Journal of the Korean Statistical Society (2004), **33**: 1, pp 59–78

NONPARAMETRIC DISCONTINUITY POINT ESTIMATION IN GENERALIZED LINEAR MODEL†

JIB HUH[1]

ABSTRACT

A regression function in generalized linear model may have a discontinuity/change point at unknown location. In order to estimate the location of the discontinuity point and its jump size, the strategy is to use a nonparametric approach based on one-sided kernel weighted local-likelihood functions. Weak convergences of the proposed estimators are established. The finite-sample performances of the proposed estimators with practical aspects are illustrated by simulated examples.

AMS 2000 subject classifications. Primary 62G08, 62J12; Secondary 62G20.
Keywords. Jump size, kernel smoothing, one-sided kernel, rate of convergence, weak convergence.

1. INTRODUCTION

Suppose that a bivariate sample (X_i, Y_i) of (X, Y) is observed, where Y_i's are real valued responses associated with covariates X_i's having density f with support $[0, 1]$, $i = 1, 2, \ldots, n$. Assume the conditional distribution of Y given $X = x$ belongs to the following one-parameter exponential family:

$$f_{Y|X}(y|x) = \exp\{y\theta(x) - b(\theta(x)) + c(y)\} \tag{1.1}$$

where b and c are some known functions. One may be interested in estimating the regression function $m(x) \equiv E(Y|X = x) = b'(\theta(x))$. In parametric generalized linear models, the function $m(x)$ is modeled linearly via a link function g by

$$\eta(x) \equiv g(m(x)) = \beta_0 + \beta_1 x.$$

Received February 2003; accepted November 2003.
†This work was supported by Korea Research Foundation Grant (KRF-2003-03-C00023).
[1]Department of Statistics, Duksung Women's University, Seoul 132-714, Korea (e-mail : jhuh@duksung.ac.kr)

If $g = (b')^{-1}$, then g is called the canonical link. See McCullagh and Nelder (1988) for details. The conditional density $f_{Y|X}(y|x)$ in (1.1) can be written in terms of $\eta(x)$ as

$$f_{Y|X}(y|x) = \exp\{y(g \circ b')^{-1}(\eta(x)) - b((g \circ b')^{-1}(\eta(x))) + c(y)\} \tag{1.2}$$

where $\circ$ denotes the composition of functions.

A more flexible approach would be to let $\eta(x)$ be a nonparametric function. Fan *et al.* (1995) investigated the extension of the nonparametric regression technique of local polynomial fitting with a kernel weight to generalized linear models and quasi-likelihood contexts. An extension of the smoothing spline methodology to generalized linear models was studied by Green and Silverman (1994). However, a problem arises when a generally smooth function has a discontinuity point in the regression function. The usual nonparametric approaches to regression modeling suffer from poor practical and theoretical performance in such situations. In this paper, a method of estimating the location and jump size of the discontinuity point in $\eta(x)$ using kernel type estimators is introduced. The proposed estimators are based on the difference of left and right one-sided local polynomial estimators using the kernel weighted local-likelihood functions. A one-sided kernel which is supported on the positive half-line has been usually chosen for detecting a discontinuity point. As in Loader (1996), the one-sided kernel which has a non-zero value at the left end of the support is used in this paper. In fact, the estimator for the location of the discontinuity point achieves the rate n^{-1} due to a property of the one-sided kernel. The following works, studied nonparametric discontinuity point estimations in the ordinary regression model case, gave us the motivation described above.

Müller (1992) developed weakly consistent estimators for the location and the jump size of a discontinuity point in the ν^{th} derivative of the regression function using the Gasser-Müller type estimator. The one-sided kernel in Müller (1992) has the zero-value at the left end of the support. Under a stronger assumption of smoothness on the regression function, this led to a slower rate of convergence in comparison with $n^{-1/(2\nu+1)}$. For the case $\nu = 0$, Loader (1996) proposed a discontinuity point estimator, based on the local polynomial fits, that attains the n^{-1} rate. It is assumed in her paper that the errors are Gaussian. Huh and Carrière (2002) used Loader's idea to detect a discontinuity point for a regression function itself or its derivatives without the assumption of Gaussian errors.

There are many another related works about the nonparametric discontinuity point detection problem. Koo (1997) used linear splines to estimate discontinuous regression functions. Wang (1995) and Raimondo (1998) followed wavelet coefficient approaches for detecting discontinuity points. Yin (1988) and Wu and Chu (1993) considered multiple discontinuity points detection problems. McDonald and Owen (1986), Hall and Titterington (1992) and Qiu and Yandell (1998) introduced smoothing algorithms to detect discontinuity points and calculate the regression estimates.

The organization of this paper as follows. Section 2 describes the generalized linear model with the discontinuity point and proposes the estimators for the location of the discontinuity point and the corresponding jump size. In Section 3, the asymptotic properties of the estimators is shown. The proposed approaches are demonstrated by two simulated numerical examples with introducing a method for bandwidth choice in Section 4. All proofs are contained in Section 5.

2. MODEL AND ESTIMATION

Write $\ell(z, y)$ for the logarithm of the conditional density in (1.2) with $\eta(x)$ replaced by z. Define $\ell_i = \partial^i \ell(z, y)/\partial z^i$, $i = 1, 2$. Note that ℓ_i is linear in y for fixed z and that

$$\ell_1(\eta(x), m(x)) = 0, \quad \ell_2(\eta(x), m(x)) = -\rho(x)$$

where $\rho(x) = v^{-1}(x)\{g'(m(x))\}^{-2}$ with $v(x) = \mathrm{Var}(Y|X = x)$. When the canonical link $g = (b')^{-1}$ is used, $\rho(x) = v(x)$.

Assume that a discontinuity point exists for the regression function η at some point τ, $0 < \tau < 1$, as given in the following assumption:

(A1) There exists a constant L_η such that

$$|\eta(x) - \eta(y)| \leq L_\eta |x - y| \quad \text{whenever} \quad (x - \tau)(y - \tau) > 0,$$

i.e. η satisfies the Lipschitz condition of order 1 over $[0, \tau)$ and $(\tau, 1]$. The jump size at the discontinuity point τ in η is given by $\Delta = \eta_+(\tau) - \eta_-(\tau)$ where $\eta_+(\tau) = \lim_{x\to\tau+} \eta(x)$, $\eta_-(\tau) = \lim_{x\to\tau-} \eta(x)$ and $\eta(\tau) = \eta_+(\tau)$. Let us assume $0 < \Delta < \infty$. The case of $-\infty < \Delta < 0$ can be treated in the same way.

Define $\widehat{\eta}_+(x) = \widehat{\alpha}_0^+$ as the right side estimator for $\eta(x)$, where the $(p+1) \times 1$ vector $\widehat{\boldsymbol{\alpha}}^+(x) = (\widehat{\alpha}_0^+, \widehat{\alpha}_1^+, \ldots, \widehat{\alpha}_p^+)^T$ maximizes the following right side kernel weighted local-likelihood function:

$$\sum_{j=1}^{n} \ell\left(\sum_{l=0}^{p} \alpha_l (X_j - x)^l, Y_j\right) K\left(\frac{X_j - x}{h}\right). \tag{2.1}$$

Here K is a one-sided kernel function with support $[0, 1]$ and $h = h_n$ is a sequence of bandwidths, which satisfy the following assumptions:

(A2) The function K satisfies $\int_0^1 K(u)du = 1$, $K(0) > 0$ and $K(u) \geq 0$ for $0 < u \leq 1$.

(A3) $h \to 0$, $nh/\log n \to \infty$ and $nh^2 \to 0$, as $n \to \infty$.

The left side estimator for $\eta(x)$ can be defined similarly. Define $\widehat{\eta}_-(x) = \widehat{\alpha}_0^-$, where the $(p+1) \times 1$ vector $\widehat{\boldsymbol{\alpha}}^- = (\widehat{\alpha}_0^-, \widehat{\alpha}_1^-, \ldots, \widehat{\alpha}_p^-)^T$ maximizes the left side kernel weighted local-likelihood function:

$$\sum_{j=1}^{n} \ell\left(\sum_{l=0}^{p} \alpha_l (X_j - x)^l, Y_j\right) K\left(\frac{x - X_j}{h}\right). \tag{2.2}$$

The local-likelihood functions in (2.1) and (2.2) are based on the one-sided data at the right and left of x, respectively. In order to guarantee that the kernel weighted log-likelihood functions in (2.1) and (2.2) are concave in $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)^T$, which ensures the uniqueness of the maximizer, the assumption $\ell_2(z, y) < 0$ for all real z and y in the range of the response variable is needed. If the canonical link is used, it is automatically satisfied. The canonical link is then chosen in this paper. In this case, one obtains

$$\ell_1(z, y) = y - g^{-1}(z) \quad \text{and} \quad \ell_2(z, y) = -(g^{-1})'(z).$$

An estimator of the jump size at a point x can be defined by taking the differences of these two estimators: $\widehat{\Delta}(x) = \widehat{\eta}_+(x) - \widehat{\eta}_-(x)$. 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 discontinuity point τ. An estimator of the jump size Δ may be obtained by

$$\widehat{\Delta}(\widehat{\tau}) = \widehat{\eta}_+(\widehat{\tau}) - \widehat{\eta}_-(\widehat{\tau}).$$

There does not exist the explicit solutions to the maximization (2.1) and (2.2) unless $p = 0$. Note that in the case $p = 0$, it is easy to show that the estimators $\widehat{\eta}_\pm(x)$ can be written as

$$\widehat{\eta}_\pm(x) = g\left(\frac{\sum_{j=1}^{n} K\left(\pm(X_j - x)/h\right) Y_j}{\sum_{j=1}^{n} K\left(\pm(X_j - x)/h\right)}\right). \tag{2.3}$$

The choice of the order p of local polynomial turns out to have little impact on the asymptotic property for $\widehat{\tau}$ derived in Section 3. The order $p = 0$ is then chosen. The assumption below is required in order to show the asymptotic result described in the following section.

(A4) The function f having support $[0,1]$ satisfies the Lipschitz condition of order 1 over $[0,1]$ and $\inf_{x\in[0,1]} f(x) > 0$.

(A5) The function v satisfies $\inf_{x\in[0,1]} v(x) > 0$. The function v satisfies the Lipschitz condition of order 1 over $[0,\tau)$ and $(\tau, 1]$.

The assumption (A5) describes that the variance function v is discontinuous at the point τ when there exists the discontinuity point in the regression function.

Note that it is easy and convenient to show the asymptotic distribution of $\widehat{\tau}$ with the estimators of m_+ and m_- instead of $\widehat{\eta}_\pm$ in (2.3), where $m_\pm(\cdot) = g^{-1}(\eta_\pm(\cdot))$. The estimators of $m_\pm(x)$ can be given by applying the inverse of the link function: $\widehat{m}_\pm(x) = g^{-1}(\widehat{\eta}_\pm(x))$. Define $\Lambda = g^{-1}(\eta_+(\tau)) - g^{-1}(\eta_-(\tau))$ and $\widehat{\Lambda}(x) = g^{-1}(\widehat{\eta}_+(x)) - g^{-1}(\widehat{\eta}_-(x))$. The estimators of τ and Λ are given by

$$\widehat{\tau} = \inf\left\{z \in Q : \widehat{\Lambda}(z) = \sup_{x\in Q} \widehat{\Lambda}(x)\right\} \text{ and } \widehat{\Lambda}(\widehat{\tau}) = g^{-1}(\widehat{\eta}_+(\widehat{\tau})) - g^{-1}(\widehat{\eta}_-(\widehat{\tau})).$$

All previous works of nonparametric discontinuity point estimations of the ordinary regression model case did not consider the existence of discontinuity point in the variance function. However, in this setting, the functions m and v are discontinuous at τ because they depend on η.

3. Asymptotic Properties

First, a weak convergence of the sequence of the process $\{\varphi_n(z) : -T \le z \le T\}$ in the following theorem is described, where

$$\varphi_n(z) = nh\left\{\widehat{\Lambda}\left(\tau + \frac{z}{n}\right) - \widehat{\Lambda}(\tau)\right\}$$

and $T < \infty$. Existence of the unique maximizer of the limit of the process φ_n will be discussed later on. The process φ_n lies in the space, denoted by $\mathcal{D}[-T,T]$, of functions having at most finitely many discontinuities defined on $[-T,T]$. To obtain the theorem, consider the following additional assumption:

(A6) $E(|Y|^{2+\epsilon}|X = x) < \infty$, for all x and some positive ϵ.

Let $\xrightarrow{\mathcal{W}}$ denote weak convergence in the space $\mathcal{D}([-T,T])$, and define $v_+(\tau) = \lim_{y\to\tau+} v(y)$ and $v_-(\tau) = \lim_{y\to\tau-} v(y)$.

THEOREM 3.1. *Suppose that assumptions (A1)–(A6) are satisfied.*

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

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

$$\sigma = 2\sqrt{\frac{\tilde{v}(\tau)+\Lambda^2}{f(\tau)}K(0)} \tag{3.1}$$

with $\tilde{v}(\tau) = v_+(\tau)$ when $z \geq 0$ and $\tilde{v}(\tau) = v_-(\tau)$ when $z < 0$.

Next, the asymptotic distribution of $\widehat{\tau}$ is described. Let Z_M denote a maximizer of the process φ on $[-M, M]$. By Remark 5.3 in Bhattacharya and Brockwell (1976) with the assumption $K(0) > 0$, Z_M is unique with probability one. Let Z_n be the location of the maximum of φ_n. By construction,

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

Theorem 3 in Bhattacharya and Brockwell (1976) then gives $Z_n \xrightarrow{\mathcal{D}} Z$, where Z is the global maximizer of φ on $(-\infty,\infty)$. Remark 5.3 in Bhattacharya and Brockwell also showed that the maximizer Z is unique with probability one. Therefore, Corollary 3.1 follows immediately.

COROLLARY 3.1. *Suppose that assumptions in Theorem 3.1 are satisfied.*

$$n(\widehat{\tau}-\tau) \xrightarrow{\mathcal{D}} \underset{z\in(-\infty,\infty)}{\operatorname{argmax}} \big\{ -\Lambda K(0)|z| + \sigma W(z)\big\}.$$

Raimondo (1998) showed that the minimax rate for the location problem is $n^{-1/(2\nu+1)}$ for the ν^{th} derivative of a class of regression function satisfying a Lipschitz condition of order 1. Although the interesting function is regression function in generalized linear model, the proposed estimator $\widehat{\tau}$ achieves the optimal rate n^{-1} according to Corollary 3.1.

The following corollary describes the asymptotic distribution of the estimator for the jump size $\widehat{\Lambda}(\widehat{\tau})$ as a consequence of Theorem 3.1.

THEOREM 3.2. *Suppose that assumptions in Theorem 3.1 are satisfied.*

$$\sqrt{nh}\big(\widehat{\Lambda}(\widehat{\tau}) - \Lambda\big) \xrightarrow{\mathcal{D}} N\Big(0, \frac{v_+(\tau) + v_-(\tau)}{f(\tau)} \int_0^1 \{K(u)\}^2 du\Big).$$

This work also applies to quasi-likelihood models, where only the relationship between the mean and the variance is specified. In this situation the proposed estimators can be achieved by replacing the log-likelihood by a quasi-likelihood.

4. NUMERICAL IMPLEMENTATIONS

An important practical problem in discontinuity point analysis is the selection of the bandwidth. Hart and Yi (1998) proposed the one-sided cross-validation to select the bandwidth for estimating the nonparametric regression function which has no discontinuity point. In all procedures to estimate the locations and jump sizes of discontinuity points in this section one-sided cross-validation is then used. In practice, Hart and Yi (1998) suggested that one could average the left and right one-sided cross-validation curves and use the minimizer of the average. Then, the following criterion is chosen

$$CV(h) = \sum_{\{i: X_i \in [h, 1-h]\}} \frac{1}{n_h} \Big[\{Y_i - \widehat{m}_{+,h,-i}(X_i)\}^2 + \{Y_i - \widehat{m}_{-,h,-i}(X_i)\}^2\Big] \quad (4.1)$$

where $\widehat{m}_{\pm,h,-i}(X_i)$ is the right and left estimators of $m(X_i)$ without using the i^{th} observation and n_h is the number of data in the interval $[h, 1-h]$ for a given h. As Müller (1992), it is desirable to choose a relatively small bandwidth for estimating location and jump size as compared to the bandwidth chosen for estimating a regression function. The bandwidth which is the smallest local minimizer of (4.1) is then taken.

The maximization of the log-likelihoods was carried out by the Newton-Raphson iteration when the order of the polynomial p is greater than 0. For example, let $\ell(\boldsymbol{\alpha})$ be the locally kernel weighted log-likelihood function in (2.1). The estimator $\widehat{\boldsymbol{\alpha}}$ satisfies the following equation $\nabla_{\boldsymbol{\alpha}} \ell(\widehat{\boldsymbol{\alpha}}) = 0$ where $\nabla_{\boldsymbol{\alpha}}$ denotes the gradient with respect to $\boldsymbol{\alpha}$. Let $\nabla^2_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha})$ be the $(p+1) \times (p+1)$ Hessian matrix of ℓ. The Newton-Raphson method for computing $\widehat{\boldsymbol{\alpha}}$ starts with an initial guesstimate $\widehat{\boldsymbol{\alpha}}^{(0)}$ and iteratively determine $\widehat{\boldsymbol{\alpha}}^{(k)}$ from the formula

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \widehat{\boldsymbol{\alpha}}^{(k)} - \big[\nabla^2_{\boldsymbol{\alpha}} \ell(\widehat{\boldsymbol{\alpha}}^{(k)})\big]^{-1} \nabla_{\boldsymbol{\alpha}} \ell(\widehat{\boldsymbol{\alpha}}^{(k)}).$$

TABLE 4.1 *The Monte Carlo estimates of the MSEs and the averages with standard errors in parentheses for the discontinuity point estimators in the case of η_1*

n	h	*Average of* $\hat{\tau}$	*MSE of* $\hat{\tau}$	*Average of* $\hat{\Delta}$	*MSE of* $\hat{\Delta}$
800	0.09995	0.49984	0.11920×10^{-3}	2.89383	0.39745
	(0.00087)	(0.00035)	(0.05474×10^{-3})	(0.01965)	(0.05745)
1200	0.09402	0.50005	0.07790×10^{-3}	2.89149	0.30498
	(0.00084)	(0.00028)	(0.05143×10^{-3})	(0.01712)	(0.06543)
1600	0.09115	0.500110	0.00230×10^{-3}	2.86634	0.19010
	(0.00082)	(0.00005)	(0.00077×10^{-3})	(0.01312)	(0.00968)

Finally, with regard to a stopping criterion, one which is employed is to stop when $\|\hat{\boldsymbol{\alpha}}^{(k)}-\hat{\boldsymbol{\alpha}}^{(k+1)}\|_2 \le \omega$ where $\|\cdot\|_2$ denotes the L_2 norm. For most implementations, choosing $\omega = 10^{-7}$ is sufficient.

To investigate the numerical performances of the proposed estimators defined in Section 2, simulation studies are carried out. Sample sizes considered here are 800, 1200 and 1600. All the results of the simulations are based on 1000 pseudo samples of each size. The predictor variable X_i's are random sample from the uniform distribution. Throughout these simulations, the one-sided kernel function is

$$K(x) = \frac{15}{8}(1-x^2)^2 1_{[0\le x\le 1]}.$$

To estimate the location of the discontinuity point, the jump sizes at $x_k = k/100$, $k = 1, \ldots, 100$, are computed first, and then choose a point which maximizes the absolute value of the calculated jump sizes over the interval Q. As suggested in Müller (1992), the interval $Q = [h, 1-h]$ for the simulation settings is taken.

First, the proposed method is applied to binary responses with the Bernoulli distribution having a discontinuity point at $\tau = 0.5$. The regression function m_1 is given by

$$m_1(x) = \frac{\exp(\eta_1(x))}{1+\exp(\eta_1(x))}$$

where $\eta_1(x) = -3x + 3 \times 1_{[0.5\le x\le 1]}$. Then, the jump sizes are $\Delta_1 = 3.0$ and $\Lambda_1 = 0.635149$. The logit link $g(u) = \log(u/(1-u))$ is canonical. In this case, the variance function $v_1(x) = \exp(\eta_1(x))/\{1+\exp(\eta_1(x))\}^2$ also has the discontinuity point at $\tau = 0.5$. In applying the proposed estimators, the local constant fitting of $p = 0$ is chosen.

The second example concerns the case of nonnegative integer responses with

TABLE 4.2 *The Monte Carlo estimates of the MSEs and the averages with standard errors in parentheses for the discontinuity point estimators in the case of η_2*

n	h	*Average of* $\widehat{\tau}$	*MSE of* $\widehat{\tau}$	*Average of* $\widehat{\Delta}$	*MSE of* $\widehat{\Delta}$
800	0.08662	0.25025	0.00750×10^{-3}	-0.98558	0.03792
	(0.00066)	(0.00009)	(0.00321×10^{-3})	(0.00614)	(0.00786)
1200	0.07497	0.25006	0.00120×10^{-3}	-0.99080	0.02222
	(0.00058)	(0.00004)	(0.00092×10^{-3})	(0.00470)	(0.00392)
1600	0.06956	0.25001	0.00010×10^{-3}	-0.99478	0.01448
	(0.00060)	(0.00001)	(0.00010×10^{-3})	(0.00380)	(0.00067)

the Poisson distribution. The regression function m_2 is given by

$$m_2(x) = \exp(\eta_2(x))$$

where $\eta_2(x) = 0.5\{\exp(2(x-0.25))1_{[0\le x<0.25]} - \exp(-2(x-0.25))1_{[0.25\le x\le 1]}\}+2.0$. The regression function has the discontinuity point at $\tau = 0.25$. The jump sizes of the discontinuity point are $\Delta_2 = -1.0$ and $\Lambda_2 = -7.700805$. The canonical link is $g(u) = \log(u)$. The variance function $v_2(x) = \exp(\eta_2(x))$ has then the discontinuity point at $\tau = 0.25$. The order of the polynomial fit $p = 1$ is chosen and maximization of the log-likelihood is carried out by the Newton-Raphson iteration described above. The C programming language has been used for these simulations.

Table 4.1 and 4.2 show the Monte Carlo estimates of the mean squared errors (MSE) and averages of the estimated the locations and jump sizes with the bandwidths selected by the criterion in (4.1). From these tables, the improvement of MSE of $\widehat{\tau}$ gets larger than that of $\widehat{\Delta}$ as sample size increases. In fact, it is needed to compare the estimate with other estimate which also reflects existence of a discontinuity point in regression function of generalized linear model. But, there is no published result in this part.

In order to display the empirical distribution of $\widehat{\tau}$ in these simulation settings, Table 4.3 and 4.4 report the frequencies with which discontinuities identified by the 1000 replications of size 800 using the bandwidths selected by the criterion in (4.1). Here the integer k denoted by the index for the point x_k which maximizes the absolute value of the estimated jump sizes.

TABLE 4.3 *Discontinuity point identification frequency in the case of* η_1

k (location)	31	35	39	48	49	50	51	52	53	68
frequency	1	1	1	10	33	899	46	6	2	1

TABLE 4.4 *Discontinuity point identification frequency in the case of* η_2

k (location)	24	25	26	29
frequency	1	985	10	4

5. PROOFS

To prove Theorem 3.1, an asymptotic expression of φ_n will be described first. Let

$$\begin{aligned} C_n^+(w,u,z) &= \frac{1}{f(\tau+z_n)}K\left(\frac{w-\tau-z_n}{h}\right)\{u-m_+(\tau+z_n)-\Lambda\,1_{[z<0]}\} \\ &\quad -\frac{1}{f(\tau)}K\left(\frac{w-\tau}{h}\right)\{u-m_+(\tau)\}, \\ C_n^-(w,u,z) &= \frac{1}{f(\tau+z_n)}K\left(\frac{\tau+z_n-w}{h}\right)\{u-m_-(\tau+z_n)+\Lambda\,1_{[z>0]}\} \\ &\quad -\frac{1}{f(\tau)}K\left(\frac{\tau-w}{h}\right)\{u-m_-(\tau)\}, \\ \phi_n(z) &= \sum_{j=1}^{n}\{C_n^+(X_j,Y_j,z)-C_n^-(X_j,Y_j,z)\} \end{aligned}$$

for $z \neq 0$ and $\phi_n(0)=0$.

LEMMA 5.1. *Suppose that the assumptions (A1)-(A4) are satisfied. Then,*

$$\varphi_n(z) = \phi_n(z)(1+o_p(1))$$

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

PROOF. Define $z_n = z/n$ and The sequence of the process $\varphi_n(z)$ can be written as follows:

$$\begin{aligned}
&\varphi_n(z)\\
&= nh\Big[\{g^{-1}(\widehat{\eta}_+(\tau+z_n)) - g^{-1}(\eta_+(\tau+z_n))\} - \{g^{-1}(\widehat{\eta}_+(\tau)) - g^{-1}(\eta_+(\tau))\}\\
&\quad - \{g^{-1}(\widehat{\eta}_-(\tau+z_n)) - g^{-1}(\eta_-(\tau+z_n))\} + \{g^{-1}(\widehat{\eta}_-(\tau)) - g^{-1}(\eta_-(\tau))\} - \Lambda\Big]\\
&= nh\Big[\{\widehat{m}_+(\tau+z_n) - m_+(\tau+z_n)\} - \{\widehat{m}_+(\tau) - m_+(\tau)\}\\
&\quad - \{\widehat{m}_-(\tau+z_n) - m_-(\tau+z_n)\} + \{\widehat{m}_-(\tau) - m_-(\tau)\} - \Lambda\Big] \qquad (5.1)
\end{aligned}$$

for all $z \neq 0$. By (2.3),

$$\begin{aligned}
&\widehat{m}_\pm(x) - m_\pm(x)\\
&= \frac{1}{nh}\sum_{j=1}^{n} K\left(\pm\frac{X_j - x}{h}\right)\{Y_j - m_\pm(x)\}\left\{\frac{1}{nh}\sum_{j=1}^{n} K\left(\pm\frac{X_j - x}{h}\right)\right\}^{-1} \qquad (5.2)
\end{aligned}$$

where $x \in Q$.

Let

$$\widehat{f}_+(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{X_j - x}{h}\right) \quad \text{and} \quad \widehat{f}_-(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right).$$

When the kernel function is positive and bounded on a compact set, Stute (1982) described the following result in Theorem 1.3:

$$\sup_{x \in Q}\left|\widehat{f}_\pm(x) - f(x)\right| = O_p\left(h + \sqrt{\frac{\log n}{nh}}\right). \qquad (5.3)$$

It follows from (5.1), (5.2) and (5.3), the result follows uniformly in $z \in [-T, T]$. □

LEMMA 5.2. *Suppose that assumptions (A1)–(A4) are satisfied. Then,*

$$E(\phi_n(z)) = -\Lambda K(0)|z| + o(1)$$

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

PROOF. Let us prove the lemma for $z > 0$, as the other case can be dealt similarly. By the assumption (A4).

$$
\begin{aligned}
&E\big[C_n^+(X_1, Y_1, z)\big] \\
&= E\left[\frac{1}{f(\tau+z_n)}K\left(\frac{X_1-\tau-z_n}{h}\right)\{Y_1 - m_+(\tau+z_n)\}\right. \\
&\quad \left. - \frac{1}{f(\tau)}K\left(\frac{X_1-\tau}{h}\right)\{Y_1 - m_+(\tau)\}\right] \\
&= h\int K(u)\big[\{m(\tau+z_n+hu) - m_+(\tau+z_n)\} \\
&\quad - \{m(\tau+hu) - m_+(\tau)\}\big]du\{1+O(h)\} \\
&\quad + h\int K(u)\left\{\frac{f(\tau+z_n+hu)}{f(\tau+z_n)} - \frac{f(\tau+hu)}{f(\tau)}\right\} \\
&\quad \times\{m(\tau+hu) - m_+(\tau)\}du \qquad (5.4)
\end{aligned}
$$

where the $O(h)$ term is uniform in $z \in [-T, T]$. Similarly,

$$
\begin{aligned}
&E\big[C_n^-(X_1, Y_1, z)\big] \\
&= h\int K(u)\big[\{m(\tau+z_{n,\nu}-hu) - m_-(\tau+z_n) + \Lambda\} \\
&\quad - \{m(\tau-hu) - m_-(\tau)\}\big]du\{1+O(h)\} \\
&\quad + h\int K(u)\left\{\frac{f(\tau+z_n-hu)}{f(\tau+z_n)} - \frac{f(\tau-hu)}{f(\tau)}\right\} \\
&\quad \times\{m(\tau-hu) - m_-(\tau)\}du \qquad (5.5)
\end{aligned}
$$

where the $O(h)$ terms are uniform in $z \in [-T, T]$. By the assumption (A1),

$$m(\tau \pm hu) - m_\pm(\tau) = \{m_\pm(\tau) - m_\pm(\tau_\pm)\}(1+o(1)) \qquad (5.6)$$

for all $u > 0$ where $\tau_\pm$ lie between τ and $\tau \pm hu$. Analogously,

$$m(\tau+z_n+hu) - m_+(\tau+z_n) = \{m_+(\tau+z_n) - m_+(\tau'_+)\}(1+o(1)) \qquad (5.7)$$

where τ'_+ lies between $\tau+z_n$ and $\tau+z_n+hu$. The second terms of (5.4) and (5.5) are $O(h^2/n)$ uniformly in z by (5.6) and (A4). Furthermore, in the case of $E\big[C_n^+(X_1, Y_1, z)\big]$, it is easy to see that the first term is also $O(h/n)$ uniformly in z since the difference between (5.6) and (5.7) is $O(1/n)$. Hence,

$$E\big[C_n^+(X_1, Y_1, z)\big] = O\left(\frac{h}{n}\right) \qquad (5.8)$$

uniformly in z.

Since the discontinuity point τ lies between $\tau+z_n-h$ and $\tau+z_n$, approximation of the first term of $E\big[C_n^-(X_1,Y_1,z)\big]$ in (5.5) is slightly different. In this case the interval of integration is divided into two parts. Note that, for $0 < u \le z/(nh)$,

$$m(\tau + z_n - hu) - m_-(\tau + z_n) = \{m_-(\tau + z_n) - m_-(\tau_-^*)\}(1 + o(1)) \quad (5.9)$$

where τ_-^* lies between $\tau + z_n$ and $\tau + z_n - hu$. However, $z/(nh) < u < 1$,

$$\begin{aligned} m(\tau + z_n - hu) - m_-(\tau) &= \{m_-(\tau'_-) - m_-(\tau)\}(1 + o(1)), & (5.10)\\ m_-(\tau + z_n) - m_+(\tau) &= \{m_+(\tau''_+) - m_+(\tau)\}(1 + o(1)) & (5.11)\end{aligned}$$

by Taylor expansions where τ'_- lies between τ and $\tau+z_n-hu$ and τ''_+ lies between τ and $\tau+z_n$. By the difference between (5.10) and (5.11),

$$\begin{aligned} & m(\tau + z_n - hu) - m_-(\tau + z_n)\\ &= [-\Lambda + \{m_-(\tau'_-) - m_-(\tau)\} - \{m_+(\tau''_+) - m_+(\tau)\}]. \end{aligned} \quad (5.12)$$

By (5.7), (5.9) and (5.12), the integral of the first term of $E\big[C_n^-(X_1,Y_1,z)\big]$ in (5.5) equals

$$\begin{aligned} &\int_0^{z/(nh)} K(u)\Lambda\, du + \int_0^{z/(nh)} K(u)\{m_-(\tau_-^*) - m_-(\tau + z_n)\}du\\ &- \int_{z/(nh)}^1 K(u)\{m_+(\tau''_+) - m_+(\tau)\}du + \int_{z/(nh)}^1 K(u)\{m_-(\tau'_-) - m_-(\tau)\}du\\ &- \int_0^1 K(u)\{m_-(\tau_-) - m_-(\tau)\}du \end{aligned}$$

uniformly in z. This with (A2) leads to

$$E\big[C_n^-(X_1,Y_1,z)\big] = h\int_0^{z/(nh)} K(u)\Lambda\, du + O\left(\frac{h}{n}\right) \quad (5.13)$$

uniformly in z. Combining (5.8) and (5.13),

$$E(\phi_n(z)) = -nh\Lambda \int_0^{z/(nh)} K(u)du + O\left(h\right)$$

uniformly in $z \in [0,T]$. Since $K(u) = K(0)(1+o(1))$ uniformly for $u \in [0,T/(nh)]$, the result follows. □

LEMMA 5.3. *Suppose that assumptions (A1)-(A5) are satisfied.*

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

where the $o(1)$ *terms hold uniformly in* $z_1,z_2\in[-T,T]$.

PROOF. The lemma for $z_1,z_2\geq 0$ will be proved first. By Lemma 5.2,

$$\begin{aligned}&\mathrm{Cov}(\phi_n(z_1),\phi_n(z_2))\\ &=n\mathrm{Cov}\Big[C_n^+(X_1,Y_1,z_1)-C_n^-(X_1,Y_1,z_1),C_n^+(X_1,Y_1,z_2)-C_n^-(X_1,Y_1,z_2)\Big]\\ &=n\Big[E\{C_n^+(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)-C_n^+(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)\\ &\quad -C_n^-(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)+C_n^-(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)\}\Big]\\ &\quad +O\Big(\frac{1}{n}\Big).\end{aligned}\tag{5.14}$$

Lemma 5.2 implies the $O(1/n)$ term directly. Define $z_{(1)}=\min(z_1,z_2)$, $z_{(2)}=\max(z_1,z_2)$, $\tau_{(1)n}=\tau+z_{(1)}/n$ and $\tau_{(2)n}=\tau+z_{(2)}/n$. Let

$$D_n^+(u,z)=\frac{1}{f(\tau+z_n)}K\left(\frac{u-\tau-z_n}{h}\right)-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right),$$
$$D_n^-(u,z)=\frac{1}{f(\tau+z_n)}K\left(\frac{\tau+z_n-u}{h}\right)-\frac{1}{f(\tau)}K\left(\frac{\tau-u}{h}\right).$$

Consider the first term in the bracket at (5.14) first. Note that

$$\sup_{u\in[x,x+h]}\big|\eta(u)-\eta_+(x)\big|\leq(\mathrm{const})h\tag{5.15}$$

for $x=\tau$ or $\tau+z_n$. By the assumptions (A2)-(A5) and (5.15),

$$\begin{aligned}&E\big[C_n^+(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)\big]\\ &=E\big[D_n^+(X_1,z_{(1)})D_n^+(X_1,z_{(2)})\{Y_1-m(X_1)\}^2\big]+O(h^3)\\ &=\Bigg[\int_\tau^{\tau_{(1)n}}\left\{\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right)\right\}^2+\int_{\tau_{(1)n}}^{\tau_{(2)n}}D_n^+(u,z_{(1)})\left\{-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right)\right\}\\ &\quad+\int_{\tau_{(2)n}}^{\tau_{(2)n}+h}D_n^+(u,z_{(1)})D_n^+(u,z_{(2)})\Bigg]vf(u)du+O(h^3)\end{aligned}$$

$$= h\frac{v_+(\tau)}{f(\tau)}\left[\{K(0)\}^2\frac{z_{(1)}}{nh} + O\left(\frac{1}{(nh)^2}\right)\right](1+O(h)) + O(h^3) \tag{5.16}$$

uniformly in z_1 and z_2 where $vf(u) = v(u)f(u)$. Next, consider the second term in the bracket at (5.14) for the case $z_{(1)} = z_1$. The other cases can be dealt in a similar way. Note that $C_n^+(w,u,z) = 0$ for $w < \tau$, and that

$$\sup_{u\in[\tau,\tau+z_n]} |\eta(u) - \eta_\pm(\tau+z_n)| \le (\text{const})n^{-1};$$
$$\sup_{u\in[\tau,\tau+z_n]} |\eta(u) - \eta_+(\tau)| \le (\text{const})n^{-1}. \tag{5.17}$$

By (5.17),

$$\begin{aligned}
&E\big[C_n^+(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)\big]\\
&= E\left[D_n^+(X_1,z_{(1)})\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-u}{h}\right)\{Y_1 - m(X_1)\}^2\right] + O\left(n^{-3}\right)\\
&= \left[\int_\tau^{\tau_{(1)n}}\left\{-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right)\right\}\left\{\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-u}{h}\right)\right\}\right.\\
&\quad\left. + \int_{\tau_{(1)n}}^{\tau_{(2)n}} D_n^+(u,z_{(1)})\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-u}{h}\right)\right] vf(u)du + O\left(n^{-3}\right)\\
&= -h\frac{v_+(\tau)}{f(\tau)}\left[\{K(0)\}^2\frac{z_{(1)}}{nh} + O\left(\frac{1}{(nh)^2}\right)\right](1+O(h)) + O\left(n^{-3}\right)
\end{aligned} \tag{5.18}$$

uniformly in z_1 and z_2. Analogously,

$$\begin{aligned}
&E\big[C_n^-(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)\big]\\
&= \int_\tau^{\tau_{(1)n}}\frac{1}{f(\tau_{(1)n})}K\left(\frac{\tau_{(1)n}-u}{h}\right)\left\{-\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right)\right\} vf(u)du + O\left(n^{-3}\right)\\
&= -h\frac{v_+(\tau)}{f(\tau)}\left[\{K(0)\}^2\frac{z_{(1)}}{nh} + O\left(\frac{1}{(nh)^2}\right)\right](1+O(h)) + O\left(n^{-3}\right)
\end{aligned} \tag{5.19}$$

uniformly in z_1 and z_2. Now, consider the last term in the bracket at (5.14). Note that

$$\sup_{u\in[\tau-h,\tau)} \Big|\eta(u) - \eta_-(\tau)\Big| \le (\text{const})h. \tag{5.20}$$

On the other hand, by (5.12)

$$\sup_{u\in[\tau+z_n-h,\tau)} \Big|m(u) - m_-(\tau+z_n) - \Lambda\Big| \le (\text{const})h. \tag{5.21}$$

By (5.20) and (5.21) with the definition of C_n^-, the last term at (5.14) equals

$$\begin{aligned}
&E[C_n^-(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)] \\
&= E\Big[D_n^-(X_1,z_{(1)})D_n^-(X_1,z_{(2)})\{Y_1 - m(X_1)\}^2 \\
&\quad + \frac{1}{f(\tau_{(1)n})}K\left(\frac{\tau_{(1)n}-X_1}{h}\right)\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-X_1}{h}\right)\Lambda^2\Big] + O(h^3) \\
&= \int_\tau^{\tau_{(1)n}} \frac{1}{f(\tau_{(1)n})}K\left(\frac{\tau_{(1)n}-u}{h}\right)\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-u}{h}\right)\{v(u)+\Lambda^2\}f(u)du \\
&\quad + \int_{\tau-h}^{\tau} D_n^-(u,z_{(1)})D_n^-(u,z_{(2)})vf(u)du + O(h^3) \\
&= h\frac{\{v_+(\tau)+\Lambda^2\}}{f(\tau)}\left[\{K(0)\}^2\frac{z_{(1)}}{nh} + O\left(\frac{1}{(nh)^2}\right)\right](1+O(h)) + O(h^3) \qquad (5.22)
\end{aligned}$$

uniformly in z_1 and z_2. Combining the first leading terms in (5.16), (5.18), (5.19) and (5.22) concludes the proof of Lemma 5.3 for the case $z_1, z_2 > 0$. It can be shown in a similar way that the lemma follows for the case $z_1, z_2 < 0$ too.

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

$$\begin{aligned}
&E\Big[C_n^+(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)\Big] \\
&= \Big[\int_\tau^{\tau_{(2)n}} D_n^+(u,z_{(1)})\left\{\frac{1}{f(\tau)}K\left(\frac{u-\tau}{h}\right)\right\} \\
&\quad + \int_{\tau_{(2)n}}^{\tau_{(2)n}+h} D_n^+(u,z_{(1)})D_n^+(u,z_{(2)})\Big]vf(u)du, \\
&E\Big[C_n^+(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)\Big] = 0, \\
&E\Big[C_n^-(X_1,Y_1,z_1)C_n^+(X_1,Y_1,z_2)\Big] \\
&= \Big[\int_{\tau_{(1)n}}^{\tau}\left\{\frac{1}{f(\tau_{(1)n})}K\left(\frac{u-\tau_{(1)n}}{h}\right)\right\}D_n^-(u,z_{(2)}) \\
&\quad + \int_\tau^{\tau_{(2)n}} D_n^+(u,z_{(1)})\left\{\frac{1}{f(\tau_{(2)n})}K\left(\frac{\tau_{(2)n}-u}{h}\right)\right\}\Big]vf(u)du,
\end{aligned}$$

$$
\begin{aligned}
&E\Big[C_n^-(X_1,Y_1,z_1)C_n^-(X_1,Y_1,z_2)\Big]\\
&=\Bigg[\int_{\tau_{(1)n}}^{\tau}\left\{\frac{1}{f(\tau)}K\left(\frac{\tau-u}{h}\right)\right\}D_n^-(u,z_{(2)})\\
&\quad+\int_{\tau_{(1)n}-h}^{\tau_{(1)n}}D_n^-(u,z_{(1)})D_n^-(u,z_{(2)})\Bigg]vf(u)du
\end{aligned}
\tag{5.23}
$$

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

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

for all w. Since $D_n^{\pm}(w,z)=O(n^{-1}h^{-1})$ uniformly in w and z, all of the leading terms in (5.23) are $O(1/(nh)^2)$. This implies the result immediately. □

LEMMA 5.4. *Suppose that the assumptions in Theorem 3.1 are satisfied. For each $z\in[-T,T]$, $\phi_n(z)$ satisfies Lyapounov's condition.*

PROOF. Let us show the lemma for $z>0$. The other case can be dealt similarly. By Lemma 5.3, $\mathrm{Var}(\phi_n(z))=O(1)$. It will be shown that, for some positive ϵ,

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

as $n\to\infty$. By the assumption (A6), $E(|\{Y_1-m(X_1)\}^2|^{2+\zeta}|X=x)<\infty$ for all x. Note that

$$
\begin{aligned}
&L_n(z)\\
&\le(\text{const})n\cdot 2^{2+\epsilon}E\Bigg[\left\{\left|D_n^+(X_1,Y_1,z)\right|^{2+\epsilon}+\left|D_n^-(X_1,Y_1,z)\right|^{2+\epsilon}\right\}\left|Y_1-m(X_1)\right|^{2+\epsilon}\\
&\quad+\left|\frac{1}{f(\tau+z_n)}K\left(\frac{\tau+z_n-X_1}{h}\right)\Lambda\right|^{2+\epsilon}\Bigg]+O\left(nh^{3+\epsilon}\right)\\
&=O\left(nh\left\{\left(\frac{1}{nh}\right)^{2+\epsilon}+\left(\frac{1}{n}\right)^{2+\epsilon}+h^{2+\epsilon}\right\}\right).
\end{aligned}
$$

By the assumption (A2), the result follows. □

LEMMA 5.5. *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_3 and a nondecreasing and continuous function F such as

$$E(\psi_n(z_1) - \psi_n(z_2))^2 \le C_3|F(z_2) - F(z_1)|^2, \tag{5.24}$$

for sufficiently large n. By Lemma 5.3, there exists a positive constant C_3 such that

$$\begin{aligned} E(\psi_n(z_1) - \psi_n(z_2))^2 &= \mathrm{Var}(\phi_n(z_1)) - \mathrm{Var}(\phi_n(z_2)) - 2\mathrm{Cov}(\phi_n(z_1), \phi_n(z_2)) \\ &\le C_3|z_2 - z_1|, \end{aligned}$$

for sufficiently large n. This concludes the proof of Lemma 5.5. □

PROOF OF THEOREM 3.1. Lemma 5.4 implies that $\psi_n(z)$, for fixed $z \in [-T, T]$, converges weakly to a normal distribution. Furthermore, by the Cramer-Wold device it may be shown that for fixed $z_1, \dots, z_l, z_i \in [-T, T]$,

$$(\psi_n(z_1), \psi_n(z_2), \dots, \psi_n(z_l)) \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

where Σ is the asymptotic covariance described in Lemma 5.3. This concludes the proof. See Theorem 8.1 and 12.3 of Billingsley (1968). □

PROOF OF THEOREM 3.2. Theorem 3.1 shows that $\sqrt{nh}\{\widehat{\Lambda}(\tau + Z_n/n) - \widehat{\Lambda}(\tau)\} \xrightarrow{p} 0$. This implies that $\sqrt{nh}\{\widehat{\Lambda}(\widehat{\tau}) - \widehat{\Lambda}(\tau)\} \xrightarrow{p} 0$. Now,

$$\sqrt{nh}\{\widehat{\Lambda}(\widehat{\tau}) - \Lambda\} = \sqrt{nh}\{\widehat{\Lambda}(\widehat{\tau}) - \widehat{\Lambda}(\tau)\} + \sqrt{nh}\{\widehat{\Lambda}(\tau) - \Lambda\}. \tag{5.25}$$

Note that $\widehat{\Lambda}(\tau) - \Lambda = \{\widehat{m}_+(\tau) - m_+(\tau)\} - \{\widehat{m}_-(\tau) - m_-(\tau)\}$. Define

$$\widetilde{\Lambda}(\tau) = \frac{1}{nhf(\tau)} \sum_{j=1}^{n} \left[K\left(\frac{X_j - \tau}{h}\right) \{Y_j - m_+(\tau)\} - K\left(\frac{\tau - X_j}{h}\right) \{Y_j - m_-(\tau)\} \right].$$

By (5.3), $\sqrt{nh}\{\widehat{\Lambda}(\tau) - \Lambda - \widetilde{\Lambda}(\tau)\} = o_p(1)$. According to (5.6),

$$\begin{aligned} &\sqrt{nh}E[\widetilde{\Lambda}(\tau)] \\ &= \frac{\sqrt{nh}}{f(\tau)} \int K(u)\big[\{m(\tau + hu) - m_+(\tau)\} - \{m(\tau - hu) - m_-(\tau)\}\big] f(\tau + hu) du \\ &= \sqrt{nh}\, O(h) \end{aligned}$$

which is $o(1)$ by the assumption $nh^2 \to 0$ in (A5). Now, since the support of K is $[0,1]$,

$$nh\,\mathrm{Var}[\widetilde{\Lambda}(\tau)] = \frac{1}{hf^2(\tau)}\left[\mathrm{Var}\left\{K\left(\frac{X_1-\tau}{h}\right)\{Y_1-m_+(\tau)\}\right\} + \mathrm{Var}\left\{K\left(\frac{\tau-X_1}{h}\right)\{Y_1-m_-(\tau)\}\right\}\right].$$

By (5.15) and (5.20),

$$nh\,\mathrm{Var}[\widetilde{\Lambda}(\tau)] = \frac{1}{f(\tau)}\int\{K(u)\}^2\{v_+(\tau)+v_-(\tau)\}du(1+O(h))+O(h)$$

The Lyapounov's condition for $\sqrt{nh}\,\widetilde{\Lambda}(\tau)$ can be easily verified. These together with (5.25) imply Theorem 3.2. □

ACKNOWLEDGEMENTS

I would like to thank the referees for their helpful comments and suggestions that improved the presentation of the paper.

REFERENCES

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

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

FAN, J., HECKMAN, N. E. AND WAND, M. P. (1995). "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions", *Journal of the American Statistical Association*, **90**, 141–150.

GREEN, P. J. AND SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, London.

HALL, P. AND TITTERINGTON, D. M. (1992). "Edge-preserving and peak-preserving smoothing", *Technometrics*, **34**, 429–440.

HART, J. D. AND YI, S. (1998). "One-sided cross-validation", *Journal of the American Statistical Association*, **93**, 620–631.

HUH, J. AND CARRIÈRE, K. C. (2002). "Estimation of regression functions with a discontinuity in a derivative with local polynomial fits", *Statistics & Probability Letters*, **56**, 329–343.

KOO, J. -Y. (1997). "Spline estimation of discontinuous regression functions", *Journal of Computational and Graphical Statistics*, **6**, 266–284.

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

MCCULLAGH, P. AND NELDER, J. A. (1988). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

MCDONALD, J. A. AND OWEN, A. B. (1986). "Smoothing with split linear fits", *Technometrics*, **28**, 195–208.

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

QIU, P. AND YANDELL, B. (1998). "A local polynomial jump detection algorithm in nonparametric regression", *Technometrics*, **40**, 142–152.

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

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

WANG, Y. (1995). "Jump and sharp cusp detection by wavelets", *Biometrika*, **82**, 385–397.

WU, J. S. AND CHU, C. K. (1993). "Kernel-type estimators of jump points and values of a regression function", *The Annals of Statistics*, **21**, 1545–1566.

YIN, Y. Q. (1988). "Detection of the number, locations and magnitudes of jumps", *Communications in Statistics-Stochastic Models*, **4**, 445–455.