측정오차가 있는 경우의 분할 퍼지회귀모형

Piecewise Fuzzy Linear Model with Measurement Error Variable

안정용*, 한범수, 최승현

ABSTRACT

In this study we present the inverse correlation method to select the exploratory variables, while Sugeno used RC method in his paper[6]. We assume linear model with measurement error variables as in Fuller's Book[9]. We provide possibilistic linear model and predict the fuzzy response variable in case of fuzzy exploratory variables. By plotting data we can divide them for piecewise plane and provide the piecewise possibilistic linear model. If the exploratory variable is fuzzy trapezoidal variable or interval variable, then we estimate fuzzy trapezoidal response variable respondent to it. We will illustrate using Nonlinear System data in Sugeno's paper

I. 서론

회귀분석은 경제학, 공학, 사회학 등에서 일어나는 현상들을 설명변수와 반응변수 사이의 관계로 설명할 때 많이 이용되고 있다. 이러한 경우 통계적 회귀분석에서는 관측값과 추정값의 편차를 관측오차 또는 모형에 포함되지 않은 변수의 영향으로 본다. 그러나 Tanaka 등이 제안한 퍼지회귀분석에서는 입출력 관계를 나타내는 시스템 구조 자체의 애매함으로 간주하고 퍼지회귀계수를 이용하여 시스템을 설명한다.

Kim[8]은 Tanaka가 제안한 대칭형 삼각퍼지계수를 확장하여 대칭형 사다리꼴 퍼지계수를 이용하여 가능성 선형시스템을 제안하였다. 본 연구에서는 관측치를 참값과 오차로 구분하고 참값들 사이의 선형관계를 구하고 하한, 상한선형모형을 구하여 가능성 선형시스템을 개선하고자한다. 설명변수가 퍼지수일 때 반응변수도 퍼지수로 표시되는 것이 타당하므로 사다리꼴 퍼지수에 대하여 고찰하고자 한다.

수정역상관행렬을 이용하여 반응변수를 잘 설명할 수 있는 설명변수를 선택하는 변수선택방법을 제안하고 변수들 사이의 선형관계를 잘 표현하기 위하여 변수구간을 분할하여 시스템구조를 표현하는 세 가지의 전형적인 예를 통하여 설명한다.

II. 가능성 선형시스템의 개선

Tanaka와 Kim은 퍼지계수를 이용하여 가능성 선형시스템을 설명할 때 대칭형 퍼지계수를 사용함으로써 반응변수의 구간이 지나치게 크게 추정되는 것이 문제점이다. 본 연구에서는 참값들 사이의 선형시스템과 하한과 상한 선형시스템을 이용하여 가능성 선형시스템을 개선하고 반응변수의 퍼지수를 구하고자 한다.

서남대학교 전산통계학과

먼저 설명변수의 수가 하나인 경우, 둘 이상인 경우의 참값들 사이의 관계를 구하는 방법과 이들을 이용하여 가능성 선형시스템을 구성하는 방법을 정리로 제안한다.

(정리1) 관찰치 (Y_t, X_t), 참값 (y_t, x_t), 측정오차 (e_t, u_t) 사이에 다음과 같은 모형을 가정한다.

$$y_t = \beta_0 + \beta_1 x_t$$

$$Z_t = (Y_t, X_t) = (y_t, x_t) + (e_t, u_t) \quad t = 1,2,\cdots,n$$

$\varepsilon = (e_t, u_t)$의 공분산행렬을

$$\Sigma_{\varepsilon\varepsilon} = \begin{pmatrix} \sigma_{ee} & \sigma_{eu} \\ \sigma_{eu} & \sigma_{uu} \end{pmatrix},$$

$$M_{ZZ} = \frac{1}{n-1} \Sigma (Y_t - \overline{Y}, X_t - \overline{X})' (Y_t - \overline{Y}, X_t - \overline{X})$$

$$= \begin{pmatrix} m_{YY} & m_{XY} \\ m_{XY} & m_{XX} \end{pmatrix}$$

그리고 $\sigma_{ee} = \delta\sigma_{uu}$ 인 관계가 있을 때 (2.1)의 각 수정량은 다음과 같다.

$$\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\overline{X}$$

$$\hat{\sigma}_{ee} = \delta\, \hat{\sigma}_{uu}$$

$$\widehat{\beta_1} = \hat{\sigma}_{xx}^{-1} m_{XY}$$

$$\hat{x}_t = X_t - \hat{\sigma}_{vv}^{-1}\, \hat{\sigma}_{uv}\, \hat{v}_t$$

여기서 $\hat{\sigma}_{vv} = \hat{\sigma}_{ee} - 2\widehat{\beta_1}\, \hat{\sigma}_{eu} + \widehat{\beta_1}^2\, \hat{\sigma}_{uu}$

$$\hat{\sigma}_{uv} = \hat{\sigma}_{eu} - \widehat{\beta}_1\, \hat{\sigma}_{uu}$$

$$\hat{v}_t = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_t$$

$$\hat{\sigma}_{xx} = (2\delta)^{-1}\{[(m_{YY} - \delta m_{XX})^2 - 4\delta m_{XY}^2]^{1/2} - (m_{YY} - \delta m_{XX})\}$$

$$\hat{\sigma}_{uu} = (2\delta)^{-1}\{m_{YY} - \delta m_{XX} - [(m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2]^{1/2}\}$$

따라서 반응변수는 $\widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1}\widehat{x_t}$로 표시된다.

(정리 2) $y_t = X_t\beta$

$$(Y_t, X_t) = (y_t, x_t) + (e_t, u_t) = z_t + \varepsilon_t \quad (2.2)$$

라 할 때

$$\varepsilon_t \sim (0,\ \Sigma_{\varepsilon\varepsilon}) \quad \Sigma_{\varepsilon\varepsilon} = \gamma_{\varepsilon\varepsilon}\sigma^2 = \begin{pmatrix} \gamma_{ee} & \gamma_{eu} \\ \gamma_{eu} & \gamma_{uu} \end{pmatrix}\sigma^2$$

이라하면

$$\hat{\beta} = (M_{XX} - \hat{\lambda}\gamma_{uu})^{-1}(M_{XY} - \hat{\lambda}\gamma_{ue}),$$

$$\hat{\sigma}^2 = \hat{\sigma}_m^2 = (k+1)^{-1}\hat{\lambda}$$

단 $M_{ZZ} = \frac{1}{n}\Sigma Z_t' Z_t$ 이고 $\hat{\lambda}$은

$|M_{ZZ} - \lambda\gamma_{\varepsilon\varepsilon}| = 0$ 의 최소고유치이다.

$$\hat{z}_t = Z_t - (Y_t - X_t\hat{\beta})[(1,\ -\hat{\beta}')\gamma_{\varepsilon\varepsilon}(1,\ -\hat{\beta}')']^{-1} \times (1,\ \hat{\beta}')\gamma_{\varepsilon\varepsilon}$$

선형하한모형과 선형상한모형을 구하므로써 가능성 선형시스템을 결정할 수 있다.

(정리 3) (1) $\min \Sigma c_j$

$$\text{s.t} \quad 0 \le Y_t - \hat{y}_t \le \Sigma c_j |\hat{x}_{tj}|$$

$$c_j \ge 0, \quad t = 1,2,\ldots,n$$

따라서 $\hat{y}_t + \Sigma c_j|\hat{x}_{tj}|$는 상한가능성모형이다.

(2) $\min \Sigma d_j$

$$\text{s.t} \quad \hat{y}_t - Y_t \le \Sigma d_j |\hat{x}_{tj}|$$

$$d_j \ge 0, \quad t = 1,2,\ldots,n$$

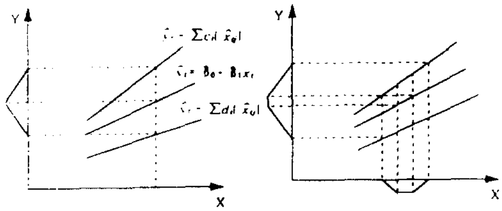
따라서 $\hat{y}_t - \Sigma d_j|\hat{x}_{tj}|$는 하한가능성모형이다.

Ⅲ. Piecewise fuzzy linear model 과 적용

선형모형에서 변수선택방법은 stepwise 방법과 모형을 가정했을 때는 RC방법 등이 사용된다. 본 연구에서는 수정된 역상관행렬을

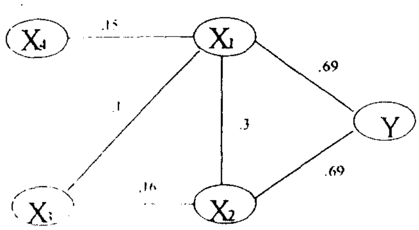
이용하여 반응변수와 관계가 있는 설명변수를 선택하고 변수들 사이의 graph 또는 change-point를 찾아 평면을 분할하여 piecewise fuzzy linear model을 적합시킨다. 퍼지 모형에서 자주 언급되는 Sugeno의 자료와 Box-Jenkins의 Gas furnace data를 이용하여 실제 응용 방법을 보이겠다.

퍼지 단순모형에서 설명변수가 퍼지수일 때 반응변수의 값을 구하는 방법은 다음과 같다.



(3.1) Sugeno data의 분석 방법

자료의 수정된 역상관행렬을 구하면 변수들 사이의 관계는 다음과 같이 요약할 수 있다.



따라서 반응변수 Y의 설명변수로 X_1, X_2를 선택하고 (Y, X_1, X_2)의 graph에 의하여 6개의 평면으로 분할하여 적합시킨 결과 중 하나만 표기해보면 다음과 같다.

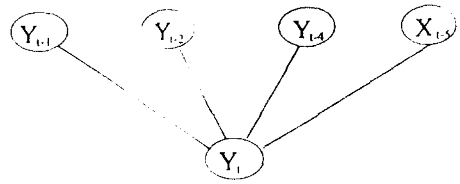
상한가능성모형 $\hat{y}_t + 0.056\ \hat{x}_{t1} + 0.235\ \hat{x}_{t2}$

적합모형 $\hat{y}_t = 42.675 - 9.582\ \hat{x}_{t1} - 14.225\ \hat{x}_{t2}$

하한가능성모형 $\hat{y}_t - 0.2479\ \hat{x}_{t1}$

(3.2) Box-Jenkins Gas furnace data

자료의 수정된 역상관행렬을 구하면 다음과 같다. 앞에서와 같은 방법으로 가능성 선형시스템을 결정할 수 있다.



Ⅳ. 결론

수정역상관행렬을 이용하여 변수를 선택하고 변수들 사이의 선형관계를 잘 표현하기 위하여 graph로 표현한 자료를 분할하여 적합하는 방법을 통해 가능성 선형시스템을 구하였다. 개선해야 될 점으로는 평면의 절단점에서의 적합이 불안정하다는 점이다.

Ⅴ. 참고문헌

[1] D.Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)

[2] Kwang Hyung Lee and Gill Rok Oh, Fuzzy Theory and Applications Volume Ⅰ, Ⅱ. Hongleung Press, Seoul (1991)

[3] Dragan A.Savic, Witold Pedrycz, Evaluation of fuzzy linear regression models. Fuzzy Setand Systems 39, 51-63 (1991)

[4] Sylvia Fruhwirth-Schnatter, On statistical inference for fuzzy data with applications to descriptive statistics. Fuzzy Set and Systems 50, 143-165 (1992)

[5] Michio Sugeno and Takahiro Yasukawa, A Fuzzy-Logic-Based Approach to Qualitative Modeling. IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL 1, NO. 1, 7-31 (1993)

[6] H. Tanaka and J.Watada, Possibibilistic

linear systems and their application to the linear regression model, Fuzzy Set and Systems 27, 257-289 (1988)

[7] L. A. Zadeh, Fuzzy Sets as a basis for a theory of possibility, Fuzzy Set and Systems 1, 3-28 (1978)

[8] Soon Ki Kim, Fuzzy Linear Regression with Interval-valued Data. fifth IFSA World Congress, 461-464 (1993)

[9] Wayne A. Fuller, MEASUREMENT ERROR MODELS. JOHN WILEY & SONS, 124-143 (1987)