The Annals of Statistics
1993, Vol. 21, No. 4, 1792–1810

ON THE LEAST SQUARES CROSS-VALIDATION BANDWIDTH IN HAZARD RATE ESTIMATION

BY P. N. PATIL

University of Missouri

It is known that the least squares cross-validation bandwidth is asymptotically optimal in the case of kernel-based density and hazard rate estimation in the settings of both complete and randomly right-censored samples. From a practical point of view, it is important to know at what rate the cross-validation bandwidth converges to the optimal. In this paper we answer this question in a general setup which unifies all four possible cases.

1. Introduction. Since Watson and Leadbetter (1964a, 1964b) introduced the kernel-based estimators of the hazard rate in an uncensored setting, it has been a topic of interest to researchers and applied statisticians. A good deal of discussion on the topic can be found in Prakasa Rao (1983). Recent work includes Tanner and Wong (1983), who prove the asymptotic normality of the kernel-based estimator of the hazard rate for censored data using Hajék projection, while Lo, Mack and Wang (1989) achieve the same result via strong representation of the Kaplan–Meier estimator. Two of the three kernel-based estimators of the hazard rate considered by the preceding authors, in either an uncensored or a censored setting, are essentially variants of the same estimator and share similar optimal properties. However, they are not comparable in the sense of mean integrated squared error (MISE). Patil, Wells and Marron (1992) give a different viewpoint for choosing between them. Although the various properties proved by the preceding researchers made the kernel estimators of the hazard rate more appealing to use, practical utility of these estimators heavily depends on the choice of the smoothing parameter.

In the context of hazard rate estimation, least squares cross-validation is employed by Sarda and Vieu (1991) to choose the smoothing parameter, which is the bandwidth in the kernel estimator of the hazard rate, and it is shown that such a choice is asymptotically optimal. For generalization of it to the censored setting and for a detailed discussion of optimality of the cross-validation bandwidth, see Patil (1993). Wells (1989) has shown the asymptotic normality of the estimator which uses the bandwidth of optimal order. Also Marron and Padgett (1987) prove that the bandwidth obtained by employing least squares cross-validation is asymptotically optimal in the case of kernel-based density estimation from the randomly right-censored samples. From the previously mentioned work [Marron and Padgett (1987), Sarda and Vieu

Received August 1990; revised May 1992.
AMS 1991 *subject classifications*. G2G05, G2P10.
Key words and phrases. Nonparametric hazard rate estimation, kernel-based estimator, censored data, bandwidth, cross-validation.

(1991) and Patil (1993)], it is known that the least squares cross-validation bandwidth is optimal (cross-validation works) in the case of kernel-based density estimation and hazard rate estimation in either setting, that is, complete or randomly right-censored samples. From a practical point of view, it is important to know at what rate the cross-validation bandwidth converges to the optimal bandwidth. In this paper our main interest is answering this question. The work of Hall and Marron (1987) is generalized through a unified approach to treat all the cases discussed previously. The bandwidth that minimizes MISE depends on an unknown functional of the parent distribution. For this choice of bandwidth to be of any practical use, one needs to estimate and study the estimate of this functional of the parent distribution. This method of bandwidth selection, generally called the plug-in method, is yet to be investigated. One of the by-products of this paper is the comparison between this bandwidth and the cross-validation bandwidth.

In Section 2 we introduce the kernel estimators and necessary notation to unify the previously discussed cases by an appropriate mathematical structure. The main theorems are given in Section 3, and Section 4 contains the proofs of the lemmas which are used in the theorems of Section 3.

2. Notation and estimator. Let $X_1^0, X_2^0, \dots, X_n^0$ denote the independent identically distributed (i.i.d.) survival times of n items or individuals that are censored on the right by i.i.d. random variables $U_1, U_2, \dots, U_n$ which are independent of the X_i's. Denote the common distribution function of the X_i^0's by F^0 and that of the U_i's by H. It is assumed that F^0 is absolutely continuous with density f^0 and that H is continuous. The observed randomly right-censored data are denoted by the pairs (X_i, Δ_i), $i = 1, 2, \dots, n$, where

$$X_i = \min\{X_i^0, U_i\} \quad \text{and} \quad \Delta_i = I_{[X_i^0 \le U_i]}.$$

The X_i's form an i.i.d. sample from a distribution F, where $1 - F = (1 - F^0)(1 - H)$. Define the empirical distribution function

$$F_n(x) = (n+1)^{-1} \sum_{i=1}^{n} I_{[X_i \le x]}.$$

Let $F_n^0(x)$ be the maximum likelihood estimator of F^0 as defined by Kaplan and Meier (1958), and let H_n be such that $(1 - F_n(x)) = (1 - F_n^0(x))(1 - H_n(x))$. A general formulation of the target function in all of the cases discussed in Section 1 is $\eta(x)$, where

$$\eta(x) = \frac{(1 - H(x))\, f^0(x)}{Q(x)} \quad \text{for } Q(x) > 0 \tag{2.1}$$

and $Q(x)$ is a nonincreasing function such that $0 \le Q(x) \le 1$, $x \in \mathscr{R}$.

Remark 2.1. (i) If $Q(x) = 1 - F(x)$, then we have the case of hazard rate estimation in the censored setting. (ii) If $Q(x) = 1 - H(x)$, then we have the case of density estimation in the censored setting.

REMARK 2.2. If the censoring random variable has all its mass at ∞, then $H(x) = 0$ for $x \in \mathscr{R}$ and $F^0(x) \equiv F(x)$. (i) If $Q(x) = 1 - F(x)$, we have the case of hazard rate estimation in the uncensored setting. (ii) If we take $Q(x) \equiv 1$, we get the usual probability density and its kernel-based estimator.

Define the estimators $\eta_n(x)$ and $\eta_n^*(x)$ of $\eta(x)$,

$$\eta_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(x-X_i)}{Q_n(X_i)}\Delta_i, \tag{2.2a}$$

$$\eta_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(x-X_i)}{Q_n(x)}\Delta_i, \tag{2.3}$$

where Q_n is an estimator of Q such that Q_n converges to Q at the rate of $n^{-\alpha}$ in sup norm where $\alpha > 2/5$. Following the recommendations of Patil, Wells and Marron (1992), we will use the estimator $\eta_n(x)$. Also define

$$\bar{\eta}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(x-X_i)}{Q(X_i)}\Delta_i, \tag{2.2b}$$

and note that under conditions B.1 and B.2 of the next section

$$\begin{aligned}|\eta_n(x) - \bar{\eta}_n(x)| &\le \frac{1}{n}\sum_{i=1}^{n}|K_h(x-X_i)\Delta_i| \sup_{t\in[0,T]}|Q_n^{-1}(t) - Q^{-1}(t)| \\ &= O_p(n^{-\alpha}).\end{aligned}$$

To assess the performance of the estimator, we consider the error criterion integrated square error (ISE),

$$\Gamma(\eta_n(x)) = \int[\eta_n(x) - \eta(x)]^2 w(x)\,dx,$$

where $w(x)$ is a weight function and will be defined later. We also denote $\Gamma(\eta_n(x))$ by $\Gamma(h)$ whenever we want to emphasize this as a function of h. For the nonrandom assessment of the estimator, we will consider the MISE, $M(h) = E[\Gamma(h)]$.

The practical problem associated with applying the estimator $\eta_n(x)$ is the choice of h. A common goal is minimization of $M(h)$. The minimizer of $M(h)$, h_0 say, depends on an unknown functional of $f^0(x)$. If the value of this functional is known, then this choice is best in an average sense over all possible data sets. But, in general, the value of the functional is not known in advance and one has to estimate it using the sample. So thinking only of the data set at hand, we propose taking the "optimal choice" to be the minimizer of $\Gamma(h)$, $\hat{h}_0$ say, as $\Gamma(h_0) > \Gamma(\hat{h}_0)$ and $E[\Gamma(h_0)] \ge E[\Gamma(\hat{h}_0)]$. Again as $\hat{h}_0$ is not available to the experimenter, we consider $\hat{h}$, a data-driven bandwidth obtained by any rational methodology, and compare it with $\hat{h}_0$. That is to say, examine the relative distance between $\hat{h}$ and $\hat{h}_0$ and examine relatively how much greater $\Gamma(\hat{h})$ is than $\Gamma(\hat{h}_0)$.

Here we consider $\hat{h} = \hat{h}_c$, the cross-validation bandwidth which minimizes the cross-validation score function,

$$CV(h) = \int \eta_n^2(x) w(x)\,dx - 2n^{-1} \sum_{i=1}^{n} \left[\frac{\eta_{ni}(X_i)}{Q_n(X_i)} \right] w(X_i) I_{[\Delta_i = 1]},$$

where η_{ni} is "leave one out" version of η_n and $\hat{h} = h_0$, the best but unachievable classical bandwidth, and show that the relative distance between $\hat{h}_c$ and $\hat{h}_0$ is of the same order of magnitude as the relative distance between h_0 and $\hat{h}_0$. We also find that the relative distance between $\Gamma(\hat{h}_c)$ and $\Gamma(\hat{h}_0)$ is of the same order of magnitude as the relative distance between $\Gamma(h_0)$ and $\Gamma(\hat{h}_0)$. Furthermore, we find that neither h_0 nor $\hat{h}_c$ consistently outperforms the other, since both probabilities, $P[\Gamma(\hat{h}_c) > \Gamma(h_0)]$ and $P[\Gamma(\hat{h}_c) < \Gamma(h_0)]$, converge to strictly positive limits. All these results extend the findings of Hall and Marron (1987) to treat all four cases, density and hazard rate estimation in complete and censored samples, simultaneously.

3. Main results. We impose the following conditions on K, f^0 and η.

B.1. K is a compactly supported, symmetric function on $\mathscr{R}$, with Hölder continuous derivative K', and satisfies

$$\int K(u)\,du = 1 \quad \text{and} \quad \int u^2 K(u)\,du = 2k \neq 0.$$

B.2. f^0 and η are bounded and twice differentiable, $f^{0'}$, η', $f^{0''}$ and η'' are bounded and integrable, and $f^{0''}$ and η'' are uniformly continuous on $[0, T]$, where

$$T = \sup\{x | Q(x) > \varepsilon\}, \varepsilon > 0 \quad \text{and} \quad w(x) = I_{[0,T]}(x).$$

B.3. K has a second derivative on $\mathscr{R}$ and K'' is Hölder continuous.

Following is the first main result which quantifies the relative distance between $\hat{h}$ and the "optimal choice", $\hat{h}_0$. The variances and other constants appearing in the theorem are defined after stating the other main result of the paper.

THEOREM 3.1. *Under conditions* B.1 *and* B.2,

$$n^{1/10} \begin{bmatrix} \dfrac{\hat{h}_0 - h_0}{\hat{h}_0} \\ \dfrac{\hat{h}_c - \hat{h}_0}{\hat{h}_0} \end{bmatrix} \to_L N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, (C_0 C_1)^{-2} \begin{bmatrix} \sigma_0^2 & \sigma_{0c} \\ \sigma_{0c} & \sigma_c^2 \end{bmatrix} \right).$$

Next is the other main result describing the relative amount by which $\hat{h}$ (h_0 or $\hat{h}_c$) fails to minimize the integrated square error.

THEOREM 3.2. *Under conditions* B.1, B.2 *and* B.3,

$$\text{(i)} \qquad n^{1/5}\left\{\frac{\Gamma(h_0)-\Gamma(\hat{h}_0)}{\Gamma(\hat{h}_0)}\right\} \to_L \frac{1}{2}(C_1C_2)^{-1}\sigma_0^2\chi_1^2$$

and

$$\text{(ii)} \qquad n^{1/5}\left\{\frac{\Gamma(\hat{h}_c)-\Gamma(\hat{h}_0)}{\Gamma(\hat{h}_0)}\right\} \to_L \frac{1}{2}(C_1C_2)^{-1}\sigma_c^2\chi_1^2.$$

To define and understand the constants, set $D(h) = \Gamma(h) - M(h)$, and for simplification of the argument, decompose $\Gamma(h)$ as

$$\Gamma(h) = \int\{\eta_n(x) - \bar{\eta}_n(x) + \bar{\eta}_n(x) - \eta(x)\}^2 w(x)\,dx$$
$$= \Gamma_1(h) + \Gamma_2(h) + 2\Gamma_3(h),$$

where

$$\Gamma_1(h) = \int\{\bar{\eta}_n(x) - \eta(x)\}^2 w(x)\,dx, \qquad \Gamma_2(h) = \int\{\eta_n(x) - \bar{\eta}_n(x)\}^2 w(x)\,dx$$

and

$$\Gamma_3(h) = \int\{\bar{\eta}_n(x) - \eta(x)\}\{\eta_n(x) - \bar{\eta}_n(x)\}w(x)\,dx.$$

Denote $E[\Gamma_i(h)] = M_i(h)$, $i = 1, 2$ and 3, so that

$$M(h) = M_1(h) + M_2(h) + 2M_3(h) \quad \text{and} \quad D_1(h) = \Gamma_1(h) - M_1(h).$$

Now after some algebra the cross-validation criterion can be written in the form

$$CV(h) = \Gamma(h) + \delta(h) - \int\eta^2(x)w(x)\,dx,$$

where $\delta(h) = \delta_1(h) + \delta_2(h)$ and

$$\delta_1(h) = 2\int\eta(x)\bar{\eta}_n(x)w(x)\,dx - 2n^{-1}\sum_{i=1}^{n}\frac{\bar{\eta}_{ni}(X_i)}{Q(X_i)}w(X_i)I_{[\Delta_1=1]},$$

$$\delta_2(h) = 2\int\eta(x)[\eta_n(x) - \bar{\eta}_n(x)]w(x)\,dx$$
$$\qquad - 2n^{-1}\sum_{i=1}^{n}\left[\frac{\eta_{ni}(X_i)}{Q_n(X_i)} - \frac{\bar{\eta}_{ni}(X_i)}{Q(X_i)}\right]w(X_i)I_{[\Delta_i=1]}.$$

Also, recall that $\hat{h}_0$, $\hat{h}_c$ and h_0 minimize $\Gamma(h)$, CV and $M(h)$, respectively. Now the contribution from $M_2(h)$ and $M_3(h)$ in $M(h)$ is negligible as compared to $M_1(h)$. So $M(h) \sim M_1(h)$ and

$$M_1(h) = a(nh)^{-1} + bh^4 + o\{(nh)^{-1} + h^4\}$$

as $h \to 0$ and $nh \to \infty$, where

$$a = \left[\int K^2(u)\,du\right]\int \frac{\eta(x)}{Q(x)}w(x)\,dx \quad \text{and} \quad b = k^2\int (\eta''(x))^2 w(x)\,dx.$$

Now the following are expressions for the first and second derivatives of $M_1(h)$ obtained by differentiating under the integral sign,

$$M_1'(h) = -a(nh^2)^{-1} + 4bh^3 + o\left\{(nh^2)^{-1} + h^3\right\},$$

$$M_1''(h) = 2a(nh^3)^{-1} + 12bh^2 + o\left\{(nh^3)^{-1} + h^2\right\}.$$

Therefore, by setting $M_1'(h) = 0$, we get $h_0 \sim C_0 n^{-1/5}$, where $C_0 = (a/4b)^{1/5}$ and

$$\begin{aligned} M_1''(h_0) &\sim 2a\left[n^{2/5}\left(\frac{a}{4b}\right)^{3/5}\right]^{-1} + 12b\left[\left(\frac{a}{4b}\right)^{2/5} n^{-2/5}\right] \\ &= \left[2a\left(\frac{4b}{a}\right)^{3/5} + 12b\left(\frac{a}{4b}\right)^{2/5}\right]n^{-2/5} \\ &= \left[2aC_0^{-3} + 12bC_0^2\right]n^{-2/5} = C_1 n^{-2/5}. \end{aligned}$$

Also $M_1(h_0) \sim C_2 n^{-4/5}$, where $C_2 = a^{4/5}b^{1/5}[4^{1/5} + 4^{-4/5}]$. Set $L(u) = -uK'(u)$,

$$\begin{aligned} \sigma_0^2 = {} & \left(\frac{2}{C_0}\right)^3\left[\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx\right]\int\left[\int K(z+u)(K(z) - L(z))\,dz\right]^2 du \\ & +(4kC_0)^2\left\{\int(\eta''(x))^2\frac{\eta(x)}{Q(x)}w^2(x)\,dx - \left[\int \eta''(x)\eta(x)w(x)\,dx\right]^2\right\}, \end{aligned}$$

$$\begin{aligned} \sigma_c^2 = {} & \left(\frac{2}{C_0}\right)^3\int\left[\frac{\eta(x)}{Q(x)}w(x)\right]^2 dx \int L^2(u)\,du \\ & +(4kC_0)^2\left\{\int(\eta''(x))^2\frac{\eta(x)}{Q(x)}w^2(x)\,dx - \left[\int \eta''(x)\eta(x)w(x)\,dx\right]^2\right\}, \end{aligned}$$

$$\begin{aligned} \sigma_{0c} = -\Bigg\{ & \left(\frac{2}{C_0}\right)^3\left[\int\left(\frac{n(x)}{Q(x)}w(x)\right)^2 dx\right] \\ & \times\int[K(u) - L(u)][K*K(u) - L*K(u)]\,du \\ & +(4kC_0)^2\left[\int(\eta''(x))^2\frac{\eta(x)}{Q(x)}w^2(x)\,dx - \left(\int \eta''(x)\eta(x)w(x)\,dx\right)^2\right]\Bigg\}, \end{aligned}$$

where $*$ denotes convolution.

REMARK 3.1 (Generalizations).

(i) A faster rate of convergence of the estimator is possible if we admit more general kernels. Suppose K is chosen so that $\int K(u)\,du = 1$ and for some integer $v \geq 2$, $\int u^j K(u)\,du = 0$ and for $1 \leq j \leq v - 1$, $\int u^v K(u)\,du \neq 0$, then the kernel L also enjoys these properties. Furthermore, if we assume f^0 and η have uniformly continuous vth derivatives on $[0, T]$, then a version of Theorem 3.1 holds with rate $n^{1/[2(2v+1)]}$, whereas Theorem 3.2 remains unchanged.

(ii) In this paper Theorems 3.1 and 3.2 are considered for one dimension because of its practical importance. Following remark (2.1) of Hall and Marron (1987), both theorems can be generalized to the multidimensional setting of Gÿorfi, Härdle, Sarda and Vieu (1989) for hazard rate estimation, leaving Theorem 3.2 unchanged and replacing $n^{1/5}$ by $n^{p/[2(p+4)]}$ in Theorem 3.1.

REMARK 3.2 (Comparison of h_0 and $\hat{h}_c$ with reference to $\hat{h}_0$).

(i) By Theorem 3.1,

$$\operatorname{cov}\left(\frac{\hat{h}_0 - h_0}{\hat{h}_0}, \frac{\hat{h}_c - \hat{h}_0}{\hat{h}_0}\right)$$
$$= -(C_0C_1)^{-2}\Bigg\{\left(\frac{2}{C_0}\right)^3\left[\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx\right]$$
$$\times\int[K(u) - L(u)][K * K(u) - L * K(u)]\,du$$
$$+(4kC_0)^2\left[\int(\eta''(x))^2\frac{\eta(x)}{Q(x)}w^2(x)\,dx - \left(\int\eta''(x)\eta(x)w(x)\,dx\right)^2\right]\Bigg\}.$$

If the kernel K has nonnegative Fourier transform, then use of the Parseval inequality and the Cauchy–Schwarz inequality, respectively, on the first and second term of the preceding covariance gives us that asymptotically $(\hat{h}_0 - h_0)/\hat{h}_0$ and $(\hat{h}_c - \hat{h}_0)/\hat{h}_0$ are negatively correlated. This means that h_0 and $\hat{h}_c$ tend to err on the same side of $\hat{h}_0$.

(ii) The Cauchy–Schwarz inequality and the fact that $\int[K(u) - L(u)]^2\,du = \int L^2(u)\,du$ give

$$\int\left[\int K(z + u)(K(u) - L(u))\,du\right]^2 dz$$
$$\leq \int\left[\int K(z + u)\,du\right]\left[\int K(z + u)(K(u) - L(u))\,du\right]dz = \int L^2(u)\,du,$$

which implies that $\sigma_0^2 \leq \sigma_c^2$. In this sense h_0 results in a marginal improvement over the cross-validation bandwidth. However, the improvement is not available with probability 1 as one of the consequences of Theorem 3.1 is $\lim_{n\to\infty} P[\Gamma(h_0) > \Gamma(\hat{h}_0)]$ exists and is strictly positive.

PROOF OF THEOREM 3.1. In the following proof we have used the lemmas which are stated and proved in the next section. Observe that

$$0 = \Gamma'(\hat{h}_0) = M'(\hat{h}_0) + D'(\hat{h}_0) = M_1'(\hat{h}_0) + D_1'(\hat{h}_0) + R_1(\hat{h}_0),$$

where $R_1(\hat{h}_0) = [\Gamma_2'(\hat{h}_0) + 2\Gamma_3'(\hat{h}_0)]$. Therefore, we get

$$0 = \Gamma'(\hat{h}_0) = (\hat{h}_0 - h_0)M_1''(h^*) + D_1'(\hat{h}_0) + R_1(\hat{h}_0), \tag{3.1}$$

where h^* lies between h_0 and $\hat{h}_0$. By Lemma 4.4, $\hat{h}_0 = h_0 + O_p(n^{-1/5-\varepsilon})$ for some $\varepsilon > 0$, and so, by Lemma 4.2 with $h_1 = h_0$, $D_1'(\hat{h}_0) = D_1'(h_0) + o_p(n^{-7/10})$. But, by Lemma 4.5,

$$n^{7/10}D_1'(h_0) \to_L N(0, \sigma_0^2).$$

So $n^{7/10}D_1'(\hat{h}_0)$ must have the same weak limit. Since $h^*/h_0 \to_P 1$, $M_1''(h^*) = C_1 n^{-2/5} + o_p(n^{-2/5})$. Thus (3.1) becomes

$$0 = (\hat{h}_0 - h_0)C_1 n^{-2/5} + D_1'(h_0) + o_p(n^{-7/10}) + R_1(\hat{h}_0). \tag{3.2}$$

So by Lemmas 4.3 and 4.5, we conclude that

$$n^{3/10}(\hat{h}_0 - h_0) \to_L N(0, C_1^{-2}\sigma_0^2). \tag{3.3}$$

For the other component in the vector $[\hat{h}_0 - h_0, \hat{h}_0 - \hat{h}_c]'$, note that

$$\begin{aligned} CV(h) &= \Gamma(h) + \delta(h) - \int \eta^2(x)w(x)\,dx \\ &= M(h) + D(h) + \delta(h) - \int \eta^2(x)w(x)\,dx. \end{aligned}$$

Therefore,

$$0 = CV'(\hat{h}_c) = M_1'(\hat{h}_c) + D_1'(\hat{h}_c) + \delta_1'(\hat{h}_c) + R_1(\hat{h}_c) + R_2(\hat{h}_c),$$

where $R_2(\hat{h}_c) = \delta_2'(\hat{h}_c)$, and we get

$$\begin{aligned} 0 = CV'(\hat{h}_c) &= (\hat{h}_c - h_0)M_1''(h^{**}) + D_1'(\hat{h}_c) + \delta_1'(\hat{h}_c) \\ &\quad + R_1(\hat{h}_c) + R_2(\hat{h}_c), \end{aligned} \tag{3.4}$$

where h^{**} lies between h_0 and $\hat{h}_c$. Now Lemma 4.4 gives $\hat{h}_0 = h_0 + O_p(n^{-1/5-\varepsilon})$ for some $\varepsilon > 0$, and Lemma 4.2 gives, with $h_1 = h_0$,

$$D_1'(\hat{h}_c) + \delta_1'(\hat{h}_c) = D_1'(h_0) + \delta_1'(h_0) + o_p(n^{-7/10}).$$

Also, by Lemmas 4.5 and 4.6, $D_1'(h_0) + \delta_1'(h_0) = O_p(n^{-7/10})$. Furthermore, since $h^{**}/\hat{h}_0 \to_P 1$, $M_1(h^{**}) = C_1 n^{-2/5} + o_p(n^{-2/5})$. So (3.4) can be expressed as

$$0 = (\hat{h}_c - h_0)C_1 n^{-2/5}[1 + o(1)] + O_p(n^{-7/10}) + R_1(\hat{h}_c) + R_2(\hat{h}_c).$$

This implies that $\hat{h}_c - h_0 = O_p(n^{-3/10})$. Therefore,

$$\begin{aligned}(\hat{h}_c - h_0)M_1(h^*) &= (\hat{h}_c - h_0)\left[C_1 n^{-2/5} + o_p(n^{-2/5})\right]\\ &= (\hat{h}_c - h_0)C_1 n^{-2/5} + (\hat{h}_c - h_0)o_p(n^{-2/5})\\ &= (\hat{h}_c - h_0)C_1 n^{-2/5} + o_p(n^{-7/10}).\end{aligned}$$

Using the last representation, refinement of (3.4) gives

$$\begin{aligned}0 = {}&(\hat{h}_c - h_0)C_1 n^{-2/5} + D_1'(h_0) + \delta_1'(h_0) + R_1(\hat{h}_c)\\ &+ R_2(\hat{h}_c) + o_p(n^{-7/10}).\end{aligned}\tag{3.5}$$

Now, subtracting (3.2) from (3.5), we get

$$0 = (\hat{h}_c - \hat{h}_0)C_1 n^{-2/5} + \delta_1'(h_0) + R_2(\hat{h}_c) + o_p(n^{-7/10}).\tag{3.6}$$

Hence, by Lemmas 4.6 and 4.3, we get

$$n^{3/10}(\hat{h}_c - \hat{h}_0) \to_L N(0, C_1^{-2}\sigma_c^2).\tag{3.7}$$

Note that, by (3.2), (3.6) and Lemma 4.3, for any $p, q \in \mathscr{R}$,

$$\begin{aligned}p(\hat{h}_0 - h_0)C_1 n^{-2/5} &+ q(\hat{h}_c - \hat{h}_0)C_1 n^{-2/5} + pD_1'(h_0)\\ &+ q\delta_1'(h_0) + o_p(n^{-7/10}) = 0.\end{aligned}$$

Therefore, by Lemma 4.7, we get

$$n^{3/10}\left[p(\hat{h}_0 - h_0) + q(\hat{h}_c - \hat{h}_0)\right] \to_L N\left(0, C_1^{-2}(p^2\sigma_0^2 + q^2\sigma_c^2 + 2pq\sigma_{0c})\right).$$

Hence we conclude that

$$n^{3/10}\begin{bmatrix}\hat{h}_0 - h_0\\ \hat{h}_c - \hat{h}_0\end{bmatrix} \to_L N\left(\begin{bmatrix}0\\0\end{bmatrix}, C_1^{-2}\begin{bmatrix}\sigma_0^2 & \sigma_{0c}\\ \sigma_{0c} & \sigma_c^2\end{bmatrix}\right).$$

Now $\hat{h}_0 \to_P h_0 \sim C_0 n^{-1/5}$ implies that

$$n^{1/10}\begin{bmatrix}\dfrac{\hat{h}_0 - h_0}{\hat{h}_0}\\[2ex] \dfrac{\hat{h}_c - \hat{h}_0}{\hat{h}_0}\end{bmatrix} \to_L N\left(\begin{bmatrix}0\\0\end{bmatrix}, (C_0C_1)^{-2}\begin{bmatrix}\sigma_0^2 & \sigma_{0c}\\ \sigma_{0c} & \sigma_c^2\end{bmatrix}\right),$$

which completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Let $\hat{h}$ denote either h_0 or $\hat{h}_c$ and consider

$$\begin{aligned}\Gamma(\hat{h}) - \Gamma(\hat{h}_0) = {}&\Gamma_1(\hat{h}) - \Gamma_1(\hat{h}_0) + \Gamma_2(\hat{h}_0) - \Gamma_2(\hat{h}_0)\\ &+ 2\left[\Gamma_3(\hat{h}) - \Gamma_3(\hat{h}_0)\right].\end{aligned}$$

Notice that Taylor's expansion of $\Gamma_1(\hat{h})$ at $\hat{h}_0$ gives

$$\Gamma(\hat{h}) - \Gamma(\hat{h}_0) = \tfrac{1}{2}(\hat{h} - \hat{h}_0)^2 \Gamma_1''(h^*) + \Gamma_2(\hat{h}) - \Gamma_2(\hat{h}_0) + 2\left[\Gamma_3(\hat{h}) - \Gamma_3(\hat{h}_0)\right],$$

where h^* lies between $\hat{h}$ and $\hat{h}_0$. By Lemma 4.8, given in the next section, and by the fact that $h^*/h_0 \to_P 1$, $\Gamma_1''(h^*) = M_1''(h^*) + o_p(n^{-2/5})$. But $M_1''(h^*) = C_1 n^{-2/5} + o_p(n^{-2/5})$ and, by (3.3) and (3.6), $\hat{h} - \hat{h}_0 = O_p(n^{-3/10})$. Therefore,

$$\begin{aligned}\Gamma(\hat{h}) - \Gamma(\hat{h}_0) = \tfrac{1}{2}(\hat{h} - \hat{h}_0)^2 C_1 n^{-2/5} + o_p(n^{-1}) + \Gamma_2(\hat{h})\\ - \Gamma_2(\hat{h}_0) + 2\left[\Gamma_3(\hat{h}) - \Gamma_3(\hat{h}_0)\right].\end{aligned}$$

Hence, by (3.3), (3.7) and Lemma 4.9, given in the next section, we get

$$n\left[\Gamma(h_0) - \Gamma(\hat{h}_0)\right] \to_L \tfrac{1}{2} C_1^{-1} \sigma_0^2 \chi_1^2$$

and

$$n\left[\Gamma(\hat{h}_c) - \Gamma(\hat{h}_0)\right] \to_L \tfrac{1}{2} C_1^{-1} \sigma_c^2 \chi_1^2.$$

Furthermore, $\Gamma(\hat{h}_0) = C_2 n^{-4/5} + o_p(n^{-4/5})$, which implies

$$n^{1/5}\left\{\frac{\mathrm{ISE}(h_0) - \mathrm{ISE}(\hat{h}_0)}{\mathrm{ISE}(\hat{h}_0)}\right\} \to_L \frac{1}{2} C_1^{-1} C_2^{-1} \sigma_0^2 \chi_1^2$$

and

$$n^{1/5}\left\{\frac{\mathrm{ISE}(\hat{h}_c) - \mathrm{ISE}(\hat{h}_0)}{\mathrm{ISE}(\hat{h}_0)}\right\} \to_L \frac{1}{2} C_1^{-1} C_2^{-1} \sigma_c^2 \chi_1^2. \qquad \square$$

4. Lemmas. In this section we will state and prove the lemmas which we have used to prove the theorems of the last section. Define $S(h) = -(h/2)D_1'(h)$,

$$K_i(x) = \frac{K\left(\frac{x - X_i}{h}\right)}{Q(X_i)} I_{[\Delta_i = 1]} - E\left[\frac{K\left(\frac{x - X_i}{h}\right)}{Q(X_i)} I_{[\Delta_i = 1]}\right],$$

$$L_i(x) = \frac{L\left(\frac{x - X_i}{h}\right)}{Q(X_i)} I_{[\Delta_i = 1]} - E\left[\frac{L\left(\frac{x - X_i}{h}\right)}{Q(X_i)} I_{[\Delta_i = 1]}\right]$$

and

$$\bar{\gamma}_n\left(\frac{x}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{L_h(X - X_i)}{Q(X_i)} \Delta_i.$$

The following decomposition of $S(h)$ [or $D_1'(h)$] is along similar lines to that of

Hall and Marron (1987):

$$S(h) = -\frac{h}{2} D_1'(h) = S_1(h) + S_2(h) + S_3(h), \tag{4.1}$$

where

$$S_1(h) = S_{11}(h) - S_{12}(h), \qquad S_2(h) = S_{21}(h) + S_{22}(h),$$
$$S_3(h) = S_{31}(h) - S_{32}(h)$$

where

$$S_{11}(h) = (nh)^{-2} \sum\sum_{1 \le i < j \le n} \int K_i(x) K_j(x) w(x)\, dx,$$

$$S_{12}(h) = (nh)^{-2} \sum\sum_{1 \le i < j \le n} \int \left[K_i(x) L_j(x) + K_j(x) L_i(x) \right] w(x)\, dx,$$

$$S_{21}(h) = (nh)^{-1} \sum_{i=1}^{n} \int K_i(x) \{ 2E[\overline{\eta}_n(x|h)] - E[\overline{\gamma}_n(x|h)] - \eta(x) \} w(x)\, dx,$$

$$S_{22}(h) = (nh)^{-1} \sum_{i=1}^{n} \int L_i(x) \{ \eta(x) - E[\overline{\eta}_n(x|h)] \} w(x)\, dx,$$

$$S_{31}(h) = (nh)^{-2} \sum_{i=1}^{n} \int \left[K_i^2(x) - E(K_i^2(x)) \right] w(x)\, dx,$$

$$S_{32}(h) = (nh)^{-2} \sum_{i=1}^{n} \int \left[K_i(x) L_i(x) - E(K_i(x) L_i(x)) \right] w(x)\, dx.$$

A similar decomposition for $\delta_1'(h)$ is obtained in (4.2). Define

$$B_1(X_i, X_j) = \frac{K\left(\frac{X_i - X_j}{h}\right)}{Q(X_i) Q(X_j)} \left[w(X_i) + w(X_j) \right] I_{[\Delta_i = 1]} I_{[\Delta_j = 1]}$$
$$= B_{11}(X_i, X_j) + B_{12}(X_i, X_j),$$

$$B_2(X_i, X_j) = \frac{L\left(\frac{X_i - X_j}{h}\right)}{Q(X_i) Q(X_j)} \left[w(X_i) + w(X_j) \right] I_{[\Delta_i = 1]} I_{[\Delta_j = 1]}$$
$$= B_{21}(X_i, X_j) + B_{22}(X_i, X_j),$$

$$b_r(X_i) = E\left[B_r(X_i, X_j) | X_i \right] = b_{r1}(X_i) + b_{r2}(X_i) \quad \text{and} \quad \mu_r = E[b_r(X_i)],$$
$$r = 1, 2.$$

$$T(h) = \frac{h}{2} \delta_1'(h) = T_1(h) + T_2(h), \tag{4.2}$$

where

$$T_1(h) = T_{11}(h) - T_{12}(h), \qquad T_2(h) = T_{21}(h) - T_{22}(h),$$

and

$$T_{1r}(h) = [n(n-1)h]^{-1} \sum\sum_{1\le i<j\le n} [B_r(X_i, X_j) - b_r(X_i) - b_r(X_j) + \mu_r],$$

$$r = 1, 2,$$

$$T_{2r}(h) = (nh)^{-1} \sum_{i=1}^{n} \left[b_{r1}(X_i) - \mu_r - \eta(X_i)w(X_i) + \int \eta(x)w(x)\,dx \right],$$

$$r = 1, 2.$$

The symbols C, C_1 and C_2 occurring in the following lemmas denote generic positive constants. In Lemmas 4.1 through 4.7 and 4.9, we assume conditions B.1 and B.2.

LEMMA 4.1. *For each* $0 < a < b < \infty$ *and all positive integers* m,

$$\sup_{n;\, a\le i\le b} E\left|n^{7/10} D_1'(n^{1/5}t)\right|^{2m} \le C_1(a, b, m), \tag{4.3}$$

$$\sup_{n;\, a\le i\le b} E\left|n^{7/10} \delta_1'(n^{-1/5}t)\right|^{2m} \le C_1(a, b, m). \tag{4.4}$$

Furthermore, there exists $\varepsilon_1 > 0$ *such that*

$$E\left|n^{7/10}[D_1'(n^{-1/5}s) - D_1'(n^{-1/5}t)]\right|^{2m} \le C_2(a, b, m)|s-t|^{\varepsilon_1 m}, \tag{4.5}$$

$$E\left|n^{7/10}[\delta_1'(n^{-1/5}s) - \delta_1'(n^{-1/5}t)]\right|^{2m} \le C_2(a, b, m)|s-t|^{\varepsilon_1 m}, \tag{4.6}$$

whenever $a \le s \le t \le b$.

PROOF. By (4.1), to prove (4.5), we shall show that for some $\varepsilon > 0$,

$$E\left|n^{9/10}[S_{11}(n^{-1/5}s) - S_{11}(n^{-1/5}t)]\right|^{2m} \le C|s-t|^{\varepsilon m}, \tag{4.7}$$

$$E\left|n^{9/10}[S_{21}(n^{-1/5}s) - S_{21}(n^{-1/5}t)]\right|^{2m} \le C|s-t|^{\varepsilon m}, \tag{4.8}$$

$$E\left|n^{13/10}[S_{31}(n^{-1/5}s) - S_{31}(n^{-1/5}t)]\right|^{2m} \le C|s-t|^{\varepsilon m}. \tag{4.9}$$

Similar inequalities may be established for the functions S_{12}, S_{22} and S_{32}.

The verification of (4.7) through (4.9) could be done along similar lines to that of Hall and Marron (1987). Here we propose to use the inequality (21.5) of Burkholder (1973) for the martingale. We will illustrate this to verify (4.8) only.

To show (4.8), note that

$$\begin{aligned} &|2E[\bar{\eta}_n(x|h)] - E[\bar{\gamma}_n(x|h)] - \eta(x)| \\ &\quad = |2(E[\bar{\eta}_n(x|h)] - \eta(x)) - (E[\bar{\gamma}_n(x|h)] - \eta(x))|. \end{aligned}$$

Now B.1, B.2, Taylor's theorem and the fact that L is also symmetric and integrates to 1 imply that for $t \in (a, b)$,

$$\begin{aligned}&\left|2E\left[\bar{\eta}_n(x|n^{-1/5}t)\right] - E\left[\bar{\gamma}_n(x|n^{-1/5}t)\right] - \eta(x)\right|\\&\quad = \left|2\left(E\left[\bar{\eta}_n(x|n^{-1/5}t)\right] - \eta(x)\right) - \left(E\left[\bar{\gamma}_n(x|n^{-1/5}t)\right] - \eta(x)\right)\right| \le Cn^{-2/5}.\end{aligned}$$

Write $S_{21}(n^{-1/5}t) = n^{-1}\Sigma_{i=1}^n V_t(i)$, where

$$E\left[V_t(i)\right] = 0 \tag{4.10}$$

and

$$\left|V_s(i) - V_t(i)\right| \le Cn^{-2/5}|s - t|^{\varepsilon}. \tag{4.11}$$

Define the sequence of sigma fields

$$\begin{aligned}\mathscr{F}_k &= \sigma\{(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_k, \Delta_k)\} \quad \text{for } k < n\\&= \sigma\{(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n)\} \quad \text{for } k \ge n.\end{aligned}$$

Now note that, by (4.10),

$$E\left[\sum_{i=1}^{n} (V_s(i) - V_t(i))|\mathscr{F}_{n-1}\right] = \sum_{i=1}^{n-1} (V_s(i) - V_t(i)),$$

that is, $\{\Sigma_{i=1}^n (V_s(i) - V_t(i))\}$ is a martingale with respect to the sequence of sigma fields $\{\mathscr{F}_k,\ k \ge 1\}$. Now we will apply the martingale inequality (21.5) of Burkholder (1973) to the finitely indexed martingale $\{\Sigma_{i=1}^n (V_s(i) - V_t(i))\}$ with $\Phi(x) = x^{2m}$,

$$f^* = \sup_{k=1,\dots,n}\left|\sum_{i=1}^{k} \left[V_s(i) - V_t(i)\right]\right|,$$

$d_k = \left[V_s(k) - V_t(k)\right]$ for $k \le n$ and 0 otherwise, and

$$s(f) = \left\{\sum_{k=1}^{n} E\left[d_k^2|\mathscr{F}_{k-1}\right]\right\}^{1/2} = \left\{\sum_{k=1}^{n} E\left[(V_s(k) - V_t(k))^2\right]\right\}^{1/2}.$$

This gives

$$\begin{aligned}&E\left|n^{9/10}\left[S_{21}(n^{-1/5}s) - S_{21}(n^{-1/5}t)\right]\right|^{2m}\\&\quad = n^{-(1/5)m}\{E[\Phi(f^*)]\}\\&\quad \le n^{-(1/5)m}\left\{CE[\Phi(s(f))] + C\sum_{i=1}^{n} E[\Phi(|d_i|)]\right\}\\&\quad \le n^{-(1/5)m}\{C|s - t|^{2\varepsilon m}n^{(1/5)m} + C|s - t|^{2\varepsilon m}n^{1-(4/5)m}\\&\quad \le C|s - t|^{2\varepsilon m},\end{aligned}$$

which completes the proof of (4.8). The proofs of (4.7) and (4.9) are based on a similar argument and so we omit the details. This completes the proof of (4.5). The same type of argument gives (4.3), (4.4) and (4.6). □

LEMMA 4.2. *For some $\varepsilon > 0$ and any $0 < a < b < \infty$,*

$$\sup_{a \le t \le b} \{|D_1'(n^{-1/5}t)| + |\delta_1'(n^{-1/5}t)|\} = O_p(n^{-3/5-\varepsilon}). \tag{4.12}$$

Furthermore, for any $\varepsilon_2 > 0$ and any nonrandom h_1, asymptotic to a constant multiple of $n^{-1/5}$,

$$\begin{aligned} \sup_{|t - n^{1/5}h_1| \le n^{-\varepsilon_2}} n^{7/10}\{&|D_1'(n^{-1/5}t) - D_1'(h_1)| \\ &+ |\delta_1'(n^{-1/5}t) - \delta_1'(h_1)|\} \to_P 0. \end{aligned} \tag{4.13}$$

PROOF. First we will prove (4.13). Note that (4.13) will be true if

$$\sup_{|t - n^{1/5}h_1| \le n^{-\varepsilon_2}} n^{7/10}|D_1'(n^{-1/5}t) - D_1'(h_1)| \to_P 0 \tag{4.13a}$$

and

$$\sup_{|t - n^{1/5}h_1| \le n^{-\varepsilon_2}} n^{7/10}|\delta_1'(n^{-1/5}t) - \delta_1'(h_1)| \to_P 0. \tag{4.13b}$$

The proofs of (4.13a) and (4.13b) basically involve two steps. In the first step the supremum over the uncountable set will be reduced to the supremum over a countable set, and in the second step the Markov inequality is used to get the required result. Illustration of these steps is given in Hall and Marron (1987), and so we omit the details here. □

LEMMA 4.3. *For some $0 < \varepsilon < 3/20$ and any $0 < a < b < \infty$,*

$$\sup_{a \le t \le b} |R_1(n^{-1/5}t)| + |R_2(n^{-1/5}t)| = O_p(n^{-3/4+\varepsilon}).$$

PROOF. We will give an argument for R_1. The proof for R_2 is the same. First note that $R_1(h) = \Gamma_2'(h) + 2\Gamma_3'(h)$ and

$$\begin{aligned} \Gamma_2'(h) = [-(2/h)]\int&(\eta_n(x|h) - \bar{\eta}_n(x|h))^2 w(x)\,dx \\ &- \int(\eta_n(x|h) - \bar{\eta}_n(x|h))(\gamma_n(x|h) - \bar{\gamma}_n(x|h))w(x)\,dx, \end{aligned}$$

$$\begin{aligned} \Gamma_3'(h) = (2/h)\int&(\eta_n(x|h) - \bar{\eta}_n(x|h))(2\bar{\eta}_n(x|h) - \bar{\gamma}_n(x|h) - \eta(x))w(x)\,dx \\ &- \int(\bar{\eta}_n(x|h) - \eta(x))(\gamma_n(x|h) - \bar{\gamma}_n(x|h))w(x)\,dx. \end{aligned}$$

Now as

$$|Q_n(X) - Q(X)| = O_p(n^{-1/2}),$$

$$\int(\eta_n(x|h) - \bar{\eta}_n(x|h))^2 w(x)\,dx = O_p(n^{-1})$$

and by the Schwarz inequality

$$\int (\eta_n(x|h) - \bar{\eta}_n(x|h))(\gamma_n(x|h) - \bar{\gamma}_n(x|h))w(x)\,dx = O_p(n^{-1}).$$

Furthermore, by Theorem 2 of Hall (1984) and Theorem 4.2.1 of Patil (1990) and the Schwarz inequality,

$$\int (\eta_n(x|h) - \bar{\eta}_n(x|h))(2\bar{\eta}_n(x|h) - \bar{\gamma}_n(x|h) - \eta(x))w(x)\,dx = O_p(n^{-19/20}),$$

and similarly

$$\int (\bar{\eta}_n(x|h) - \eta(x))(\gamma_n(x|h) - \bar{\gamma}_n(x|h))w(x)\,dx = O_p(n^{-19/20}).$$

Therefore, $\Gamma_2'(h_0) = O_p(n^{-4/5})$ and $\Gamma_3'(h_0) = O_p(n^{-3/4})$; hence $R_1(h_0) = O_p(n^{-3/4})$. Now, to get the magnitude of $R_1(n^{-1/5}t)$ uniform in t, consider $\sup_{a \le t \le b}|R_1(n^{-1/5}t)|$. Then, reducing the supremum to a countable set as done in Lemma 4.2, we get the required result by an argument similar to that used in the previous lemma. □

LEMMA 4.4. *For some* $\varepsilon > 0$,

$$|\hat{h}_0 - h_0| + |\hat{h}_c - h_0| = O_p(n^{-\varepsilon-1/5}).$$

PROOF. First we treat $|\hat{h}_0 - h_0|$. Since $\hat{h}_0/h_0 \to_P 0$, by Lemmas 4.2 and 4.3,

$$\begin{aligned}
\Gamma'(h_0) &= \Gamma'(h_0) - \Gamma'(\hat{h}_0)\\
&= M_1'(h_0) - M_1'(\hat{h}_0) + D_1'(h_0) - D_1'(\hat{h}_0) + R_1(h_0) - R_1(\hat{h}_0)\\
&= M_1'(h_0) - M_1'(\hat{h}_0) + O_p(n^{-\varepsilon-3/5}).
\end{aligned}$$

Also, $\Gamma'(h_0) = D_1'(h_0) + R_1(h_0)$, and, by Lemma 4.2, $D_1'(h_0) = O_p(n^{-\varepsilon-3/5})$, and, by Lemma 4.3, $R_1(h_0) = O_p(n^{-3/4})$. So $\Gamma'(h_0) = O_p(n^{-\varepsilon-3/5})$. Therefore,

$$O_p(n^{-\varepsilon-3/5}) = M_1'(h_0) - M_1'(\hat{h}_0) + O_p(n^{-\varepsilon-3/5}),$$

that is,

$$O_p(n^{-\varepsilon-3/5}) = M_1'(h_0) - M_1'(\hat{h}_0) = (h_0 - \hat{h}_0)M_1''(h^*),$$

where h^* lies between h_0 and $\hat{h}_0$. As in Section 3, $M_1''(h^*) = C_1 n^{-2/5} + o_p(n^{-2/5})$. Using this estimate in the preceding equation, we conclude that $|\hat{h}_0 - h_0| = O_p(n^{-\varepsilon-1/5})$ as required. The proof of $|\hat{h}_c - h_0| = O_p(n^{-\varepsilon-1/5})$ is exactly along similar lines, and for the details refer to Patil (1990). □

LEMMA 4.5. $n^{7/10}D_1'(h_0) \to_L N(0, \sigma_0^2)$.

PROOF. We shall start from the decomposition (4.1) and prove that

$$n^{9/10}S(h_0) \to_L N(0, \sigma_0^2).$$

Now the argument leading to (4.9) gives $E[S_3^2(h_0)] = O(n^{-13/5})$ and so $S_3(h_0) = o(n^{-9/10})$. Therefore, it suffices to show that

$$(n^{9/10}S_1, n^{9/10}S_2) \to_L (Z_1, Z_2), \tag{4.14}$$

where $S_i = S_i(h_0)$, $i = 1, 2$, and Z_1 and Z_2 are independent normal variables with zero mean and variances adding up to $(C_0^2/4)\sigma_0^2$. The argument leading to (4.14) is a key step in the proof of Theorem 2 of Hall (1984) and Theorem 4.2.1 of Patil (1990), and a detailed treatment can be found there. Therefore, to complete the proof, we need only to show that

$$\begin{aligned} n^{9/5}\operatorname{Var}(S_1) \to 2C_0^{-1}&\left[\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx\right] \\ &\times\int\left[\int K(z+u)(K(z)-L(z))\,dz\right]^2 du, \end{aligned} \tag{4.15}$$

$$\begin{aligned} n^{9/5}\operatorname{Var}(S_2) \to C_0^4 4k^2&\left\{\int(n''(x))^2\frac{\eta(x)}{Q(x)}w^2(x)\,dx\right. \\ &\left.-\left[\int n''(x)\eta(x)w(x)\,dx\right]^2\right\}. \end{aligned} \tag{4.16}$$

Again, to compute the preceding variances, one can mimic the steps of Hall and Marron (1987). We will illustrate those steps only for the variance of S_1.

Let

$$\alpha_1(x,y) = E[K_i(x)K_i(y)], \qquad \alpha_2(x,y) = E[L_i(x)L_i(y)],$$
$$\alpha_3(x,y) = E[K_i(x)L_i(y)], \qquad \alpha_4(x,y) = \alpha_3(x,y)$$

and consider

$$\operatorname{Var}(S_1) = \operatorname{Var}(S_{11}) + \operatorname{Var}(S_{12}) - 2\operatorname{Cov}(S_{11}, S_{12}),$$

where $S_{ij} = S_{ij}(h_0)$. Now

$$\operatorname{Var}(S_{11}) = 2(nh_0)^{-4}n(n-1)\iint\alpha_2^2(x,y)w(x)w(y)\,dx\,dy,$$

$$\begin{aligned} \operatorname{Var}(S_{12}) = (nh_0)^{-4}&n(n-1) \\ &\times\iint[\alpha_1(x,y)\alpha_2(x,y)+\alpha_4(x,y)\alpha_3(x,y)]w(x)w(y)\,dx\,dy, \end{aligned}$$

and

$$\begin{aligned} \operatorname{Cov}(S_{11}, S_{12}) = (nh_0)^{-4}2\sum\sum_{1\le i<j\le n}\sum\sum_{1\le r<s\le n}&\operatorname{Cov}\left\{\int K_i(x)K_j(x)w(x)\,dx\right. \\ &\left.\times\int[K_r(x)L_s(x)+L_r(x)K_s(x)]w(x)\,dx\right\}. \end{aligned}$$

But as

$$\operatorname{Cov}\left\{\int K_i(x)K_j(x)w(x)\,dx, \int K_r(x)L_s(x)w(x)\,dx\right\} = 0 \quad \text{for } \begin{cases} r \neq i,\ s = j, \\ r \neq i,\ s \neq j, \\ r = i,\ s \neq j, \end{cases}$$

$$= \iint \alpha_1(x,y)\alpha_3(x,y)w(x)w(y)\,dx\,dy \qquad \text{for } r = i,\ s = j,$$

$$= \iint \alpha_1(x,y)\alpha_4(x,y)w(x)w(y)\,dx\,dy \qquad \text{for } r = j,\ s = i,$$

we get

$$\operatorname{Cov}(S_{11}, S_{12}) = (nh_0)^{-4} n(n-1) \iint [\alpha_1(x,y)\alpha_3(x,y) + \alpha_1(x,y)\alpha_4(x,y)]\, w(x)w(y)\,dx\,dy.$$

Therefore,

$$\operatorname{Var}(S_1) = (nh_0)^{-4} n(n-1) \iint (2\alpha_1^2 + \alpha_1\alpha_2 + \alpha_3\alpha_4 - 2\alpha_1\alpha_3 - 2\alpha_1\alpha_4)(x,y)\,w(x)w(y)\,dx\,dy.$$

Note that

$$\iint \alpha_1^2(x,y)w(x)w(y)\,dx\,dy \sim h_0^3 \int \left(\frac{\eta(x)}{Q(x)} w(x)\right)^2 dx \int \beta_1^2(u)\,du,$$

where $\beta_1(u) = \int K(z)K(u+z)\,dz$, and a similar computation gives

$$\iint \alpha_1\alpha_j(x,y)w(x)w(y)\,dx\,dy \sim h_0^3 \int \left(\frac{\eta(x)}{Q(x)} w(x)\right)^2 dx \int \beta_j(u)\,du, \qquad j = 1,2,3,$$

and

$$\iint \alpha_3\alpha_4(x,y)w(x)w(y)\,dx\,dy \sim h_0^3 \int \left(\frac{\eta(x)}{Q(x)} w(x)\right)^2 dx \int \beta_3(u)\beta_4(u)\,du,$$

where

$$\beta_2(u) = \int L(z)L(u+z)\,dz$$

and

$$\beta_3(u) = \int K(z)L(u+z)\,dz = \int L(z)K(u+z)\,dz = \beta_4(u).$$

Thus

$$\operatorname{Var}(S_1) \sim n^{-2}h_0^{-1}\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx$$
$$\times\int(2\beta_1^2+\beta_1\beta_2+\beta_3\beta_4-2\beta_1\beta_3-2\beta_1\beta_4)(u)\,du.$$

Now, using the fact that $\beta_3(u)=\beta_4(u)$ and $\int\beta_1\beta_2(u)\,du=\int\beta_3^2(u)\,du$,

$$\operatorname{Var}(S_1) \sim n^{-2}h_0^{-1}\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx\int 2(\beta_1-\beta_3)^2(u)\,du$$
$$= 2n^{-9/5}C_0^{-1}\int\left(\frac{\eta(x)}{Q(x)}w(x)\right)^2 dx$$
$$\times\int\left[\int K(z+u)(K(z)-L(z))\,dx\right]^2 du,$$

and hence the proof of (4.15) is complete. □

Lemma 4.6. $n^{7/10}\delta_1'(h_0)\to_L N(0,\sigma_c^2)$.

Proof. For the details refer to Patil (1990).

Lemma 4.7. *For any* $p,q\in\mathscr{R}$,

$$n^{7/10}(pD_1'(h_0)+q\delta_1'(h_0))\to_L N(0,\sigma^2),$$

where $\sigma^2=p^2\sigma_0^2+q^2\sigma_c^2+2pq\sigma_{0c}$.

Proof. Again we refer the reader to Patil (1990) for the proof.

Lemma 4.8. *Under conditions* B.1, B.2 *and* B.3 *and for any* $0<a<b<\infty$,

$$\sup_{a\le t\le b}\left|D_1''(n^{-1/5}t)\right|=o_p(n^{-2/5}).$$

Proof. Again the proof will follow by arguments similar to those of Lemma 4.2. But to use such an argument, we first have to prove the analog of (4.3),

$$\sup_{n;\,a\le t\le b}E\left[n^{1/2}\left|D_1''(n^{-1/5}t)\right|^{2l}\right]\le C(a,b,l). \tag{4.17}$$

Since the proof of (4.17) is almost identical to (4.3), we omit the details. □

Lemma 4.9. *For any* $\varepsilon>0$ *and any nonrandom* h_1, *asymptotic to a constant multiple of* $n^{-1/5}$,

$$\sup_{|t-n^{1/5}h_1|\le n^{-\varepsilon_2}} n\left\{\left|\Gamma_2(n^{-1/5}t)-\Gamma_2(h_1)\right|+\left|\Gamma_3(n^{-1/5}t)-\Gamma_3(h_1)\right|\right\}\to_P 0.$$

PROOF. Since the proof is based on steps identical to those of (4.13), we refer the reader to Patil (1990) for details. □

Acknowledgments. This paper is part of the author's Ph.D. dissertation. He would like to thank Professor J. S. Marron for suggesting the problem and thereafter for continuing guidance, encouragement and helpful discussions during the course of his Ph.D. thesis at the University of North Carolina, Chapel Hill.

REFERENCES

BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Statist.* **1** 19–42.

GYÖRFI, L., HÄRDLE, W., SARDA, P. and VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series*. Springer, Berlin.

HALL, P. (1984). Central limit theorem for integrated square of multivariate nonparametric density estimator. *J. Multivariate Anal.* **14** 1–16.

HALL, P. and MARRON, J. S. (1987). Extent to which least squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* **74** 567–581.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

LO, S. H., MACK, Y. P. and WANG, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probab. Theory Related Fields* **80** 461–473.

MARRON, J. S. and PADGETT, W. J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples. *Ann. Statist.* **15** 1520–1535.

PATIL, P. N. (1990). Automatic smoothing parameter selection in hazard rate estimation. Ph.D. dissertation, Univ. North Carolina, Chapel Hill.

PATIL, P. N. (1993). Bandwidth choice for nonparametric hazard rate estimation. *J. Statist. Plann. Inference*. **35** 15–30.

PATIL, P. N., WELLS, M. T. and MARRON, J. S. (1992). Some heuristics of kernel based estimators of ratio functions. *Journal of Nonparametric. Statistics*. To appear.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic, New York.

SARDA, P. and VIEU, P. (1991). Smoothing parameter selection in hazard estimation. *Statist. Probab. Lett.* **11** 429–434.

TANNER, M. A. and WONG, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11** 989–993.

WATSON, G. S. and LEADBETTER, M. R. (1964a). Hazard analysis. I. *Biometrika* **51** 175–184.

WATSON, G. S. and LEADBETTER, M. R. (1964b). Hazard analysis. II. *Sankhyā Ser. A* **26** 101–116.

WELLS, M. T. (1989). On the estimation of hazard rates and their extrema from general randomly censored data. Unpublished manuscript.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65211