

# Discontinuous log-variance function estimation with log-residuals adjusted by an estimator of jump size

Hyeseon Hong<sup>a</sup> · Jib Huh<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Duksung Women's University

(Received January 25, 2017; Revised February 22, 2017; Accepted February 27, 2017)

---

## Abstract

Due to the nonnegativity of variance, most of nonparametric estimations of discontinuous variance function have used the Nadaraya-Watson estimation with residuals. By the modification of Chen *et al.* (2009) and Yu and Jones (2004), Huh (2014, 2016a) proposed the estimators of the log-variance function instead of the variance function using the local linear estimator which has no boundary effect. Huh (2016b) estimated the variance function using the adjusted squared residuals by the estimated jump size in the discontinuous variance function. In this paper, we propose an estimator of the discontinuous log-variance function using the local linear estimator with the adjusted log-squared residuals by the estimated jump size of log-variance function like Huh (2016b). The numerical work demonstrates the performance of the proposed method with simulated and real examples.

Keywords: discontinuity point, jump size, local linear estimator, log-squared residual, variance function

---

## 1. 서론

표본  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ 은 랜덤표본으로 확률벡터  $(X, Y)$ 로부터 추출되었고, 공변량  $X$ 의 확률밀도함수는  $f(x)$ 이고 토대(support)는  $[0, 1]$ 이라 하자. 회귀함수와 분산함수를 각각  $m(x) = E(Y|X = x)$ 와  $v(x) = \text{Var}(Y|X = x)$ 라 두면, 회귀모형은 다음과 같이 표현된다.

$$Y_i = m(X_i) + v^{\frac{1}{2}}(X_i)\epsilon_i, \quad i = 1, 2, \dots, n. \quad (1.1)$$

여기서 서로 독립인 오차항  $\epsilon_i$ 는  $X_1, \dots, X_n$ 과 독립이며 평균과 분산은 각각 0과 1이다.

임의의 점  $x$ 에서 분산함수의 좌극한값과 우극한값을 각각  $v_-(x)$ 와  $v_+(x)$ 라 하면, 분산함수가 한 점  $\tau$ 에서 불연속인 경우 점프크기는

$$\Delta = v_+(\tau) - v_-(\tau) \quad (1.2)$$

가 된다. 분산함수가 불연속점을 가진다면  $|\Delta| > 0$ 이고, 그렇지 않다면  $\Delta = 0$ 이다. Huh (2014) 등에서 분산함수가 불연속인 이유는 회귀함수가 불연속인 경우와 회귀함수는 연속이고 분산함수 자체가 불

---

This research was supported by the Duksung Women's University Research Grants 2016.

<sup>1</sup>Corresponding author: Department of Statistics, Duksung Women's University, 33, Samyang-ro 144-gil, Dobong-gu, Seoul 01369, Korea. E-mail: [jhuh@duksung.ac.kr](mailto:jhuh@duksung.ac.kr)

연속인 경우로 구분하여 설명하고, 전자의 경우는 회귀함수의 불연속점의 추정으로 분산함수의 불연속점을 추정할 수 있음을 언급하였다. 본 연구에서는 식 (1.1)의 회귀모형에서 회귀함수는 연속이며 분산함수는 한 점  $\tau$ 에서 불연속점을 가지는 후자의 경우를 고려하고자 한다.

회귀모형에서 연속인 분산함수의 커널함수를 이용한 비모수적 추정 연구는 Chen 등 (2009), Hall과 Carroll (1989), Hall 등 (1990), Gasser 등 (1986), Müller와 Stadtmüller (1987), Rice (1984), Ruppert 등 (1997)과 Yu와 Jones (2004) 등이 있다. 대부분의 연구들은 회귀함수의 추정량을 이용한 잔차제곱들을 이용하여 분산함수를 추정하였고, Hall 등 (1990)은 반응변수의 표본들의 차의 제곱들을 활용하여 분산함수를 비모수적으로 추정하였다. Chen 등 (2009)과 Yu와 Jones (2004)은 로그분산함수를 추정 대상으로 삼았다.

분산함수가 불연속인 경우에 회귀함수의 커널추정량에 의한 잔차제곱의 국소상수항추정량(local constant estimator)인 Nadaraya-Watson 추정량으로 Kang과 Huh (2006)가 불연속인 분산함수의 추정에 대해 연구하였다. 회귀함수가 연속인 경우에, 회귀함수의 추정 정도(precision)에 의존하는 잔차를 이용하지 않기 위해 Huh (2005)는 분산함수의 불연속점의 추정을 이차적률함수의 불연속점의 추정으로 제안하였다. Huh (2009, 2016a)는 Yu와 Jones (2004)의 카이제곱분포의 가능도함수를 이용한 국소선형추정량(local linear estimator)의 로그분산함수의 추정법을 활용하여 분산함수의 불연속점 추정에 대한 연구를 하였다. 한편, Chen 등 (2009)은 분산함수가 연속일 때, 잔차제곱들의 로그변환을 이용한 국소선형추정량으로 로그분산함수의 커널추정량을 제안하였다. Huh (2014)는 Chen 등 (2009)이 제안한 방법인 로그잔차제곱들을 이용하여 불연속인 분산함수의 커널추정량을 제안하여 Huh (2016a)의 방법과 비교 연구하였다.

회귀함수가 불연속일 때, 불연속점의 점프크기추정량으로 반응변수의 표본들을 연속인 회귀함수로부터 추출된 표본인 것처럼 수정하여 Kang 등 (2000)은 불연속 회귀함수의 추정을 연구하였다. Huh (2016b)는 Kang 등 (2000)의 방법을 이용하여 분산함수의 불연속점의 점프크기추정량으로 잔차제곱을 수정하여 불연속 분산함수의 추정을 연구하였다. 본 연구에서는 회귀함수의 추정량에 의한 로그잔차제곱들을 Kang과 Huh (2006)가 제안한 불연속점의 점프크기추정량으로 로그잔차제곱을 연속인 로그분산함수로부터의 로그잔차제곱인 것처럼 수정하여 불연속인 로그분산함수를 국소선형추정량으로 추정하고자 한다.

2절에서는 로그분산함수의 불연속점 추정량과 점프크기추정량을 이용한 로그잔차제곱의 수정 방법과 이를 이용한 불연속 로그분산함수의 추정법을 제안하고, 3절에서는 모의실험을 통하여 Huh (2014)의 불연속 로그분산함수의 추정량과 비교하고자 한다. 분산함수 혹은 불연속 분산함수의 비모수적 추정 연구에 흔히 쓰였던 실제자료에 적용한 결과를 4절에 제시하고자 한다.

## 2. 로그잔차제곱의 수정과 불연속 로그분산함수의 추정

본 연구에서는 불연속 로그분산함수의 추정에 대한 연구이기에, 식 (1.2)의 불연속 분산함수의 점프크기 대신 로그분산함수의 불연속점에서의 점프크기를 다음과 같이 고려한다. 식 (1.1)의 불연속 분산함수의 로그 변환된 함수를  $s(x) = \log v(x)$ 라 하고, 로그분산함수  $s(x)$ 는 다음을 만족한다고 가정하자.

$$s(x) = w(x) + \delta \times I[\tau \leq x \leq 1]. \quad (2.1)$$

이때 함수  $w(x)$ 는 연속이다. 분산함수와 로그분산함수는 로그함수의 연속 성질에 의해 분산함수와 로그분산함수는 동일한 점  $\tau$ 에서 불연속점을 가지게 된다. 불연속점  $\tau$ 에서 로그분산함수  $s(x)$ 의 점프크기는  $\delta$ 라 하자. 즉,

$$\delta = s_+(\tau) - s_-(\tau) \quad (2.2)$$

가 되고, 분산함수의 점프크기  $\Delta$ 가 0이 아니면 로그분산함수의 점프크기  $\delta$ 도 0이 아니다. 여기서  $s_+(\tau)$ 와  $s_-(\tau)$ 는 불연속점  $\tau$ 에서 각각 로그분산함수  $s(x)$ 의 우극한값과 좌극한값이다.

식 (2.2)의 불연속점의 점프크기  $\delta$ 를 추정하기 위하여 Kang과 Huh (2006)가 제안한 다음의 분산함수의 추정량  $\hat{v}_+(x)$ 과  $\hat{v}_-(x)$

$$\hat{v}_{\pm}(x) = \frac{\sum_{i=1}^n K\left(\pm \frac{X_i - x}{h_1}\right) \hat{R}_i^2}{\sum_{i=1}^n K\left(\pm \frac{X_i - x}{h_1}\right)} \quad (2.3)$$

을 고려하였다. Kang과 Huh (2006)은 국소상수항추정량  $\hat{m}$ 으로 잔차  $\hat{R}_i = Y_i - \hat{m}(X_i)$ 를 구하였다. 식 (2.3)의  $h_1$ 은 평활량인 띠폭이며  $K$ 는 토대가  $[0, 1]$ 인 한쪽방향커널함수(one-sided kernel function)이다.  $\hat{v}_+(x)$ 과  $\hat{v}_-(x)$ 은 국소상수항추정량인 Nadaraya-Watson 추정량으로 어떤 점  $x$ 에서 각각  $v(x)$ 의 오른쪽 추정량과 왼쪽 추정량으로 정의한 것이다. 이 추정량들은 먼저 회귀함수를 추정한 후 잔차제곱들을 이용하여 토대가  $[0, 1]$ 인 한쪽방향커널함수로 어떤 점  $x$ 에서 오른쪽의 표본과 왼쪽의 표본을 각각 이용한 것이다.

Kang과 Huh (2006)는 식 (2.3)의 추정량들을 이용하여 점  $x$ 의 점프크기추정량을  $\hat{\Delta}(x) = \hat{v}_+(x) - \hat{v}_-(x)$ 라 정의하여  $|\hat{\Delta}(x)|$ 가 최대가 되는 점  $x$ 를 불연속점 추정량  $\hat{\tau}$ 으로 제안하였고 추정된 불연속점  $\hat{\tau}$ 에서 점프크기  $\hat{\Delta}(\hat{\tau})$ 를 식 (1.2)의 점프크기  $\Delta$ 의 추정량으로 제안하였다. 본 연구에서는 로그분산함수의 점프크기인 식 (2.2)의  $\delta$ 의 추정량으로

$$\hat{\delta}(\hat{\tau}) = \hat{s}_+(\hat{\tau}) - \hat{s}_-(\hat{\tau}) \quad (2.4)$$

을 고려하였다. 여기서  $\hat{s}_{\pm}(\hat{\tau})$ 은  $\log \hat{v}_{\pm}(\hat{\tau})$ 이다.

Huh (2014)는 Chen 등 (2009)의 방법론을 활용하여 로그잔차제곱  $\log \hat{R}_i^2$ 로 국소선형추정량을 이용하여 한 점  $\tau$ 에서 불연속인 로그분산함수의 추정을 다음과 같이 제안하였다. 먼저, 잔차제곱  $\hat{R}_i^2$ 이 0에 매우 근접한 경우의 문제를 해결하기 위하여 식 (2.5)에는 다음과 같이 주어진 어떤 작은 실수  $c$ 에 대하여

$$\log \tilde{R}_i^2 = \begin{cases} \log \hat{R}_i^2, & \hat{R}_i^2 > c, \\ \log c, & \text{그 외} \end{cases} \quad (2.5)$$

를 고려하였다. Chen 등 (2009)은 잔차제곱  $\hat{R}_i^2$ 이 0에 매우 근접한 경우의 문제를 해결하기 위하여  $\log(\hat{R}_i^2 + n^{-1})$ 을 사용하였다. 여기서  $n$ 은 표본의 수이다. 로그분산함수가 연속일 때,  $E(\log(\varepsilon_i^2/d)) = 0$ 가 되게 하는 상수  $d$ 를 이용하여  $\log(Y_i - m(X_i))^2 = \log(dv(x)) + \log(\varepsilon_i^2/d)$ 라 표현하고  $\log(dv(x))$ 를 국소선형적합으로 다음의 식

$$\sum_{i=1}^n \left\{ \log \tilde{R}_i^2 - \beta_0 - \beta_1(X_i - x) \right\}^2 L_h^*(X_i - x; \hat{\tau}) \quad (2.6)$$

를 최소로 하여 추정하였다. 즉, 식 (2.6)을 최소로 하는 해를  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ 라 하면  $dv(x)$ 의 추정을  $\widehat{dv}(x) = e^{\hat{\beta}_0}$ 로 정의하여 분산함수를  $\hat{v}_{\text{CCP}}(x; \hat{\tau}) = e^{\hat{\beta}_0}/\hat{d}$ 로 추정하였다. 여기서,  $d$ 의 추정  $\hat{d}$ 은 Chen 등 (2009)이 제안한 것으로

$$\hat{d} = \left[ n^{-1} \sum_{i=1}^n \tilde{R}_i^2 \exp(-\widehat{dv}(X_i)) \right] \quad (2.7)$$

이다. 식 (2.6)에서  $h$ 는 식 (2.3)의 띠폭  $h_1$ 과는 다른 띠폭이며  $L_h^*(u-x; t)$ 는 다음과 같다.

$$L_h^*(u-x; t) = \begin{cases} \frac{1}{h} L\left(\frac{u-x}{h}\right) I[x-h \leq u < t], & t-h \leq x < t, \\ \frac{1}{h} L\left(\frac{u-x}{h}\right) I[t \leq u < x+h], & t \leq x < t+h, \\ \frac{1}{h} L\left(\frac{u-x}{h}\right), & \text{그 외.} \end{cases}$$

함수  $L$ 은 토대  $[-1, 1]$ 를 가지는 커널함수이고,  $I$ 는 표시함수(indicator function)이다. 불연속 로그분산함수의 추정으로는 분산함수의 추정량  $\hat{v}_{\text{CCP}}(x; \hat{\tau})$ 에 로그를 취한 형태로

$$\hat{s}_{\text{CCP}}(x; \hat{\tau}) = \hat{\beta}_0 - \log \hat{d} \quad (2.8)$$

로 정의하였다. 식 (2.6)의 해  $\hat{\beta}_0$ 는 명시적 형태로 다음과 같이 표현됨을 쉽게 알 수 있다.

$$\hat{\beta}_0 = n^{-1} \sum_{i=1}^n \frac{\{u_2(x; \hat{\tau}) - u_1(x; \hat{\tau})(X_i - x)\} L_h^*(X_i - x; \hat{\tau}) \log \tilde{R}_i^2}{u_2(x; \hat{\tau}) u_0(x; \hat{\tau}) - \{u_1(x; \hat{\tau})\}^2}, \quad (2.9)$$

여기서  $u_k(x; \hat{\tau}) = n^{-1} \sum_{i=1}^n (X_i - x)^k L_h^*(X_i - x; \hat{\tau})$ ,  $k = 0, 1, 2$ 이다.

점프크기추정량을 이용한 수정된 로그잔차제곱으로 불연속 로그분산함수의 추정량을 제안하기 위하여, 식 (2.5)의 로그잔차제곱  $\log \tilde{R}_i^2$ 를 추정된 점프크기  $\hat{\delta}(\hat{\tau})$ 로 다음과 같이 수정된 로그잔차제곱인  $\log \tilde{R}_i^{*2}$ 을 생각하자.

$$\log \tilde{R}_i^{*2} = \log \tilde{R}_i^2 - \hat{\delta}(\hat{\tau}) \times I[\hat{\tau} \leq X_i \leq 1], \quad i = 1, \dots, n. \quad (2.10)$$

식 (2.10)의 수정된 로그잔차제곱  $\log \tilde{R}_i^{*2}$ 는 식 (2.1)의 연속인 함수  $w(x)$ 로부터 얻어진 로그잔차제곱인 것으로 간주할 수 있기에 로그분산함수를 추정하기 위하여 국소선형적합을 위해 다음과 같은 식

$$\sum_{i=1}^n \left\{ \log \tilde{R}_i^{*2} - \beta_0 - \beta_1(X_i - x) \right\}^2 L_h(X_i - x) \quad (2.11)$$

을 생각할 수 있다. 식 (2.11)을 최소로 하는  $\beta_0$ 의 해  $\hat{\beta}_0^*$ 는 명시적 형태로 다음과 같이 표현된다.

$$\hat{\beta}_0^* = n^{-1} \sum_{i=1}^n \frac{\{u_2(x) - u_1(x)(X_i - x)\} L_h(X_i - x) \log \tilde{R}_i^{*2}}{u_2(x) u_0(x) - \{u_1(x)\}^2}, \quad (2.12)$$

여기서  $u_k(x) = n^{-1} \sum_{i=1}^n (X_i - x)^k L_h(X_i - x)$ ,  $k = 0, 1, 2$ 이고,  $L_h(x) = h^{-1} L(x/h)$ 이다.

식 (2.1)의 연속함수인  $w$ 에 대하여  $v^*(x) = e^{w(x)}$ 라 하면, 식 (2.12)의  $\hat{\beta}_0^*$ 를 이용하여  $dv^*(x)$ 의 추정으로  $\widehat{dv^*(x)} = e^{\hat{\beta}_0^*}$ 로 정의하여  $v^*$ 를  $\hat{v}_{\text{ALR}}^*(x; \hat{\tau}) = e^{\hat{\beta}_0^*} / \hat{d}^*$ 로 추정한다. 여기서,  $\hat{d}^*$ 은 식 (2.7)과 유사한 형태인

$$\hat{d}^* = \left[ n^{-1} \sum_{i=1}^n \tilde{R}_i^{*2} \exp\left(-dv^*(X_i)\right) \right]$$

을 사용한다. 불연속 로그분산함수  $s(x)$ 의 추정은 불연속 분산함수  $s(x)$ 과  $w(x)$ 의 관계 (2.1)을 이용하여 제안한 분산함수의 추정량  $\hat{v}_{\text{ALR}}^*(x; \hat{\tau})$ 에 로그를 취하고 식 (2.4)의 점프크기추정량을 더해준 다음의

$$\hat{s}_{\text{ALR}}(x; \hat{\tau}) = \hat{\beta}_0^* - \log \hat{d}^* + \hat{\delta}(\hat{\tau}) \times I[\hat{\tau} \leq x \leq 1] \quad (2.13)$$

로 정의한다. 식 (1.1)의 불연속 분산함수의 추정량으로는  $s(x)$ 와의 관계를 이용하여  $\hat{v}_{\text{ALR}}(x; \hat{\tau}) = e^{\hat{s}_{\text{ALR}}(x; \hat{\tau})}$ 로 추정할 수 있다.

제한한 위 추정량  $\hat{s}_{\text{ALR}}(x; \hat{\tau})$ 은 식 (2.8)의 추정량  $\hat{s}_{\text{CCP}}(x; \hat{\tau})$ 과 달리 로그잔차제곱을 두 부분으로 분리하지 않고 불연속점 추정량 부근에서 양쪽의 수정된 로그잔차제곱을 모두 이용하게 되는 장점을 가지게 되어 추정의 정도가 좋을 것이다.

### 3. 모의실험

본 절에서는 2절에서 제안한 식 (2.13)의 불연속 로그분산함수의 커널추정량  $\hat{s}_{\text{ALR}}$ 과 Chen 등 (2009)의 추정량을 이용한 식 (2.8)의 Huh (2014)의 불연속 로그분산함수 추정량인  $\hat{s}_{\text{CCP}}$ 의 소표본 추정 정도를 비교해보기 위하여 모의실험을 실시하였다.

설명변수  $X$ 의 분포는 토대  $[0, 1]$ 인 균등분포 선택하였고, 회귀함수는 Huh (2014, 2016b)에서 고려하였던

$$m(x) = 4x + 4e^{-100(x-0.5)^2}, \quad 0 \leq x \leq 1$$

를 선택하였다. 분산함수가 한 점에서 불연속인 두 가지 경우를 다음과 같이 고려하였다.

$$v_1(x) = \begin{cases} x^2, & 0 \leq x \leq 0.65, \\ 25(1-x)^2, & 0.65 < x \leq 1, \end{cases}$$

$$v_2(x) = \begin{cases} 25x^2, & 0 \leq x \leq 0.35, \\ (1-x)^2, & 0.35 < x \leq 1. \end{cases}$$

첫 번째 분산함수  $v_1$ 은  $\tau = 0.65$ 에서 불연속점을 가지고 이때 점프크기는  $\Delta = 2.64$ 이다. 로그분산함수  $s_1(x) = \log v_1(x)$ 의 점프크기는  $\tau = 0.65$ 에서  $\delta = 1.980797$ 이다. 첫 번째 분산함수와 로그분산함수  $v_1$ 과  $s_1$ 의 점프크기는 양의 값을 가지기에, 점프크기가 음수가 되는 경우를 두 번째 분산함수에서 고려해보았다. 두 번째 분산함수와 로그분산함수인  $v_2(x)$ 와  $s_2(x) = \log v_2(x)$ 는  $\tau = 0.35$ 에서 불연속이고 점프크기는 각각  $\Delta = -2.64$ 와  $\delta = -1.980797$ 이다.

식 (1.1)의 회귀모형의 오차  $\varepsilon_i$ 의 분포는 표준정규분포를 선택하였고, 표본의 수를 500으로 하고 반복은 1,000회 실시하였다. 두 로그분산함수 추정량  $\hat{s}_{\text{CCP}}$ 과  $\hat{s}_{\text{ALR}}$ 은 2절에서 설명하였듯이 같은 불연속점 추정량  $\hat{\tau}$ 를 사용하기에 두 추정량의 비교를 위한 모의실험이므로  $s_1$ 과  $s_2$ 의 불연속점  $\tau$ 들은 참값을 이용하였다. 비교하고자 하는 두 추정량  $\hat{s}_{\text{CCP}}(x; \tau)$ 와  $\hat{s}_{\text{ALR}}(x; \tau)$ 에 쓰이는 식 (2.5)의 로그잔차제곱의 상수  $c$ 는 Huh (2014)에서와 같이  $c = 0.001$ 을 선택하였다.

식 (2.3)의 Kang과 Huh (2006)가 제안한 불연속점의 점프크기의 추정을 위해 쓰이는 한쪽방향커널함수로는 Epanechnikov 커널을 토대가  $[0, 1]$ 이 되도록 다음

$$K(x) = \frac{3}{2} (1-x^2) \times I[0 \leq x \leq 1]$$

을 선택하였다. 회귀함수와 분산함수를 추정하기 위한 식 (2.9)와 (2.12)의 커널함수는 다음과 같이 Epanechnikov 커널

$$L(x) = \frac{3}{4} (1-x^2) \times I[-1 \leq x \leq 1]$$

을 선택하였다.

**Table 3.1.** The minimum estimated ISE over  $h$  with the standard errors given in parentheses and the minimizing bandwidth  $h$  for the case of  $s_1$

$h_m$	$\hat{s}_{CCP}(x; \tau)$		$\hat{s}_{ALR}(x; \tau)$		
	$h$	ISE	$h_1$	$h$	ISE
0.05	0.08	0.1695 (0.0014)	0.05	0.07	0.1572 (0.0013)
			0.10	0.07	0.1565 (0.0013)
			0.15	0.07	0.1562 (0.0013)
0.10	0.08	0.2451 (0.0019)	0.05	0.08	0.2340 (0.0018)
			0.10	0.08	0.2332 (0.0018)
			0.15	0.08	0.2337 (0.0018)
0.15	0.10	0.6272 (0.0033)	0.05	0.09	0.6226 (0.0032)
			0.10	0.09	0.6213 (0.0032)
			0.15	0.09	0.6257 (0.0032)

ISE = integrated squared error.

**Table 3.2.** The minimum estimated ISE over  $h$  with the standard errors given in parentheses and the minimizing bandwidth  $h$  for the case of  $s_2$

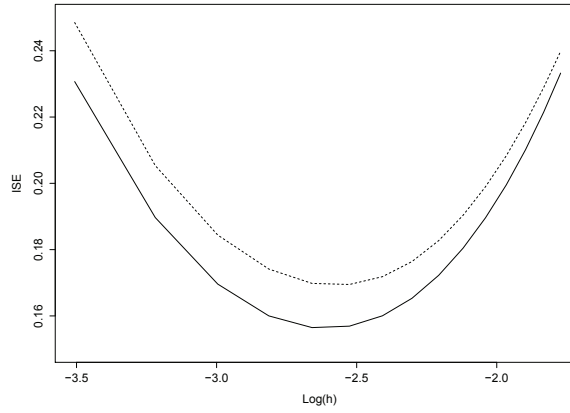
$h_m$	$\hat{s}_{CCP}(x; \tau)$		$\hat{s}_{ALR}(x; \tau)$		
	$h$	ISE	$h_1$	$h$	ISE
0.05	0.08	0.1680 (0.0014)	0.05	0.07	0.1560 (0.0013)
			0.10	0.07	0.1553 (0.0013)
			0.15	0.07	0.1551 (0.0013)
0.10	0.08	0.2450 (0.0019)	0.05	0.08	0.2344 (0.0018)
			0.10	0.08	0.2338 (0.0018)
			0.15	0.08	0.2343 (0.0018)
0.15	0.10	0.6302 (0.0031)	0.05	0.09	0.6261 (0.0031)
			0.10	0.09	0.6248 (0.0031)
			0.15	0.09	0.6291 (0.0031)

ISE = integrated squared error.

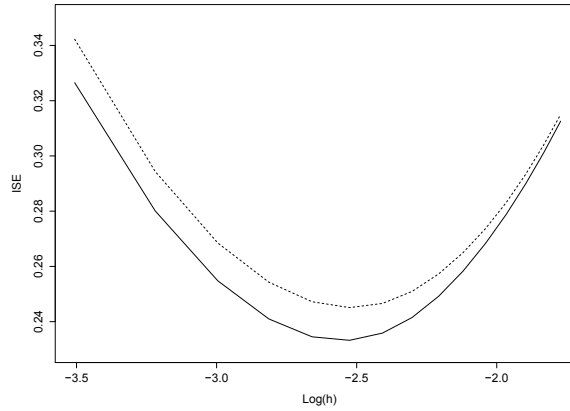
식 (2.5)의 로그잔차제곱을 구하기 위하여 회귀함수  $m$ 은 국소선형추정량을 이용하였으며, 이때 쓰이는 띠폭  $h_m$ 은 0.05, 0.10과 0.15를 고려하여 다양한 띠폭에 의해 구해진 로그잔차제곱이 로그분산함수의 두 추정량  $\hat{s}_{CCP}(x; \tau)$ 와  $\hat{s}_{ALR}(x; \tau)$ 의 정도에 어떤 영향을 미치는지를 살펴보았다. 한편, 제안한 추정량  $\hat{s}_{ALR}(x; \tau)$ 을 위한 수정 로그잔차제곱을 구하기 위하여 식 (2.4)의 점프크기추정량  $\hat{\delta}(\tau)$ 에 쓰이는 띠폭  $h_1$ 도 0.05, 0.10과 0.15를 사용하였다. 불연속 로그분산함수의 추정에 쓰이는 띠폭  $h$ 를 다양하게 변화를 주면서 로그분산함수의 두 추정량  $\hat{s}_{CCP}(x; \tau)$ 와  $\hat{s}_{ALR}(x; \tau)$ 를 비교하기 위하여 적분제곱오차(integrated squared error)들을 계산하였다.

로그분산함수 모형  $s_1$ 과  $s_2$ 에 대한 모의실험 결과를 각각 Tables 3.1과 3.2에서 보여주고 있다. 1,000번의 반복으로 계산된 Huh (2014)의  $\hat{s}_{CCP}(x; \tau)$ 와 제안한 추정량  $\hat{s}_{ALR}(x; \tau)$ 의 추정된 적분제곱오차를 최소로 하는 띠폭  $h$ 와 그 때의 추정된 적분제곱오차를 보여주고 있다. 괄호 안은 추정된 적분제곱오차들 각각의 표준오차들이다. Tables 3.1과 3.2는 로그잔차제곱을 구하기 위한 회귀함수의 추정에 쓰이는 띠폭  $h_m$ 이 커짐으로써 두 추정량  $\hat{s}_{CCP}(x; \tau)$ 와  $\hat{s}_{ALR}(x; \tau)$ 의 추정된 적분제곱오차들도 증가하는 경향을 보여주고 있다. 로그잔차제곱을 구하기 위한 회귀함수의 추정에 쓰인 띠폭  $h_m$  세 가지 모두에서 점프크기추정량  $\hat{\delta}(\tau)$ 에 쓰이는 띠폭  $h_1$ 의 선택에 상관없이 제안한 추정량  $\hat{s}_{ALR}(x; \tau)$ 의 최소의 추정된 적분제곱오차들이  $\hat{s}_{CCP}(x; \tau)$ 의 것들보다 작은 경향을 보여주고 있다.

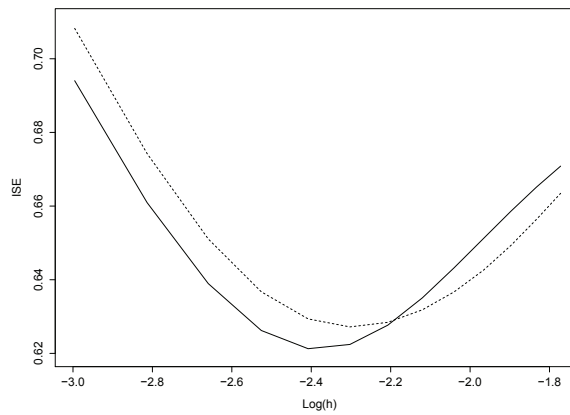
로그분산모형  $s_1$ 에 대하여, Figure 3.1에서는 회귀함수의 추정에 쓰인 띠폭  $h_m$ 인 0.05, 0.01과 0.15일 때 로그분산함수의 추정을 위한 다양한 띠폭  $h$ 에 대하여 구한 추정된 적분제곱오차들의 변화를 보여



(a)  $h_m = 0.05$  and  $h_1 = 0.10$



(b)  $h_m = 0.10$  and  $h_1 = 0.10$



(c)  $h_m = 0.15$  and  $h_1 = 0.10$

**Figure 3.1.** The ISEs as function of log-bandwidth  $h$  for the case of  $s_1$ . The ISEs of  $\hat{s}_{CCP}(x; \tau)$  and  $\hat{s}_{ALR}(x; \tau)$  are represented by the dotted and the solid line respectively. ISE = integrated squared error.

주고 있다. Figure 3.1에서는 제안한 추정량  $\hat{s}_{ALR}(x; \tau)$ 을 위한 수정 로그잔차제곱을 구하기 위하여 식 (2.4)의 점프크기추정량  $\hat{\delta}(\tau)$ 에 쓰이는 띠폭  $h_1$ 은 0.10인 경우의 결과만 제시하였다. 그 외의  $h_1$ 이 0.05와 0.15인 경우의 결과들은  $h_1$ 이 0.10인 경우와 유사하기에 생략하였다. 또한, 점프크기가 음의 값을 가지는 로그분산모형  $s_2$ 에 대해서도 띠폭의 변화에 따른 추정된 적분제곱오차의 변화도 모형  $s_1$ 의 결과들과 비슷하여 생략하였다.

고려한 로그분산함수 모형  $s_1$ 과  $s_2$ 에 대한 모의실험 결과에 의하면 점프크기추정량에 의해 수정된 로그잔차제곱의 국소선형추정량  $\hat{s}_{ALR}(x; \tau)$ 가 Chen 등 (2009)이 제안한 로그잔차제곱을 이용한 불연속 로그분산함수 추정량  $\hat{s}_{CCP}(x; \tau)$ 보다 우수함을 알 수 있다.

#### 4. 실제자료분석

Ruppert 등 (1997)은 커널함수를 이용한 국소선형추정량을 이용하여 분산함수의 비모수적 추정법을 제안하고 light detection and range (LIDAR) 자료에 적용하여 분산함수를 추정하였고, 이 과정에서 LIDAR의 분산함수의 불연속점의 존재에 대하여 언급하였다. Huh (2016a)와 Kang과 Huh (2006)는 분산함수의 불연속점과 불연속 분산함수를 추정에 대한 실제자료분석으로 LIDAR 자료를 사용하였다. 자료의 수가 221개인 LIDAR 자료는 R 소프트웨어의 SemiPar library에 공개되어 있는 자료이다.

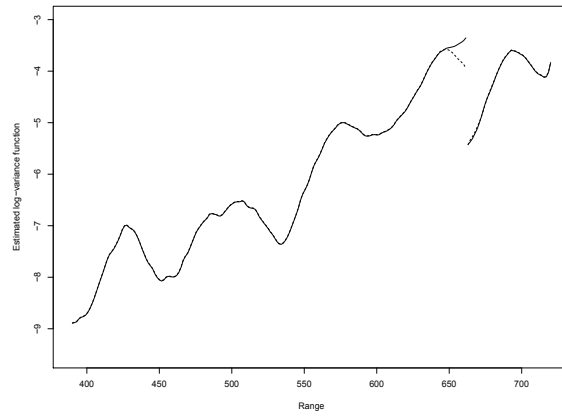
불연속 분산함수 추정량  $\hat{s}_{CCP}(x; \hat{\tau})$ 과  $\hat{s}_{ALR}(x; \hat{\tau})$ 의 계산을 위하여 필요한 잔차를 구하기 위하여  $m$ 은 국소선형추정량을 이용하였다. 이때 쓰인 띠폭  $h_m$ 은 Huh (2016a)의 5.2절에 설명하였듯이 교차타당성(cross-validation)에 의하여 구해진 15를 사용하였다. 한편, 추정량  $\hat{s}_{ALR}(x; \hat{\tau})$ 에 사용되는 수정된 로그잔차제곱을 구하기 위한 점프크기추정치인 띠폭  $h_1$ 은 Huh (2016a)의 5.2절의 식 (22)의 교차타당성을 최소로 하는 27을 사용하였다. 이때 추정된 불연속점과 점프크기는 각각  $\hat{\tau} = 661.5$ 와  $\hat{\delta}(\hat{\tau}) = -2.115286$ 이다. LIDAR 자료와 추정된 회귀함수 및 잔차제곱은 Huh (2016a)의 Figures 7과 8을 참조 하길 바란다.

아래 Figure 4.1은 추정량  $\hat{s}_{CCP}(x; \hat{\tau})$ 과  $\hat{s}_{ALR}(x; \hat{\tau})$ 를 추정된 로그분산함수의 부드러움 정도에 따라 비교해보기 위하여 띠폭  $h$ 를 20, 40, 60을 선택하여 추정한 결과들을 각각 보여주고 있다. 이들 Figure 내의 실선들은 수정된 로그잔차제곱을 이용한  $\hat{s}_{ALR}(x; \hat{\tau})$ 이고 점선은  $\hat{s}_{CCP}(x; \hat{\tau})$ 을 보여주고 있다. 이들 추정량은 추정된 불연속점인  $\hat{\tau} = 661.5$ 를 기준으로 좌우 띠폭의 크기 내의 구간인  $[\hat{\tau} - h, \hat{\tau} + h]$ 에서만 추정치가 차이가 있다. 그 외의 구간에서는 두 추정치는 동일하다. 이러한 현상은  $\hat{s}_{CCP}(x; \hat{\tau})$ 는 실질적으로  $\hat{\tau}$ 를 기준으로 로그잔차제곱들을 좌우로 분리하여 독립적으로 로그분산함수를 추정하지만  $\hat{s}_{ALR}(x; \hat{\tau})$ 는  $\hat{\tau}$ 와 상관없이 수정된 로그잔차제곱들 전체를 사용하여 로그분산함수를 추정하기 때문이다. 따라서  $\hat{s}_{ALR}(x; \hat{\tau})$ 은 마치 연속인 로그분산함수를 추정하고 추정된 로그분산함수를  $\hat{\tau}$ 를 기준으로 오른쪽 부분을  $\hat{\delta}(\hat{\tau}) = -2.115286$ 만큼 이동한 것이기에  $\hat{\tau}$  주변의 로그분산함수의 추정치가 LIDAR 자료의 로그분산함수를 좀더 잘 설명해주고 있다고 판단된다.

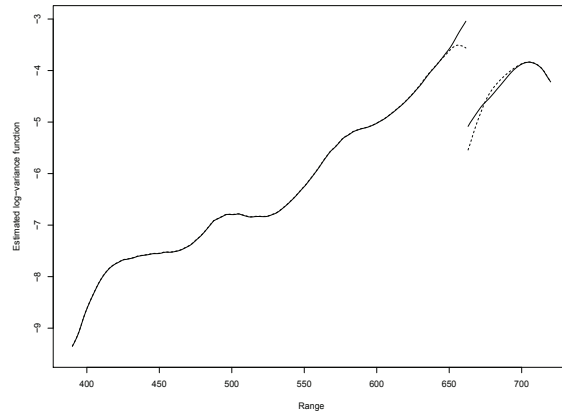
#### References

- Chen, L., Chen, M., and Peng, M. (2009). Conditional variance estimation in heteroscedastic regression models, *Journal of Statistical Planning and Inference*, **139**, 236–245.
- Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression, *Biometrika*, **73**, 625–634.
- Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean, *Journal of the Royal Statistical Society B*, **51**, 3–14.
- Hall, P., Kay, J. W., and Titterton, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika*, **77**, 521–528.

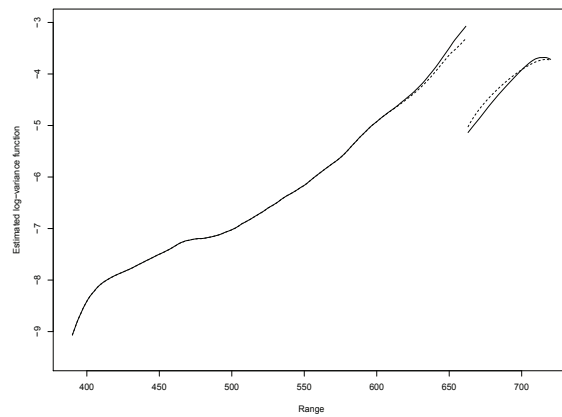




(a) Bandwidth 20



(b) Bandwidth 40



(c) Bandwidth 60

**Figure 4.1.** The estimated log-variance functions of LIDAR. The estimated log-variance functions of  $\hat{s}_{CCP}(x; \tau)$  and  $\hat{s}_{ALR}(x; \tau)$  are represented by the dotted and the solid line respectively. LIDAR = light detection and range.

- Huh, J. (2005). Nonparametric detection of a discontinuity point in the variance function with the second moment function, *Journal of the Korean Data & Information Science Society*, **16**, 591–601.
- Huh, J. (2009). Testing a discontinuity point in the log-variance function based on likelihood, *Journal of the Korean Data & Information Science Society*, **20**, 1–9.
- Huh, J. (2014). Comparison study on kernel type estimators of discontinuous log-variance, *Journal of the Korean Data & Information Science Society*, **25**, 87–95.
- Huh, J. (2016a). Estimation of a change point in the variance function based on the  $x^2$ -distribution, *Communications in Statistics - Theory and Methods*, **45**, 4937–4968.
- Huh, J. (2016b). Nonparametric estimation of the discontinuous variance function using adjusted residuals, *Journal of the Korean Data & Information Science Society*, **27**, 111–120.
- Kang, K. H. and Huh, J. (2006). Nonparametric estimation of the variance function with a change point, *Journal of the Korean Data & Information Science Society*, **35**, 1–24.
- Kang, K. H., Koo, J. Y., and Park, C. W. (2000). Kernel estimation of discontinuous regression functions, *Statistics and Probability Letters*, **47**, 277–285.
- Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis, *Annals of Statistics*, **15**, 610–625.
- Rice, J. (1984). Bandwidth choice for nonparametric regression, *Annals of Statistics*, **12**, 1215–1230.
- Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997). Local polynomial variance-function estimation, *Technometrics*, **39**, 262–273.
- Yu, K. and Jones, M. C. (2004). Likelihood-Based Local Linear Estimation of the Conditional Variance Function, *Journal of the American Statistical Association*, **99**, 139–144.

# 점프크기추정량에 의한 수정된 로그잔차를 이용한 불연속 로그분산함수의 추정

홍혜선<sup>a</sup>, 허집<sup>a,1</sup>

<sup>a</sup>덕성여자대학교 정보통계학과

(2017년 1월 25일 접수, 2017년 2월 22일 수정, 2017년 2월 27일 채택)

---

## 요약

분산함수가 불연속점을 가지는 경우, 대부분의 비모수적 함수 추정 연구에서 분산함수가 음수 값을 갖지 않기에 잔차제곱을 이용한 Nadaraya-Watson 추정량인 국소상수항추정량을 이용하였다. 한편, Huh (2014, 2016a)는 Chen 등 (2009)과 Yu와 Jones (2004)의 연구를 바탕으로 불연속 분산함수를 로그 변환한 로그분산함수를 추정 대상으로 삼아 잔차제곱이나 로그잔차제곱으로 경계점 문제를 가지지 않는 국소선형추정량을 이용하여 비모수적으로 추정하였다. Huh (2016b)는 불연속점에서 점프크기추정량을 활용하여 잔차제곱을 분산함수가 연속인 회귀모형에서 얻어진 잔차제곱인 것처럼 수정한 후 이들을 이용하여 불연속 분산함수의 추정을 연구하였다. 본 연구에서는 불연속 로그분산함수의 점프크기추정량을 이용하여 로그잔차제곱을 수정하고 불연속 로그분산함수를 국소선형추정량을 이용하여 추정하고자 한다. 제안된 추정량의 우수성을 모의실험을 통하여 Chen 등 (2009)의 로그분산함수 추정량을 이용한 Huh (2014)의 불연속 로그분산함수 추정량과 비교하고 실제자료에 적용하고자 한다.

주요용어: 국소선형추정량, 분산함수, 로그잔차제곱, 불연속점, 점프크기

---

본 연구는 덕성여자대학교 2016년도 교내연구비 지원에 의해 수행되었음.

<sup>1</sup>교신저자: (01369) 서울시 도봉구 삼양로 144길 13, 덕성여자대학교 정보통계학과.

E-mail: jhuh@duksung.ac.kr