Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 3, pp. 669 ~ 678

Nonparametric detection algorithm of discontinuity points in the variance function[1)]

Jib Huh[2)]

Abstract

An algorithm to detect the number of discontinuity points of the variance function in regression model is proposed. The proposed algorithm is based on the left and right one-sided kernel estimators of the second moment function and test statistics of the existence of a discontinuity point coming from the asymptotic distribution of the estimated jump size. The finite sample performance is illustrated by simulated example.

Keywords : Asymptotic Distribution, Kernel Type Estimator, Second Moment Function, Uniform Consistency;

1. 서론

비모수적 회귀모형에서 회귀함수/평균함수의 불연속점(discontinuity point)/변화점(change point)의 추론은 널리 연구되어져 왔다. 또한, 커널형 함수 추정에서 평활량을 결정하는 띠폭의 선택(bandwidth selection), 가중최소제곱추정, 회귀함수에 대한 신뢰구간 혹은 예측구간의 추정, 품질관리 등에서 사용되는 분산함수(variance function)의 불연속점의 추론도 최근 연구되어지고 있다. 분산함수가 미지의 한 점에서 불연속점을 가질 때 그 불연속점의 위치(location)와 점프크기(jump size)에 대한 비모수적 추정법의 연구는 평균함수가 불연속일 때 비모수적 추정에 대한 연구에 비해 상대적으로 활발히 연구되어 오지는 않았다. 최근의 분산함수의 불연속점에 대한 비모수적 추정 연구는 Delgado and Hidalgo (2000)에 의해 연구되었고, Perron (2001)은 Delgado and Hidalgo (2000)의 분산함수의 불연속점의 비모수적 추정을 개선하여 연구하였고, 또한 점프크기에 대한 추정량도 제안하였다. 시계열 모형에서 Chen, Choi and Zhou (2004)는 최소제곱법으로 불연속점의 추정법과 가설검정법을 제시한 바 있다. Kang and Huh (2006)는 Perron의 추정량이 가지는 수렴속도를 개선할 수 있는 추정량을

---

1) 본 연구는 2006학년도 덕성여자대학교 연구비지원으로 이루어졌음.
2) 서울특별시 도봉구 쌍문동 419번지 덕성여자대학교 정보통계학전공 조교수
E-mail : jhuh@duksung.ac.kr

제시하였다. 이러한 분산함수의 불연속점 추정은 잔차제곱(squared residual)을 이용하고 있다.

한편, 분산함수의 불연속은 회귀함수에 기인한 것일 수도 있으며 회귀함수가 연속이고 분산함수 자체의 불연속에 기인 할 수도 있다. 전자의 경우 한 점에서 두 함수 모두 불연속이므로 분산함수의 불연속점의 위치와 점프크기 추정량은 회귀함수의 불연속점의 위치와 점프크기 추정량으로 사용하는 것이 타당하고, 후자의 경우는 회귀함수가 연속이므로 분산함수의 불연속점의 위치와 점프크기는 이차적률함수(second moment function)의 불연속점의 위치와 점프크기 추정량을 사용하는 것이 이론적으로나 실용적으로나 편리하다는 것이 Huh (2005)에 의해 연구되어졌다. 즉, 기존 연구들에서 이루어졌던 분산함수의 불연속점의 추정은 잔차제곱들의 커널추정량으로 제시되었기에 잔차제곱들을 구하기 위한 회귀함수의 커널추정량의 띠폭 선택이 선행되어야 하는 번거로움이 있으며, 그럼으로 인하여 추정의 정도(precision)가 떨어지는 현상이 생길 수 있으나, 이차적률함수의 불연속점의 추정은 반응변수의 관측치의 제곱들을 이용하여 계산되어짐으로 회귀함수를 우선적으로 추정하지 않아도 되는 용이성이 있다. 이러한 이차적률함수를 이용한 불연속점의 존재 유무에 대한 가설검정법은 Huh (2006)에 의해 연구되었다.

분산함수가 다중(multiple) 불연속점을 가지고 불연속점들의 수가 알려져 있지 않다면, 그 수를 추정하고 각각의 불연속점의 점프크기를 추정하는 것도 실제 구현에서 매우 중요하다. 이 논문에서는 분산함수의 불연속점의 수, 위치들 그리고 점프크기들을 추정하는 알고리듬을 Huh (2005, 2006)가 제시한 이차적률함수의 커널추정량을 이용한 추정과 가설검정법을 이용하여 제시하고자 한다. 지금까지 연구된 회귀함수의 불연속점들의 추정 알고리듬은 Yin (1988), Wu and Chu (1993), Qiu (1994), Qiu and Yandell (1998) 그리고 Kim, Choi and Huh (2003) 등에 의해 연구되었다. 다음 절에서는 한쪽방향커널함수(one-sided kernel function)를 이용한 이차적률함수 커널추정량으로 분산함수의 불연속점의 위치와 점프크기 추정량을 소개하고, 점프크기 추정량의 점근분포(asymptotic distribution)를 이용한 불연속점의 존재 유무에 대한 가설검정법을 소개한다. 3절에서는 2절에서 소개한 가설검정법으로 분산함수가 가지는 알려져 있지 않은 불연속점의 수를 추정하는 알고리듬을 소개할 것이다. 모의실험을 통하여 4절에서는 알고리듬의 타당성을 보이고자 한다.

2. 이차적률함수를 이용한 불연속점의 추론

다음 $\{(X_i, Y_i) : i = 1, \cdots, n\}$은 이변량 확률벡터 (X, Y)로부터의 랜덤표본이라 하고 $f(x)$를 토대(support)가 [0,1]인 X의 확률밀도함수라 하자. 이때, 회귀모형은 다음의 식

$$Y_i = m(X_i) + v^{1/2}(X_i)\epsilon_i, \quad i = 1, \cdots, n \tag{1}$$

으로 제시된다. 여기서 $m(x) = E(Y|X = x)$는 회귀함수이고 $v(x) = Var(Y|X = x)$는 분산함수이다. 오차항 ϵ_i는 $X_1, \cdots, X_n$와 독립이며 평균과 분산은 각각 0과 1이다. 식 (1)의 분산함수 v는 알려져 있지 않은 불연속점의 수 q개를 가지며 그 불연속의 위치는 각각 $\tau_j \in (0,1)$, $(j = 1, \cdots, q)$라고 가정하자. 그리고 $\Delta(\tau_j)$는 각각 τ_j에서 불연속점의 점프크기라고 하자. 또한, 분산함수 v는 두 번 미분한 함수가 연속인 $g \in C^2([0,1])$

를 이용하여 다음의 식

$$v(x) = g(x) + \sum_{j=1}^{q} \Delta(\tau_j)\, 1_{[x \geq \tau_j]}$$

으로 표현되어진다고 가정하자. 함수 a에 대하여 임의의 점 x를 중심으로 오른쪽과 왼쪽의 함수극한값들을 각각 $a_+(x) = \lim_{y \to x+} a(y)$와 $a_-(x) = \lim_{y \to x-} a(y)$으로 정의하자. 그러면, 불연속점들 $\tau_j,\ j = 1, \cdots, q,$에서 각 불연속점들의 점프크기$\Delta(\tau_j)$는 다음의 식

$$\Delta(\tau_j) = v_+(\tau_j) - v_-(\tau_j) \tag{2}$$

으로 표현된다. 식 (2)에서 점 t에서 $\Delta(t) = 0$ 이면 t에서는 분산함수는 연속이다.

분산함수의 한 점에서의 불연속은 회귀함수에 기인한 것일 수도 있으며 회귀함수는 연속이고 분산함수 자체가 불연속일 수도 있다. 전자는 회귀함수와 분산함수가 모두 같은 한 점에서 불연속이므로 분산함수의 불연속점의 위치추정량은 회귀함수의 위치추정량으로 사용하는 것이 편리하다. 이차적률함수를 다음과 같이

$$s(x) = E(Y^2 | X = x)$$

두면 후자의 경우는 회귀함수가 연속이므로 다음의 분산함수

$$v(x) = s(x) - \{m(x)\}^2$$

에 의하면 분산함수의 한 점에서의 불연속점의 위치는 분산함수 뿐만 아니라 이차적률함수를 이용하여 추정할 수도 있다. 또한 분산함수의 불연속점의 점프크기의 추정에서도 회귀함수가 연속인 경우에 식 (2)에 의하면 분산함수와 이차적률함수의 점프크기는 같음으로 이차적률함수의 점프크기의 추정으로 분산함수의 점프크기의 추정이 가능하다. 이러한 연구는 Huh (2005, 2006)에 의해 연구되었다.

어떤 점 x에서 이차적률함수 s의 왼쪽과 오른쪽 커널추정량들을 자연스럽게 각각 다음의 식들

$$\hat{s}_+(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) Y_i^2 \Big/ \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right), \tag{3}$$

$$\hat{s}_-(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i^2 \Big/ \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{4}$$

로 제시할 수 있다. 여기서 K는 확률밀도함수로서 토대가 $[0,1]$인 한쪽방향커널함수이고 h는 띠폭이며, 식 (3)와 (4)는 반응변수의 표본들의 제곱을 x를 기준으로 왼쪽과 오른쪽의 표본을 각각 평활하여 이차적률함수를 추정하고 있다. 불연속점의 위치 추정량을 제시하기 위하여 다음의 표현

$$\hat{\Delta}(x) = \hat{s}_+(x) - \hat{s}_-(x) \tag{5}$$

을 임의의 점 x에서의 분산함수의 점프크기의 추정량이라 정의할 수 있다. 어떤 개구간 $Q \subset [0,1]$에서 분산함수가 하나의 불연속점을 가진다면 그 불연속점의 위치추정량은 다음과 같이

$$\hat{\tau} = \inf\{z \in Q : |\hat{\Delta}(z)| = \sup|\hat{\Delta}(x)|\}$$

제안될 수 있다. 또한, 추정된 불연속점에서 점프크기 추정량은 자연스럽게 다음의 식으로 정의할 수 있다.

$$\widehat{\Delta}(\hat{\tau}) = \hat{s}_{+}(\hat{\tau}) - \hat{s}_{-}(\hat{\tau})$$

임의의 $x \in Q$에서 분산함수의 불연속점의 존재유무에 대한 귀무가설과 대립가설을 다음과 같이

$$H_0 : \Delta(x) = 0, \quad H_1 : \Delta(x) \neq 0 \tag{6}$$

설정할 수 있고, 귀무가설 하에서 다음 Theorem 1은 점프크기 추정량 (5)의 점근분포가 $x \in Q$에 대해 균일하게(uniformly in $x \in Q$) 성립함을 보였다. 먼저, 몇 가지 조건을 제시하고자 한다.

(A1) $\inf_{x \in [0,1]} f(x) > 0$이고 f는 구간 $[0,1]$에서 Lipschitz 1차 연속조건(Lipschitz condition of order 1)을 만족
(A2) 사차적률함수(fourth moment function) $\kappa(x) = E(Y^4 | X = x)$는 구간 $[0,1]$에서 Lipschitz 1차 연속조건을 만족
(A3) K는 구간 $[0,1]$에서 Lipschitz 1차 연속조건을 만족하고, $K(0) > 0$이고 $0 < u \leq 1$에 대하여 $K(u) \geq 0$를 만족
(A4) $n \to \infty$일 때 $h \to 0$, $nh/\log n \to \infty$, $nh^3 \to 0$을 만족
(A5) 임의의 $\zeta > 0$에 대해, $E[|Y|^{4+\zeta} | X = x] < \infty$을 만족

Theorem 1 식 (6)의 귀무가설 하에서 위 조건 (A1)-(A5)가 만족되면 다음의 결과

$$\sqrt{nh}\,(\widehat{\Delta}(x)) \xrightarrow{d} N(0, 2\frac{\kappa(x)}{f(x)} \int_0^1 \{K(u)\}^2 du)$$

가 $x \in Q$에 대해 균일하게 성립한다.

증명) Stute (1982)의 Theorem 1.2에 의하여 다음이 성립한다.

$$\sup_{x \in Q} \left| \frac{1}{nh} \sum_{i=1}^{n} K(\pm \frac{X_i - x}{h}) - f(x) \right| = O_p\left(\sqrt{\frac{\log n}{nh}} + h \right).$$

그러므로, x에 대해 균일하게 다음의 표현

$$\sqrt{nh}\,(\widehat{\Delta}(x)) = \sqrt{nh}\,\widetilde{\Delta}(x)(1 + o_p(1))$$

이 성립한다. 여기서

$$\widetilde{\Delta}(x) = \frac{1}{nhf(x)} \sum_{i=1}^{n} \left\{ K(\frac{X_i - x}{h}) - K(\frac{x - X_i}{h}) \right\} Y_i^2$$

이다. 한편, x에 대해 균일하게 다음의 결과들

$$\sqrt{nh}\,E(\widetilde{\Delta}(x)) = 0$$

과

$$nh\,Var(\widetilde{\Delta}(x)) = \frac{1}{h\{f(x)\}^2} E\left[\left\{ K(\frac{X_1 - x}{h}) - K(\frac{x - X_1}{h}) \right\} \kappa(X_1) \right] \tag{7}$$

$$= 2\frac{\kappa(x)}{f(x)} \int_0^1 \{K(u)\}^2 du (1 + o(1))$$

이 성립한다. 마지막으로 x에 대해 균일하게 $\sqrt{nh}\,\widetilde{\Delta}(x)$가 Lyapounov의 조건을 만족함을 보이고자 한다. 위 결과 (7)로부터 x에 대해 균일하게 $Var(\sqrt{nh}\,\widetilde{\Delta}(x)) = O(1)$

이므로, 임의의 $\zeta > 0$에 대해 $n \to \infty$일 때 다음의 식

$$L_n(x) = \sum_{i=1}^{n} E\left[\left|\frac{1}{f(x)}\left\{K(\frac{X_i - x}{h}) - K(\frac{x - X_i}{h})\right\} Y_i^2\right|^{2+\zeta}\right] \to 0$$

이 성립함을 보여주면 된다. 가정 (A5)에 의하면, x에 대해 균일하게 다음의 결과

$$L_n(x) \le nE\left[\left|\frac{1}{f(x)}\left\{K(\frac{X_1 - x}{h}) - K(\frac{x - X_1}{h})\right\}\right|^{2+\zeta} |Y_1^2|^{2+\zeta}\right]$$

$$= O\left(nh(\frac{1}{nh})^{2+\zeta}\right)$$

가 성립된다. 따라서, 가정 (A4)에 의하여 Lyapounov의 조건이 만족된다. ■

위 Theorem 1은 어떤 개구간 Q의 임의의 점 x에 대해 가설 (6)를 검정하기 위한 검정통계량으로 다음의 통계량

$$\sqrt{nh}\frac{\sqrt{f(x)}\hat{\Delta}(x)}{\sqrt{2\kappa(x)\int_0^1 \{K(u)\}^2 du}} \tag{8}$$

을 제시할 수 있고, 이 검정통계량의 점근분포는 Theorem 1에 의해 표준정규분포 $N(0,1)$이 된다. 검정통계량 (8)에 내재해 있는 장애모수(nuisance parameter)인 $f(x)$와 $\kappa(x)$를 커널추정량으로 다음과 같이

$$\hat{f}(x) = \frac{1}{nb}\sum_{i=1}^{n} L(\frac{X_i - x}{b})$$

$$\hat{\kappa}(x) = \frac{1}{nb}\sum_{i=1}^{n} L(\frac{X_i - x}{b}) Y_i^4 / \frac{1}{nb}\sum_{i=1}^{n} L(\frac{X_i - x}{b})$$

제시하자. 여기서 커널함수 L은 확률밀도함수로 토대가 [-1, 1]이고 b는 h와 다른 띠폭이다. 한편, Stute (1986)에 의해 $\hat{f}(x)$는 x에 대해 균일일치성(uniform consistency)이 규명되어졌고, Huh (2006)는 $\hat{\kappa}(x)$의 x에 대한 균일일치성을 보였다. 이 결과에 의하여 다음의 Corollary가 성립된다.

Corollary 1 Theorem 1의 조건을 만족하고 커널함수 L의 1차 연속조건을 만족하며 띠폭 b가 $n \to \infty$일 때 $b \to 0$, $nb/\log n \to \infty$을 만족할 때, (6)의 귀무가설 하에서 다음의 결과

$$T(x) = \sqrt{nh}\,\frac{\sqrt{\hat{f}(x)}\,\hat{\Delta}(x)}{\sqrt{2\hat{\kappa}(x)\int_0^1 \{K(u)\}^2 du}} \xrightarrow{d} N(0,1) \tag{9}$$

가 성립한다.

3. 가설검정을 통한 불연속점의 수 추정

이 절에서는 2절에서 소개한 점프크기 추정량과 그 점근성질을 이용한 불연속점의 존재유무에 대한 가설검정법으로 분산함수가 가지고 있는 알려져 있지 않은 불연속점의 수 q를 추정하는 알고리듬을 생각해 보자.

설명의 편의를 위하여 점프크기 $\Delta(\tau_j), j=1,\cdots,q,$ 들이 다음의 조건

$$|\Delta(\tau_1)| \geq |\Delta(\tau_2)| \geq \cdots \geq |\Delta(\tau_q)|$$

을 만족한다고 가정하자. 직관적인 관점에서, 점프크기 추정값 $\hat{\Delta}(x)$의 절대값이 가장 크다면, 점 x에서 가장 큰 점프크기를 가지는 불연속점이라고 추정할 수 있다. 즉, 집합 $Q_1 = \{x: h \leq x_k \leq 1-h\}$에서 가장 큰 점프크기를 가지는 불연속의 위치 τ_1의 추정량 $\hat{\tau}_1$은 다음과 같이

$$\hat{\tau}_1 = \arg\max_{Q_1} |\hat{\Delta}(x)|$$

정의할 수 있다. 여기서 $Q_1 = [h, 1-h]$으로 선택한 것은 토대 $[0,1]$에서 0과 1은 또 다른 형태의 불연속점으로 이해될 수 있고 이러한 이유로 여러 연구 논문(Müller (1992), Loader (1996), Jose and Ismail (1999), Huh and Carrière (2002), Huh and Park (2004))에서 불연속점의 존재 구간으로 $[h, 1-h]$가 사용되었기 때문이다.

두 번째로 큰 점프크기를 가지는 불연속점의 위치 τ_2의 추정량은 다음과 같이

$$\hat{\tau}_2 = \arg\max_{Q_2} |\hat{\Delta}(x)|$$

정의할 수 있다. 여기서 $Q_2 = \{x: h \leq x \leq 1-h,\ \ |x-\hat{\tau}_1| \geq 2h\}$이다. 한 불연속점의 점프크기 추정치들이 인접해 있는 x의 점프크기에 영향을 주기 때문에 하나의 불연속점 주변의 점들에서도 불연속점으로 판정될 가능성이 있다. 따라서, Jose와 Ismail (1999)이 언급하였던 것처럼 근접해 있는 두 개의 불연속점들 사이의 거리들은 $2h$보다 크다는 가정이 필요하게 된다. 따라서 Q_2는 $\hat{\tau}_1$으로부터 $2h$ 떨어진 점들로 구성된 집합이다. 이와 같은 방법으로 계속해서 다음으로 큰 점프크기를 가지는 τ_3, τ_4 등의 추정량으로 $\hat{\tau}_3$, $\hat{\tau}_4$ 등을 순차적으로 다음과 같이

$$\begin{aligned}\hat{\tau}_3 &= \arg\max_{Q_3} |\hat{\Delta}(x)|,\\ \hat{\tau}_4 &= \arg\max_{Q_4} |\hat{\Delta}(x)|,\\ &\vdots\end{aligned}$$

제시할 수 있다. 여기서 $Q_j = \{x: h \leq x \leq 1-h,\ \ |x-\hat{\tau}_i| \geq 2h,\ i=1,\cdots,j-1\}$, $j=1,2,\cdots$는 j번째로 점프크기가 큰 불연속점의 존재영역에 대한 구간이다.

불연속점들의 위치 추정량 $\hat{\tau}_j$, $j=1,2,\cdots$들을 기초로 하여 불연속점의 개수 q를 다음의 통계적 가설검정법의 결과에 의해 추정할 수 있다. 먼저, 어떤 불연속점 τ_j에서 불연속점의 존재여부에 대한 가설검정을 하여 보자. 다음의 가설

$$H_0: \Delta(\tau_j) = 0, \quad H_1: \Delta(\tau_j) \neq 0 \tag{10}$$

에 대해 τ_j의 추정량 $\hat{\tau}_j$을 이용하여 검정통계량 (9)을 바탕으로 위 가설에 대한 검정통계량으로 다음과 같이

$$T(\hat{\tau}_j) = \sqrt{nh}\frac{\sqrt{\hat{f}(\hat{\tau}_j)}\hat{\Delta}(\hat{\tau}_j)}{\sqrt{2\hat{\kappa}(\hat{\tau}_j)\int_0^1 \{K(u)\}^2 du}} \tag{11}$$

제시할 수 있다. 2절의 Corollary 1에 의하여 귀무가설 $\Delta(\tau_j) = 0$하에서 $T(\hat{\tau}_j) \xrightarrow{d} N(0,1)$이다. 그러므로 가설 (10)에 대한 기각역은 유의수준 α에서 다음과 같이

$$|T(\hat{\tau}_j)| > z_{\alpha/2} \tag{12}$$

주어진다. 여기서 z_α는 표준정규분포의 (1-α) 분위수이다. 한편, τ_j에서 점프크기 $\Delta(\tau_j)$에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이

$$\hat{\Delta}(\hat{\tau}_j) \pm z_{\alpha/2}\left(2\frac{\hat{\kappa}(\hat{\tau}_j)}{\hat{f}(\hat{\tau}_j)}\int_0^1 \{K(u)\}^2 du\right)^{1/2}$$

주어진다. 기각역 (12)을 이용하여 불연속점으로 추정된 $\hat{\tau}_j$, $j = 1,2,\cdots$ 에서 불연속점의 존재여부에 대한 가설검정을 할 수 있다. 즉, $\hat{\tau}_j$에 대해 순차적으로 ($\hat{\Delta}(\hat{\tau}_j)$의 크기 순으로) 가설검정을 해 나가면서 귀무가설이 기각되지 않은 단계가 $j = r$이라면 미지의 불연속점의 수 q의 추정량을 $\hat{q} = r - 1$로 정의할 수 있다.

4. 모의실험

이 절에서는 3절에서 제안한 분산함수의 불연속점의 수의 추정에 대한 알고리듬을 모의실험을 통하여 구현해 보고자 한다. 이 모의실험을 위하여 설명변수인 X_i의 확률밀도함수가 [0, 1]구간에서의 균등분포인 경우를 생각해 보자.

회귀함수는 m은 다음과 같이

$$m(x) = x(2x^2 - 3x + 1)$$

선택하였다. 한편, 오차항 ϵ은 표준정규분포 $N(0,1)$를 따르고 불연속점을 가지는 분산함수는 v는 다음과 같이

$$v(x) = \begin{cases} 0.32, & x < 0.35 \\ 0.02, & 0.35 \le x < 0.75 \\ 0.62, & x \ge 0.75 \end{cases} \tag{13}$$

주어졌다. 위 분산함수의 불연속점의 수 $q = 2$이고 각 불연속점의 위치는 $\tau_1 = 0.75$, $\tau_2 = 0.35$이며 이 불연속점에서의 점프크기는 $\Delta_1 = 0.6$, $\Delta_2 = -0.3$이다. 이 모형에서 회귀함수는 연속이고 분산함수만 불연속이기에 분산함수의 불연속점의 위치와 점프크기는 이차적률함수의 불연속점의 위치와 점프크기는 같다. 위 (13)의 분산함수의 불연속점의 수의 추정과 그 위치의 추정의 정도를 보기 위하여 표본 $n = 1000$으로 반복 1000회를 하였다.

식 (3)와 (4)의 왼쪽과 오른쪽 이차적률함수의 커널추정량에 사용된 한쪽방향커널함수는 biweight 커널함수를 이용하여 만든 것으로 다음의 커널

$$K(x) = \frac{15}{8}(1-x^2)^2,\ 0 < x < 1$$

을 선택하였다. 또한 (11)의 검정통계량에 내에 있는 추정된 불연속점에서의 커널형 확률밀도함수와 사차적률함수의 추정량에 이용된 커널은 biweight 커널로서 다음과 같이

$$L(x) = \frac{15}{16}(1-x^2)^2,\ |x| < 1$$

선택하였다. 한편, 불연속점의 위치와 점프크기 추정에 쓰인 띠폭 h와 확률밀도함수와 사차적률함수의 커널 추정에 사용된 띠폭 b는 각각 0.15로 선택하였다. 띠폭들의 선택은 커널형 불연속점의 추정에서 중요한 부분이지만 본 논문에서는 추가적인 연구로 남겨두고자 한다.

불연속점의 위치추정량을 구하기 위하여 3절에서 언급한 것처럼 구간 Q_1은 $[h, 1-h]$로 선택하였고, 불연속점의 위치를 추정하기 위하여 구간 $[0,1]$을 100등분한 $t_k = k/100,\ k = 1, \cdots, 100$에서 각각 점프크기를 구하고 3절의 불연속점의 수를 추정하는 방법을 적용하였다. 유의수준은 $\alpha = 0.05$를 선택하였다.

다음은 위 모형에 의한 모의실험의 결과들이다. 먼저, 불연속점의 수 $q=2$에 대한 추정치의 평균과 표준오차는 각각 2.017과 0.004088이다. 불연속점의 수의 추정량 $\hat{q}$의 모의실험 결과의 분포를 보기 위해 다음의 <표 1>에서 추정된 $\hat{q}$의 빈도를 제시하였다.

<표 1> 불연속점의 수의 추정치 $\hat{q}$의 빈도

$\hat{q}$의 추정치	2	3
빈도	983	17

불연속점의 수뿐만 아니라 그 위치의 추정의 정도도 중요하다. 다음의 <표 2>와 <표 3>은 첫 번째 위치 추정량 $\hat{\tau}_1$과 두 번째 위치 추정량 $\hat{\tau}_2$의 모의실험 결과의 분포를 보기 위해 추정된 각 t_k에서의 빈도를 보여주고 있다. 이 분포를 보면 추정량 $\hat{\tau}_1$의 추정 정도가 $\hat{\tau}_2$의 추정 정도 보다는 우수해 보이고 있으며 두 추정량이 편의를 보이고 있지만 각각의 참값인 0.75와 0.35를 잘 추정하고 있다.

<표 2> 첫 번째 불연속점의 위치 추정치 $\hat{\tau}_1$의 t_k점 위에서의 빈도

k	74	75	76	77	78	79	80	81
빈도	1	856	112	23	5	1	1	1

<표 3> 두 번째 불연속점의 위치 추정치 $\hat{\tau}_2$의 t_k점 위에서의 빈도

k	22	25	26	28	29	30	31	32	33	34	35
빈도	1	1	1	2	3	5	5	10	44	125	803

5. 결론

회귀모형에서 회귀함수의 불연속점의 추정은 지금까지 많은 연구가 되어 왔고, 분산함수의 추정은 회귀함수의 추정과 더불어 그 중요함에 비하여 상대적으로 불연속점의 추정의 연구가 많이 이루어지지 않았다. 이 논문에서는 지금까지 분산함수가 하나의 불연속점을 가질 때 연구되어진 커널형 불연속점의 추정과 가설검정 연구를 이용하여 하나 이상의 불연속점을 가지는 분산함수의 불연속점의 수를 추정하는 알고리듬을 제시하였다. 이 알고리듬에서 불연속점의 위치와 점프크기 추정량에 사용된 띠폭과 제안된 검정통계량에서 불연속점의 위치 추정량의 확률밀도함수와 사차적률함수의 커널추정량에 사용되는 띠폭의 선택에 대해서는 차후에 연구될 필요가 있는 부분이다.

참고문헌

1. Chen, G., Choi, Y. K. and Zhou, Y. (2004). Nonparametric estimation of structural change points in volatility models for time series, *J. Econometrics, forthcoming.*
2. Delgado, M. A. and Hidalgo, J. (2000). Nonparametric inference on structural breaks, *J. Econometrics*, 96, 113-144.
3. Huh, J. (2005). Nonparametric detection of a discontinuity point in the variance function with the second moment function, *J. Korean Data & Information Science Society*, 16, 591-601.
4. Huh, J. (2006). Testing the Existence of a Discontinuity Point in the Variance Function, *J. Korean Data & Information Science Society*, 17, 707-716.
5. Huh, J. and Carrière, K. C. (2002). Estimation of regression functions with a discontinuity in a derivative with local polynomial fits, *Statist. Probab*. Lett., 56, 329-343.
6. Huh, J. and Park, B. U. (2004). Detection of change point with local polynomial fits for random design case, *Australian and New Zealand J. Statist.*, 46, 425-441.
7. Jose, C. T. and Ismail, B. (1999). Change points in nonparametric regression functions, *Commun. Statist.-Theory Meth.*, 28, 1883-1902.
8. Kang, K. C. and Huh, J. (2006). Nonparametric estimation of the variance function with a change point, *J. Korean Statist*. Soc., 35, 1-24.

9. Kim, J. T., Choi, H. and Huh, J. (2003). Detection of change-points by local linear regression fit, *The Korean Commun. Statist.*, 10, 31-38.
10. Loader, C. R. (1996). Change point estimation using nonparametric regression, *Ann. Statist.*, 24, 1667-1678.
11. Müller, H G. (1992). Change-points in nonparametric regression analysis, *Ann. Statist.*, 20, 737-761.
12. Perron, B. (2001). Jumps in the volatility of financial markets, *Mimeo*, available at http://mapageweb.umontreal.ca/perrob.
13. Qiu, P. (1994). Estimation of the number of jumps of the jump regression functions, *Commun. Statist.-Theory Meth.* 23, 2141-2155.
14. Qiu, P. and Yandell, B. (1998). A local polynomial jump detection algorithm in nonparametric regression, *Technometrics*, 40, 142-152.
15. Stute, W. (1982). A law of the logarithm for kernel density estimators, *Ann. Probab.*, 10, 414-422.
16. Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function, *Ann. Statist.*, 21, 1545-1566.
17. Yin, Q. (1988). Detection of the number, locations and magnitudes of jumps, *Commun. Statist. Stoch. Models*, 4, 445-455.

[2007년 7월 접수, 2007년 8월 채택]