Journal of the KSQC
Vol. 13, No. 1, 1985

多元回歸 MODEL에 있어서 構造變化에 관한 研究

A Study on Structural Change in the Multivariate Regression Model

趙 嚴*

ABSTRACT

There are several approaches for dealing with the structural change in regression model, but by introducing a concept of Spline, the structural change can be expressed more clearly. This makes it possible not only to know the location where the structural change happens and the total number, but also to derive posterior distribution from anterior-posterior distribution when the probability of the judgement anterior for entire combination was given to each model, by which, the model that has the highest posterior probability is the method which realizes the structural change.

The purpose of this study is to find a peculiarity of the posterior probability on the occasion of anterior information acquired and of not acquired with Baysian approach.

1. 序　論

回歸 Model에 있어서 構造變化에 대처하는 方法은 여러가지 생각할 수 있으나 Spline의 개념을 도입하는데 따라 構造變化를 보다 明確히 표현할 수 있다. 이것은 構造變化가 일어나는 位置와 全體의 個數를 알 수 있으며 全體的인 組合에 대하여 個個의 Model에 事前의 判斷確率을 주어 事前事後分布를 통해 事後分布를 導出하고 이것에 의해 事後確率이 가장 높은 Model이 構造變化를 具現하는 Model이라고 하는 方法이다. 本 研究는 Baysian 어프로치의 方法을 使用하여 事前情報를 얻은 경우와 그렇지 않은 경우에 대하여 事後確率의 特性을 찾아내는데 目的이 있다.

*東國大學校 工科大學 產業工學科(專任講師)

2. 構造變化

部分 部分이 적당한 유연성을 가지고 서로가 結合되어 있는 部分函數인 Spline은 보통 個個의 部分에서는 多項式인 것이 要望되며, 유연성은 Spline과 그 導函數의 連續性에 의해 표시되어진다.

定義 1 : n次의 多項 Spline 函數는 Spline

과 $(n-1)$次까지의 導函數가 연속이나 n次의 多項式부터 部分多項函數이다.[1]

連續性에 관해서는 y가 x의 線形 Spline 인 것을 의미한다. 또한 y가 x의 部分線形函數이며는 y는 어떤 x의 線形函數에서 他의 線形函數에 轉換하는 것은 構造變化라 할 수 있다.

3. 構造變化와 Spline 函數

j번째의 部分을 表示하면

$$S_{\Delta}(x) = \left[\frac{y_{j-1}\bar{x}_j - y_j\bar{x}_{j-1}}{\bar{x}_j - \bar{x}_{j-1}}\right]$$

$$+ \left[\frac{y_j - y_{j-1}}{\bar{x}_j - \bar{x}_{j-1}}\right] x \qquad (2.1)$$

단 : $\bar{x}_{j-1} \leq x \leq \bar{x}_j$

$S_{\Delta}(x)$: 線形 Spline

파라메타化의 方法에서 k個의 變換된 變數를 定義하면

$$w_1 = x$$

$$w_j = (x-\bar{x}_{j-1})_+ = \max(x-\bar{x}_{j-1}, 0)$$

$$= \begin{pmatrix} x-\bar{x}_{j-1}, & x > \bar{x}_{j-1} \\ 0 & , \ x \leq x_{j-1} \end{pmatrix} \qquad (2.2)$$

단 : $j = 2, 3, \cdots, k$ w : 變換變數

이때 어떤 x에 대해서도 線形 Spline $S_{\Delta}(x)$는 다음과 같이 表示된다.

$$S_{\Delta}(x) = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \cdots\cdots + \beta_k w_k \qquad (2.3)$$

β : Spline 傾斜

만약 β_j가 0에서 동떨어진 값이면 이때는 $\bar{x}_{j-1}$에 있어서 어떤 種類의 "構造變化"가 생겼다고 볼 수 있다.

代替的 表現은 區間$(j+1)$이 構造變化를 測定하기 위한 基本으로서 有效하다.

註1) 高橋崇弘., 多項式における構造変化" 至文堂(1978年) p. 13

$$y = r_0 + r_1 w_1 + r_2 w_2 + \cdots\cdots + r_k w_k + \varepsilon \qquad (2.4)$$

變換된 變數는 다음과 같다.

$$w_i = (x-\bar{x}_i)_- = \min(x-\bar{x}_i, 0)$$

$$= \begin{pmatrix} x-\bar{x}_i & , \ x < \bar{x}_{i-1} \\ 0 & , \ x \geq \bar{x}_{i-1} \end{pmatrix} \qquad (2.5)$$

단 : $i = 1, 2, \cdots\cdots j$

$w_i = (x-\bar{x}_{i-1})_+$

$(i = j+1, \ j+2, \cdots\cdots, k)$

區間i의 기울기 : $(r_i + r_{i+1} \cdots + r_j)$

4. Baysian 과 事前分布 및 事後分布

事前分布가 $N(m_0, \sigma_0{}^2)$이며, σ^2이 알고 있는 正規分布 $N(m, \sigma^2)$으로부터 크기 n의 랜덤標本에 의한 標本平均値가 $\bar{x}$일때 未知母數 n의 事後分布는 다음과 같다.

$$m_n = (m_0 v_0 + \bar{x} v_n)/(v_0 + v_n) \qquad (3.1)^{(2)}$$

여기서 事前分布가 연결된 v_0와 尤度에 연결된 v_n을 비교할 때 v_0가 v_n에 비하여 작다.

이와 같은 경우 尤度函數는 事前分布를 지배한다.

$k(x|\theta)$를 尤度函數의 核이라 할때 事後分布 $\xi(\theta|x)$는 다음과 같다.

$$\xi(\theta|x) \propto \xi(\theta)\, k(x|\theta) \qquad (3.2)^{(3)}$$

그리고 $\tilde{x}(n) = x(n)$이 주어진 後의 $\tilde{\theta}$의 事

註2) $v_0 = \sigma_0^{-2}$, $v_n = n\sigma^{-2}$, 分散의 逆數 $\sigma_n^{-2} = v_0 + v_n$
$N(\bar{x}, \sigma^2/h)$는 $\sigma_0^2 \to \infty$
$v_0 \to 0$일때는 極根分布
m_n : 期待値

註3) 任意로 $x \in X$를 固定하고 $f(x|\theta)$을 θ의 函數로 할때 $\theta \in \Theta$, $x \in X$에 對해 $L(x|\theta) = k(x|\theta)\rho x$.
$\rho(x)$: 殘餘因數. 즉 $L(x|\theta) \propto k(x|\theta)$

後分布 $\xi(\theta \mid x(n))$은 다음 式에서 얻어진다. 이때 $\tilde{x}=x$이다.

$$\xi(\theta \mid x(n))=\xi(\theta \mid y) \propto k(y' \mid \theta) \\ k(y \mid \theta) \qquad (3.3).$$

5. Baysian 論的 回歸分析

Baysian 論에 있어서 B는 未知라 하더라도 決定者는 B에 對한 判斷確率로서 事前分布 $\xi(B)$을 가지고 있는 것을 出發點으로 한다. B의 事前分布가 變則的인 事前分布인 경우에는 다음 式으로 된다.

$$\xi(B \mid = X, y) \propto \exp\left\{-\frac{1}{2}(B-\hat{B})' \frac{X'X}{\sigma^2}(B-\hat{B})\right\} \qquad (4.1)^{(4)}$$

6. 事後分布의 導出

H_m이 주어진 B_m과 σ^2의 同時事前分布는 多變量正規 감마分布이므로 다음 式과 같다.

$$\xi(B_m, h \mid \mu'_m, \Sigma'_m, \upsilon'_m, v'_m, H_m) \propto h^{\frac{1}{2}(r(m)+2)} \exp\left\{-\frac{h}{2}(B_m-\mu'_m)^t (\Sigma'_m)^{-1}(B_m-\mu'_m)\right\} \times h^{\frac{1}{2}\upsilon'_m-1} \exp\left(-\frac{1}{2}\upsilon'_m v'_m h\right) \qquad (5.1)^{(5)}$$

따라서 事後確率 $P(H_m \mid y)$는 다음 式으로 나타난다.

$$P(H_m \mid y) \propto P(H_m) \frac{P(\upsilon''_{m/2}) \mid \Sigma''_m \mid^{\frac{1}{2}}}{\left(\frac{1}{2}\upsilon''_m v''_m\right)^{\frac{1}{2}\upsilon''_m} \mid \Sigma''_m \mid^{\frac{1}{2}}} \qquad (5.2)^{(6)}$$

7. Simulation Model

Simulation Model 로서 $Z_1 < Z_2 < Z_3$ 라는 세개의 Knots 의 設定한다. 그리고 實際로 構造變化가 생기는 Knots 를 Z_1, Z_3라 하자. 이때 생각되는 假設은 다음의 8가지가 있다.

$$\begin{aligned}
H_1 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\varepsilon_i \\
H_2 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_2 w_{i2}+\varepsilon_i \\
H_3 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_3 w_{i3}+\varepsilon_i \\
H_4 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_4 w_{i4}+\varepsilon_i \\
H_5 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_2 w_{i2}+\beta_3 w_{i3}+\varepsilon_i \\
H_6 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_2 w_{i2}+\beta_4 w_{i4}+\varepsilon_i \\
H_7 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_3 w_{i3}+\beta_4 w_{i4}+\varepsilon_i \\
H_8 &: y_i=\beta_0 w_{i0}+\beta_1 w_{i1}+\beta_2 w_{i2}+\beta_3 w_{i3}+\beta_4 w_{i4}+\varepsilon_i
\end{aligned}$$

Simulation 은 期間을 $n=20$으로 하고 Knots 를 $Z_1=30.0$, $Z_2=50.0$, $Z_3=75.0$으로 한다. date를 發生시키기 위하여 選擇된 Model 은 $Z_1=30.0$, $Z_3=75.0$으로 Knots를 가진다. 回歸直線의 파라메타로서는 $\beta_0=25.0$, $\beta_1=2.0$, $\beta_2=-1.5$, $\beta_3=0.0$, $\beta_4=1.0$을 주고 攪亂項의 標準偏差는 $\sigma=3$, 7, 11, 15, 19, 23, 27, 31의 8개 경우로 나누어 date를 發生시킨다. Simulation 은 B_m에 대한 條件付 事前分布가 變則的이 아닐때 즉 $B_m \sim \mathrm{MVN}(\mu'_m, \sigma^2\Sigma'_m)$

註4) B의 分布는 償乱項의 分布와는 獨立임.
　條件 1. $\rho(X)=k(\leq T)$
　條件 2. $E\{\bar{\varepsilon}_t\}=0$
　條件 3. 償乱項 $\bar{\varepsilon}_t$는 正規分布 $N(0, \sigma^2)$
　條件 4. 償乱項 $\bar{\varepsilon}_t$의 共通分散 σ^2은 알고 있는 것.

註5) H_m이 주어진 σ^2의 周邊事前分布는 h가 파라메타 υ_m, v_m($\upsilon_m>0$, $v_m>0$)을 가진 감마分布임.
　藤本熙, 松原望 共著., 「決定の数理」(1976), 筑摩書房, pp. 235.

註6) $\Sigma''_m=\{(\Sigma'_m)^{-1}+B^t_m W^t W B_m\}$
　$\mu''_m=\Sigma''_m\{(\Sigma'_m)^{-1}\mu'_m+B^t_m W^t y\}$
　$\upsilon''_m=\upsilon'_m+n$
　$v''_m=1/\upsilon''_m\{\upsilon'_m v'_m-(\mu''_m)^t(\Sigma''_m)^{-1}\mu''_m+(\mu'_m)^t(\Sigma'_m)^{-1}\mu'_m+y^t y\}$

일때[7] μ'_m에 관하여 正確한 情報로서 파라메타의 最小二乘推定値를 준다.

이를 계산하면

$\mu'_1=(43.8632,\ 1.0537)^t$

$\mu'_2=(43.1920,\ 1.0798,\ -0.0278)^t$

$\mu'_3=(53.6838,\ 0.7611,\ 0.3844)^t$

$\mu'_4=(55.2981,\ 0.7508,\ 0.3844)^t$

$\mu'_5=(19.7280,\ 2.2665,\ -2.2731,\ 1.2208)^t$

$\mu'_6=(25.0000,\ 2.0000,\ -1.5000,\ 1.0000)^t$

$\mu'_7=(46.1649,\ 1.0549,\ -0.7372,\ 1.1982)^t$

$\mu'_8=(25.0000,\ 2.0000,\ -1.5000,\ 0.0000,\ 1.0000)^t$

期待値 $E(\sigma^2{}_{ii})$ $(i=1,2,\cdots\cdots r(m)+2$ 를 가지고 標準偏差가 期待値의 5%가 되게끔 設定하며, 共分散에 관해서는 $\rho=0.5$로 하여

$E(\sigma^2{}_{ii})=0.2,\ 0.6,\ 1.0,\ 1.4,\ 1.8,\ 2.2,\ 2.6,\ 3.0,\ 3.4,\ 3.8,\ 4.2$의 11개로 Simulation 행한다.

事前의 判斷確率 $P(H_m)$에 관해서는 表1에 있는 바와같이 各 Knots에 대하여는

i) 構造變化가 發生하는 確率과 생기지 않는 確率을 같다는 判斷의 基本에서 $P(H_m)$을 決定하는 경우.

ii) Z_1에 대하여 構造變化가 생기는 確率이 생기지 않는 確率의 2倍라는 判斷의 基本에서 $P(H_m)$을 決定하는 경우.

iii) Z_2에 대하여 構造變化가 생기는 確率이 생기지 않는 때의 確率보다 2倍라는 判斷의 경우.

iv) Z_3에 대하여 構造變化가 생기는 確率이 發生하지 않는 確率의 2倍라고 判斷하는 경우.

4가지를 Simulation 행한다.

	i			ii			iii			iv		
	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
構造變化가 생기는 確率	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$
構造變化가 생기지 않는 確率	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$

〈表1〉 各 knots 에 대한 確率

	$P(H_1)$	$P(H_2)$	$P(H_3)$	$P(H_4)$	$P(H_5)$	$P(H_6)$	$P(H_7)$	$P(H_8)$
(i)	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
(ii)	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
(iii)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$
(iv)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

〈表2〉 各 경우에 있어서의 $P(H_m)$

註 7) B_m과 σ^2이 주어진 n차 다변량 正規分布

事前의 判斷確率 $P(H_m)$은 表2에 나타난다. Simulation 은 上記 (i), (ii), (iii), (iv) 에 대하여 다음과 같이 調査한다.

(I) B_m과 σ^2의 양쪽 條件付事前分布가 變則的인 σ의 크기에 대하여 事後確率 $P(H_m|y)$의 變化.

(II) σ^2의 條件付事前分布가 變則的이나 B_m이 變則的이 아닌 μ'_m에 관하여 正確한 情報가 얻어진다는 것에서 $\sigma^2\sum'_m$ 個個의 分散크기가 事後確率 $P(H_m|y)$을 補正하는 範圍와 σ와의 關係.

(III) μ'_m에 관해 틀린 情報를 얻은 경우의 事後確率 $P(H_m|y)$에 영향을 미치는 (i), (ii) (iii), (iv)에 대하여 調査한다.

8. 結果 및 考察

1) B_m과 σ^2의 양쪽 條件付事前分布가 함께 變則的인 경우 : σ값의 크기에 따라 μ''_m이 分散되어 있는 것을 알 수 있으며, $\sigma=15$ 까지는 構造變化를 나타내는 Model 의 事後確率 $P(H_6|y)$가 가장 높은 確率을 나타내며 $P(H_8|y)$가 다음으로 높은 確率을 나타낸다. Z_1, Z_3에 實際의 Knots를 設定하고 있으므로 이 두개의 Knots를 包含한 Model 의 事後確率 $P(H_6|y)$, $P(H_8|y)$가 다른 Model 의 事後確率보다 높이 나타난다. 또한 σ값이 크게 됨에 따라 $P(H_6|y)$ 값이 작게 되며, $\sigma=19$에 대해서는 $P(H_8|y)$ 값이 $P(H_6|y)$ 값을 上回한다. $\sigma=23$에서는 Z_3에서 構造變化의 點을 가진 $P(H_4|y)$이나 $\sigma=27$에서는 Z_1, Z_2의 Knots를 가진 $P(H_5|y)$가 가장 높은 確率을 나타내고 있다. 이와같은 結果가 생기는 原因으로서 생각되는 것은 date 發生의 단계에 있어서 한쪽으로 기울어지기 때문이라고 생각된다.

무엇보다도 σ값의 크기에 따라서 date 分散이 커지며 그것이 構造變化를 나타내는 Model 이되나 識別이 어려웠다.

2) 事後의 回歸係數 μ''_m 값은 B_m에 대한 條件付事前分布가 變則的인 경우에 비하여 分散이 대단히 작다. 이는 μ'_m에 관해 正確한 情報가 μ''_m에 주는 영향이 상당히 크다는 것을 나타낸다고 생각된다. 이에 비하여 各 σ는 $E(\sigma^2_{ii})=0.2$일때, $E(\sigma^2_{ii})=1.8$인 경우에도 거의 變化가 보이지 않는 것은 μ''_m 값에 대해 $E(\sigma^2_{ii})$의 크기가 거의 영향을 주지 않는 것으로 생각된다. $E(\sigma^2_{ii})$가 커짐에 따라 事後確率은 서서히 平均化되어진다. 이에 비해 各 σ에 있어서 $E(\sigma^2_{ii})$의 영향은 반드시 $E(\sigma^2_{ii})$가 커짐에 따라 構造變化를 나타내는 $P(H_6|y)$가 작아지지 않음을 알았으며, $E(\sigma^2_{ii})$의 크기가 事後確率을 補正시키는 範圍를 넘으면 事後確率은 極端的으로 기울어지는 結果를 나타낸다.

3) B_m에 대한 條件付事前分布가 變則的이 아닌 경우 : μ''_m에 관해 틀린 情報를 얻었다고 한다면 事後確率 $P(H_m|y)$에 어떤 영향이 미치는가의 結果는 (a), (b)로 나누어 나타내어 본다.

(a) μ'_m에 관해 틀린 情報가 $\mu'_m=A'_m+\eta'_m$의 경우나 $\mu'_m=A'_m-\eta'_m$의 경우나 事後確率의 變化는 대단히 비슷하다.

(b) Model 의 事後確率 $P(H_6|y)$가 다른 事後確率보다 낮아지는 結果에서 事前의 判斷確率이 事後確率에 미치는 영향이 커지며 事前判斷確率의 決定에 신중을 기할 필요가 있다고 생각된다.

REFERENCES

1. D.J. Poirier. "The Econometrics of structural change" (1976). North-Holland, American, Elsevier.
2. E.L. Halpern., "Bayesian spline Regression when the number of knots is unknown" (1973). Journal of the Royal statistical society, B., 35, 347-360.
3. B.W. Silverman., "A fast and efficient crossvalidation method for smoothing parameter choide in spline Regression (1984). Journal of American statistical Association, SEP., 1984. Volume 79, Number 387, Theory and methods section.
4. William E. Wecker and Craig F. Ansley., "The signal extraction approach to nonlinear Regression and spline smoothing" Journal of American statistical Association, Mar. 1983. Volume 78, Number 381. Theory and methods section.
5. 宮澤光一 著,「情報・決定理論」(1971), 岩波書店
6. 藤本熙, 松原望 共著,「決定の數理」(1976), 筑摩書房.