Trend in Fuzzy Regression Model

최승회1), 김해경2), 정은경3)

요 약

종속변수와 독립변수 사이의 통계적인 관계를 설명하기 위해 사용되는 회귀모형을 분석하는 방법을 회귀분석이라 한다. 독립변수와 종속변수가 퍼지수인 퍼지회귀모형을 추정하기 위해 최소절대편차추정량을 제시하고. 예제를 이용하여 퍼지최소절대편차회귀모형과 퍼지최소자승회귀모형의 효율성을 평가한다.

주요용어 : Fuzzy Regression Model, Least Square Method, Least Absolute Deviation Method.

1. 서론

자연적인 현상이나 사회적인 현상을 추상화하여 수학적인 문제로 만드는 과정에서 발생하는 문제는 불확실성(uncertainty)이다. 불확실성에는 시간이 흘러가거나 실험을 통하여 해결되는 확률적인 불확실성과 실험이나 시간과는 무관한 퍼지적 불확실성이 있다.

Zadeh는 실험이나 시간과는 무관한 퍼지적 불확실성을 애매함(ambiquity)과 모호함(vagueness)으로 설명하고, 애매하고 모호한 문장이나 정보들을 처리하기 위하여 필요한 시스템을 구현하기 위하여 퍼지이론을 소개하였다 [1].

종속변수에 대한 확률적인 불확실성은 확률론을 이용한 연구가 활발하게 진행되어 왔으나, 퍼지적 불확실성에 대한 연구는 최근에 시작되었다. 지금까지 회귀분석에서 발생하는 오차는 측정 오차로 설명하였으나, Tanaka는 종속변수의 퍼지적 불확실성을 회귀분석에서 사용하는 반응함수나 혹은 회귀계수에 대한 퍼지성으로 설명하고 퍼지회귀모형을 처음으로 소개하였다 [9, 10].

본 논문에서는 퍼지회귀모형에 대한 퍼지최소절대편차추정량을 제안하고, 퍼지최소절대편차모형과 퍼지최소자승모형의 효율성을 비교하기 위하여 두 가지 예를 제시한다.

2. 퍼지선형회귀모형

퍼지회귀모형은 종속변수와 독립변수의 조건에 따라서 다음과 같이 세 가지로 분류할 수 있다.

i) 종속변수와 독립변수가 모두 실수인 경우.
ii) 독립변수는 실수이고, 종속변수는 퍼지수인 경우.
iii) 종속변수와 독립변수가 모두 퍼지수인 경우.

---

1) 한국항공대학교, 교양학부, 부교수, 경기도 고양시 덕양구 화전동 2001-1.
2) 연세대학교, 수학과, 교수, 서울시 서대문구 신촌동, 120-749.
3) 연세대학교, 수학과, 석사과정, 서울시 서대문구 신촌동, 120-749.

먼저, 독립변수는 실수이고 종속변수는 퍼지수인 다음과 같은 퍼지회귀모형을 생각하자.

$$Y_i = A_1 x_{i1} + \cdots + A_p x_{ip} \tag{1}$$

여기서, A_i는 퍼지수로 표현되는 퍼지회귀계수이고, x_{ij}는 실수이고, Y_i는 퍼지수이다.

모형 (1)과 같은 퍼지모형을 추정하기 위하여 Savic과 Pedrycz는 퍼지회귀계수에 대한 폭의 합을 최소화 하는 퍼지최소제곱방법을 연구하였고, 추정된 퍼지수와 관측된 퍼지수의 차의 제곱을 최소로 하는 방법은 여러 사람들에 의하여 연구되었다 [3, 7]. 최근에는 Kim과 Bishu는 퍼지회귀계수를 추정하기 위하여 퍼지소속함수의 차의 합을 최소로 하는 FMLS(Fuzzy Membership Least Squares)방법을 제안하였다 [6].

모형 (1)에서 주어진 퍼지수를 각 각 $A_i = (m_i, l_i, r_i)_{LR}$,와 $Y_i = (y_i, l_{y_i}, r_{y_i})_{LR}$라 하자. 그러면, 모형 (1)과 퍼지수의 연산을 이용하여 종속변수 Y_i에 대한 소속함수를 구할 수 있다. 만약 관측된 퍼지수에 대한 α-절단과 퍼지회귀계수의 지지(Support)가 같다고 하면, 다음과 같은 식을 얻을 수 있다.

(i) $y_i - |L^{-1}(\alpha)| \times (y_i - l_{y_i}) = \sum_{j=1}^{p} l_i x_{ij}$,

(ii) $y_i = \sum_{j=1}^{p} m_i x_{ij}$,

(iii) $y_i - |L^{-1}(\alpha)| \times (r_{y_i} - y_i) = \sum_{j=1}^{p} r_i x_{ij}$,

여기서, $|L^{-1}(\alpha)|$ 은 소속함수 $L(x)$의 α-절단에 의한 역상이다. 퍼지회귀계수의 중심인 m_i에 대한 최소절대편차추정치 $\widehat{m_i}$ 는 절대편차

$$\sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} m_i x_{ij} \right|$$

를 최소화하는 값으로 구할 수 있다.

비슷한 방법으로 회귀계수의 왼쪽과 오른쪽의 폭인 l_i 와 r_i에 대한 최소절대편차추정치 $\hat{l_i}$ 와 $\hat{r_i}$ 를 얻을 수 있다. 따라서 퍼지소속함수의 차를 최소로 하는 FMLAD(Fuzzy Membership Least Absolute Deviation)모형은 다음과 같이 얻을 수 있다.

$$\widehat{Y_i} = \widehat{A}_1 x_{i1} + \cdots + \widehat{A}_p x_{ip},$$

여기서, $\widehat{A_i} = (\hat{l_i}, \widehat{m_i}, \hat{r_i})$이다.

다음으로, 독립변수와 종속변수가 모두 퍼지수인 다음과 같은 퍼지회귀모형을 생각하자.

$$Y_i = a_1 X_{i1} + \cdots + a_p X_{ip} \tag{2}$$

여기서 a_i는 실수이고, $X_{ij} = (l_{x_{ij}}, x_{ij}, r_{x_{ij}})_{LR}$와 $Y_i = (l_{y_i}, y_i, r_{y_i})_{LR}$는 퍼지수이다.

퍼지회귀모형 (2)의 회귀계수를 추정하기 위하여 Diamond는 최소제곱방법을 연구하였고, Kao와 Chyu는 퍼지수의 중심점에 대한 최소제곱법과 퍼지오차를 구하기 위하여 퍼지소속함수의 차를 최소화하는 방법을 소개하였다 [3, 5].

최소절대편차추정량을 이용하여 모형 (2)을 추정하기 위하여 다음과 같은 두 단계 절차를 생각하자.

Step Ⅰ. 최소절대편차추정량을 이용하여 회귀계수의 추정치 $\hat{a_i}$을 찾는다.
비퍼지화한 실수값 (x_{ijc}, y_{ic})을 이용하여 회귀계수 a_i의 추정치는 다음과 같은 절대편차

$$\sum_{i=1}^{n} \left| y_{ic} - \sum_{j=1}^{p} a_j x_{ijc} \right|,$$

를 최소화하는 값으로 정의하자. 여기서,

$$x_{ijc} = \frac{\int_{-\infty}^{\infty} x \mu_{X_{ij}}(x)dx}{\int_{-\infty}^{\infty} \mu_{X_{ij}}(x)dx} = \frac{1}{3}\left(l_{x_{ij}} + x_{ij} + r_{x_{ij}}\right),$$

이고

$$y_{ic} = \frac{\int_{-\infty}^{\infty} y \mu_{Y_i}(y)dy}{\int_{-\infty}^{\infty} \mu_{Y_i}(y)dy} = \frac{1}{3}\left(l_{y_i} + y_i + r_{y_i}\right).$$

Step Ⅱ. 퍼지회귀모형의 퍼지오차를 구하기 위하여 $l_{\min}$과 $r_{\min}$를 각 각 관측된 퍼지수의 가장 작은 왼쪽과 오른쪽의 폭이라 하자. 퍼지회귀모형의 퍼지오차를 구하기 위하여 Kim과 Bishu이 제안한 소속함수의 차를 이용하여 다음과 같은 식을 생각하자.

$$\sum_{i=1}^{n} \int_{S_{Y_i} \cup S_{\hat{Y}_i}} \left| \mu_{Y_i}(y) - \mu_{\hat{Y}_i}(y) \right| dy = Min!$$

여기서 $\hat{Y_i} = \hat{a_1} X_{i1} + \cdots + \hat{a_p} X_{ip} + (-l, 0, r)_{LR}$, $\left(\hat{y_i} - l_{\hat{y_i}}\right) \ge l_{\min}$, $\left(r_{\hat{y_i}} - \hat{y_i}\right) \ge r_{\min}$이고, S_{Y_i} 과 $S_{\hat{Y}_i}$는 각 각 μ_{Y_i} 와 $\mu_{\hat{Y}_i}$ 의 지지(support)이다.

3. 예제 및 결론

최소자승법을 이용한 퍼지회귀모형과 최소절대편차추정량을 이용하여 추정한 퍼지회귀모형의 효율성을 조사하기 위하여 다음과 같은 두 가지 예제를 생각하자. 첫 번째 예제는 종속변수가 퍼지수인 경우이고, 두 번째 예제는 독립변수와 종속변수가 모두 퍼지수인 경우이다.

예제1. Tanaka가 제시한 퍼지회귀모형에 대한 각 방법의 추정식은 다음과 같다.
(ⅰ) Y_{FMLS} = $(3.11, 4.95, 6.84)$ + $(1.55, 1.71, 1.82)x$.
(ⅱ) Y_{KC} = $4.95 + 1.71x + (-3.01, 0, 1.80)$.
(ⅲ) Y_{FMLAD} = $(3.46, 5.56, 7.71)$ + $(1.43, 1.49, 1.54)x$.

표 3.1은 세 가지 방법에 대한 오차를 보여준다. FMLAD모형에 대한 오차의 합은 9.147이고, FMLS모형과 Kao-Chyu모형에 대한 오차의 합은 각 각 10.026와 9.678이므로 FMLAD모형이

더 효율적임을 알 수 있다.

표 3.1:

Independent variable	Response variable	Errors in estimation FMLS	Kao-Chyu	FMLAD
1	$(6.2, 8.0, 9.8)$	2.207	2.7826	1.6850
2	$(4.2, 6.4, 8.6)$	3.025	2.5951	3.2818
3	$(6.9, 9.5, 12.1)$	1.042	0.5559	0.9949
4	$(10.9, 13.5, 16.1)$	2.902	3.3569	3.1849
5	$(10.6, 13.0, 15.4)$	0.850	0.3874	0.0000
Total error		10.026	9.6779	9.1467

예제2. Sakawa와 Yano가 연구한 자료에 대한 퍼지회귀모형의 추정치는 다음과 같다.

(i) $Y_{KC} = 3.5724 + 0.5193X + (-0.24, 0, 0.24)$

(ii) $Y_{Diamond} = (3.2636, 3.5632, 3.8628) + 0.5206X$

(iii) $Y_{LAD} = 3.94 + 0.44X + (-0.28, 0, 0.28)$.

표 3.2는 세 가지 방법에 대한 오차를 보여준다. Diamond, Kao-Chyu, 그리고 LAD 방법에 대한 각 각의 오차 합은 9.431, 7.4881, 6.2362이다. 따라서 LAD 방법이 효율적임을 알 수 있다.

표 3.2 :

Independent variable	Response variable	Errors in estimation Diamond	Kao-Chyu	FLAD
$(1.5, 2.0, 2.5)$	$(3.5, 4.0, 4.5)$	0.633	0.8488	0.9722
$(3.0, 3.5, 4.0)$	$(5.0, 5.5, 6.0)$	0.453	0.2076	0.0000
$(4.5, 5.5, 6.5)$	$(6.5, 7.5, 8.5)$	1.613	1.4903	1.5054
$(6.5, 7.0, 7.5)$	$(6.0, 6.5, 7.0)$	1.165	0.9147	0.8025
$(8.0, 8.5, 9.0)$	$(8.0, 8.5, 9.0)$	0.770	0.7630	0.9506
$(9.5, 10.5, 11.5)$	$(7.0, 8.0, 9.0)$	1.977	1.4535	1.0054
$(10.5, 11.0, 11.5)$	$(10.0, 10.5, 11.0)$	1.368	1.0001	1.0000
$(12.0, 12.5, 13.0)$	$(9.0, 9.5, 10.0)$	1.452	0.8102	0.0000
Total error		9.431	7.4881	6.2362

일반회귀모형과 같이 퍼지회귀모형에서도 최소자승법을 이용한 방법보다 최소절대편차추정량을 이용한 방법이 더 효율적일 수 있음을 예제를 통하여 확인하였다. 따라서 퍼지회귀분석에 있어서 이상치를 포함한 경우에는 최소절대편차추정량을 이용한 퍼지회귀모형이 최소제곱방법을 이용한 퍼지회귀모형보다 더욱 효율적일 수 있다.

참고문헌

(1) R. E. Bellman, L. A. Zadeh, 1970, Decision-making in a fuzzy environment, *Manage. Sci.* 17B, 141-164.

(2) S. J. Chen, C. L. Hwang, 1992, Fuzzy Multiple Attribute Decision Making, *Springer*, NY.

(3) P. Diamond, 1988, Fuzzy least squares, *Inform. Sci.* 46, 141-157.

(4) H. Hellendoorn, C. Thomas, 1993, Defuzzification in fuzzy controllers, *J. Intell. Fuzzy Systems* 1, 109-123.

(5) C. Kao, C. Chyu, 2002, A fuzzy linear regression model with better explanatory power, *Fuzzy Sets and Systems* 126, 401-409.

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

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

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

(9) H. Tanaka, I. Hayashi, J. Watada, 1989, Possibilistic linear regression analysis for fuzzy data, *Eur. J. Oper. Res.* 40, 389-396.

(10) H. Tanaka, S. Uejima, K. Asai, 1982, Linear regression analysis with fuzzy model, *IEEE Trans. Systems, Man Cybernet.* 12, 903-907.

(11) R. R. Yager, 1980, On a general class of fuzzy connectives, *Fuzzy Sets and Systems* 4 254-242.