Bernoulli **11**(4), 2005, 715–735

Empirical likelihood based inference for the derivative of the nonparametric regression function

GENGSHENG QIN[1] and MIN TSAO[2]

[1]*Department of Mathematics and Statistics, Georgia State University, University Plaza, Atlanta, Georgia 30303, USA. E-mail: gqins@gu.edu*
[2]*Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada V8N 3P4. E-mail: tsao@math.uvic.ca*

We study statistical inference for the derivative of the nonparametric regression function using local linear model based empirical likelihood. We first derive a normal equation for the derivative through the local linear model and use this equation to construct an empirical likelihood for the derivative. We show that the limiting distribution of the empirical likelihood ratio is a scaled χ_1^2 distribution rather than the usual (unscaled) χ_1^2 distribution. We use this limiting distribution to construct pointwise confidence intervals for the derivative. Such empirical likelihood ratio confidence intervals are easier to obtain than the normal approximation based confidence intervals. A small simulation study also suggests that they are more accurate.

Keywords: derivative function; empirical likelihood; local linear fitting; nonparametric regression function; normal approximation

1. Introduction

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be independently and identically distributed copies of a bivariate random variable (X, Y). The nonparametric regression function $m(x)$ is

$$m(x) = \mathrm{E}(Y|X = x).$$

Statistical inference for the regression function $m(x)$ is an important problem and there is a considerable amount of work in the literature on this topic. This work includes Mack and Silverman (1982), Müller (1988), Härdle (1990) and Wand and Jones (1995) on kernel methods; Stone (1977), Cleveland (1979), Tsybakov (1986) and Fan and Gijbels (1996) on local polynomial methods; Wahba (1977), Eubank (1988), Green and Silverman (1994) and Stone *et al.* (1997) on spline methods; and Donoho and Johnstone (1994), Ogden (1997), Efromovich (1999), Vidakovic (1999) and Luan and Xie (2001) on Fourier methods and wavelet methods.

The derivatives of the regression function are also of interest for many reasons. For example, in fitting growth curves, Müller (1988) has argued that the first derivative (speed) and the second derivative (spurt) of the height as functions of age are important quantities to be studied. Also, the optimal bandwidth for estimating the regression function depends on higher derivatives of $m(x)$ when using the 'plug-in' rules for the bandwidth selection. Estimation of the derivatives has been studied by a number of authors: Gasser and Müller (1984) and Georgiev (1984) considered estimating the derivatives using the usual kernel method; Ruppert and Wand (1994) estimated the derivatives based on locally weighted least-squares fitting; Welsh (1996) proposed a robust estimation method for the derivatives in a general heteroscedastic regression model; Efromovich and Samarov (2000) considered the estimation of the integral of squared regression derivatives; Zhou and Wolfe (2000) provided a spline estimation for the regression derivatives; and Lai and Chu (2001) proposed a new kernel estimator for the derivatives by using global smoothing parameters.

It is desirable to have a confidence interval to accompany a point estimator, and usually a confidence interval is easily obtained through the asymptotic normality of the point estimator. For the nonparametric regression function and its derivative, however, the construction of confidence intervals is difficult. The local linear estimators for $m(x)$ and $m'(x)$, for example, have many advantages and are easy to use (Fan and Gijbels 1996). Nevertheless, it is difficult to construct confidence intervals using their asymptotic normality, although in theory this is possible (Masry and Fan 1997). Indeed, their asymptotic normal distributions involve the marginal density function of X, the conditional variance of Y given x, and the derivatives of the regression function $m(x)$, all of which need to be estimated before normal approximations can be applied to construct the confidence intervals. The need to estimate these functions makes the computation quite involved. Recent work by Chen and Qin (2000) has provided a practical method for computing the confidence interval for $m(x)$ at any fixed x. Their approach combines the method of empirical likelihood with the local linear model to yield an empirical likelihood ratio confidence interval for $m(x)$. The computation of this interval does not require the estimation of the density function, conditional variance function and derivative functions. The method of empirical likelihood was first introduced by Owen (1988; 1990). It has many advantages over other nonparametric methods. The most important is that it studentizes internally, thereby eliminating the need for a pivot. This makes the empirical likelihood ratio confidence interval for $m(x)$ much easier to use than the normal approximation based interval. Chen and Qin (2000) also found that the coverage error of the empirical likelihood ratio confidence interval for $m(x)$ is of the same order throughout the support of $m(x)$. This is a significant improvement over the normal approximation based confidence interval which has a larger order of coverage error near the boundary.

The construction of the confidence interval for $m'(x)$ is also a challenging problem due to the difficulties mentioned above. In spite of the fact that there are now multiple methods for point estimation of $m'(x)$, there seems to be no easy method for constructing confidence intervals for $m'(x)$. Inspired by Chen and Qin (2000), this paper proposes an empirical likelihood ratio confidence interval for $m'(x)$ based on the local linear model. It is not obvious how one should formulate an empirical likelihood for $m'(x)$. Through local linear fitting, we first show that $m'(x)$ satisfies a normal equation whose components are readily available (see Section 2). This allows us to construct an empirical likelihood for $m'(x)$ by using these components which do not need to be estimated. The limiting distribution of the resulting empirical likelihood ratio for $m'(x)$ is shown to be a scaled χ^2_1

instead of the usual χ_1^2. The limiting distribution is then used to construct the empirical likelihood confidence interval for $m'(x)$. The resulting confidence interval for $m'(x)$ is practical and easy to compute, complementing the point estimator given by the local linear model.

The rest of this paper is organized as follows. In Section 2, we describe the method of local linear fitting and derive the normal equation for $m'(x)$. Then we construct the empirical likelihood for $m'(x)$ based on the normal equation and give the limiting distribution of the empirical likelihood ratio. In Section 3, we present some simulation results to compare the empirical likelihood method with the normal approximation based method. Proofs of the main theorems are given in Section 4.

2. Empirical likelihood for the derivative of the regression function via local linear fitting

2.1. Local linear regression

A regression function is commonly used to describe a general relationship between an explanatory variable X and a response variable Y. The basic idea of smoothing is to use a local average of the data near x to construct estimators for $m(x)$ at x. Because the regression function $m(x)$ satisfies

$$m(x) = \arg\min_a \mathrm{E}\big((Y_i - a)^2 | X = x\big),$$

it can be estimated by

$$\widehat{m}_K(x) = \arg\min_a \sum_{i=1}^n (Y_i - a)^2 K\left(\frac{X_i - x}{h}\right),$$

where $K(\cdot)$ denotes a non-negative weight function and h the smoothing parameter which determines the size of the neighbourhood of x. This leads to the *kernel estimator* of $m(x)$:

$$\widehat{m}_K(x) = \frac{\sum_{i=1}^n Y_i K((X_i - x)/h)}{\sum_{i=1}^n K((X_i - x)/h)}.$$

If $K(\cdot)$ is sufficiently smooth, the derivative of $\widehat{m}_K(x)$ may be used as an estimator for $m'(x)$.

Although the kernel estimators are known to have nice properties, such as simplicity, flexibility, consistency and asymptotic normality, they have some disadvantages – for example, boundary problems (Fan 1993; Fan and Gijbels 1996) and the dependence of the asymptotic bias on the derivative of the marginal density. To remedy these problems, Fan (1993) proposed a local linear fit to the regression function. Assuming that $m'(x)$ exists and noting that $m(X) \approx m(x) + m'(x)(X - x) \equiv a + b(X - x)$ in a small neighbourhood of a point x, the local linear estimators for $m(x)$ and $m'(x)$ at a fixed x are defined as the solutions to the following problem:

$$\min_{a,b} \sum_{i=1}^{n} (Y_i - a - b(X_i - x))^2 K\left(\frac{X_i - x}{h}\right).$$

It follows that the local linear estimators $\widehat{a} = \widehat{m}(x)$ and $\widehat{b} = \widehat{m}'(x)$ for $m(x)$ and $m'(x)$ satisfy the normal equations

$$(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{y}, \tag{1}$$

where $\boldsymbol{\beta} = (a, b)^{\mathrm{T}}$ and

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x) \\ \vdots & \vdots \\ 1 & (X_n - x) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{W} = \mathrm{diag}(K_h(X_i - x)), \ K_h(t) = K(t/h)/h.$$

Upon solving (1) for a and b, we obtain the following equivalent normal equations:

$$\frac{1}{n}\sum_{i=1}^{n}\left(s_{n2} - \frac{X_i - x}{h} s_{n1}\right) K_h(X_i - x)(Y_i - a) = 0, \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(s_{n1} - \frac{X_i - x}{h} s_{n0}\right) K_h(X_i - x)(Y_i - b(X_i - x)) = 0, \tag{3}$$

where

$$s_{nj} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^j K_h(X_i - x), \qquad j = 0, 1, \cdots$$

Noting that equation (3) for $b = m'(x)$ does not involve unknown quantities except for $m'(x)$ itself, it can be used to construct an empirical likelihood for $m'(x)$.

2.2. Local linear model based empirical likelihood for $m'(x)$

To construct an empirical likelihood, let $\mathbf{p}_b = (p_{b1}, \ldots, p_{bn})$ be a probability vector, that is, $\sum_{i=1}^{n} p_{bi} = 1$ and $p_{bi} \geqslant 0$ for all i. For $1 \leqslant i \leqslant n$, we define

$$\widetilde{W}_{bi} = \left(s_{n1} - \frac{X_i - x}{h} s_{n0}\right) K_h(X_i - x), \tag{4}$$

$$W_{bi} = \widetilde{W}_{bi}(Y_i - b(X_i - x)). \tag{5}$$

Based on normal equation (3), we define the empirical likelihood for the derivative of the regression function evaluated at the true value $b = m'(x)$ as

$$L(b) = \sup\left\{\prod_{i=1}^{n} p_{bi} : \sum p_{bi} = 1, \sum_{i=1}^{n} p_{bi} W_{bi} = 0\right\}.$$

By the Lagrange multiplier, we have

$$p_{bi} = \frac{1}{n}\{1 + \lambda_b W_{ai}\}^{-1}, \qquad i = 1, \ldots, n,$$

where λ_b is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{W_{bi}}{1 + \lambda_b W_{bi}} = 0. \tag{6}$$

Noting that $\prod_{i=1}^{n} p_{bi}$, subject to $\sum_{i=1}^{n} p_{bi} = 1$, attains its maximum n^{-n} at $p_{bi} = n^{-1}$, we define the empirical likelihood ratio at b as

$$R(b) = \prod_{i=1}^{n}(np_{bi}) = \prod_{i=1}^{n}\{1 + \lambda_b W_{bi}\}^{-1},$$

and define the corresponding empirical log-likelihood ratio as

$$l(b) = 2\sum_{i=1}^{n} \log\{1 + \lambda_b W_{bi}\}, \tag{7}$$

where λ_b is the solution of (6).

Let $f(x)$ be the marginal density function of X, and

$$\sigma^2(x) = \mathrm{var}(Y|X = x), \qquad \mu_j = \int_{-\infty}^{\infty} u^j K(u)\mathrm{d}u, \qquad \nu_j = \int_{-\infty}^{\infty} u^j K^2(u)\mathrm{d}u.$$

In order to derive the asymptotic distribution of $l(b)$, we need the following conditions:

Condition 1. *The kernel function $K(\cdot)$ is bounded with a bounded support* $[-1, 1]$.

Condition 2. *$f(\cdot)$ and $\sigma^2(\cdot)$ are continuous at x, and $m(\cdot)$ has a continues second derivative at x.*

Theorem 1. *Assume that* $\mathrm{E}|Y|^s < \infty$ *for some* $s > 2$ *and that Conditions* 1 *and* 2 *hold. If* $n \to \infty$, $h \to 0$, *and* $nh^5 \to 0$, *then the limiting distribution of* $l(m'(x))$ *is a scaled chi-square distribution with one degree of freedom; that is,*

$$r \cdot l(m'(x)) \xrightarrow{\mathcal{L}} \chi_1^2, \tag{8}$$

where the scaling constant r is

$$r = 1 + \frac{m^2(x)}{\sigma^2(x)}.$$

We remark that the main condition, $nh^5 \to 0$, is also necessary for (8) when

$$C_0(x) \equiv m^{(2)}(x) f^2(x)(\mu_1\mu_2 - \mu_0\mu_3) \neq 0.$$

This can be readily seen from the proof of Theorem 1. Further, in deriving the empirical likelihood confidence region for the regression function $m(x)$, Chen and Qin (2000) defined the empirical log-likelihood ratio at the true value $a = m(x)$ as

$$l(a) = 2\sum_{i=1}^{n} \log\{1 + \lambda_a W_{ai}\},$$

where λ_a is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{W_{ai}}{1 + \lambda_a W_{ai}} = 0,$$

with

$$W_{ai} = \widetilde{W}_{ai}(Y_i - a), \qquad \widetilde{W}_{ai} = \left(s_{n2} - \frac{X_i - x}{h} s_{n1}\right) K_h(X_i - x).$$

We would arrive at the same formulation for $l(a)$ if we were to define the empirical likelihood for $m(x)$ based on the normal equation (2) for a. Under the conditions that $\mathrm{E}|Y|^s < \infty$ for a real number $s \geqslant 4$ and $h = o\left(n^{-1/5}\right)$, they obtained the limiting distribution of $l(a)$, which is a chi-square distribution with one degree of freedom.

The empirical likelihood defined above is for $\mathrm{E}(\widehat{m}'(x)) = m'(x) + \text{bias}$, rather than for $m'(x)$. To convert to the empirical likelihood for $m'(x)$, one can either correct the bias explicitly by direct estimation, or reduce the bias by undersmoothing. Hall (1992) showed that better coverage accuracy is achieved by the undersmoothing method (see also Chen and Qin 2000). We use the latter method in this paper. Instead of the usual chi-square distribution, the limiting distribution of $l(b)$ for $m'(x)$ is a scaled chi-square distribution. The scaling constant is the consequence of the difference between $\tilde{\sigma}_b^2(x)$ and $\sigma_b^2(x)$ (see Lemmas 1 and 2 in Section 3). In order to apply Theorem 1, we need to find a consistent estimate of the scaling constant r. In principle, plugging any consistent estimators for $m(x)$ and $\sigma^2(x)$ into the expression for r will produce a consistent estimator. Here, we use

$$\widetilde{r} = 1 + \frac{\widehat{m}^2(x)}{\hat{\sigma}^2(x)},$$

where $\widehat{m}(x)$ is the local linear estimator of $m(x)$, and $\hat{\sigma}^2(x)$ is the local linear estimator $\hat{\alpha}$ of $\sigma^2(x)$ defined by

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} (\widehat{r}_i - \alpha - \beta(X_i - x))^2 W\left(\frac{X_i - x}{h_b}\right),$$

where $\widehat{r}_i = (Y_i - \widehat{m}(X_i))^2$, $W(\cdot)$ is a known weight function, and h_b is a bandwidth. $\widehat{m}(x)$ and $\hat{\sigma}^2(x)$ are efficient estimators for $m(x)$ and $\sigma^2(x)$, respectively. See Fan (1993), Fan and Yao (1998) and Ruppert *et al.* (1997). Thus $\widetilde{r}$ is a suitable estimator for r.

We now examine the convergence property of $\widetilde{r}$. We need the following conditions.

Condition 3. *$f(x)\cdot\sigma^2(x) > 0$ for a given x, and the function $\mathrm{E}(Y^k|X = x)$ is continuous at x for $k = 3, 4$. Further, $m^{(2)}(z)$ and $\sigma^{(2)}(z) \equiv \mathrm{d}^2\{\sigma^2(z)\}/\mathrm{d}z^2$ are uniformly continuous on an open set containing the point x.*

Condition 4. *$\mathrm{E}\left(Y^{4(1+\delta)}\right) < \infty$, where $\delta \in [0, 1)$ is a constant.*

Condition 5. *$K(\cdot)$ is a symmetric density function; $W(\cdot)$ is a symmetric density function with a bounded support on the real line; $K(\cdot)$, $W(\cdot)$ and $f(\cdot)$ are Lipschitz of order 1.*

Conditions 3–5 are the conditions Fan and Yao (1998) used to derive the asymptotic distribution of $\hat{\sigma}^2(x)$.

Theorem 2. *Assume Conditions* 1–5 *hold. If we select $h = o\left(n^{-1/5}\right)$ and $h_b = O\left(n^{-1/5}\right)$, then*

$$\widetilde{r} - r = O_p\left(n^{-2/5}\right).$$

To compute $\widetilde{r}$, we need to first estimate the variance function $\sigma^2(x)$. A referee pointed out that care must be taken to ensure that the estimate is positive. In practice, we can use the algorithm proposed by Fan and Yao (1998) to select the bandwidth h_b for the variance function estimation, which usually leads to reliable, positive estimates. Further, to guarantee the positivity, we can take $\hat{\sigma}^2(x) = \max(\hat{\alpha}, n^{-1})$. It is easily seen from the proof of Theorem 2 in Section 4 that this adjustment does not change the conclusion of Theorem 2.

Finally, the empirical likelihood ratio confidence interval for $m'(x)$ at any fixed x can be constructed as follows:

$$R_\alpha(b) = \{b : \widetilde{r}\cdot l(b) \leqslant c_\alpha\}, \tag{9}$$

where c_α is the $100(1-\alpha)$th quantile of χ^2_1. By Theorems 1 and 2, $R_\alpha(b)$ is an approximate confidence interval for $m'(x) = b$ with an asymptotically correct coverage probability $1 - \alpha$. That is,

$$P(m'(x) \in R_\alpha(b)) = 1 - \alpha + o(1).$$

3. Simulation study

In this section, we conduct two simulation studies to investigate the finite-sample performance of the empirical likelihood (EL) intervals for $m'(x)$. For comparison, the following normal approximation (NA) based intervals for $m'(x)$ (Masry and Fan 1997) is also included in the study:

$$\left(\widehat{m}'(x) + \frac{1}{2}\widehat{m}^{(2)}(x)B_1 h\right) \pm z_{1-\alpha/2}\left(\frac{V_1\hat{\sigma}^2(x)}{nh^3\widehat{f}(x)}\right)^{1/2},$$

where $\widehat{m}'(x)$, $\widehat{m}^{(2)}(x)$, $\hat{\sigma}^2(x)$ and $\widehat{f}(x)$ are consistent estimates for $m'(x)$, $m^{(2)}(x)$, $\sigma^2(x)$ and $f(x)$, respectively. For example, $m'(x)$ and $m^{(2)}(x)$ can be estimated by the local polynomial regression method (Fan and Gijbels 1996), and $f(x)$ can be estimated by the usual kernel density estimation method. $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution, and

$$B_1 = \frac{\mu_0\mu_3 - \mu_1\mu_2}{\mu_0\mu_2 - \mu_1^2}, \qquad V_1 = \frac{\mu_0(\mu_0\nu_2 - \mu_1\nu_1) - \mu_1(\mu_0\nu_1 - \mu_1\nu_0)}{(\mu_0\mu_2 - \mu_1^2)^2}.$$

The regression model considered in the simulation is given by

$$Y_i = \sin(X_i) + \sigma(X_i) \cdot \varepsilon_i, \; i = 1, 2, \ldots, n.$$

where the X_i are drawn from the uniform distribution $U[0, 1]$, the conditional standard deviation function $\sigma(x)$ is chosen to be $0.5\exp(x)$, and the errors ε_i are generated from a standard normal distribution. The biweight kernel

$$K(x) = \frac{15}{16}\left(1 - x^2\right)^2, \qquad |x| \leqslant 1,$$

is chosen as the kernel function. Two bandwidths, $h = 5(n\log n)^{-1/5}$ and $h = 3n^{-1/3}$, are selected as the reference bandwidths for comparing the coverage probabilities of the EL intervals and the NA intervals for $m'(x)$; in order to compare the coverage properties over a wide range of bandwidth, 40 equally spaced bandwidth values centred around each reference bandwidth were used to calculate the two intervals.

In the first simulation study, the sample size n ranges from 50 to 200 and x is fixed at 0.5. Figures 1 and 2 display the graphs for the coverage probability and interval length (width) of the EL and NA intervals for $m'(x)$. The two curves in, say, the upper left-hand plot in Figure 1 were generated as follows. For each simulated random sample of size $n = 50$, 95% EL and NA intervals were computed using all 40 values of the bandwidth, resulting in 40 confidence intervals of each type. This was repeated 1000 times using 1000 random samples, which led to 1000 95% confidence intervals of each type at each of the 40 values of the bandwidth. Then, at each value of the bandwidth, the coverage probability of the EL interval was approximated using the proportion of the 1000 EL intervals constructed with this bandwidth that contain $m'(x)$. This proportion, when plotted against the value of the bandwidth, gave rise to the EL curve in the plot. The NA curve was generated similarly. The average length of the 1000 confidence intervals was also computed and plotted against the bandwidth.

On the relative coverage accuracy, we note that for $h = 5(n\log n)^{-1/5}$ (Figure 1), the coverage probabilities of the EL intervals are quite close to the nominal level 0.95, while those for the NA intervals are substantially lower, even for samples of size 200. The EL intervals performed much better. The poor accuracy of the NA intervals may have partly been the consequence of having to estimate $m'(x)$, $m^{(2)}(x)$, $\sigma^2(x)$ and $f(x)$. For $h = 3n^{-1/3}$ (Figure 2), the coverage probabilities of NA intervals are substantially lower than the nominal level for samples of size 50 but they become more accurate for larger samples. Nevertheless, the EL intervals still outperform the NA intervals.

The relative (average) length of the two types of confidence intervals is also of interest. For $h = 5(n\log n)^{-1/5}$ (Figure 1), the EL intervals are only slightly longer than the NA intervals but their coverage probabilities are substantially better. For $h = 3n^{-1/3}$ (Figure 2),

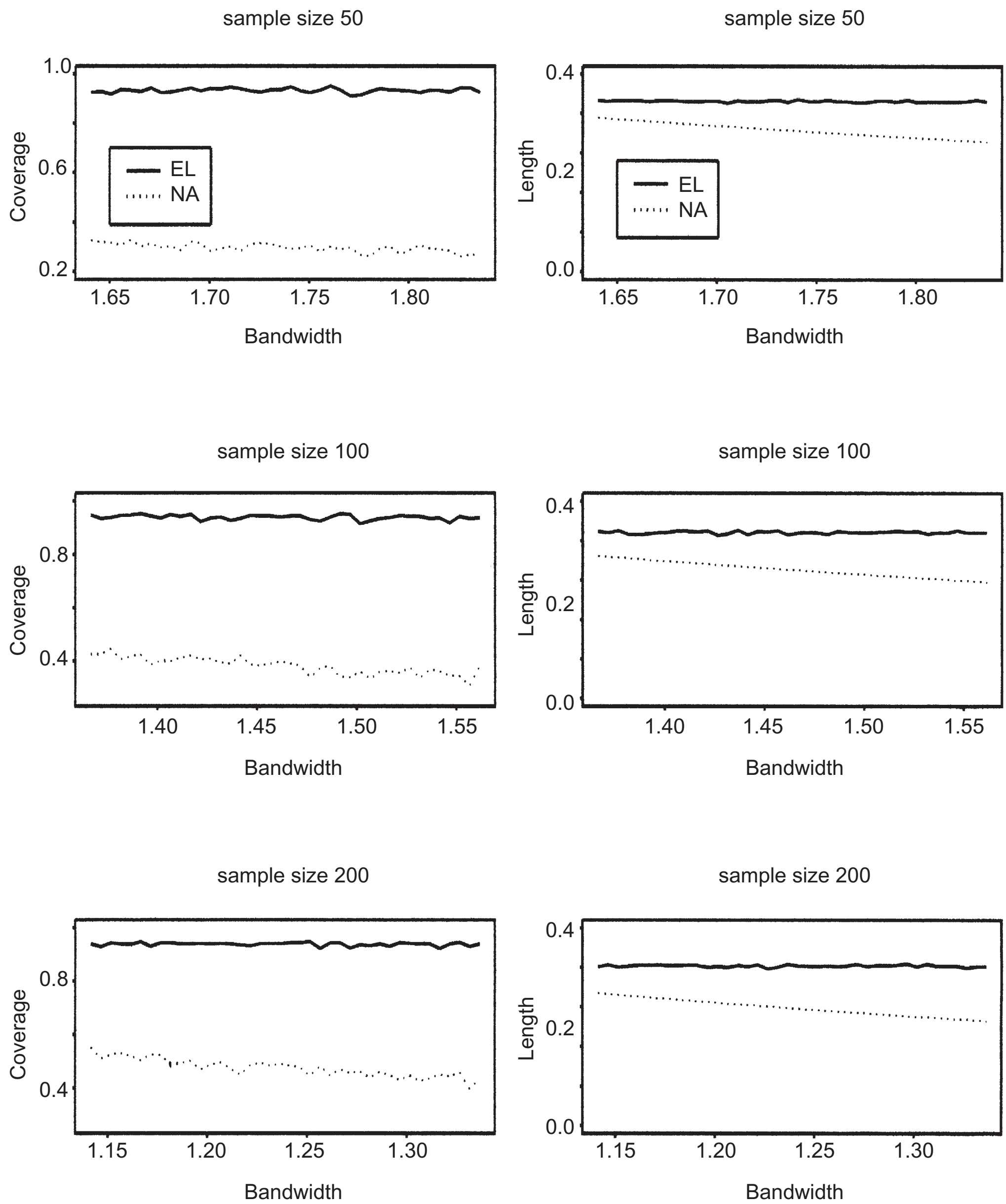

Figure 1. Coverage probability and interval length of 95% confidence intervals for $m'(x)$ at $x = 0.5$; $h = 5(n \log n)^{-1/5}$.

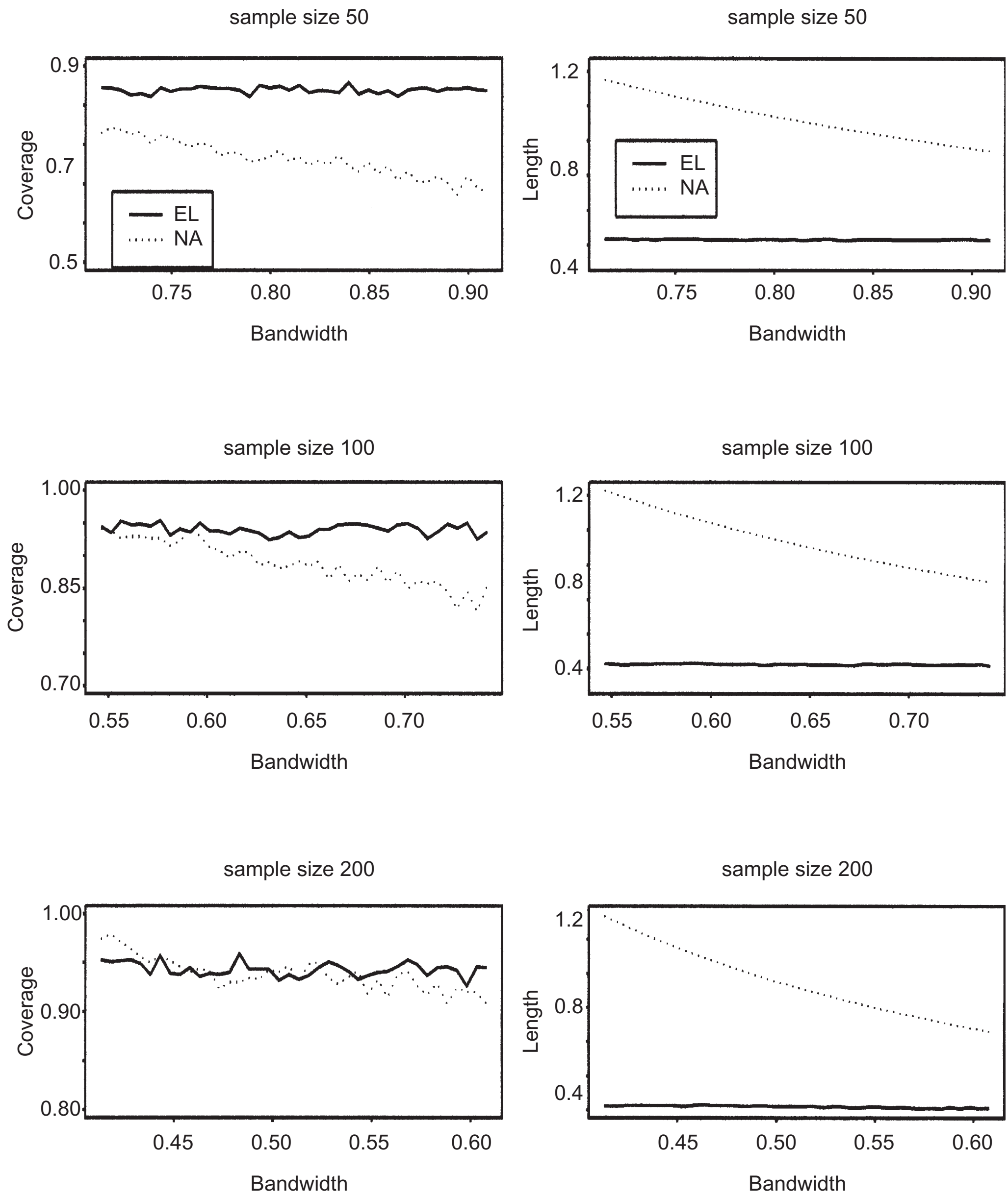

Figure 2. Coverage probability and interval length of 95% confidence intervals for $m'(x)$ at $x = 0.5$; $h = 3n^{-1/3}$.

the NA intervals are much longer but their coverage probabilities are only comparable to or lower than that of the EL intervals. This suggests that the EL intervals are superior even when both the length and coverage probability are considered. Figure 3 further confirms this observation.

To examine the boundary effects, in the second simulation study, x is fixed at the boundary point 1 and the sample size is set to 100. Figure 3 displays the graphs for the simulated coverage probability and average interval length of the EL and NA intervals for $m'(x)$. For this example, the EL intervals do not seem to have a boundary effect; their coverage probabilities are slightly lower at the boundary than at the middle of the range, but the difference is quite small. The average length of the intervals is also about the same at both locations. The NA intervals, on the other hand, show a clear boundary effect; their interval length curves at the boundary are quite different from those for the middle of the range. Other cases involving different boundary point and sample size combinations have also been examined. They also revealed little boundary effect for the EL intervals and a clear boundary effect for the NA intervals.

Finally, bandwidth selection for confidence intervals for $m'(x)$ is also an interesting problem, and currently work is still continuing to identify the optimal bandwidth. Nevertheless, it is worth noting that while the performance of the NA intervals depends critically on the bandwidth, the EL intervals are less sensitive to the bandwidth selection.

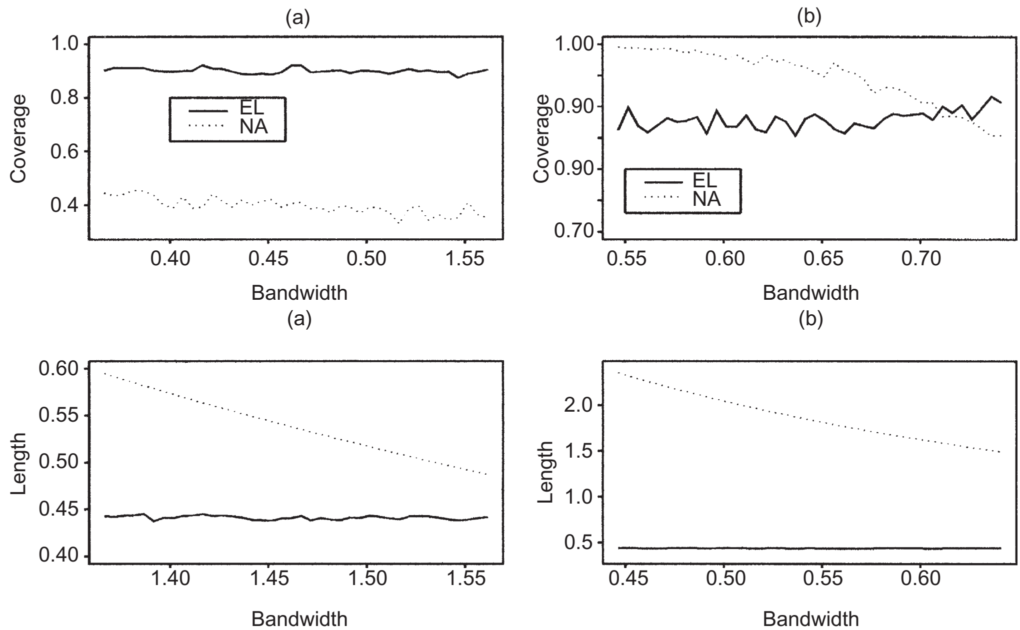

Figure 3. Coverage probability and interval length of 95% confidence intervals for $m'(x)$ at $x = 1$; Sample size $n = 100$. (a) $h = 5(n \log n)^{-1/5}$. (b) $h = 3n^{-1/3}$.

This is a particularly important advantage in the absence of a method for optimal bandwidth selection for confidence intervals.

4. Proofs of theorems

In this section, we give the proofs of Theorems 1 and 2. We need the following lemmas.

Lemma 1. *Assume that Conditions* 1 *and* 2 *hold. If* $h = O_p(n^{-1/5})$, *then*

$$(nh)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} W_{bi} - \Delta\right) \xrightarrow{\mathcal{L}} N(0, \sigma_b^2(x)),$$

where

$$\Delta = \frac{h^2}{2} C_0(x)$$

and

$$\sigma_b^2(x) = \sigma^2(x) f^3(x)\big(\mu_1^2\nu_0 - 2\mu_0\mu_1\nu_1 + \mu_0^2\nu_2\big).$$

Proof. Let

$$A_{nj} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^j K_h(X_i - x)(Y_i - m(X_i)), \qquad j = 0, 1.$$

We have the following decompositions:

$$\frac{1}{n}\sum_{i=1}^{n} W_{bi} = s_{n1}\cdot A_{n0} - s_{n0}\cdot A_{n1} + s_{n1}\cdot\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\big(m(X_i) - a - b(X_i - x)\big)$$
$$- s_{n0}\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{X_i - x}{h} K_h(X_i - x)\big(m(X_i) - a - b(X_i - x)\big).$$

Since the regression is conducted in the neighbourhood of $|X_i - x| \leqslant h$, by Taylor's expansion,

$$m(X_i) = a + b(X_i - x) + \frac{1}{2} m^{(2)}(x)(X_i - x)^2 + o_p(h^2). \tag{10}$$

Hence we have

$$\frac{1}{n}\sum_{i=1}^{n} W_{bi} = (s_{n1}, -s_{n0})\cdot(A_{n0}, A_{n1})^{\mathrm{T}} + \frac{h^2}{2} m^{(2)}(x)\big(s_{n1}s_{n2} - s_{n0}s_{n3}\big) + o_p(h^2), \tag{11}$$

where $(A_{n0}, A_{n1})^{\mathrm{T}}$ stands for the transpose of (A_{n0}, A_{n1}). By the central limit theorem, any linear combination of A_{n0} and A_{n1} is asymptotically normal distributed. In particular,

$$\sqrt{nh}(A_{n0},\ A_{n1})^{\mathrm{T}} \xrightarrow{\mathcal{L}} N(0,\ \sigma^2(x)f(x)\mathbf{S}) \tag{12}$$

where

$$\mathbf{S} = \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}.$$

Noting that $s_{nj} \xrightarrow{p} f(x)\mu_j$ for $j = 0,\ 1,\ 2$, then Lemma 1 follows from Slutsky's theorem, (11) and (12). □

Lemma 2. *Under Conditions* 1 *and* 2, *we have*

$$\frac{h}{n}\sum_{i=1}^{n} W_{bi}^2 \xrightarrow{p} \tilde{\sigma}_b^2(x),$$

where $\tilde{\sigma}_b^2(x) = (\sigma^2(x) + m^2(x))f^3(x)\big(\mu_1^2\nu_0 - 2\mu_0\mu_1\nu_1 + \mu_0^2\nu_2\big)$.

Proof. The following decomposition is straightforward:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} W_{bi}^2 &= \frac{1}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2(Y_i - m(X_i))^2 + \frac{1}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2(m(X_i) - a - b(X_i - x))^2 \\
&\quad + \frac{a^2}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2 + \frac{2a}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2(m(X_i) - a - b(X_i - x)) \\
&\quad + \frac{2}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2(Y_i - m(X_i))(m(X_i) - b(X_i - x)) \\
&\equiv J_1 + J_2 + a^2J_3 + 2aJ_4 + 2J_5.
\end{aligned} \tag{13}$$

J_1 can be further decomposed into

$$\begin{aligned}
J_1 &= \frac{1}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2(Y_i - m(X_i))^2 \\
&= s_{n1}^2 \cdot \frac{1}{n}\sum_{i=1}^{n} K_h^2(X_i - x)(Y_i - m(X_i))^2 \\
&\quad + s_{n0}^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \left(\frac{X_i - x}{h}\right)^2 K_h^2(X_i - x)(Y_i - m(X_i))^2 \\
&\quad - 2s_{n0}s_{n1} \cdot \frac{1}{n}\sum_{i=1}^{n} \left(\frac{X_i - x}{h}\right) K_h^2(X_i - x)(Y_i - m(X_i))^2 \\
&\equiv J_{11} + J_{12} - 2J_{13}.
\end{aligned}$$

Using $s_{nj} \xrightarrow{p} f(x)\mu_j$ ($j = 0, 1, 2$) once again and noticing that $K_h(\cdot) = K(\cdot/h)/h$, we obtain

$$\begin{aligned}
hJ_{11} &= \mu_1^2 f^2(x) \cdot \mathrm{E}\left[hK_h^2(X_1 - x)(Y_1 - m(X_1))^2\right] + o_p(1) \\
&= \mu_1^2 f^2(x) \cdot \mathrm{E}\left[hK_h^2(X_1 - x)\sigma^2(X_1)\right] + o_p(1) \\
&= \mu_1^2 \sigma^2(x) f^3(x) \int_{-\infty}^{\infty} K^2(u)\mathrm{d}u + o_p(1) \\
&= \sigma^2(x) f^3(x) \mu_1^2 \nu_0 + o_p(1), \\
hJ_{12} &= \mu_0^2 f^2(x) \cdot \mathrm{E}\left[h\left(\frac{X_1 - x}{h}\right)^2 K_h^2(X_1 - x)(Y_1 - m(X_1))^2\right] + o_p(1) \\
&= \mu_0^2 f^2(x) \cdot \mathrm{E}\left[h\left(\frac{X_1 - x}{h}\right)^2 K_h^2(X_1 - x)\sigma^2(X_1)\right] + o_p(1) \\
&= \mu_0^2 \sigma^2(x) f^3(x) \int_{-\infty}^{\infty} u^2 K^2(u)\mathrm{d}u + o_p(1) \\
&= \sigma^2(x) f^3(x) \mu_0^2 \nu_2 + o_p(1).
\end{aligned}$$

Similarly,

$$hJ_{13} = \sigma^2(x) f^3(x) \mu_0 \mu_1 \nu_1 + o_p(1).$$

Hence,

$$hJ_1 = \sigma^2(x) f^3(x)\left(\mu_1^2 \nu_0 - 2\mu_0 \mu_1 \nu_1 + \mu_0^2 \nu_2\right) + o_p(1). \tag{14}$$

For J_2, we have

$$\begin{aligned}
J_2 &= \frac{1}{n}\sum_{i=1}^n \left(s_{n1} - \frac{X_i - x}{h} s_{n0}\right)^2 K_h^2(X_i - x)\left(m(X_i) - a - b(X_i - x)\right)^2 \\
&= s_{n1}^2 \cdot \frac{1}{n}\sum_{i=1}^n K_h^2(X_i - x)\left(m(X_i) - a - b(X_i - x)\right)^2 \\
&\quad + s_{n0}^2 \cdot \frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - x}{h}\right)^2 K_h^2(X_i - x)\left(m(X_i) - a - b(X_i - x)\right)^2 \\
&\quad - 2s_{n0}s_{n1} \cdot \frac{1}{n}\sum_{i=1}^n \left(\frac{X_i - x}{h}\right) K_h^2(X_i - x)\left(m(X_i) - a - b(X_i - x)\right)^2.
\end{aligned}$$

Using (10), we obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)^{j} K_h^2(X_i-x)(m(X_i)-a-b(X_i-x))^2 \\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)^{j} K_h^2(X_i-x)\left(\frac{1}{2}m^{(2)}(x)(X_i-x)^2+o_p(h^2)\right)^2 \\
&\leqslant \frac{(m^{(2)}(x))^2h^4}{2n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)^{j+4} K_h^2(X_i-x) \\
&\quad + o_p(h^4)\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)^{j} K_h^2(X_i-x) \\
&=\frac{(m^{(2)}(x))^2h^3}{2}\cdot\left\{\mathrm{E}\left[h\left(\frac{X_1-x}{h}\right)^{j+4} K_h^2(X_1-x)\right]+o_p(1)\right\} \\
&\quad + o_p(h^3)\left\{\mathrm{E}\left[h\left(\frac{X_1-x}{h}\right)^{j} K_h^2(X_1-x)\right]+o_p(1)\right\} \\
&=\frac{h^3}{2}\nu_{j+4}(m^{(2)}(x))^2f(x)+o_p(h^3) \\
&=O_p(h^3), \qquad j=0,\,1,\,2.
\end{aligned}$$

By $s_{nj}\xrightarrow{p} f(x)\mu_j$ $(j=0,\,1,\,2)$, we have

$$J_2=O_p(h^3). \tag{15}$$

For J_3, we have

$$\begin{aligned}
J_3 &= s_{n1}^2\cdot\frac{1}{n}\sum_{i=1}^{n}K_h^2(X_i-x)+s_{n0}^2\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)^2K_h^2(X_i-x) \\
&\quad -2s_{n0}s_{n1}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-x}{h}\right)K_h^2(X_i-x).
\end{aligned}$$

Using $s_{nj}\xrightarrow{p} f(x)\mu_j$ $(j=0,\,1,\,2)$, we have

$$\begin{aligned}
hJ_3 &= f^2(x)\mu_1^2\mathrm{E}\big(hK_h^2(X_1-x)\big)+f^2(x)\mu_0^2\mathrm{E}\left(h\left(\frac{X_1-x}{h}\right)^2K_h^2(X_1-x)\right) \\
&\quad -2f^2(x)\mu_1\mu_0\mathrm{E}\left(h\left(\frac{X_1-x}{h}\right)K_h^2(X_1-x)\right)+o_p(1) \\
&=f^3(x)\big[\mu_1^2\nu_0-2\mu_1\mu_0\nu_1+\mu_0^2\nu_2\big]+o_p(1).
\end{aligned} \tag{16}$$

By (10) and (16),

$$J_4 = \frac{1}{n}\sum_{i=1}^{n} \widetilde{W}_{bi}^2 \left(\frac{1}{2} m^{(2)}(x)(X_i - x)^2 + o_p(h^2)\right)$$

$$\leqslant \frac{1}{2}|m^{(2)}(x)|h^2 \cdot J_3 + o_p(h^2) \cdot J_3$$

$$= O_p(h). \tag{17}$$

As for J_5, we have

$$J_5 = \frac{1}{n}\sum_{i=1}^{n}\left(s_{n1} - \frac{X_i - x}{h} s_{n0}\right)^2 K_h^2(X_i - x)(Y_i - m(X_i))(m(X_i) - b(X_i - x))$$

$$= s_{n1}^2 \cdot \frac{1}{n}\sum_{i=1}^{n} K_h^2(X_i - x)(Y_i - m(X_i))(m(X_i) - b(X_i - x))$$

$$+ s_{n0}^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^2 K_h^2(X_i - x)(Y_i - m(X_i))(m(X_i) - b(X_i - x))$$

$$- 2s_{n0}s_{n1} \cdot \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right) K_h^2(X_i - x)(Y_i - m(X_i))(m(X_i) - b(X_i - x)).$$

Noting that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^j K_h^2(X_i - x)(Y_i - m(X_i))(m(X_i) - b(X_i - x))$$

$$= \mathrm{E}\left[\left(\frac{X_1 - x}{h}\right)^j K_h^2(X_1 - x)(Y_1 - m(X_1))(m(X_1) - b(X_1 - x))\right] + o_p(1)$$

$$= \mathrm{E}\left[\left(\frac{X_1 - x}{h}\right)^j K_h^2(X_1 - x)(m(X_1) - b(X_1 - x))(\mathrm{E}(Y_1|X_1) - m(X_1))\right] + o_p(1)$$

$$= o_p(1), \qquad j = 0, 1, 2,$$

we obtain that

$$J_5 = o_p(1). \tag{18}$$

Finally, it follows from (13)–(18) that

$$\frac{h}{n}\sum_{i=1}^{n} W_{bi}^2 = hJ_1 + hJ_2 + a^2 hJ_3 + 2ahJ_4 + 2hJ_5$$
$$= \left(\sigma^2(x) + m^2(x)\right) f^3(x)\left[\mu_1^2\nu_0 - 2\mu_1\mu_0\nu_1 + \mu_0^2\nu_2\right] + o_p(h)$$
$$= \tilde{\sigma}_b^2(x) + o_p(1).$$

The proof of Lemma 2 is thus complete. □

Proof of Theorem 1. From (6),

$$0 = \frac{1}{n}\sum_{i=1}^{n} W_{bi} - \frac{\lambda_b}{n}\sum_{i=1}^{n}\frac{W_{bi}^2}{1+\lambda_b W_{bi}}.$$

By Lemma 1,

$$\frac{|\lambda_b|}{1+|\lambda_b|\max_i|W_{bi}|}\cdot\frac{h}{n}\sum_{i=1}^{n} W_{bi}^2 \leqslant \left|\frac{h}{n}\sum_{i=1}^{n} W_{bi}\right| = O_p\left((n^{-1}h)^{1/2} + h^3\right).$$

Note that the condition $\mathrm{E}|Y|^s < \infty$ $(s > 2)$ implies that

$$\max_i|W_{bi}| = \max_i|\widetilde{W}_{bi}(Y_i - b(X_i - x))| = o_p\left(n^{1/s}\right). \tag{19}$$

Then from Lemma 2 and a similar argument used in Owen (1991), it follows that

$$|\lambda_b| = O_p\left((n^{-1}h)^{1/2} + h^3\right) = O_p\left(n^{-3/5}\right). \tag{20}$$

Noticing that

$$\sum_{i=1}^{n}\frac{W_{bi}}{1+\lambda_b W_{bi}} = \sum_{i=1}^{n} W_{bi}\left[1 - \lambda_b W_{bi} + \frac{(\lambda_b W_{bi})^2}{1+\lambda_b W_{bi}}\right]$$
$$= \sum_{i=1}^{n} W_{bi} - \lambda_b\sum_{i=1}^{n} W_{bi}^2 + \sum_{i=1}^{n}\frac{W_{bi}(\lambda_b W_{bi})^2}{1+\lambda_b W_{bi}},$$

by (6), (19) and (20), we have that

$$\lambda_b = \left(\sum_{i=1}^{n} W_{bi}^2\right)^{-1} \sum_{i=1}^{n} W_{bi} + O_p(\lambda_b^2 \max_i |W_{bi}|)$$

$$= \left(\sum_{i=1}^{n} W_{bi}^2\right)^{-1} \sum_{i=1}^{n} W_{bi} + o_p(n^{-7/10})$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} W_{bi}^2\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} W_{bi} - \Delta\right) + \frac{\Delta}{n^{-1}\sum_{i=1}^{n} W_{bi}^2} + o_p(n^{-7/10}). \quad (21)$$

Again by (6), we obtain that

$$0 = \sum_{i=1}^{n} \frac{\lambda_b W_{bi}}{1+\lambda_b W_{bi}} = \sum_{i=1}^{n} (\lambda_b W_{bi}) - \sum_{i=1}^{n} (\lambda_b W_{bi})^2 + \sum_{i=1}^{n} \frac{(\lambda_b W_{bi})^3}{1+\lambda_b W_{bi}}. \quad (22)$$

By (19), (20) and Lemma 2,

$$\sum_{i=1}^{n} \frac{(\lambda_b W_{bi})^3}{1+\lambda_b W_{bi}} = o_p(1). \quad (23)$$

It follows from (22) and (23) that

$$\sum_{i=1}^{n} \lambda_b W_{bi} = \sum_{i=1}^{n} (\lambda_b W_{bi})^2 + o_p(1).$$

Then applying Taylor's expansion to (7), we have

$$l(b) = 2\sum_{i=1}^{n} \log\{1+\lambda_b W_{bi}\} = 2\sum_{i=1}^{n} \left(\lambda_b W_{bi} - \frac{1}{2}(\lambda_b W_{bi})^2\right) + r_n$$

$$= \lambda_b^2 \sum_{i=1}^{n} W_{bi}^2 + r_n,$$

with

$$|r_n| \leqslant C|\lambda_b|^3 \max_i |W_{bi}| \sum_{i=1}^{n} W_{bi}^2 = o_p\left(n^{-9/5+1/s} \cdot (nh^{-1})\right) = o_p(1).$$

Therefore, by (21) and Lemmas 1 and 2, we have

$$
\begin{aligned}
r \cdot l(b) &= r \cdot \lambda_b^2 \sum_{i=1}^{n} W_{bi}^2 + o_p(1) \\
&= \left(\frac{(nh)^{1/2} \left(n^{-1} \sum_{i=1}^{n} W_{bi} - \Delta \right)}{\sigma_b(x)} \right)^2 \cdot \frac{r \cdot \sigma_b^2(x)}{n^{-1} h \sum_{i=1}^{n} W_{bi}^2} + \frac{nh^5 C_0^2(x)}{4\tilde{\sigma}_b^2(x)} \\
&\quad + \frac{(nh^5)^{1/2} \sigma_b(x) C_0(x)}{\tilde{\sigma}_b^2(x)} \cdot \left(\frac{(nh)^{1/2} \left(n^{-1} \sum_{i=1}^{n} W_{bi} - \Delta \right)}{\sigma_b(x)} \right) + o_p(1) \\
&= \chi_1^2 + \frac{nh^5 C_0^2(x)}{4\tilde{\sigma}_b^2(x)} + \frac{(nh^5)^{1/2} \sigma_b(x) C_0(x)}{\tilde{\sigma}_b^2(x)} \cdot N + o_p(1)
\end{aligned}
$$

where χ_1^2 is a chi-square random variable with one degree of freedom and N is a standard normal random variable. Clearly, if $h = o(n^{-1/5})$, then $r \cdot l(b) \xrightarrow{\mathcal{L}} \chi_1^2$. When $C_0(x) \neq 0$, $r \cdot l(b) \xrightarrow{\mathcal{L}} \chi_1^2$ if and only if $h = o(n^{-1/5})$. The proof is thus complete. □

Proof of Theorem 2. Let

$$
d = \frac{h^2}{2} m^{(2)}(x) B_0, \qquad \sigma_W^2 = \int t^2 W(t) \mathrm{d}t,
$$

$$
\theta_n = \frac{h_b^2}{2} \sigma_W^2 \left(\sigma^{(2)}(x) \right)^2 + o\left(h^2 + h_b^2 \right).
$$

We have

$$
\begin{aligned}
\widetilde{r} - r &= \frac{\sigma^2(x) \widehat{m}^2(x) - \hat{\sigma}^2(x) m^2(x)}{\sigma^2(x) \hat{\sigma}^2(x)} \\
&= \frac{\sigma^2(x) \widehat{m}^2(x) - \hat{\sigma}^2(x) m^2(x)}{\sigma^4(x)} \left(1 + \frac{\sigma^2(x) - \hat{\sigma}^2(x)}{\hat{\sigma}^2(x)} \right).
\end{aligned}
$$

By the consistency of $\hat{\sigma}^2(x)$,

$$
\frac{\sigma^2(x) - \hat{\sigma}^2(x)}{\hat{\sigma}^2(x)} = o_p(1),
$$

so we only need to show that $\sigma^2(x) \widehat{m}^2(x) - \hat{\sigma}^2(x) m^2(x) = O_p\left(n^{-2/5} \right)$. Note that

$$\begin{aligned}
&\sigma^2(x)\widehat{m}^2(x) - \hat{\sigma}^2(x)m^2(x) \\
&\quad = \sigma^2(x)(\widehat{m}(x) - m(x) - d)^2 + \sigma^2(x)d(d + 2(\widehat{m}(x) - m(x) - d)) \\
&\quad\quad + 2\sigma^2(x)m(x)(\widehat{m}(x) - m(x) - d) + 2\sigma^2(x)m(x)d \\
&\quad\quad - m^2(x)\big(\hat{\sigma}^2(x) - \sigma^2(x) - \theta_n\big) - m^2(x)\theta_n. \qquad (24)
\end{aligned}$$

Under Conditions 1–5, by Theorem 1 in Fan and Yao (1998),

$$\hat{\sigma}^2(x) - \sigma^2(x) - \theta_n = O_p\Big((nh_b)^{-1/2}\Big) = O_p\Big(n^{-2/5}\Big).$$

By the asymptotic normality of $\widehat{m}(x)$,

$$\widehat{m}(x) - m(x) - d = O_p\Big((nh)^{-1/2}\Big) = o_p\Big(n^{-2/5}\Big)$$

(Tsybakov 1986; Masry and Fan 1997). Also, noting that $\theta_n = O_p\big(n^{-2/5}\big)$ and $d = o_p\big(n^{-2/5}\big)$, it follows from (24) that

$$\hat{\sigma}^2(x)m^2(x) - \sigma^2(x)\widehat{m}^2(x) = O_p\Big(n^{-2/5}\Big).$$

The proof is thus complete. □

Acknowledgement

The authors are very grateful to an anonymous referee and the Editor, Professor Peter McCullagh, for many very helpful comments which have led to a much improved paper. The work of the second author was partially supported by the National Science and Engineering Research Council of Canada.

References

Chen, S.X. and Qin, Y.S. (2000) Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, **87**, 946–953.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatter plots. *J. Amer. Statist. Assoc.*, **74**, 829–836.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Efromovich, S. (1999) *Nonparametric Curve Estimation: Methods, Theory and Applications*. New York: Springer-Verlag.

Efromovich, S. and Samarov, A. (2000) Adaptive estimation of the integral of squared regression derivatives. *Scand. J. Statist.*, **27**, 335–351.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

Fan, J. (1993) Local linear regression smoothers and their minimax efficiency, *Ann. Statist.*, **21**, 196–216.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645–660.

Gasser, T. and Müller, H.-G. (1984) Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171–185.

Georgiev, A.A. (1984) Speed of convergence in nonparametric kernel estimation of a regression function and its derivatives. *Ann. Inst. Statist. Math.*, **36**, 455–462.

Green, P.J. and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. London: Chapman & Hall.

Hall, P. (1992) On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, **20**, 695–711.

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

Lai, C.-J. and Chu, C.K. (2001) Triple smoothing estimation of the regression function and its derivatives in nonparametric regression. *J. Statist. Plann. Inference*, **98**, 157–175.

Luan, Y.H. and Xie, Z.J. (2001) The wavelet identification for jump points of derivative in regression model. *Statist. Probab. Lett.*, **53**, 167–180.

Mack, Y.P. and Silverman, B.W. (1982) Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **61**, 405–415.

Masry, E. and Fan, J. (1997) Local polynomial estimation of regression functions for mixing processes, *Scand. J. Statist.*, **24**, 165–179.

Müller. H.-G. (1988) *Nonparametric Regression Analysis of Longitudinal Data*, Lecture Notes in Statist. 46. Berlin: Springer-Verlag.

Ogden, T. (1997) *Essential Wavelets for Statistical Applications and Data Analysis*. Boston: Birkhäuser.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (1990) Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. (1991) Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725–1747.

Ruppert, D. and Wand, M.P. (1994) Multivariate locally weighted least squares regression. *Ann. Statist.*, **22**, 1346–1370.

Ruppert, D., Wand, M.P., Holst, U. & Hössjer, O. (1997) Local polynomial variance function estimation. *Technometrics*, **39**, 262–273.

Stone, C.J. (1977) Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–645.

Stone, C.J., Hansen, M., Kooperberg, C. and Truong, Y.K. (1997) Polynomial splines and their tensor products in extended linear modelling. With discussion and rejoinder by the authors and Jianhua Z. Huang. *Ann. Statist.*, **25**, 1371–1470.

Tsybakov, A.B. (1986) Robust reconstruction of functions by the local-approximation method. *Problems Inform. Transmission*, **22**, 133–146.

Vidakovic, B. (1999) *Statistical Modeling by Wavelets*. New York: Wiley.

Wahba, G. (1977) A survey of some smoothing problems and the method of generalized cross-validation for solving them. In P.R. Krisnaiah (ed.) *Application of Statistics*, pp. 507–523. Amsterdam: North-Holland.

Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing.* London: Chapman & Hall.

Welsh, A.H. (1996) Robust estimation of smooth regression and spread functions and their derivatives. *Statist. Sinica*, **6**, 347–366.

Zhou, S.G. and Wolfe, D.A. (2000) On derivative estimation in spline regression. *Statist. Sinica*, **10**, 93–108.

Received February 2004 and revised May 2004