 퍼지회귀모형은 아래와 같다.

$$\begin{aligned} \widehat{Y}_i^1 &= 1.052 + 0.147X_i, \\ \widehat{Y}_i^2 &= (1.201, 0.18, 0.18)_T + 0.136X_i. \end{aligned}$$

Diamond (1988)의 결과는 학생성적에 대한 오른쪽과 왼쪽 폭, 그리고 중심에 대한 기울기가 일치하는 것을 의미한다.

표 3.1: Diamond의 자료

$X_i = (x_i, l_{x_i}, r_{x_i})_T$	$Y_i = (y_i, l_{y_i}, r_{y_i})_T$
$(21, 4.2, 2.1)_T$	$(4, 0.6, 0.8)_T$
$(15, 2.25, 2.25)_T$	$(3, 0.3, 0.3)_T$
$(15, 1.5, 2.25)_T$	$(3.5, 0.35, 0.35)_T$
$(9, 1.35, 1.35)_T$	$(2, 0.4, 0.4)_T$
$(12, 1.2, 1.2)_T$	$(3, 0.3, 0.45)_T$
$(18, 3.6, 1.8)_T$	$(3.5, 0.35, 0.7)_T$
$(6, 0.6, 1.2)_T$	$(2.5, 0.25, 0.38)_T$
$(12, 1.8, 2.4)_T$	$(2.5, 0.5, 0.5)_T$

표 3.2: 추정된 퍼지회귀모형의 정확성

y_i	정확성 척도		
	$m(Y_i, \widehat{Y}_i^1)$	$m(Y_i, \widehat{Y}_i^2)$	$m(Y_i, \widehat{Y}_i)$
4	0.484	0.283	0.475
3	1.377	1.37	1.059
3.5	0.972	1.099	1.042
2	1.286	1.545	1.437
3	0.89	0.873	0.783
3.5	0.359	0.564	0.24
2.5	1.413	1.887	1.442
2.5	1.094	1.173	0.927
Total	7.875	8.794	7.405

본 논문에서 제시된 독립퍼지회귀모형의 결과는 학생성적에 대한 폭과 중심의 기울기는 일치하지 않음을 확인할 수 있다. 표 3.1에 주어진 자료의 중심에 대한 최소제곱추정량 $(\widehat{c}_0, \widehat{c}_1)$은 다음과 같다.

$$y_i = 1.375 + 0.12x_i.$$

추정된 퍼지수에 대한 양쪽 폭에 대한 결과는 다음과 같다.

$$\begin{aligned} l_{\widehat{y}_i} &= -\min\{0, k_i\} + 0.259 + 0.06 l_{x_i}, \\ r_{\widehat{y}_i} &= \max\{0, k_i\} + 0.401 + 0.046 r_{x_i}. \end{aligned}$$

여기서 $k_i = y_i - 1.375 - 30.12x_i$이다.

표 3.1에 주어진 자료를 이용하여 추정한 퍼지회귀모형에 대한 정확성을 비교하기 위해 척도 (3.1)를 이용하자. 표 3.2는 척도 (3.1)를 이용하여 비교한 독립퍼지회귀모형과 Diamond (1988)의 퍼지회귀모형에 대한 정확성을 보여준다.

Diamond (1988)가 최소제곱추정량을 이용하여 추정한 퍼지회귀모형 (1.2)의 정확성은 각각 7.875와 8.794이고, 최소제곱추정량을 이용하여 추정한 독립퍼지회귀모형의 정확성은 7.405이다. 이 결과는 퍼지수인 종속변수의 중심과 폭에 대한 반응함수를 독립적으로 추정하는 것이 더 정확할 수 있음을 보여준다.

예제 3.2 표 3.3는 Sakawa와 Yano (1992)가 독립변수와 종속변수가 삼각퍼지수인 회귀모형을 설명하기 위해 제시한 자료이다. 표 3.3를 이용하여 Sakawa와 Yano (1992)가 추정한 퍼지회귀모형은

$$\widehat{Y}_i^s = (3.201, 0.17, 0.17)_T + (0.579, 0.081, 0.08)_T X_i$$

와 같다. Kao와 Chyu (2003)는 표 3.3에 제시된 자료에 대한 회귀모형을 추정하기 위하

표 3.3: Sakawa와 Yano의 자료

$X_i = (x_i, l_{x_i}, r_{x_i})_T$	$Y_i = (y_i, l_{y_i}, r_{y_i})_T$
$(2.0, 0.5, 0.5)_T$	$(4, 0.5, 0.5)_T$
$(3.5, 0.5, 0.5)_T$	$(5.5, 0.5, 0.5)_T$
$(5.5, 1.0, 1.0)_T$	$(7.5, 1.0, 1.0)_T$
$(7.0, 0.5, 0.5)_T$	$(6.5, 0.5, 0.5)_T$
$(8.5, 0.5, 0.5)_T$	$(8.0, 0.5, 0.5)_T$
$(10.5, 0.5, 0.5)_T$	$(8.0, 1.0, 1.0)_T$
$(11.0, 0.5, 0.5)_T$	$(10.5, 0.5, 0.5)_T$
$(12.5, 0.5, 0.5)_T$	$(9.5, 0.5, 0.5)_T$

여 퍼지수의 곱에 대한 하계와 상계를 이용하였다. Kao와 Chyu (2003)가 추정한 퍼지회귀모형

$$\widehat{Y}_i^k = 3.565 + 0.522X_i + (-0.011, 0.951, 0.949)_T$$

은 추정된 퍼지수 $\widehat{Y}_i^k$의 중심과 폭에 대한 기울기는 일치하는 것을 의미한다. 그러나 표 3.3에 제시된 퍼지수는 폭에 대한 회귀계수와 중심에 대한 기울기가 일치하지 않음을 보여준다.

Sakawa와 Yano (1992), Kao와 Chyu (2003)가 이용한 자료에 대한 회귀모형을 추정하기 위해 독립퍼지회귀모형을 이용하자. 최소제곱법을 이용하여 추정한 독립퍼지회귀모형의 중심에 대한 회귀모형은

$$y_i = 3.572 + 0.519x_i$$

와 같다. 또한 추정된 퍼지수에 대한 양쪽 폭에 대한 결과는 다음과 같다.

$$\begin{aligned} l_{\widehat{y}_i} &= -\min\{0, k_i\} + l_{x_i}, \\ r_{\widehat{y}_i} &= \max\{0, k_i\} + r_{x_i}. \end{aligned}$$

여기서 $k_i = y_i - 1.375 - 30.12x_i$이다.

표 3.4은 척도 (3.1)를 이용하여 독립퍼지회귀모형의 정확성을 Sakawa와 Yano (1992), Kao와 Chyu (2003)가 추정한 퍼지회귀모형와 비교한 결과이다. 최소제곱법을 사용하여 추정한 독립퍼지회귀모형의 정확성은 6.833이고, 퍼지회귀모형을 이용한 Sakawa와 Yano (1992), Kao와 Chyu (2003)의 모형에 대한 정확성은 각각 9.431과 9.363이다. 퍼지수인 종속변수의 중심과 폭에 대한 회귀모형을 각각 독립적으로 분리하여 회귀모형을 추정하는 것이 일반적인 퍼지회귀모형보다 더 정확성할 수 있음을 예제로부터 확인할 수 있다. 예제 3.1과 예제 3.2를 통하여 확인된 독립퍼지회귀모형의 결과를 모든 퍼지회귀모형으로 일반화하기 위해서는 본 논문에서 제시된 독립퍼지회귀모형에 대한 통계적인 성질을 연구하여야 한다.

표 3.4: 추정된 퍼지회귀모형의 정확성

y_i	정확성 척도		
	$m(Y_i, \widehat{Y}_i^s)$	$m(Y_i, \widehat{Y}_i^k)$	$m(Y_i, \widehat{Y}_i)$
4.0	0.633	0.974	0.684
5.5	0.453	0.702	0.154
7.5	1.613	1.683	1.645
6.5	1.165	1.111	0.767
8.5	0.770	0.848	0.596
8.0	1.977	1.603	1.189
10.5	1.368	1.508	1.156
9.5	1.452	0.934	0.642
Total	9.431	9.363	6.833

4. 결론

본 논문에서는 독립변수와 종속변수가 퍼지수인 퍼지회귀모형에서 관찰된 퍼지수의 중심과 폭에 대한 반응함수가 일치하지 않을 수 있음을 확인하였다. 그리고 종속변수의 소속함수가 독립변수의 소속함수에 독립적이고, 추정된 퍼지수의 중심과 폭에 대한 반응함수가 서로 구분된 독립퍼지회귀모형을 소개하였다. 독립퍼지회귀모형을 추정하기 위하여 최소제곱법을 이용하였고, 본 논문에서 제시된 퍼지회귀모형이 Diamond (1988)와 Kao와 Chyu (2003)가 추정한 퍼지회귀모형보다 더 정확할 수 있음을 예제를 통하여 확인하였다. 앞으로 일반적인 퍼지회귀모형에 대한 정확성을 유도하기 위해 독립퍼지회귀모형에 대한 통계적인 성질을 연구할 필요가 있다.

참고문헌

Chang, Y. O. and Ayyub, B. M. (2001). Fuzzy regression methods-a comparative assessment. *Fuzzy Sets and Systems*, **119**, 187–203.

Chen, Y. (2001). Outlier detection and confidence interval modification in fuzzy regression, *Fuzzy Sets and Systems*, **119**, 259–272.

Diamond, P. (1988). Fuzzy least squares. *Information Sciences*, **46**, 141–157.

Diamond, P. and Korner, R. K. (1997). Extended fuzzy linear models and least-squares estimates. *Journal of Computational and Applied Mathematics*, **9**, 15–32.

Freund, J. E. (1992). *Mathematical Statistics*, Prentice-Hall International, Inc.

Kao, C. and Chyu, C. (2003). Least Squares estimates in fuzzy regression analysis. *European Journal of Operational Research*, **148**, 420–435.

Kim, B. and Bishu, R. R. (1998). Evaluation of fuzzy linear regression models by comparing membership functions. *Fuzzy Sets and Systems*, **100**, 343–352.

Sakawa, M. and Yano, M. (1992). Multiobjective fuzzy linear regression analysis for fuzzy input-output data. *Fuzzy Sets and Systems*, **47**, 173–181.

Savic, D. and Pedryzc, W. (1991). Evaluation of fuzzy linear regression models. *Fuzzy Sets and Systems*, **39**, 51–63.

Tanaka, H., Hayashi, I. and Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research*, **40**, 389–396.

Tanaka, H., Uejima, S. and Asai, K. (1980). Fuzzy linear regression model. *International Congress Applied Systems and Cybernetics*, **4**, 2933–2938.

Tanaka, H., Uejima, S. and Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transaction on Systems, Man, and Cybernetics*, **12**, 903–907.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.

[Received December 2006, Accepted January 2007]