Journal of the Korean Statistical Society (2006), **35**: *1, pp 1–23*

NONPARAMETRIC ESTIMATION OF THE VARIANCE FUNCTION WITH A CHANGE POINT

KEE-HOON KANG[1] AND JIB HUH[2]

ABSTRACT

In this paper we consider an estimation of the discontinuous variance function in nonparametric heteroscedastic random design regression model. We first propose estimators of the change point in the variance function and then construct an estimator of the entire variance function. We examine the rates of convergence of these estimators and give results for their asymptotics. Numerical work reveals that using the proposed change point analysis in the variance function estimation is quite effective.

AMS 2000 subject classifications. Primary 62G08; Secondary 62G20.
Keywords. Discontinuity point, jump size, nonparametric regression, variance estimation, one-sided kernel, rate of convergence.

1. INTRODUCTION

In most nonparametric regression function estimation, the variance of errors is assumed to be a homogeneous or a heterogeneous smooth function. Estimation of this variance function has great meaning because it is important in its own right and in various applications. An estimation of the variance function is needed in some bandwidth selection procedures, weighted least squares estimation, constructions of confidence and prediction intervals for mean functions, quality control, *etc.* These applications are discussed in Carroll (1982), Carroll and Ruppert (1988) and Müller (1988). Variance function estimation is important for inference purposes as well as risk management based on second moments of financial data. Volatility in financial markets corresponds to variance in conventional statistical research areas. Recently, relative high attention is paid to capture and explain the variation of the second moments of financial data.

Received October 2005; accepted December 2005.

[1]Corresponding author. Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Korea (e-mail : khkang@hufs.ac.kr)

[2]Department of Statistics, Duksung Women's University, Seoul 132-714, Korea

Extensive literature exists on nonparametric methods for variance function estimation in nonparametric regression, much of which is based on using squared residuals from a nonparametric fit of the mean function. For the homogeneous error variance, these methods include those of Rice (1984), Gasser *et al.* (1986) and Hall *et al.* (1990). Their estimates are based on the squared differences of the data of various orders. Müller and Stadtmüller (1987) provided an estimation of the heteroscedasticity by using Gasser-Müller type kernel weighted averages of initial local variance estimates and showed that the estimator is uniformly consistent. Hall and Carroll (1989) studied the influence of the smoothness of the mean function on the convergence rate of the nonparametric variance estimator. Their variance function estimator is a Nadaraya-Watson type estimator based on squared residuals and results reveal that the accuracy of estimating the variance function depends on the information of the mean function. Ruppert *et al.* (1997) considered a local polynomial estimator for the variance function and gave results for the bias and the variance of the estimator.

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

$$Y_i = m(X_i) + v^{1/2}(X_i)\varepsilon_i, \ i = 1, \ldots, n, \tag{1.1}$$

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 with support $[0, 1]$. Here, the difference from the former works is we do not assume that the variance function is continuous. That is, we assume that a change point exists for the variance function v at some point τ in the interior of the support of f. One should note that relatively little attention has been given to the fact that the variance function may not be continuous, compared to its importance.

Our approach on variance function estimation is similar to the one for estimating the discontinuous regression function, which was discussed in Müller (1992) and in Loader (1996). One-sided kernel regression estimates based on squared residuals are used to estimate the location of a change point and the jump size. The resulting estimator of the change point is shown to be consistent with convergence rate of order n^{-1} since we use the one-sided kernel which has a non-zero value at the left end of the support.

For estimating the variance function itself, we use a Nadaraya-Watson type estimator for the data set splitted by the estimated location of the change point.

In terms of integrated squared error, we show that the convergence rate of the estimated variance function does not depend on the rate of convergence of the estimated change point.

This paper is organized as follows. In Section 2, the assumptions used in this paper are stated. The estimators of the change point of the variance function and the variance function itself are proposed. Section 3 investigates the theoretical properties of these estimators. Their numerical properties are examined in Section 4. Technical arguments are deferred to Section 5.

2. ASSUMPTIONS AND ESTIMATORS

We begin by stating a set of assumptions for the unknown functions in the model (1.1).

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

$$|v(x) - v(y)| \le C|x - y| \text{ whenever } (x-\tau)(y-\tau) > 0, \tag{2.1}$$

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 $v_+(\tau) = \lim_{x\to\tau+} v(x)$ and $v_-(\tau) = \lim_{x\to\tau-} v(x)$. 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 over $[0,1]$.

(A.3) The regression function m satisfies the Lipschitz condition of order 1 over $[0,1]$.

We apply a Nadaraya-Watson smoother with a one-sided kernel function to the squared residuals to detect the location of the change point of the variance function. Härdle (1990, 1991) also estimated the variance function $v(x)$ by using a Nadaraya-Watson kernel smoother with different types of the squared residuals to get the confidence intervals for $m(x)$. To get the 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: L and h_1 satisfy

(A.4) and (A.5), respectively.

(A.4) The function L is symmetric and satisfies the Lipschitz condition of order 1 over $[-1, 1]$.

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

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

$$\widehat{v}_+(x) = \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_2}\right)\widehat{R}_i \Big/ \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_2}\right), \tag{2.2}$$

$$\widehat{v}_-(x) = \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_2}\right)\widehat{R}_i \Big/ \frac{1}{nh_2}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_2}\right). \tag{2.3}$$

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 over $[0, 1]$ and $K(0) > 0, K(u) \geq 0$ for $0 < u \leq 1$ with $\int_0^1 K(u)du = 1$.

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

The estimators $\widehat{v}_+(x)$ and $\widehat{v}_-(x)$ are based only on the residuals at the right side and the left side of x, respectively. We estimate the jump size at a point x by taking the difference 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

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

for the location of the change point τ. Here, we have assumed continuity of the regression function m in (A.3). When both the regression function and the variance function have the same change point, τ can be estimated based on an estimation of the regression function.

Now, we propose an estimator for the variance function by 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 over $[-1, 1]$.

Using the squared residuals $\widehat{R}_i$, $i = 1, \ldots, n$, we define

$$\widehat{v}(x;\widehat{\tau}) = \frac{1}{nh}\sum_{i=1}^{n} W^*\left(\frac{X_i - x}{h};\widehat{\tau}\right)\widehat{R}_i \Big/ \frac{1}{nh}\sum_{i=1}^{n} W^*\left(\frac{X_i - x}{h};\widehat{\tau}\right). \tag{2.4}$$

Here, W^* is defined by

$$W^*\left(\frac{u-x}{h};\widehat{\tau}\right) = \begin{cases} W\Big(\frac{u-x}{h}\Big)I(x-h \le u \le \widehat{\tau}), & \widehat{\tau} - h \le x \le \widehat{\tau}, \\ W\Big(\frac{u-x}{h}\Big)I(\widehat{\tau} \le u \le x+h), & \widehat{\tau} \le x \le \widehat{\tau} + h, \\ W\Big(\frac{u-x}{h}\Big), & \text{otherwise}, \end{cases} \tag{2.5}$$

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

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

3. ASYMPTOTIC PROPERTIES

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

$$\varphi_n(z) = nh_2\left\{\widehat{\Delta}\Big(\tau + \frac{z}{n}\Big) - \widehat{\Delta}(\tau)\right\} \tag{3.1}$$

and $M < \infty$. The process φ_n lies in the space, denoted by $\mathcal{D}([-M, M])$, of functions defined on $[-M, M]$ having, at most, finitely many discontinuities. Let $\xrightarrow{\mathcal{W}}$ denote weak convergence in the space $\mathcal{D}([-M, M])$ and let $\kappa(x) = E\left[\{Y - m(X)\}^4 \mid X = x\right]$. To obtain the theorem below, consider the following additional assumptions.

(A.10) The function κ satisfies the Lipschitz condition of order 1 over $[0, 1]$.

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

THEOREM 3.1. *Suppose that assumptions (A.1)-(A.7), (A.10) and (A.11) are satisfied. Then,*

$$\varphi_n(z) \xrightarrow{\mathcal{W}} \varphi(z) = -\Delta K(0)|z| + \sigma W(z), \tag{3.2}$$

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

$$\sigma = \sqrt{\frac{4\kappa(\tau)}{f(\tau)}}K(0). \tag{3.3}$$

REMARK 3.1. 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 (3.2) 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 (3.3) 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, we first discuss that the maximizer (minimizer) of the limit of the process φ_n when $\Delta > 0$ ($\Delta < 0$) exists with probability one. In the case where $\Delta > 0$, that follows directly from Remark 5.3 in Bhattacharya and Brockwell (1976), where it is argued that the maximizer of a two-sided Brownian motion with an additional drift is unique with probability one. The other case where $\Delta < 0$ is analogous. The following corollary describes the asymptotic distribution of $\widehat{\tau}$:

COROLLARY 3.1. *Suppose that the assumptions in Theorem* 3.1 *are satisfied. Then,*

$$n(\widehat{\tau} - \tau) \xrightarrow{\mathcal{D}} \begin{cases} \underset{z\in(-\infty,\infty)}{\operatorname{argmax}}\ \varphi(z), \text{ when } \Delta > 0, \\ \underset{z\in(-\infty,\infty)}{\operatorname{argmin}}\ \varphi(z), \text{ when } \Delta < 0. \end{cases}$$

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 the variance, our proposed estimator provides the rate of convergence n^{-1} according to Corollary 3.1.

We now turn to the asymptotic property of the estimator of the variance function. Theorem 3.2 gives the rate of global L_p convergence of the estimator $\widehat{v}(\cdot;\widehat{\tau})$ in (2.4).

THEOREM 3.2. *Suppose that assumptions (A.1)–(A.11) are satisfied. Then, for* $p \geq 1$,

$$\begin{aligned} \int_0^1 & |\widehat{v}(x;\widehat{\tau}) - v(x)|^p dx \\ &= O_P\left\{h_1^{2p} + \left(\frac{\log n}{nh_1}\right)^p\right\} + O_P\left\{\left(\frac{\log n}{n\sqrt{h_1 h}}\right)^p + \left(h_1\sqrt{\frac{\log n}{nh}}\right)^p\right\} \\ &\quad + |\widehat{\tau} - \tau| O_P(1) + |\widehat{\tau} - \tau|^p O_P(h^{-p}) + O_P\left\{h^p + \left(\sqrt{\frac{\log n}{nh}}\right)^p\right\}. \end{aligned} \tag{3.4}$$

The first two terms in (3.4) depend on the rate of global L_∞ convergence of the estimator $\widehat{m}$. See (5.20) for more details. Since a Nadaraya-Watson type smoother guarantees the nonnegativity of the estimated variance function based on squared residuals, we proposed that type of smoother in (2.4). As a result, the order of the 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 (3.4) is then $O_P\{h^2+\log n/(nh)\}$. Thus minimizing the terms in (3.4) does not depend on the rate of the estimator $\widehat{\tau}$. This implies that the estimator $\widehat{v}(\cdot;\widehat{\tau})$ has exactly the same rate as $\widehat{v}(\cdot;\tau)$, when the location of the change point is known.

4. NUMERICAL PROPERTIES

To investigate the practical performance of the proposed estimator defined in Section 3, we carried out a simulation study. For that, response-predictor pairs (X_i, Y_i), $i = 1, \dots, n$, are generated according to the prescription (1.1) for various m and v. Kernel functions $L(u) = (15/16)(1-u^2)^2 I(|u| \le 1)$ and $K(u) = (15/8)(1-u^2)^2 I(0 \le u \le 1)$ are used to estimate the mean regression function and the change point, respectively. For the estimation of the variance function, the kernel function W is taken to be the same as L.

In this section, we shall report only results for the following four settings. We obtained very similar results for other cases. Cases (a), (b) and (c) represent typical types of regression functions and have homogeneous variance functions with a change point. Case (d) represents a nonhomogeneous variance function with a change point.

(a) $m_1(x) = x$, $v_1(x) = 0.01\, I(x \le 0.5) + 0.09\, I(x > 0.5)$.

(b) $m_2(x) = 4x(1-x)$, $v_2(x) = 0.01\, I(x \le 0.75) + 0.16\, I(x > 0.75)$.

(c) $m_3(x) = 5x(2x^2 - 3x + 1)$, $v_3(x) = 0.49\, I(x \le 0.5) + 2.25\, I(x > 0.5)$.

(d) $m_4(x) = 4x + 4\exp\{-100(x-0.5)^2\}$, $v_4(x) = (x^2/9)\, I(x \le 0.55) + 4(1-x)^2\, I(x > 0.55)$.

Throughout, the distribution of the predictor variable is assumed to be uniform in $(0,1)$ and $\hat{\tau}$ is estimated on $[h, 1-h]$. The integrated squared error (ISE), a measure of performance, is estimated on the interval $[0,1]$ by using the trapezoidal rule. Average values are obtained from 1000 simulations.

Table 4.1 presents the mean integrated squared errors (MISE) for two types of variance function estimates. One is based on our approach with change point estimation and the other is nonparametric estimation without any change point estimation. The sample sizes considered here are 500 and 1000. From this table, in terms of the MISE, our estimates dominate the variance estimates without change point estimation and the improvement gets larger as the sample size is increased. In fact, our estimate needs to be compared with other estimates that also reflect the existence of a change point in the variance function. However, to the best of our knowledge, such estimates have not yet been published.

TABLE 4.1 *Simulation results*

Model	n	*(i)*	*(ii)*	*(iii)*
(a)	500	0.0187 (0.0002)	0.0163 (0.0003)	0.5212 (0.0016)
	1000	0.0132 (0.0001)	0.0089 (0.0001)	0.5086 (0.0011)
(b)	500	0.0439 (0.0006)	0.0432 (0.0010)	0.7671 (0.0010)
	1000	0.0305 (0.0003)	0.0244 (0.0005)	0.7586 (0.0007)
(c)	500	10.8682 (0.1181)	9.3584 (0.2066)	0.5255 (0.0018)
	1000	7.8781 (0.0726)	5.0637 (0.1161)	0.5114 (0.0013)
(d)	500	0.8914 (0.0072)	0.6569 (0.0129)	0.5561 (0.0004)
	1000	0.6412 (0.0039)	0.3457 (0.0071)	0.5526 (0.0002)

NOTE : *Average integrated squared errors over* 500 *runs from four regression curves. Column (i) corresponds to average integrated squared errors of the variance function estimates over* 500 *runs from each model without estimating the change point and the jump size. Column (ii) corresponds to the estimates with a change-point analysis. Columns (i) and (ii) are multiplied by* 10^2. *Column (iii) contains average values of* $\hat{\tau}$. *In all columns, the standard errors are in brackets.*

The average values of the estimated change point $\hat{\tau}$ and their standard errors are also summarized in Table 4.1. Figure 4.1 illustrates frequency plots for the 1000 values of $\hat{\tau}$ for models (b) and (d). For model (b), the proposed procedure results in a value of $\hat{\tau}$ that is somewhat larger than the true value of 0.75. The distribution of $\hat{\tau}$ for model (d) is more concentrated toward its target value, which makes the whole procedure more significant.

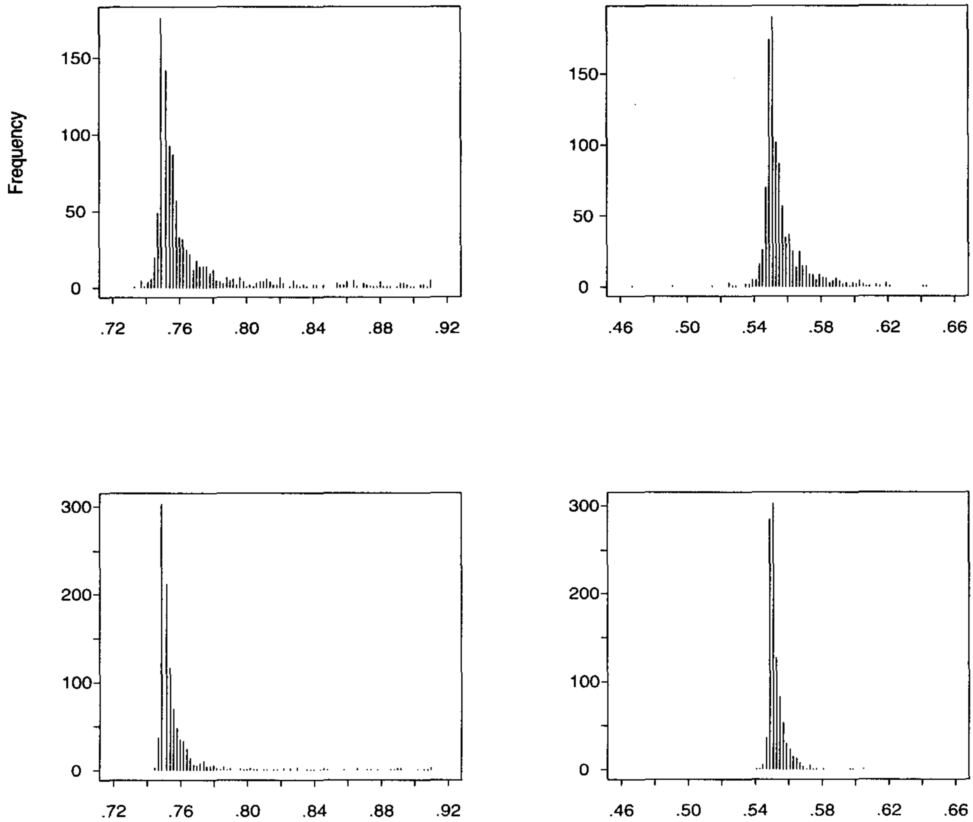

FIGURE 4.1 *Distributions of $\hat{\tau}$ from models (b) and (d).*

NOTE : *Plots in the left column come from model (b) and those in the right column from model (d). In each column, the upper panel corresponds to $n = 500$ and the lower panel corresponds to $n = 1000$. The vertical axis represents the frequency.*

For model (d), Figure 4.2 depicts raw data, estimates of the mean function and estimates of the variance function with and without an estimated change point. The data set used here is the one which gives the 75 percentile of the integrated squared errors of the estimates without change point estimation among 1000 runs. Note that the proposed estimates of the variance function are the same as those without the change point except for the area close to the change point. Plots for the other models give the similar results, so they are not included here.

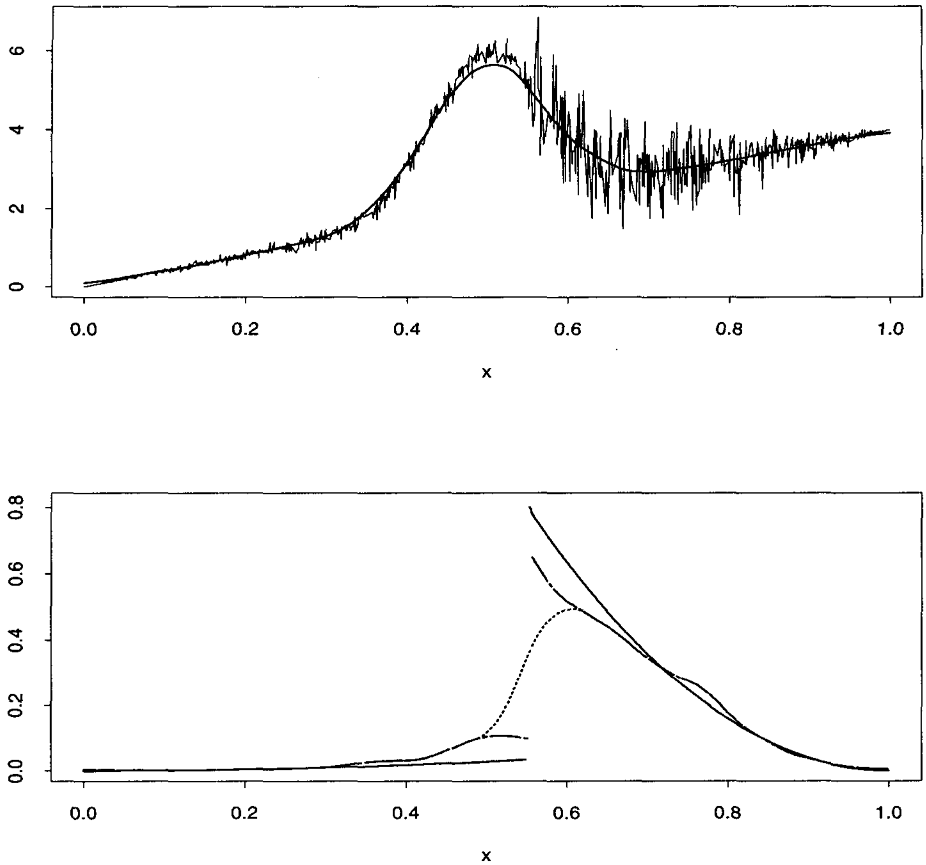

FIGURE 4.2 *Plots from a data set of model (d).*

NOTE : *The data set corresponds to the one which gives the* 75 *percentile of the integrated squared errors of the estimates without any change-point estimation from among* 500 *simulations. The upper panel depicts the raw data (solid) and estimated mean function (bold solid) while the lower panel illustrates the true variance function (solid), the estimated variance function with a change-point estimation (dot-and-dashed) and the estimated variance function without change-point estimation (dotted).*

We apply our approach to the real data set called LIDAR (LIght Detection And Ranging), which was analyzed by Ruppert *et al.* (1997). They used local polynomial fit to estimate the variance function. Figure 1(b) in Ruppert *et al.* (1997) indicates that a change point, which might be close to 650 or 680, may exist in the variance function. Figure 4.3 presents the data and the resulting variance estimates, which has a shape that is slightly different from the shape of Figure 1(c) in Ruppert *et al.* (1997). Our approach results in a change point estimate of 682.9 and a corresponding jump size of 0.0158. This change point estimate seems plausible from the plot of the raw data although Gibbs phenomenon seems to exist near the change point.

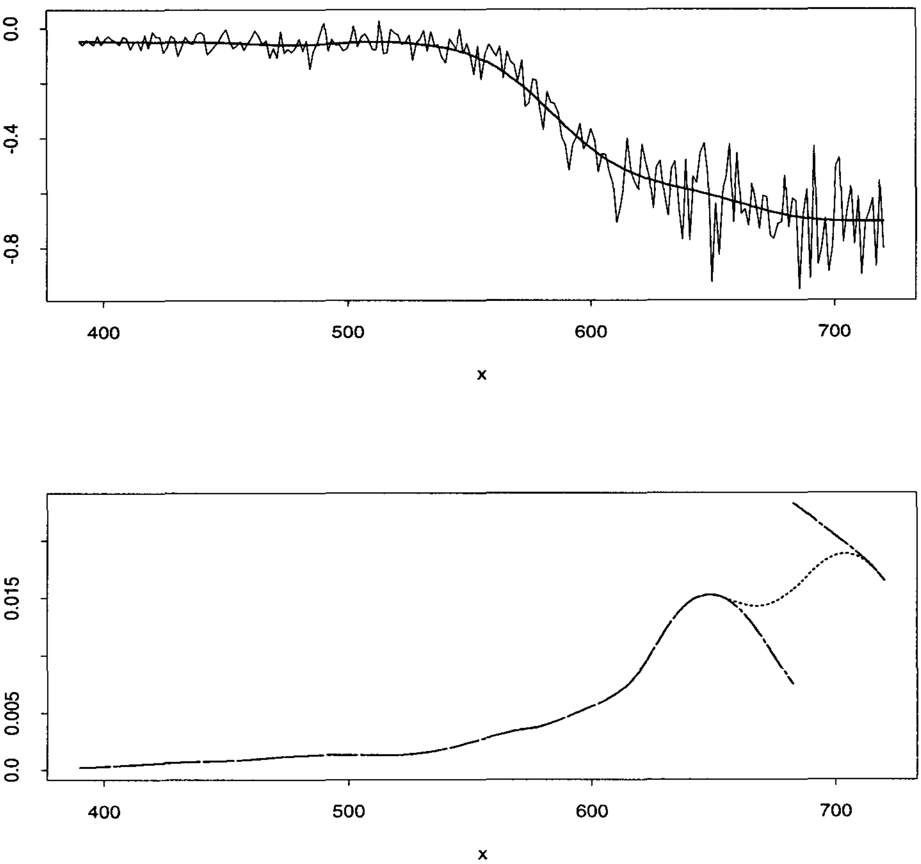

FIGURE 4.3 *Plots for LIDAR data.*

NOTE : *The upper panel depicts the raw data (solid) and estimated mean function (bold solid) while the lower panel illustrates the estimated variance function with a change-point estimation (dot-and-dashed) and the estimated variance function without change-point estimation (dotted).*

An important practical problem in change point analysis is the selection of the bandwidths. However, we do not have any optimal theory for that yet. In all the procedures to estimate the regression function and the variance function in this section, we used a cross-validation criterion over a fine grid of bandwidths, which does not require any auxiliary stage of estimation.

5. PROOFS

LEMMA 5.1. *Assume that a bandwidth b satisfies the requirement that $b \to 0$ and $b/\log n \to \infty$ as $n \to \infty$. Let $\eta(X_j, \varepsilon_j)$, which depend on X_j and ε_j, be independent random variables with mean 0 and bounded second moments and let $\mathcal{K}$ be a kernel function satisfying the Lipschitz condition of order 1 over a compact set $[r, s]$ where $r \le 0$, $s \ge 0$ and $r < s$. Then,*

$$\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nb}\sum_{j=1}^{n}\mathcal{K}\left(\frac{X_j-x}{b}\right)\eta(X_j,\varepsilon_j)\Big/\frac{1}{nb}\sum_{j=1}^{n}\mathcal{K}\left(\frac{X_j-x}{b}\right)\right| = O_P\left(\sqrt{\frac{\log n}{nb}}\right),$$

where $\mathcal{S}_f = [0,1]$ is the support of f.

PROOF. According to Theorem 1.2 in Stute (1982),

$$\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nb}\sum_{j=1}^{n}\mathcal{K}\left(\frac{X_j-x}{b}\right) - \mu_{\mathcal{K},x}f(x)\right| = O_P\left(\sqrt{\frac{\log n}{nb}}+b\right), \tag{5.1}$$

where

$$\mu_{\mathcal{K},x} = \begin{cases} \displaystyle\int_{cr}^{s}\mathcal{K}(u)du, & 0 \le x < -crb, \\ \displaystyle\int_{r}^{s}\mathcal{K}(u)du, & -crb \le x \le 1-csb, \\ \displaystyle\int_{r}^{cs}\mathcal{K}(u)du, & 1-csb < x \le 1, \end{cases}$$

with $0 \le c < 1$. By (5.1), it is enough to show that

$$\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nb\mu_{\mathcal{K},x}f(x)}\sum_{j=1}^{n}\mathcal{K}\left(\frac{X_j-x}{b}\right)\eta(X_j,\varepsilon_j)\right| = O_P\left(\sqrt{\frac{\log n}{nb}}\right).$$

Define

$$g_n(x) = \frac{1}{nb\mu_{\mathcal{K},x}f(x)}\sum_{j=1}^{n}\mathcal{K}\left(\frac{X_j-x}{b}\right)\eta(X_j,\varepsilon_j).$$

Let D_n be a discretized grid of $\mathcal{S}_f$, which is given by $D_n = \{j\delta_n \mid j = 0,\dots,1/\delta_n\}$ where $\delta_n = O(n^{-\alpha})$ for some positive α. Then, we obtain

$$\sup_{x\in\mathcal{S}_f}|g_n(x)| \le \sup_{y\in D_n}|g_n(y)| + \sup_{x,y:|x-y|\le\delta_n}|g_n(x)-g_n(y)|. \tag{5.2}$$

We take α to be large enough to ensure that the second term in (5.2) is negligible compared to the first one. Now, we will show that the first term is $O_P\{\sqrt{\log n/(nb)}\}$. If we define $\bar{\eta}(X_j, \varepsilon_j) = \eta(X_j, \varepsilon_j) I(|\eta(X_j, \varepsilon_j)| \leq \sqrt{nb/\log n}\,)$, it follows that

$$\sup_{y \in D_n} |g_n(y)| \leq \sup_{y \in D_n} |g_n(y) - \bar{g}_n(y)| + \sup_{y \in D_n} |\bar{g}_n(y)|, \tag{5.3}$$

where $\bar{g}_n(x) = \sum_{i=1}^n W_{n,j}(x)$ with

$$W_{n,j}(x) = \frac{1}{nb\mu_{\mathcal{K},x} f(x)} \mathcal{K}\left(\frac{X_j - x}{b}\right) \bar{\eta}(X_j, \varepsilon_j), \; j = 1, \ldots, n.$$

By using Borel-Cantelli lemma, one can show the first term on the right hand side in (5.3) is $O\{\sqrt{\log n/(nb)}\}$ with probability 1. Using Bernstein's inequality and some probability inequalities, we obtain $\sup_{y \in D_n} |\bar{g}_n(y)| = O_P\{\sqrt{\log n/(nb)}\}$. This implies the result. □

For the proof of Theorem 3.1, an asymptotic expression of φ_n will be described. Let

$$D_n^+(u, z) = \left\{ \frac{1}{f(\tau + z_n)} K\left(\frac{u - \tau - z_n}{h_2}\right) - \frac{1}{f(\tau)} K\left(\frac{u - \tau}{h_2}\right) \right\},$$
$$D_n^-(u, z) = \left\{ \frac{1}{f(\tau + z_n)} K\left(\frac{\tau + z_n - u}{h_2}\right) - \frac{1}{f(\tau)} K\left(\frac{\tau - u}{h_2}\right) \right\},$$
$$\phi_n(z) = \sum_{j=1}^n \left\{ D_n^+(X_j, z) - D_n^-(X_j, z) \right\} R_j^2,$$

where $R_j = Y_j - m(X_j) = v^{1/2}(X_j)\varepsilon_j$, $j = 1, \ldots, n$ and $z_n = z/n$, $z \in [-M, M]$.

LEMMA 5.2. *Suppose that assumptions (A.1)–(A.7) are satisfied. Then,*

$$\sup_{z \in [-M, M]} |\varphi_n(z) - \phi_n(z)| = o_P(1).$$

PROOF. Let the denominators in (2.2) and (2.3) be $\widehat{f}_+(x)$ and $\widehat{f}_-(x)$, respectively, as follows

$$\widehat{f}_\pm(x) = \frac{1}{nh_2} \sum_{j=1}^n K\left(\pm \frac{X_j - x}{h_2}\right).$$

Note that

$$\sup_{x \in Q} \left| \frac{1}{nh_2 \widehat{f}_\pm(x)} \sum_{j=1}^n K\left(\pm \frac{X_j - x}{h_2}\right) \widehat{R}_j^2 - \frac{1}{nh_2 f(x)} \sum_{j=1}^n K\left(\pm \frac{X_j - x}{h_2}\right) R_j^2 \right|$$

$$\leq \sup_{x\in Q}\left|\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)\left[\{\widehat{m}(X_j)-m(X_j)\}^2 -2R_i\{\widehat{m}(X_j)-m(X_j)\}\right]\right|$$
$$+\sup_{x\in Q}\left|\frac{1}{nh_2\widehat{f}_\pm(x)}\left(1-\frac{\widehat{f}_\pm(x)}{f(x)}\right)\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)R_j^2\right|. \tag{5.4}$$

Although the variance function v is not continuous at τ, since v is bounded, it is easy to show that

$$\sup_{x\in S_f}|\widehat{m}(x)-m(x)| = O_P\left(\sqrt{\frac{\log n}{nh_1}}+h_1\right) \tag{5.5}$$

by using Theorem B in Mack and Silverman (1982). By Lemma 5.1 and (5.5), the first term on the right-hand side of (5.4) is $o_P(1)$ since R_i has mean 0 and bounded second moments. In the second term on the right-hand side of (5.4), $\{1-\widehat{f}_\pm(x)/f(x)\}$ is $o_P(1)$ due to (5.1). With the definition $\xi_j=\varepsilon_j^2-1$, $j=1,\dots,n$, note that

$$\sup_{x\in Q}\left|\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)R_j^2\right|$$
$$\leq \sup_{x\in Q}\left|\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)v(X_j)\xi_j\right|$$
$$+\sup_{x\in Q}\left|\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)v(X_j) -E\left\{\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)v(X_j)\right\}\right|$$
$$+\sup_{x\in Q}\left|E\left\{\frac{1}{nh_2\widehat{f}_\pm(x)}\sum_{j=1}^n K\left(\pm\frac{X_j-x}{h_2}\right)v(X_j)\right\}\right|. \tag{5.6}$$

The first two terms on the right-hand side of (5.6) are $O_P\left(\sqrt{\log n/nh_2}\right)$ by Lemma 5.1. According to assumption (A.1), the variance function v is bounded, which implies the last term in (5.6) is $O(1)$. These imply that the second term on the right-hand side of (5.4) is $o_P(1)$. The result follows immediately. □

We need the following four lemmas to prove Theorem 3.1.

LEMMA 5.3. *Suppose that assumptions (A.1)–(A.7) are satisfied. Then,*

$$E\{\phi_n(z)\} = -\Delta K(0)|z| + o(1)$$

uniformly in $z \in [-M, M]$.

PROOF. We prove the lemma for $z > 0$, as the other case can be dealt with similarly. By assumption (A.2),

$$\begin{aligned}
&E\{D_n^{\pm}(X_1, z)R_1^2\}\\
&= E\left[\left\{\frac{1}{f(\tau+z_n)}K\left(\pm\frac{X_1-\tau-z_n}{h_2}\right) - \frac{1}{f(\tau)}K\left(\pm\frac{X_1-\tau}{h_2}\right)\right\}R_1^2\right]\\
&= h_2\int K(u)\Big\{v(\tau+z_n\pm h_2u) - v(\tau\pm h_2u)\Big\}du\{1+O(h_2)\}\\
&\quad + h_2\int K(u)\left\{\frac{f(\tau+z_n\pm h_2u)}{f(\tau+z_n)} - \frac{f(\tau\pm h_2u)}{f(\tau)}\right\}v(\tau\pm h_2u)du, \qquad (5.7)
\end{aligned}$$

where the $O(h_2)$ term is uniform in $z \in [-M, M]$. In the case of $E\{D_n^+(X_1, z)R_1^2\}$, by (A.1) and (A.2), it is easy to see that the terms in (5.7) are $O(h_2/n)$ uniformly in z.

As in the case of $E\{D_n^+(X_1, z)R_1^2\}$, the second term of $E\{D_n^-(X_1, z)R_1^2\}$ is $O(h_2/n)$. The approximation of the first term in (5.7) is slightly different in the case of $E\{D_n^-(X_1, z)R_1^2\}$ because the change point τ lies between $\tau + z_n - h_2$ and $\tau + z_n$. In this case, we divide the interval of integration into two parts. Note that, for $z/(nh_2) < u \le 1$,

$$v(\tau + z_n - h_2u) - v(\tau - h_2u) = O(h_2) \qquad (5.8)$$

uniformly in z. However, for $0 \le u \le z/(nh_2)$,

$$\begin{aligned}
&v(\tau + z_n - h_2u) - v(\tau - h_2u)\\
&= \{v(\tau + z_n - h_2u) - v_+(\tau)\} - \{v(\tau - h_2u) - v_-(\tau)\} + \Delta\\
&= \Delta + O(h_2) \qquad (5.9)
\end{aligned}$$

uniformly in z. By (5.8) and (5.9), the leading term of $E\{D_n^-(X_1, z)R_1^2\}$ equals

$$h_2\Delta\int_0^{z/nh_2} K(u)du$$

uniformly in z. Since $K(u) = K(0)\{1 + o(1)\}$ uniformly in $u \in [0, M/(nh_2)]$, the leading term of $E\{D_n^-(X_1, z)R_1^2\}$ is equal to $\Delta K(0)z_n$. This implies the result. □

LEMMA 5.4. *Suppose that assumptions (A.1)-(A.7) and (A.10) are satisfied. Then,*

$$\mathrm{Cov}\{\phi_n(z_1), \phi_n(z_2)\} = \begin{cases} \dfrac{4\kappa(\tau)}{f(\tau)} \min(|z_1|, |z_2|)\{K(0)\}^2 + o(1), & z_1 z_2 \geq 0, \\ o(1), & \text{elsewhere}, \end{cases}$$

uniformly in $z_1, z_2 \in [-M, M]$.

PROOF. We prove the lemma for $z_1, z_2 > 0$ first. By Lemma 5.3,

$$\begin{aligned} &\mathrm{Cov}\{\phi_n(z_1), \phi_n(z_2)\} \\ &= n\mathrm{Cov}\Big[\{D_n^+(X_1, z_1) - D_n^-(X_1, z_1)\}R_1^2,\ \{D_n^+(X_1, z_2) - D_n^-(X_1, z_2)\}R_1^2\Big] \\ &= nE\Big[\{D_n^+(X_1, z_1)D_n^+(X_1, z_2) - D_n^+(X_1, z_1)D_n^-(X_1, z_2) \\ &\quad - D_n^-(X_1, z_1)D_n^+(X_1, z_2) + D_n^-(X_1, z_1)D_n^-(X_1, z_2)\}R_1^4\Big] + O\Big(\frac{1}{n}\Big). \end{aligned} \tag{5.10}$$

The first equality holds since R_i^2 and R_j^2 are independent when $i \neq j$. Define $z_{\min} = \min(z_1, z_2)$, $z_{\max} = \max(z_1, z_2)$, $\tau_n^{\min} = \tau + z_{\min}/n$ and $\tau_n^{\max} = \tau + z_{\max}/n$. Consider the first term in the square brackets in (5.10) first. By assumptions (A.2), (A.6), (A.7) and (A.10),

$$\begin{aligned} &E\Big\{D_n^+(X_1, z_1)D_n^+(X_1, z_2)R_1^4\Big\} \\ &= \Bigg[\int_\tau^{\tau_n^{\min}} \left\{\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h_2}\right)\right\}^2 + \int_{\tau_n^{\min}}^{\tau_n^{\max}} D_n^+(u, z_{\min})\left\{-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h_2}\right)\right\} \\ &\quad + \int_{\tau_n^{\max}}^{\tau_n^{\max}+h_2} D_n^+(u, z_{\min})D_n^+(u, z_{\max})\Bigg] \kappa(u)f(u)du \\ &= h_2\frac{\kappa(\tau)}{f(\tau)}\left[\{K(0)\}^2\frac{z_{\min}}{nh_2} + O\left(\frac{1}{(nh_2)^2}\right)\right]\{1 + O(h_2)\} \end{aligned} \tag{5.11}$$

uniformly in z_1 and z_2. The $O\{1/(nh_2)^2\}$ term in (5.11) follows from $D_n^{\pm}(u, z)$ being equal to $O\{1/(nh_2)\}$ uniformly in u and z due to (A.2) and (A.6). Next, consider the second term in the square brackets in (5.10) for the case $z_{\min} = z_1$. The other cases can be dealt with in a similar way. We note that $D_n^+(u, z) = 0$ for $u < \tau$ and $D_n^-(u, z) = 0$ for $u > \tau_n^{\max}$. Then,

$$\begin{aligned} &E\Big\{D_n^+(X_1, z_1)D_n^-(X_1, z_2)R_1^4\Big\} \\ &= \int_\tau^{\tau_n^{\min}} \left\{-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h_2}\right)\right\}\left\{\frac{1}{f(\tau_n^{\max})}K\left(\frac{\tau_n^{\max}-u}{h_2}\right)\right\}\kappa(u)f(u)du \end{aligned}$$

$$
\begin{aligned}
&+ \int_{\tau_n^{\min}}^{\tau_n^{\max}} D_n^+(u, z_{\min}) \left\{ \frac{1}{f(\tau_n^{\max})} K\left(\frac{\tau_n^{\max} - u}{h_2} \right) \right\} \kappa(u) f(u) du \\
&= -h_2 \frac{\kappa(\tau)}{f(\tau)} \left[\{K(0)\}^2 \frac{z_{\min}}{nh_2} + O\left(\frac{1}{(nh_2)^2} \right) \right] \{1 + O(h_2)\} \qquad (5.12)
\end{aligned}
$$

uniformly in z_1 and z_2. Analogously,

$$
\begin{aligned}
&E\left\{ D_n^-(X_1, z_1) D_n^+(X_1, z_2) R_1^4 \right\} \\
&= -h_2 \frac{\kappa(\tau)}{f(\tau)} \left[\{K(0)\}^2 \frac{z_{\min}}{nh_2} + O\left\{ \frac{1}{(nh_2)^2} \right\} \right] \{1 + O(h_2)\} \qquad (5.13)
\end{aligned}
$$

uniformly in z_1 and z_2 and

$$
\begin{aligned}
&E\left\{ D_n^-(X_1, z_1) D_n^-(X_1, z_2) R_1^4 \right\} \\
&= h_2 \frac{\kappa(\tau)}{f(\tau)} \left[\{K(0)\}^2 \frac{z_{\min}}{nh_2} + O\left\{ \frac{1}{(nh_2)^2} \right\} \right] \{1 + O(h_2)\} \qquad (5.14)
\end{aligned}
$$

uniformly in z_1 and z_2. Combining the first leading terms in (5.11), (5.12), (5.13) and (5.14) concludes the proof of Lemma 5.4 for the case $z_1, z_2 > 0$.

Now, consider the case of $z_1 > 0$, $z_2 < 0$. Following the proof for the case $z_1, z_2 > 0$, we obtain

$$
\begin{aligned}
E\left\{ D_n^+(X_1, z_1) D_n^+(X_1, z_2) R_1^4 \right\} &= \left[\int_{\tau}^{\tau_n^{\max}} D_n^+(u, z_{\min}) \left\{ \frac{1}{f(\tau)} K\left(\frac{u - \tau}{h_2} \right) \right\} \right. \\
&\quad \left. + \int_{\tau_n^{\max}}^{\tau_n^{\max} + h_2} D_n^+(u, z_{\min}) D_n^+(u, z_{\max}) \right] \kappa(u) f(u) du, \\
E\left\{ D_n^+(X_1, z_1) D_n^-(X_1, z_2) R_1^4 \right\} &= 0, \\
E\left\{ D_n^-(X_1, z_1) D_n^+(X_1, z_2) R_1^4 \right\} &= \left[\int_{\tau_n^{\min}}^{\tau} \left\{ \frac{1}{f(\tau_n^{\min})} K\left(\frac{u - \tau_n^{\min}}{h_2} \right) \right\} D_n^-(u, z_{\max}) \right. \\
&\quad \left. + \int_{\tau}^{\tau_n^{\max}} D_n^+(u, z_{\min}) \left\{ \frac{1}{f(\tau_n^{\max})} K\left(\frac{\tau_n^{\max} - u}{h_2} \right) \right\} \right] \kappa(u) f(u) du, \\
E\left\{ D_n^-(X_1, z_1) D_n^-(X_1, z_2) R_1^4 \right\} &= \left[\int_{\tau_n^{\min}}^{\tau} \left\{ \frac{1}{f(\tau)} K\left(\frac{\tau - u}{h_2} \right) \right\} D_n^-(u, z_{\max}) \right. \\
&\quad \left. + \int_{\tau_n^{\min} - h_2}^{\tau_n^{\min}} D_n^-(u, z_{\min}) D_n^-(u, z_{\max}) \right] \kappa(u) f(u) du \qquad (5.15)
\end{aligned}
$$

uniformly in z_1 and z_2. Here, the second identity follows from the fact that

$$D_n^+(u, z_1) D_n^-(u, z_2) = 0$$

for all u. Since $D_n^{\pm}(u,z) = O\{1/(nh_2)\}$ uniformly in u and z, all of the leading terms in (5.15) are $O\{1/(nh_2)^2\}$. This implies the result immediately. □

LEMMA 5.5. *Suppose that the assumptions in Theorem* 3.1 *are satisfied. For each* $z \in [-M, M]$, $\phi_n(z)$ *satisfies Lyapounov's condition.*

PROOF. We will show the lemma for $z > 0$. The other case can be dealt with similarly. By Lemma 5.4, $Var(\phi_n(z)) = O(1)$. We will show that, for some positive ζ,

$$L_n(z) = \sum_{j=1}^{n} E\left[\left|\{D_n^+(X_j,z) - D_n^-(X_j,z)\}R_j^2\right|^{2+\zeta}\right] \longrightarrow 0$$

as $n \to \infty$. By assumption (A.10), $E(|R_1^2|^{2+\zeta}|X = x) < \infty$ for all x. Note that

$$\begin{aligned} L_n(z) &\le n \cdot 2^{2+\zeta} E\left[\left\{\left|D_n^+(X_1,z)\right|^{2+\zeta} + \left|D_n^-(X_1,z)\right|^{2+\zeta}\right\} |R_1^2|^{2+\zeta}\right] \\ &= O\left\{nh_2\left(\frac{1}{nh_2}\right)^{2+\zeta}\right\}. \end{aligned}$$

By assumption (A.7), the result follows. □

LEMMA 5.6. *Suppose that the assumptions in Theorem* 3.1 *are satisfied. Then, the sequence of the process* $\psi_n(\cdot) = \phi_n(\cdot) - E(\phi_n(\cdot))$ *is tight.*

PROOF. By Theorem 12.3 in Billingsley (1968), it is enough to show that there exist a positive constant C_1 and a nondecreasing and continuous function F such that

$$E(\psi_n(z_1) - \psi_n(z_2))^2 \le C_1|F(z_2) - F(z_1)|^2 \qquad (5.16)$$

for sufficiently large n. By Lemma 5.4, there exists a positive constant C_1 such that

$$\begin{aligned} E(\psi_n(z_1) - \psi_n(z_2))^2 &= \mathrm{Var}(\phi_n(z_1)) + \mathrm{Var}(\phi_n(z_2)) - 2\mathrm{Cov}(\phi_n(z_1), \phi_n(z_2)) \\ &\le C_1|z_2 - z_1| \end{aligned}$$

for sufficiently large n and this concludes the proof of Lemma 5.6. □

Proof of Theorem 3.1.

Lemma 5.5 implies that $\psi_n(z)$, for fixed $z \in [-M, M]$, converges weakly to a normal distribution. Furthermore, by the Cramer-Wold device, we may show that for fixed $z_1, \ldots, z_l, z_i \in [-M, M]$,

$$\big(\psi_n(z_1), \ldots, \psi_n(z_l)\big) \xrightarrow{\mathcal{D}} N(0, \Sigma),$$

where Σ is the asymptotic covariance described in Lemma 5.4. This concludes the proof. See Theorems 8.1 and 12.3 of Billingsley (1968). □

Proof of Corollary 3.1.

According to Theorem 5.1 in Billingsley (1968), we have

$$\operatorname*{argmax}_{z\in[-M,M]} \varphi_n(z) \xrightarrow{\mathcal{D}} \operatorname*{argmax}_{z\in[-M,M]} \varphi(z) \tag{5.17}$$

for any $M > 0$. If we prove

$$\sup_{x\in Q,|x-\tau|\geq(M/n)} \widehat{\Delta}(x) = o_P(1)$$

for any $M > 0$, the result in (5.17) can be extended to the entire real line $(-\infty, \infty)$. Note that by (5.5), $|\widehat{R}_j^2 - R_j^2| = o_P(1)$ for all j. Therefore, it is enough to show that

$$\sup_{x\in Q,|x-\tau|\geq(M/n)} \left| \frac{1}{nh_2\widehat{f}_\pm(x)} \sum_{j=1}^{n} K\Big(\pm\frac{X_j - x}{h_2}\Big) R_j^2 - v(x) \right| = o_P(1). \tag{5.18}$$

This term is bounded by (5.6) with replacing $\sup_{x\in Q}$ and the last term in (5.6) by $\sup_{x\in Q,|x-\tau|\geq(M/n)}$ and

$$\sup_{x\in Q,|x-\tau|\geq(M/n)} \left| E\left[\frac{1}{nh_2\widehat{f}_\pm(x)} \sum_{j=1}^{n} K\Big(\pm\frac{X_j - x}{h_2}\Big) v(X_j)\right] - v(x) \right|,$$

respectively. This bias term is $O(h_2)$ due to the interval $|x - \tau| \geq (M/n)$ not having the change point. This implies the result. □

Proof of Theorem 3.2.

Defining $\widetilde{v}(x; w) = \frac{1}{nh}\sum_{j=1}^{n} V_j(x; w) R_j^2$, where

$$V_j(x; w) = W^*\left(\frac{X_j - x}{h}; w\right) \Big/ \frac{1}{nh}\sum_{i=1}^{n} W^*\left(\frac{X_i - x}{h}; w\right), \; j = 1, \ldots, n,$$

one obtains

$$\sup_{x\in\mathcal{S}_f}\left|\widehat{v}(x;\widehat{\tau})-v(x)\right|\le\sup_{x\in\mathcal{S}_f}\left|\widehat{v}(x;\widehat{\tau})-\widetilde{v}(x;\widehat{\tau})\right|+\sup_{x\in\mathcal{S}_f}\left|\widetilde{v}(x;\widehat{\tau})-v(x)\right|. \quad (5.19)$$

Consider the first term on the right-hand side in (5.19). Note that

$$\begin{aligned}\sup_{x\in\mathcal{S}_f}\left|\widehat{v}(x;\widehat{\tau})-\widetilde{v}(x;\widehat{\tau})\right| &\le \sup_{x\in\mathcal{S}_f}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;\widehat{\tau})\right|\left\{\sup_{x\in\mathcal{S}_f}\left|\widehat{m}(x)-m(x)\right|\right\}^2\\ &+2\sup_{w\in Q}\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;w)R_j\right|\sup_{x\in\mathcal{S}_f}\left|\widehat{m}(x)-m(x)\right|. \quad (5.20)\end{aligned}$$

By Lemma 5.1,

$$\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;w)R_j\right|=O_P\left(\sqrt{\frac{\log n}{nh}}\right). \quad (5.21)$$

Note that the result in Lemma 5.1 does not depend on the endpoints of the interval $[r,s]$. Then, the result in (5.21) is uniform for $w\in Q$, where Q is the left or the right endpoint of the support of W^*. The results in (5.5) and (5.21) imply that

$$\sup_{x\in\mathcal{S}_f}\left|\widehat{v}(x;\widehat{\tau})-\widetilde{v}(x;\widehat{\tau})\right|=O_P\left(h_1^2+\frac{\log n}{nh_1}\right)+O_P\left(\frac{\log n}{n\sqrt{h_1h}}+h_1\sqrt{\frac{\log n}{nh}}\right). \quad (5.22)$$

Now, we consider the second term on the right-hand side of (5.19). This proof is similar to that for Theorem 4.1 in Müller (1992). Defining $A=(\min(\tau,\widehat{\tau}),\max(\tau,\widehat{\tau}))$ and $B=\mathcal{S}_f-A$ and observing that

$$\begin{aligned}&\sup_{x\in B}\left|\widetilde{v}(x;\widehat{\tau})-v(x)\right|\\ &\le\sup_{x\in B}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;\tau)v(X_j)\xi_j\right|+\sup_{x\in B}\left|\frac{1}{nh}\sum_{j=1}^{n}\{V_j(x;\widehat{\tau})-V_j(x;\tau)\}v(X_j)\xi_j\right|\\ &+\sup_{x\in B}\left|\frac{1}{nh}\sum_{j=1}^{n}\{V_j(x;\widehat{\tau})-V_j(x;\tau)\}v(X_j)\right|+\sup_{x\in B}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;\tau)v(X_j)-v(x)\right|,\end{aligned} \quad (5.23)$$

one notes that $E(\xi_j)=0$ for all j. By Lemma 5.1, one obtains

$$\sup_{w\in Q}\sup_{x\in\mathcal{S}_f}\left|\frac{1}{nh}\sum_{j=1}^{n}V_j(x;w)v(X_j)\xi_j\right|=O_P\left(\sqrt{\frac{\log n}{nh}}\right), \quad (5.24)$$

which implies that the first two terms on the right-hand side of (5.23) are $O_P\{\sqrt{\log n/(nh)}\}$. Let

$$\widetilde{V}_j(x;w) = \frac{1}{\mu_{x,w} f(x)} W^*\left(\frac{X_j - x}{h}; w\right),\ j = 1, \ldots, n,$$

where

$$c = \begin{cases} (w-x)/h, & |w-x| < h, \\ -1, & w - x \leq -h, \\ 1, & w - x \geq h, \end{cases} \quad \mu_{x,w} = \begin{cases} \int_{-1}^{c} W(u)du, & w - x \geq 0, \\ \int_{c}^{1} W(u)du, & w - x < 0. \end{cases}$$

Since the result in (5.1) does not depend on the support $[r, s]$ of the kernel function, we have $\sup_{w\in Q} \sup_{x\in \mathcal{S}_f} |V_j(x;w) - \widetilde{V}_j(x;w)| = o_P(1)$ for all j. The function $\mu_{x,w}$ satisfies the Lipschitz condition of order 1 for w and x because the kernel function W is symmetric and satisfies the Lipschitz condition. For fixed x, the kernel function W^* is constant for w. These facts imply that the function $\widetilde{V}_i(x;w)$ satisfies the Lipschitz condition of order 1 for w as well as for x. Then,

$$\sup_{x\in B} \left| \frac{1}{nh}\{\widetilde{V}_j(x;\widehat{\tau}) - \widetilde{V}_j(x;\tau)\}\right| = |\widehat{\tau} - \tau| O_P\left(\frac{1}{nh^2}\right)$$

by following the proof of Lemma 7.1 in Müller (1992). Considering the third term on the right-hand side of (5.23), one obtains

$$\sup_{x\in B} \sum_{j=1}^{n} \left| \frac{1}{nh}\{\widetilde{V}_j(x;\widehat{\tau}) - \widetilde{V}_j(x;\tau)\}\right| |v(X_j)| \leq |\widehat{\tau} - \tau| O_P\left(\frac{1}{nh^2}\right) \sum_{j=1}^{n} |v(X_j)| I(X_j \in H), \tag{5.25}$$

where $H = \{X_j |\ 1 \leq j \leq n, |X_j - \tau| \leq h \text{ or } |X_j - \widehat{\tau}| \leq h\}$ for which cardinality is $O_P(nh)$ uniformly in τ and $\widehat{\tau}$. Since v is bounded according to (A.1) and $0 < \Delta < \infty$, the right-hand side of (5.25) is $|\widehat{\tau} - \tau| O_P(1/h)$.

Next, the last term on the right-hand side of (5.23) is bounded by

$$\sup_{x\in B} \left| \frac{1}{nh} \sum_{j=1}^{n} V_j(x;\tau) v(X_j) - E\left\{\frac{1}{nh} \sum_{j=1}^{n} V_j(x;\tau) v(X_j)\right\}\right| + \sup_{x\in B} \left| E\left\{\frac{1}{nh} \sum_{j=1}^{n} V_j(x;\tau) v(X_j)\right\} - v(x)\right|. \tag{5.26}$$

By Lemma 5.1, the first term in (5.26) is $O_P\{\sqrt{\log n/(nh)}\}$. Consider the second term in (5.26) for $x \le \tau$. The other case is analogous. Observing that

$$\begin{aligned}
&E\left\{\frac{1}{nh}\sum_{j=1}^{n} V_j(x;\tau)v(X_j)\right\} - v(x) \\
&= \int_{-1}^{c} \frac{1}{\mu_{x,w} f(x)} W(u)v(x+hu)f(x+hu)dx\{1+o(1)\} - v(x) \\
&= O(h)
\end{aligned}$$

uniformly in x, one finds the last term on the right-hand side in (5.23) to be $O_P\{\sqrt{\log n/(nh)} + h\}$. Therefore,

$$\sup_{x\in B} |\widetilde{v}(x;\widehat{\tau}) - v(x)| = O_P\left(\sqrt{\frac{\log n}{nh}} + h\right) + |\widehat{\tau} - \tau| O_P(h^{-1}). \tag{5.27}$$

On the interval A, one obtains

$$\begin{aligned}
&\sup_{x\in A} |\widehat{v}(x;\widehat{\tau}) - v(x)| \\
&\le \sup_{w\in Q}\sup_{x\in \mathcal{S}_f} \left|\frac{1}{nh}\sum_{j=1}^{n} V_j(x;w)v(X_j) - v(x)\right| + \sup_{w\in Q}\sup_{x\in \mathcal{S}_f}\left|\frac{1}{nh}\sum_{j=1}^{n} V_j(x;w)v(X_j)\xi_j\right|.
\end{aligned} \tag{5.28}$$

In the first term on the right-hand side of (5.28), the term $\sup_{x\in\mathcal{S}_f} |\frac{1}{nh}\sum_{j=1}^{n} V_j(x;\tau)v(X_j) - v(x)|$ is bounded by (5.26) with $\sup_{x\in B}$ being replaced by $\sup_{x\in\mathcal{S}_f}$. In this case, the bias term

$$\sup_{x\in\mathcal{S}_f}\left|E\left[\frac{1}{nh}\sum_{j=1}^{n} V_j(x;\tau)v(X_j)\right] - v(x)\right|$$

is $O(1)$ due to the existence of the change point. By Lemma 5.1, the second term of (5.28) is $O_P\{\sqrt{\log n/(nh)} + h\}$. Then the right-hand side of (5.28) is $O_P(1)$. This implies

$$\int_A |\widetilde{v}(x;\widehat{\tau}) - v(x)|^p dx = |\widehat{\tau} - \tau| O_P(1). \tag{5.29}$$

Combining (5.22), (5.27) and (5.29), we get the result immediately. □

ACKNOWLEDGEMENTS

This work was supported by grant No. R05-2002-000-00216-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

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 und Verw. Gebiete*, **37**, 51–75.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York.

CARROLL, R. J. (1982). "Adapting for heteroscedasticity in linear models", *The Annals of Statistics*, **10**, 1224–1233.

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

GASSER, T., SROKA, L. AND JENNEN-STEINMETZ, C. (1986). "Residual variance and residual pattern in nonlinear regression", *Biometrika*, **73**, 625–633.

HALL, P. AND CARROLL, R. J. (1989). "Variance function estimation in regression: The effect of estimating the mean", *Journal of the Royal Statistical Society*, Ser. B, **51**, 3–14.

HALL, P., KAY, J. W. AND TITTERINGTON, D. M. (1990). "Asymptotically optimal difference-based estimation of variance in nonparametric regression", *Biometrika*, **77**, 521–528.

HÄRDLE, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

HÄRDLE, W. (1991). *Smoothing Techniques, With Implementation in S.* Springer-Verlag, New York.

LOADER, C. R. (1996). "Change point estimation using nonparametric regression", *The Annals of Statistics*, **24**, 1667–1678.

MACK, Y. P. AND SILVERMAN, B. W. (1982). "Weak and strong uniform consistency of kernel regression estimates", *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **61**, 405–415.

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

MÜLLER, H.-G. (1992). "Change-points in nonparametric regression analysis", *The Annals of Statistics*, **20**, 737–761.

MÜLLER, H.-G. AND STADTMÜLLER, U. (1987). "Estimation of heteroscedasticity in regression analysis", *The Annals of Statistics*, **15**, 610–625.

RAIMONDO, M. (1998). "Minimax estimation of sharp change points", *The Annals of Statistics*, **26**, 1379–1397.

RICE, J. (1984). "Bandwidth choice for nonparametric regression", *The Annals of Statistics*, **12**, 1215–1230.

RUPPERT, D., WAND, M. P., HOLST, U. AND HÖSSJER, O. (1997). "Local polynomial variance-function estimation", *Technomtrics*, **39**, 262–273.

STUTE, W. (1982). "A law of the logarithm for kernel density estimators", *The Annals of Probability*, **10**, 414–422.

WHITT, W. (1970). "Weak convergence of probability measures on the function space $C[0, \infty)$", *Annals of Mathematical Statistics*, **41**, 939–944.