2002 Proceedings of the Autumn Conference, Korean Statistical Society

Nonparametric Estimation of Discontinuous Variance Function in Regression Model

Kee-Hoon Kang[1] Jib Huh[2]

ABSTRACT

We consider an estimation of discontinuous variance function in nonparametric heteroscedastic random design regression model. We first propose estimators of a change point and jump size in variance function and then construct an estimator of entire variance function. We examine the rates of convergence of these estimators and give results on their asymptotics. Numerical work reveals that the effectiveness of change point analysis in variance function estimation is quite significant.

KEY WORDS. *Discontinuity point; Jump size; Nonparametric regression; One-sided kernel; Rate of convergence.*

1. Introduction

In most nonparametric regression function estimation, the variance of errors is assumed to be homogeneous or heterogeneous with smooth function. It is of great meaningful to consider estimation of these variance functions since it is important in its own right and in various applications. The estimation of variance function is needed in some bandwidth selection procedures, weighted least squares estimation, the construction of confidence and prediction intervals for mean function and quality control, etc. These applications are discussed in Carroll (1982), Carroll and Ruppert (1988) and Müller (1988).

In this paper we consider an approach for estimating variance function in the following random design regression model:

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

where $m(x) = E(Y|X = x)$ is the mean regression function, $v(x)$ is the conditional variance of Y given $X = x$ and conditional on $X_1, \ldots, X_n$, ε_i's are independent random variables with mean 0 and variance 1. Let f be the design density of X. Difference from former works is we do not assume that the variance function is continuous. That is, it is assumed that a change point exists for the variance function v at some point τ in the interior of the support of f. In fact, relatively little attention is paid in this problem compared with its importance.

Our approach on this problem is similar to one for estimating discontinuous regression function, which was discussed in Müller (1992) and Loader (1996), etc. One-sided kernel

[1]Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Korea.
[2]Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.

regression estimates based on squared residuals are used to estimate the location of a change point and jump size. It is shown that the resulting estimator of the change point is consistent with n^{-1} convergence rate and jump size estimator has asmyptotic normality. For estimating variance function itself, we use Nadaraya-Watson type estimator with the data set splitted by the estimated location of the change point. We show that the rate of convergence of the integrated squared error of the estimated variance function does not depend on the rate of the estimated location of the change point.

2. Estimators and theoretical properties

We begin by stating a set of assumptions on unknown functions in the model (1):

(A.1) There exists a constant C such that

$$|v(x) - v(y)| \leq C|x - y| \text{ whenever } (x - \tau)(y - \tau) > 0, \tag{2}$$

i.e. v satisfies the Lipschitz condition of order 1 over $[0, \tau)$ and $(\tau, 1]$. The jump size at the change point τ in v is given by $\Delta = v_+(\tau) - v_-(\tau)$, where $v_+(\tau) = \lim_{x \to \tau+} v(x)$, $v_-(\tau) = \lim_{x \to \tau-} v(x)$. Without loss of generality we may assume $0 < \Delta < \infty$, since the case of $-\infty < \Delta < 0$ can be treated in the same way.

(A.2) The design density f is supported on $[0, 1]$ with $\inf_{x \in [0,1]} f(x) > 0$, and satisfies the Lipschitz condition of order 1.

(A.3) The regression function m satisfies the Lipschitz condition of order 1.

We apply Nadaraya-Watson smoother with a one-sided kernel function on squared residuals to detect the location of a change point and the jump size in variance function. To get residuals, we first define

$$\widehat{m}(x) = \frac{1}{nh_1} \sum_{j=1}^{n} L\left(\frac{X_j - x}{h_1}\right) Y_j \Big/ \frac{1}{nh_1} \sum_{j=1}^{n} L\left(\frac{X_j - x}{h_1}\right)$$

as the estimator of m, where L is a nonnegative kernel function with support $[-1, 1]$ and $h_1 = h_{1n}$ is a sequence of bandwidths satisfying:

(A.4) The function L is symmetric and satisfies the Lipschitz condition of order 1.

(A.5) $h_1 \to 0$, and $nh_1/\log n \to \infty$, as $n \to \infty$.

Using the squared residuals $R_i = (Y_i - \widehat{m}(X_i))^2$, $i = 1, \cdots, n$, we define

$$\widehat{v}_+(x) = \frac{1}{nh_2} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_2}\right) R_i \Big/ \frac{1}{nh_2} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_2}\right) \tag{3}$$

$$\widehat{v}_-(x) = \frac{1}{nh_2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_2}\right) R_i \Big/ \frac{1}{nh_2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_2}\right). \tag{4}$$

Here, K is a kernel function with support $[0, 1]$ and $h_2 = h_{2n}$ is a sequence of bandwidths, which satisfy the following assumptions:

(A.6) The function K satisfies the Lipschitz condition of order 1, and $K(0) > 0, K(u) \geq 0$ for $0 < u \leq 1$ with $\int_0^1 K(u)du = 1$.

(A.7) $h_2 \to 0$, $nh_2/\log n \to \infty$, and $nh_2^3 \to 0$, as $n \to \infty$.

The estimators $\widehat{v}_+(x)$ and $\widehat{v}_-(x)$ are based on residuals only at the right and the left side of x, respectively. We estimate the jump size at a point x by taking the differences of these two estimators: $\widehat{\Delta}(x) = \widehat{v}_+(x) - \widehat{v}_-(x)$. A reasonable estimator $\widehat{\tau}$ of τ is the value of x that maximizes $\widehat{\Delta}(x)$. Let $Q \subset (0,1)$ be a closed interval such that $\tau \in Q$. Define

$$\widehat{\tau} = \inf\left\{ z \in Q : \widehat{\Delta}(z) = \sup_{x \in Q} \widehat{\Delta}(x) \right\}$$

for the location of the change point τ. We assumed continuity of regression function in (A.3). When both of regression function and variance function have the same change point, τ can be estimated based on the estimation of regression function. An estimator of the jump size Δ may be obtained by

$$\widehat{\Delta}(\widehat{\tau}) = \widehat{v}_+(\widehat{\tau}) - \widehat{v}_-(\widehat{\tau}). \tag{5}$$

Now, we propose an estimator for the variance function using the estimated location of the change point $\widehat{\tau}$. Let W be a kernel function with support $[-1,1]$ satisfying the following condition.

(A.8) The function W is a symmetric probability density function, and satisfies the Lipschitz condition of order 1.

Using the squared residuals R_i, $i = 1, \cdots, n$, we define

$$\widehat{v}(x;\widehat{\tau}) = \frac{1}{nh}\sum_{i=1}^{n} W^*\left(\frac{X_i - x}{h};\widehat{\tau}\right) R_i \bigg/ \frac{1}{nh}\sum_{j=1}^{n} W^*\left(\frac{X_i - x}{h};\widehat{\tau}\right). \tag{6}$$

Here W^* is defined by

$$W^*\left(\frac{u-x}{h};\widehat{\tau}\right) = \begin{cases} W\left(\frac{u-x}{h}\right) 1_{[x-h,\widehat{\tau}]}(u), & \widehat{\tau} - h \leq x \leq \widehat{\tau}, \\ W\left(\frac{u-x}{h}\right) 1_{[\widehat{\tau},x+h]}(u), & \widehat{\tau} \leq x \leq \widehat{\tau} + h, \\ W\left(\frac{u-x}{h}\right), & \text{otherwise} \end{cases}$$

with a sequence of bandwidths $h = h_n$ satisfying the following condition.

(A.9) $h \to 0$, and $nh/\log n \to \infty$, as $n \to \infty$.

The following theorem describes weak convergence of the sequence of the process $\{\varphi_n(z) : -M \leq z \leq M\}$ where

$$\varphi_n(z) = nh_2\left\{\widehat{\Delta}\left(\tau + \frac{z}{n}\right) - \widehat{\Delta}(\tau)\right\}, \tag{7}$$

and $M < \infty$. The process φ_n lies in the space, denoted by $C([-M,M])$, of continuous functions defined on $[-M,M]$. Let $\xrightarrow{\mathcal{W}}$ denote weak convergence in the space $C([-M,M])$ and $\kappa(x) = E[(Y - m(X))^4 | X = x]$. To obtain the theorem, consider the following additional assumptions:

(A.10) The function κ satisfies the Lipschitz condition of order 1.

(A.11) $E(|Y|^{4+\zeta}|X=x)<\infty$, for all x and some positive ζ.

Theorem 1. *Suppose that the assumptions (A.1)-(A.7), (A.10) and (A.11) are satisfied. Then,*

$$\varphi_n(z) \xrightarrow{\mathcal{W}} \varphi(z) = -\Delta K(0)|z| + \sigma W(z) \tag{8}$$

where $W(z)$ is a two-sided Brownian motion defined in Bhattacharya and Brockwell (1976), and

$$\sigma = \sqrt{\frac{4\kappa(\tau)}{f(\tau)}}K(0). \tag{9}$$

Remark. *Since the conditional central fourth moment κ depends on v, it is highly possible that κ also has a change point at τ. In that case, the asymptotic variance part of the limit process φ in (8) is slightly changed. Define $\kappa_+(\tau) = \lim_{x\to\tau+}\kappa(x)$ and $\kappa_-(\tau) = \lim_{x\to\tau-}\kappa(x)$. Then, $\kappa(\tau)$ in (9) is replaced by $\kappa_+(\tau)$ when $z\geq 0$, and by $\kappa_-(\tau)$ when $z<0$.*

Next, we describe the asymptotic distribution of $\widehat{\tau}$. For doing this, let Z be the maximizer of the process φ. By $K(0)>0$, the limit process φ in (8) has a unique maximizer with probability one. See Remark 5.3 in Bhattacharya and Brockwell (1976) for more detail. Let Z_n be the maximizer of φ_n. By construction,

$$\widehat{\tau} = \tau + \frac{Z_n}{n}.$$

By Theorem 5 in Whitt (1970), the weak convergence in Theorem 1 can be extended to the space $\mathcal{C}(-\infty,\infty)$. Theorem 3 in Bhattacharya and Brockwell (1976) then gives $Z_n \xrightarrow{\mathcal{D}} Z$, where Z_n is the global maximizer of φ_n on $(-\infty,\infty)$. Therefore, we have the following corollary.

Corollary 1. *Suppose that the assumptions in Theorem 1 are satisfied. Then,*

$$n(\widehat{\tau}-\tau) \xrightarrow{\mathcal{D}} \text{argmax}_{z\in(-\infty,\infty)}\{-\Delta K(0)|z| + \sigma W(z)\}.$$

Raimondo (1998) showed that the minimax optimal rate for the location problem is n^{-1} for a class of regression functions. Although the interesting function is variance, our proposed estimator gets the rate of convergence n^{-1} according to Corollary 1. The asymptotic variance depends on $f(\tau)$. The corollary tells us that the change point estimator gets more stable as the density at the change point increases. As another consequence of Theorem 1, the following corollary describes the asymptotic distribution of the estimator $\widehat{\Delta}(\widehat{\tau})$ for the jump size as defined at (5).

Corollary 2. *Under the assumptions of Theorem 1,*

$$\sqrt{nh_2}(\widehat{\Delta}(\widehat{\tau})-\Delta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\frac{\kappa(\tau)}{f(\tau)}\int_0^1 \{K(u)\}^2 du\right).$$

If κ has a change point at τ, it is easy to see that the asymptotic variance in Corollary 2 is replaced by $\{f(\tau)\}^{-1}\{\kappa_+(\tau) + \kappa_-(\tau)\}\int_0^1 \{K(u)\}^2 du$.

We now turn to the asymptotic property of the estimator of the variance function. Theorem 2 gives the rate of global L_p convergence of the estimator $\widehat{v}(\cdot;\widehat{\tau})$ in (6).

Theorem 2. *Suppose that the assumptions (A.1)-(A.11) are satisfied. Then, for* $p \geq 1$,

$$\int_0^1 \left|\widehat{v}(x;\widehat{\tau}) - v(x)\right|^p dx = O_P\left(h_1^{2p} + \left(\frac{\log n}{nh_1}\right)^p\right) + O_P\left(\left(\frac{\log n}{n\sqrt{h_1 h}}\right)^p + \left(h_1\sqrt{\frac{\log n}{nh}}\right)^p\right)$$
$$+\left|\widehat{\tau} - \tau\right| O_P(1) + \left|\widehat{\tau} - \tau\right|^p O_P(h^{-p}) + O_P\left(h^p + \left(\sqrt{\frac{\log n}{nh}}\right)^p\right). \quad (10)$$

The first two terms in (10) depend on the rate of global L_∞ convergence of the estimator $\widehat{m}$. Since Nadaraya-Watson type smoother guarantees the nonnegativity of the estimated variance function based on squared residuals, we proposed that type smoother in (6) and results in the order of squared bias in the last term is h^p rather than h^{2p}. In the L_2 sense, if we choose two bandwidths as $h_1 \sim h$, the rate in (10) is then $O_P(h^2 + \log n/(nh))$ which means that minimizing the terms in (10) does not depend on the rate of the estimator $\widehat{\tau}$. It implies that the estimator $\widehat{v}(\cdot;\widehat{\tau})$ has the exactly same rate as for $\widehat{v}(\cdot;\tau)$, where the location of the change point is known.

3. Numerical properties

To investigate the practical performance of the proposed estimator defined in Section 3, a simulation study is carried out. For this, response-predictor pairs (X_i, Y_i), $i = 1, \ldots, n$ are generated according to the prescription (1) for various m and v. Kernel functions $L(u) = \frac{15}{16}(1-u^2)^2 I(|u| \leq 1)$ and $K(u) = \frac{15}{8}(1-u^2)^2 I(0 \leq u \leq 1)$ are used to estimate the mean regression function and the change point, respectively. For the estimation of variance function, kernel function W is taken to the same as L.

In this section, we shall report only results for the following four settings. We get very similar impression for the other cases. Cases (a), (b) and (c) represent typical types of regression function and have homogeneous variance function with a change point. Case (d) represent the case which has nonhomogeneous variance function with a change point. Throughout, the distribution of predictor variable is assumed to be uniform in $(0,1)$ and $\widehat{\tau}$ is estimated on $[h, 1-h]$. Integrated squared error(ISE), the measure of performance, is estimated on the interval $[0,1]$, using the trapezoidal rule. Average values are obtained from 500 simulations.

(a) $m_1(x) = x$, $v_1(x) = 0.01\, I(x \leq 0.5) + 0.09\, I(x > 0.5)$.
(b) $m_2(x) = 4x(1-x)$, $v_2(x) = 0.01\, I(x \leq 0.75) + 0.16\, I(x > 0.75)$.
(c) $m_3(x) = 5x(2x^2 - 3x + 1)$, $v_3(x) = 0.49\, I(x \leq 0.5) + 2.25\, I(x > 0.5)$.
(d) $m_4(x) = 4x + 4\exp\{-100(x-0.5)^2\}$, $v_4(x) = \frac{x^2}{9} I(x \leq 0.55) + 4(1-x)^2 I(x > 0.55)$.

From the examination of mean integrated squared error(MISE), our estimates dominate variance estimates without change point estimation and the improvement gets larger as sample size increases. Some plots of estimated variance functions show that the proposed estimator works quite significantly.

REFERENCES

BHATTACHARYA, P. K. AND BROCKWELL, P. J. (1976), The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **37**, 51–75.

CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10**, 1224–1233.

CARROLL, R. J. AND RUPPERT, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, London.

HÄRDLE, W. (1990). *Applied nonparametric regression.* Cambridge University Press, London.

LOADER, C. R. (1996), Change point estimation using nonparametric regression. *Ann. Statist.* **24**, 1667–1678.

MÜLLER, H G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer, Berlin.

MÜLLER, H G. (1992), Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737–761.

RAIMONDO, M. (1998), Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379–1397.

WHITT, W. (1970), Weak convergence of probability measures on the function space $C([0,\infty])$. *Ann. Math. Statist.* **41**, 939–944.