Ann. Inst. Statist. Math.
Vol. 56, No. 3, 475–496 (2004)

CONFIDENCE BANDS IN NONPARAMETRIC REGRESSION WITH LENGTH BIASED DATA

J. A. CRISTÓBAL, J. L. OJEDA AND J. T. ALCALÁ

Department of Statistical Methods, University of Zaragoza, 50009 Zaragoza, Spain

(Received March 18, 2002; revised September 29, 2003)

Abstract. In this paper we deduce a confidence bands construction for the nonparametric estimation of a regression curve from length biased data, where a result from Bickel and Rosenblatt (1973, *The Annals of Statistics*, **1**, 1071–1095) is adapted to this new situation. The construction also involves the estimation of the variance of the local linear estimator of the regression, where we use a finite sample modification in order to improve the performance of these confidence bands in the case of finite samples.

Key words and phrases: Confidence bands, length-biased data, nonparametric regression.

1. Introduction

One of the main aims of nonparametric regression is to highlight an important structure in the data without strong assumptions being made about the particular relationship between the variables. For this reason, together with the development of current computer-intensive methods, an enormous amount of effort has been devoted to smoothing methods, and a prolific literature has recently appeared on this topic. However, the study of confidence bands for the regression function has been somewhat slower to develop. Methods for constructing such confidence bands in a fixed design regression model can be found in Hall and Titterington (1988), who present an approach based on the discretization method introduced in Knafl *et al.* (1985). Härdle and Marron (1991) use the wild bootstrap technique to construct variability bands by first establishing simultaneous confidence intervals for a set of grid points. In a random design, Johnston (1982) (with known marginal density) and Härdle (1989) (by way of M-smoothers) develop the confidence bands, based on the extreme value theory of Gaussian processes. Eubank and Speckman (1993) study the properties of a bias-corrected method for constructing data-driven confidence bands. Diebolt (1995) links techniques for confidence bands with the elaboration of a test for the regression function. Neumann and Polzehl (1998) obtain confidence bands based on a local polynomial estimator, via wild bootstrap. Other authors have extended these results to dependent data (see Xia (1998), among others). Another useful way to construct confidence bands in both the parametric and nonparametric setting is the "tube formula", see Sun and Loader (1994). This method is based on an important result from Hotelling with respect to tube volume.

Against this background, our aim is to deduce confidence bands for the regression function in a nonparametric setting when the data in the response variable is drawn in a length-biased sampling. This kind of sampling appears naturally in many fields of research. Thus, if we record an observation by nature according to a certain stochastic model, such an observation will not have the original distribution unless every observation is given an equal chance of being recorded. In some situations, the probability of selecting any individual from within a population is proportional to its length. Thus, in analyzing discovery data, the size-bias phenomenon, in which the larger units tend to be discovered first, is a common problem; in aerial census data collected for estimating wildlife population characteristics, visibility bias is generally present, because groups with a larger number of animals are more likely to be sighted; the sojourn time of tourists in a country, as reported by those contacted at their hotels, has a size biased distribution, since the longer the time spent by the tourist at the hotel, the more probability there is of this being collected in the sample. Models of this kind arise quite naturally in econometric and epidemiological contexts and, more generally, in problems related to renewal processes. A wide range of examples are analyzed in Patil and Rao (1978), Patil *et al.* (1988), Rao (1997), and Cristóbal and Alcalá (2001) among others. Of course, from the point-of-view of practical applications, the most interesting problems appear when covariables are involved and where the response is length or size biased. In these cases, if the sampling bias in the response variable is ignored, a distortion is caused in the determination of the regression function and the direct application of kernel regression estimators produces inconsistencies (see Cristóbal and Alcalá (2000)).

In this paper, we analyze the relationship between a response variable Y, such that $Y > C > 0$ (a.s.), and a covariable X, where the data is observed by way of length biased sampling in the variable Y. For simplicity, we assume that $X \in [0,1]$ with density function f_X, and if f_{XY} denotes the joint density of (X,Y), then the density associated with the length biased vector is given by:

$$f_{XY}^w(x,y) = \frac{y f_{XY}(x,y)}{\mu_Y} \tag{1.1}$$

where $\mu_Y = \int y f_{XY}(x,y)dxdy$ is the mean of Y, which is assumed to be finite. Let us also denote by $\boldsymbol{E}^w[\cdot]$ and $\mathbf{Var}^w[\cdot]$ the expectation and variance, respectively, when these are calculated with density f_{XY}^w, in order to distinguish them from $\boldsymbol{E}[\cdot]$ and $\mathbf{Var}[\cdot]$, which are obtained from f_{XY}.

It is worth considering some simple consequences of equation (1.1). First, note that the regression function of the biased data does not agree with the regression function $m(x) = \boldsymbol{E}[Y \mid X = x]$. In fact, we have that:

$$\boldsymbol{E}^w[Y \mid X = x] = m(x)(1 + c^2(x)),$$

where $c^2(x)$ is $\sigma^2(x)/m^2(x)$, the squared conditional coefficient of variation. This is a simple consequence of $f^w(y \mid x)$, the observed response conditional density, being equal to $yf(y \mid x)m(x)^{-1}$. On the other hand, the marginal density function f_X^w of the observed X is different from the unobserved one f_X:

$$f_X^w(x) = \frac{f_X(x)m(x)}{\mu_Y}.$$

Hence, the direct application of kernel estimation methods to length biased data for both density and regression leads to inconsistencies.

Cristóbal and Alcalá (2000) propose estimating $m(x)$ from an i.i.d. sample (x_i, y_i) by adapting the local polynomial fitting technique (see Fan and Gijbels (1996)) to length biased data. Since the empirical distribution function has a mass proportional to $\frac{1}{y_i}$ at each sampling point (x_i, y_i) they minimize:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \cdots - \beta_p(x_i - x)^p)^2 \frac{1}{y_i} K_h(x_i - x) \tag{1.2}$$

and demonstrate that $\hat{m}_h(x) = \hat{\beta}_0$ is consistent, also studying its asymptotic mean squared error. A paper by Wu (2000), dealing with general weighted distributions, appeared at about the same time as Cristóbal and Alcalá (2000). Previously, Ahmad (1995) had already considered Nadaraya-Watson type kernel regression estimators in this framework, and later, Sköld (1999), introduced a local linear estimator.

Our main goal here is to obtain a function $l_n^\alpha(x)$ such that:

$$\lim_{n\to\infty} \boldsymbol{P}\{|m(x) - \hat{m}_h(x)| \le l_n^\alpha(x),\ \forall x \in [0,1]\} \ge 1 - \alpha \tag{1.3}$$

for $\hat{m}_h$ the local linear estimator, i.e.: the first order local polynomial estimator, $p = 1$ in equation (1.2). To that end, we consider a suitable estimation error process, proving that its path is close to a Gaussian process with known covariance structure. An asymptotic uniform confidence band can then be constructed from the distribution of the supremum of such a Gaussian process (by extending the result in Bickel and Rosenblatt (1973)). Because the data are length biased, to proceed in this way requires some inverse moment conditions on Y.

We also make the following set of assumptions, that we will refer to as the A conditions:

A1. The kernel K is a symmetric function, twice continuously differentiable in the interior of its compact support $[-a, a]$.

A2. Let h_n, the bandwidth used in the nonparametric estimation, be $O(n^{-\eta})$, where $1/5 < \eta < 1/2$.

A3. We assume that there are constants C_1 and C_2, such that $0 < C_1 < f_X(x) < C_2 < \infty$ for all $x \in [0,1]$.

A4. Let $G(x) = \boldsymbol{E}^w[(\frac{Y-m(x)}{Y})^2 \mid X = x]$ and further, let there be constants C_3 and C_4, such that $0 < C_3 < G(x) < C_4 < \infty$ for all $x \in [0,1]$.

A5. $m(x)$ and $G(x)$ are twice continuously differentiable, while $f'_X(x)$ is continuous in $(0,1)$.

Hypothesis A1 is usually made in nonparametric estimations. For its part K support compactness holds for simplicity in technical developments, but can be relaxed using more general assumptions on the integrability for K. Hypothesis A2 ensures consistency. Hypotheses A3–A5 are technical assumptions to obtain the uniform and almost sure convergence and the error term convergence.

The rest of the paper is organized as follows. In Section 2 we derive the strong convergence of the proposed estimator, besides an almost sure representation for the error process. As a consequence, we provide the confidence bands for the regression curve. Section 3 is devoted to the different functionals that appear in the confidence bands. In particular, a finite sample adapted estimator is proposed to overcome what is a cumbersome direct implementation. In Section 4 we report a brief simulation study with a view to analyzing the behavior of the proposed confidence bands. Finally, the proofs of the main results are given in the Appendix.

2. Estimation from length biased data and the error process

As is shown in Cristóbal and Alcalá (2000), one of the methods that can be used to overcome the problem raised by the length biased mechanism is the compensation of the length bias. While this is not the only technique available to solve the problem, it is general enough to provide estimations not only for the regression curve, but also for its derivatives. Such a compensation can be carried out by balancing the effect that large data values exhibit due to the length bias sampling. This is achieved using $\frac{1}{y_i}K(\frac{x_i-x}{h})$ instead of $K(\frac{x_i-x}{h})$ as weight for the nonparametric estimation. The use of this "modified kernel" in the local linear estimation of the regression curve leads to the following estimator:

$$\hat{m}_h(x) = \sum_{i=1}^{n} \frac{w_{ih}^w(x)}{\sum_{i=1}^{n} w_{ih}^w(x)} y_i \tag{2.1}$$

where:

$$w_{ih}^w(x) = \frac{1}{y_i}\left(s_2^w(x;h)K\left(\frac{x_i-x}{h}\right) - s_1^w(x;h)K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)\right), \tag{2.2}$$

and where

$$s_j^w(x;h) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{y_i} K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j.$$

As a consequence, we can also write $\hat{m}_h$ as

$$\hat{m}_h(x) = \frac{s_2^w(x;h)t_0^w(x;h) - s_1^w(x;h)t_1^w(x;h)}{s_2^w(x;h)s_0^w(x;h) - s_1^w(x;h)s_1^w(x;h)},$$

with t_j^w being the functions given by

$$t_j^w(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j.$$

Cristóbal and Alcalá (2000) also provide expressions for the asymptotic bias and variance of such an estimator. Moreover, they derive a criterion for bandwidth selection based on AMISE minimization. As a consequence:

$$\hat{m}_{h_n}(x) = m(x)(1 + o_p(1))$$

and, therefore, the consistency of the proposed estimator is proved and asymptotic normal distribution can be used to obtain confidence interval for $m(x)$.

Assuming that the regression curve is differentiability up to a large enough order, it is possible to develop confidence bands by means of a pointwise confidence intervals using the Bonferroni technique over a grid of points in $[0, 1]$ (see for example Knafl *et al.* (1985)). Since the use of the Bonferroni method on a large number of points involves the use of large normal quantile values, these bands are too wide and, hence, although they can be useful, they are not sufficiently accurate. In Bickel and Rosenblatt (1973), they proposed a new method for constructing confidence bands for the density curve estimation that relies on the supremum distribution of Gaussian processes. Other authors have also taken advantage of this technique: see Härdle (1989), Johnston (1982), among others. Although these confidence bands are much more accurate, they require more technical results about the supremun of processes and in this context they have to deal with ratios of random variables.

In order to take full advantage of such a method, we must be able to deal with the supremum distribution of an error process. This implies that simple consistency is not enough and that a stronger result: i.e. the uniform strong consistency is essential. Following Masry (1996), and using the assumptions referred to in the previous section, we can prove the following theorem, which will be used as the basis to establish the strong uniform consistency and weak convergence required in order to apply the Bickel and Rosenblatt result. As is usual, we will denote $\int u^i K(u)du$ by μ_i and $\int u^i K^2(u)du$ by ν_i. $R(g)$ will denote $\int g^2(u)du$, ($\nu_0 = R(K)$), while δ_n stands for $\sqrt{\frac{\log n}{nh_n}}$, which will appear as the order in the remainder term in the strong convergence terms. Let us also introduce the quantities $e_j^w(x; h_n)$ that will play a central role in our further developments:

$$e_j^w(x; h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} \left(\frac{y_i - m(x_i)}{y_i} \right) K\left(\frac{x_i - x}{h_n} \right) \left(\frac{x_i - x}{h_n} \right)^j.$$

THEOREM 2.1. *If conditions* A1–A5 *are satisfied and the kernel* K *is chosen such that* $\mu_0 = \mu_2 = 1$, *and* $\mu_1 = \mu_3 = 0$ *then:*

$$\hat{m}_{h_n}(x) = m(x) + \frac{\mu_Y e_0^w(x; h_n)}{f_X(x)} + \frac{h_n^2}{2} m''(x) + O(h_n \delta_n) \tag{2.3}$$

uniformly in $[0, 1]$ *and almost surely.*

This theorem, together with the fact that $e_j^w(x; h_n) = o(\delta_n)$ uniformly in $[0, 1]$ and almost surely, as is proven in Proposition A.4 in the Appendix, leads to the following corollary where the strong uniform consistency is established.

COROLLARY 2.1. *Under conditions in Theorem* 2.1 *we have that:*

$$\hat{m}_{h_n}(x) = m(x) + O(h_n^2 + \delta_n)$$

uniformly in $[0, 1]$ *and almost surely.*

Thus, as a consequence of Theorem 2.1, we can ensure that there is a constant $C > 0$, such that:

$$\boldsymbol{P}\{\|\hat{m}_{h_n}(x) - m(x)\| \leq C(h_n^2 + \delta_n) \text{ a.e.}\} = 1,$$

where $\|\cdot\|$ denotes the supremum norm, and *a.e.* means that there is an n_0 such that the property holds for every $n > n_0$. This, of course, means that with probability one the supremum over all x in the unit segment does not grow faster than $C(h_n^2 + \delta_n)$, a quantity that tends to 0 as n increases.

Moreover, Theorem 2.1 leads to the following expansion for the error process, i.e. the difference $\hat{m}_{h_n}(x) - m(x)$:

$$\hat{m}_{h_n}(x) - m(x) = \frac{\mu_Y e_0^w(x; h_n)}{f_X(x)} + O(h_n^2 + h_n \delta_n)$$

uniformly in $[0,1]$ and almost surely. The relevance of the error process lies in the fact that its distribution provides us with the quantiles required in order to construct the confidence band. Thus, being able to obtain a distribution for the error process is crucial to our objectives. It is easy to see that the asymptotic conditional variance of $e_0^w(x;h_n)$ is given by:

$$\mathbf{AVar}^w[e_0^w(x;h_n)] = \frac{1}{nh_n}G(x)f_X^w(x)R(K) = \frac{1}{nh_n}S(x)R(K) \tag{2.4}$$

where G is introduced in assumption A4 and $S(x) = G(x)f_X^w(x)$. Therefore, Theorem 2.1 ensures that the standard error process can be represented by some summation process (i.e.: $e_0^w(x;h_n)$), in such a way that the difference between both process paths is uniformly and almost surely bounded by a quantity that decreases to 0 as n grows to infinity. Hence, and under condition A2, we have that for $\gamma = (5\eta - 1)/2 > 0$:

$$\sqrt{\frac{nh_n}{S(x)}}\frac{f_X(x)}{\mu_Y}(\hat{m}_{h_n}(x) - m(x)) = \frac{e_0^w(x;h_n)}{\sqrt{\frac{S(x)}{nh_n}}} + O(n^{-\gamma}) \tag{2.5}$$

uniformly over $[0,1]$ and almost surely. Notice that in case $\eta = 1/5$ the distributional behavior of $e_0^w(x;h_n)$ will not be sufficent to address the asymptotic behavior of the left term in equation (2.5).

Then, to obtain the confidence bands, we will approximate uniformly the path of this standardized version of $e_0^w(x;h_n)$ by way of a Gaussian process path. Specifically, it is possible to bound the paths of the standardized error process by the supremum of the following Gaussian process:

$$Z_n^*(x) = \frac{1}{\sqrt{h_n}}\int_{[0,1]} K\left(\frac{z-x}{h_n}\right) dW(z), \tag{2.6}$$

where $W(\cdot)$ is a standard Brownian Motion. Because of this approximation, the error process path is uniformly close to a process path whose confidence bands can easily be constructed. As a consequence, and applying a result in Bickel and Rosenblatt ((1973), p. 1084) to the Z_n^* process, we find that the confidence bands can be constructed at a level α using the following theorem:

THEOREM 2.2. *Let $L_{h_n}^\alpha$ be*

$$\sqrt{-2\log h_n} + \frac{1}{\sqrt{-2\log h_n}}\left\{\log\frac{1}{2\pi}\left(\frac{R(K')}{R(K)}\right)^{1/2} - \chi_\alpha\right\},$$

where $\chi_\alpha = \log(\frac{-\log(1-\alpha)}{2})$. *Then, under assumptions* A1–A5, *we have*:

$$\lim_{n\to\infty} \boldsymbol{P}\left\{\left\|\sqrt{\frac{nh_n}{S(x)}}\frac{f_X(x)}{\mu_Y}(\hat{m}_{h_n}(x) - m(x))\right\| \le \sqrt{R(K)}L_{h_n}^\alpha\right\} = 1-\alpha.$$

Hence, for $l_n^\alpha(x)$ in (1.3), we can use:

$$\sqrt{\frac{S(x)}{nh_n}}\frac{\mu_Y}{f_X(x)}\sqrt{R(K)}L_{h_n}^\alpha. \tag{2.7}$$

Note that in the previous expression there are several unknown functionals that should be estimated.

From equations (2.3) and (2.4) in Theorem 2.1, we can find the AMISE, whose minimization leads to a bandwidth selector. The AMISE bandwidth selector development can be found in Cristóbal and Alcalá (2000) and Wu (2000). While the former bandwidth selector is entirely based on asymptotics, Wu (2000) also proposes a bandwidth selector based on Cross Validation. Nervertheless usual methods can also be applied to obtain a suitable bandwidth, given that:

$$\boldsymbol{E}^w[(\hat{m}_h(X) - m(X))^2] = \int (\hat{m}_h(x) - m(x))^2 f_X^w(x)dx.$$

This equality shows that using Cross Validation criteria with weight $w(x) = 1$ is equivalent to using ISE with weight $w(x) = f_X^w(x)$, which is a common weight function, useful in order to avoid the contribution of isolated points to the ISE.

We should also note that using the expansion for m_{h_n} given in Theorem 2.1, and once $m''(x)$ is estimated, we can use it to perform bias correction (see Xia (1998)). Note that this estimation has to be performed separately from that carried out for $m(x)$, given that optimal bandwidth selection in local polynomials depends on what is going to be estimated (see Fan and Gijbels (1996)). It should further be recalled that these results are asymptotic and, therefore, that it is convenient to adapt the confidence bands construction to finite samples so as to obtain a better performance.

3. Finite sample construction

The previous sections have been concerned with the theory required in order to construct the confidence bands. As we have seen, the results are asymptotic, i.e. they have to do with sample sizes that grow to infinity. Hence, there are unknown functionals, and we will have to provide estimations for them using sample data, adapting such theoretical results to finite samples.

These unknown functionals are $S(x)$, $f_X(x)$, and, μ_Y. Because of the length bias, the estimation of μ_Y can be carried out by means of the harmonic mean, which, in this case, is a root n consistent estimator. Nevertheless, there is no need to estimate it, given that for a suitable bandwidth h'_n

$$\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{y_i}\frac{1}{h'_n} K\left(\frac{x_i - x}{h'_n}\right),$$

is a consistent estimator for $g(x) = f_X(x)/\mu_Y$. In fact, under our assumptions, this estimator is a uniform strong consistent estimator for g. Note that in the last expression we use $1/y_i$ because of length bias in the response.

The estimation of $S(x)$ can be carried out in various ways, but since $S(x) = G(x)f_X^w(x)$, the estimation of $G(x)$ is required. In order to perform the estimation of $G(x)$, we have chosen methods analogous to those proposed in Härdle and Tsybakov (1997) or Fan and Yao (1998) to estimate the variance function given that $G(x) = \boldsymbol{E}^w[(Y^{-1}(Y - m(x)))^2 \mid X = x]$. Hence, the following estimator

$$\hat{G}_n(x) = \sum_{i=1}^{n} \left(\frac{y_i - \hat{m}_n^{(i)}(x_i)}{y_i}\right)^2 \frac{w_{ih''_n}(x)}{\sum_{i=1}^{n} w_{ih''_n}(x)} \tag{3.1}$$

where $\hat{m}_n^{(i)}(x_i)$ is the estimator for the regression function when the data are length biased, as was proposed in Section 2, evaluated in $x = x_i$ but without using the observation (x_i, y_i). Furthermore, $w_{ih''}(x)$ are the weights given in (2.2), but without the length bias compensation, that is to say:

$$w_{ih}(x) = s_2(x;h)K\left(\frac{x_i - x}{h}\right) - s_1(x;h)K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)$$

where in this case:

$$s_j(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j.$$

The consistency of this estimator follows in the same way as was proved in the case of the local linear estimator for the regression function, taking into account the consistency for $\hat{m}_n$ that was given in Corollary 2.1.

THEOREM 3.1. *Under assumptions* A1–A5, *if the bandwidth* h''_n *verifies condition* A2, *we have that:*

$$\hat{G}_n(x) = G(x) + O(h''^2_n + \delta''_n)$$

uniformly in $[0,1]$ *and almost surely, where* $\delta''_n = \sqrt{\frac{\log n}{nh''_n}}$.

Note that the strong uniform consistency rate can be written in terms of h_n, that is to say the bandwidth for the regression function, because both estimators are based on the same number of observations

We are now able to estimate $S(x)$ by means of a Parzen-Rosenblatt estimation for $f_X^w(x)$, and then to construct the confidence bands by means of the direct use of (2.7). Nevertheless, we instead propose using a finite sample modification that is closer to the error process variance, and furthermore, it is simpler than the given construction in terms of computational cost. This modification takes the form:

$$\hat{l}_n^\alpha(x) = C_n^w(x)\sqrt{R(K)}L_n^\alpha,$$

where $C_n^w(x)$ is defined as:

$$C_n^w(x)^2 = \sqrt{\mu_2}\frac{1}{\sum_{i=1}^n w_{ih_n}^w(x)}\sum_{i=1}^{n}\frac{w_{ih_n}(x)}{\sqrt{\sum_{i=1}^n w_{ih_n}(x)}}\hat{G}(x_i). \tag{3.2}$$

Note the use of weights w_{ih} and w_{ih}^w. The first is used to achieve the estimation of $S(x)$, given that we are smoothing the variance estimator, and the square root on the denominator leads to the appearance of a f_X^w factor on the numerator. The second weights are used simply to obtain f_X, see (2.7). It is worth noting that the use of bandwidth h_n makes this estimation closely resemble the behavior of the local lineal estimator of the regression function. The following result proves the strong consistency of the proposed estimator.

THEOREM 3.2. *Under assumptions* A1–A5, *we have that:*

$$C_n^w(x)^2 = \mu_Y^2\frac{G(x)f_X^w(x)}{nh_nf_X(x)^2} + O(h_n^2 + \delta_n).$$

uniformly in $[0,1]$ *and almost surely.*

The confidence bands construction can also be improved by means of bias correction for the estimation of the regression curve (see Xia (1998)). This improvement can be obtained using a higher degree local polynomial modified estimator for length biased data (see Cristóbal and Alcalá (2000)).

4. A brief simulation study

In this section we describe a brief simulation study with the aim of better appreciating the behavior of the proposed confidence bands. We will first consider these bands in a simple and known situation with unbiased data. Such an approach will help us to focus on the main facts that arise in the construction of the confidence bands using the distribution of the Gaussian processes supremum. In a second step, we will consider the performance of such a construction for length biased data in a more realistic case where, besides the use of estimators given in the second section, $f_X(x)$ and $S(x)$ also have to be estimated.

For the first study, let us consider the following process:

$$\sqrt{nh_n}e(x,h_n) = \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right)\epsilon_i$$

where the kernel function K is the so called Epanechnikov kernel

$$K(u) = \frac{3}{4}(1-u^2)\mathbf{1}_{[-1,1]}(u),$$

and $\{(x_i,\epsilon_i)\}_{i=1}^n$ is a sample from the random variable (X,E), and with X being uniformly distributed over $[0,1]$ and independent from E, a Gaussian standard random variable. Apart from its simplicity, this process has several useful features related to the local polynomial estimation. Thus, $e(x,h_n)$ estimates $\boldsymbol{E}[E \mid X=x]f_X(x)$, which, in this case, is the null function over $[0,1]$. Moreover, as

$$\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right) = f(x) + O(h_n^2 + \delta_n)$$

uniformly in $[0,1]$ and almost surely because of Proposition A.3, the Nadaraya-Watson estimator in the example can be expressed using the following asymptotic expansion:

$$\hat{m}_n^{NW}(x) = e(x,h_n) + O(h_n^2)$$

uniformly in $[0,1]$ and almost surely. The same kind of expansion has been shown to hold for the local linear estimator when the data is length biased and, of course, it is also valid in the unbiased case. As a consequence, this process represents the main stochastic features of these estimators. Using Proposition A.5, it can be seen that the limit process for $\sqrt{nh_n}e(x,h_n)$ agrees with that for $Z_n^*(x)$, and thus $\sqrt{R(K)}L_h^\alpha$ are asymptotic confidence bands for that process.

Tables 1 and 2 include the number of times that

$$S_{nh} = \max_x \sqrt{nh}e(x,h)$$

Table 1. S_{nh} simulation. $\alpha = 0.1$.

$n\backslash h$	0.005	0.0125	0.025	0.05	0.1	0.25
25	0.36	0.27	0.25	0.18	0.12	0.07
50	0.37	0.30	0.21	0.14	0.10	0.08
100	0.40	0.25	0.17	0.11	0.08	0.06
200	0.36	0.16	0.11	0.08	0.07	0.06
400	0.30	0.18	0.11	0.08	0.07	0.06
800	0.23	0.13	0.10	0.09	0.10	0.08
1600	0.17	0.13	0.10	0.09	0.08	0.08

Table 2. S_{nh} simulation. $\alpha = 0.05$.

$n\backslash h$	0.005	0.0125	0.025	0.05	0.1	0.25
25	0.29	0.20	0.14	0.07	0.03	0.02
50	0.29	0.19	0.13	0.07	0.05	0.02
100	0.29	0.17	0.10	0.06	0.04	0.02
200	0.23	0.10	0.04	0.04	0.03	0.02
400	0.19	0.09	0.04	0.03	0.03	0.02
800	0.15	0.08	0.05	0.04	0.03	0.02
1600	0.10	0.05	0.04	0.04	0.03	0.03

is larger than $\sqrt{R(K)}L_h^\alpha$, i.e. the uncovered proportion of samples, for a given confidence level $1-\alpha$, where the maximum is computed over a grid on $[0,1]$ with 1600 equidistant points. 500 observations for $S_{n,h}$ have been simulated for different samples sizes n and bandwidths h in order to obtain the percentage of times that $S_{n,h} > \sqrt{R(K)}L_h^\alpha$.

We can note that for a fixed n, as h increases, so the uncovered percentage of samples decreases to 0. This agrees with the behavior of the local polynomial estimator when this estimator is under-smoothed, i.e. when the bandwidth is relatively small with respect to the sample size n. As the bandwidth gets close to zero, the value of the estimators at x_i becomes simply ϵ_i but, in other points, it is null if defined and, therefore, exhibits characteristic peaks. On the other hand, as h gets larger when the estimation is over-smoothed, the local polynomial estimator tends to assign the same weight to every point in the sample. Thus, it behaves approximately as an average of the data, producing a flat and smooth estimation. Note that this behavior is shared by the process $Z_n^*(x)$.

Consequently, in order to obtain valid confidence bands, it is not possible to use any value of h for any sample size. Rather, this bandwidth depends on a well known fact for local linear estimators, namely that, with a view to achieving consistency, it is required that the bandwidths decrease to zero more slowly than $1/n$. Moreover, to obtain the minimum possible squared error, h_n should be an $O(n^{1/(2p+3)})$ for the order p local polynomial estimator. That is the reason why we should consider only a certain range of bandwidths(about 0.05–0.125) where the confidence level is approximately achieved depending on n.

Now, and with the aim of providing an additional perspective of such a construction from the practical and visual point of view, we will consider a more realistic case, where the data is affected by length bias and where we also have to estimate some other

Table 3. Empirical coverage.

α	A	n	Emp. Cov.	α	A	n	Emp. Cov.	α	A	n	Emp. Cov.
0.10	1	100	0.86	0.05	1	100	0.92	0.01	1	100	0.98
		200	0.88			200	0.91			200	0.97
		400	0.88			400	0.96			400	0.99
	2	100	0.89		2	100	0.92		2	100	0.96
		200	0.90			200	0.94			200	0.98
		400	0.91			400	0.96			400	0.99
	3	100	0.85		3	100	0.93		3	100	0.98
		200	0.91			200	0.95			200	1.00
		400	0.92			400	0.96			400	0.99

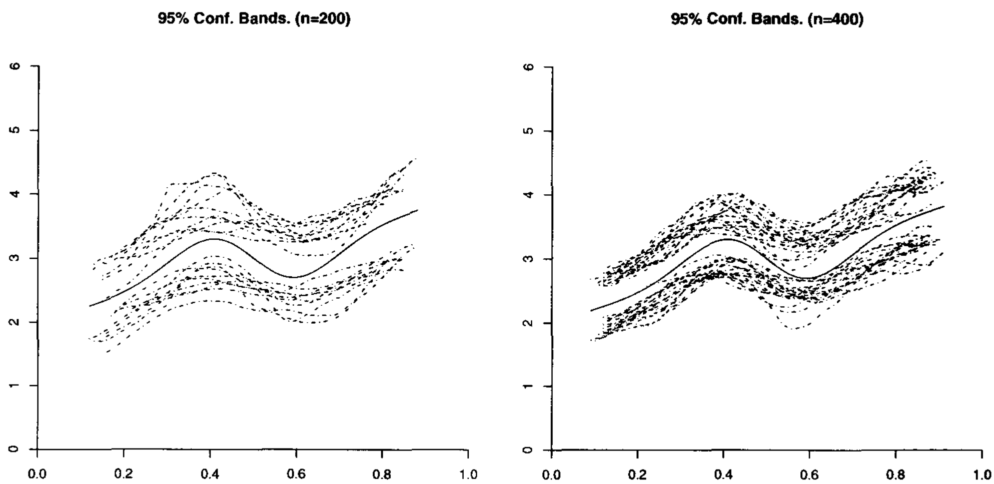

Fig. 1. Samples of confidence bands with a 95% level, for a sample size $n = 200$ (left-hand panel) and $n = 400$ (right panel) for the model (4.1) with $A = 1$.

functional data. Let us suppose we are interested in $\boldsymbol{E}[Y \mid X = x]$, where X is distributed uniformly in $[0,1]$ and $Y = m(x)(1 + 0.35\epsilon)$, with ϵ being a uniform random variable over $[-\sqrt{3}, \sqrt{3}]$. The regression function m is given by:

$$m(x) = 2 + 2x + A(\exp(-(x-0.45)^2/0.025) - \exp(-(x-0.55)^2/0.025)). \tag{4.1}$$

In Table 3 we have added the empirical coverage for the proposed confidence bands in 500 simulations of the previous model, but now with the observed data being affected by length bias. This has been done for different confidence levels $1 - \alpha$, and for different sample size values n. In order to study the behavior in the circumstances where the regression function has peaks, we have considered different values for A in (4.1). As we can appreciate, better results are achieved in the case of a larger n, while in the case of shorter sample sizes the coverage is a little less than the expected nominal value. Note that this has to do with the large bias that results from the ample crests in the

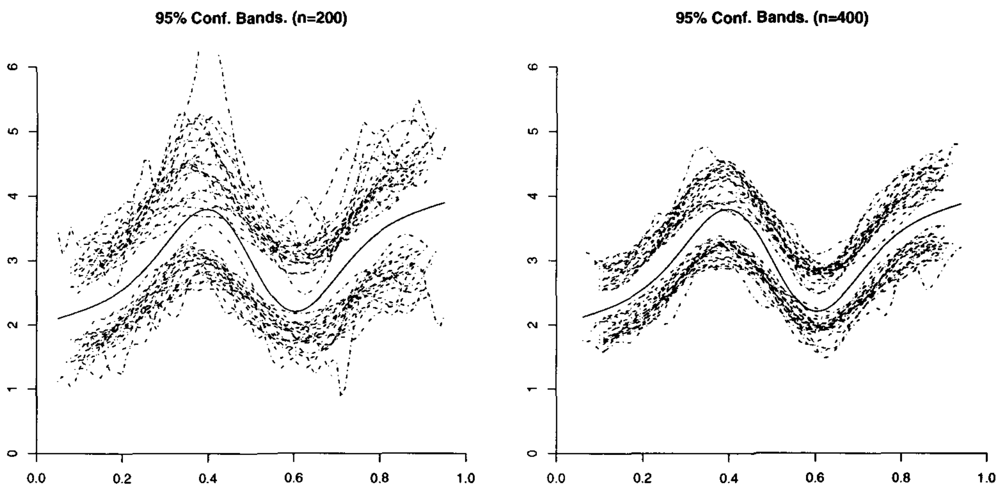

Fig. 2. Samples of confidence bands with a 95% level, for a sample size $n = 200$ (left-hand panel) and $n = 400$ (right panel) for the model (4.1) with $A = 2$.

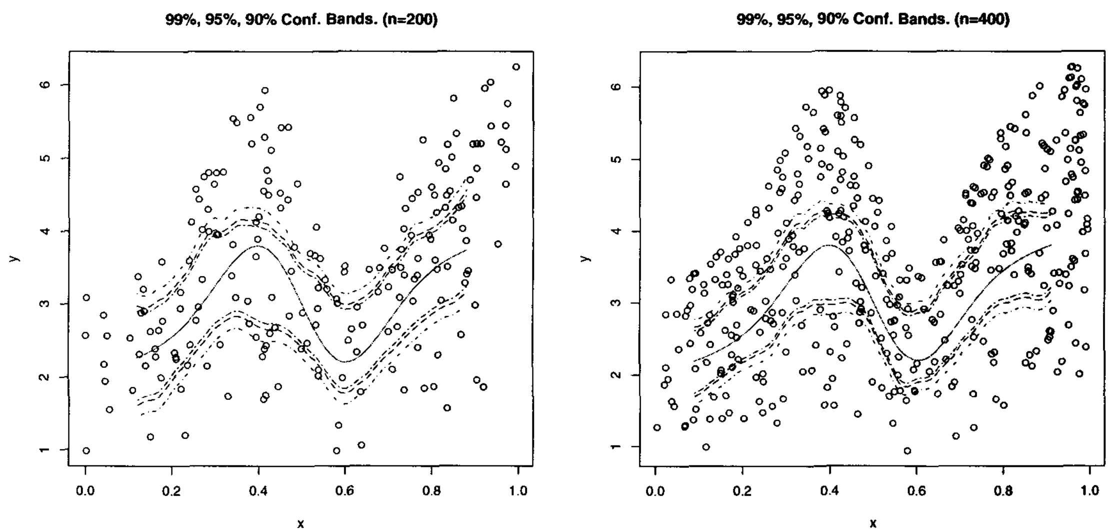

Fig. 3. Two typical length biased samples of sizes $n = 200$ (left-hand panel) and $n = 400$ (right-hand panel) with the true regression function (solid line) and estimated confidence bands with 90%, 95% and 99% levels of confidence (dotted lines). The simulated model is given by (4.1) with $A = 2$.

regression function, whose effect is much more noticeable when the sample size is small. In every simulation, we have considered bandwidths h_n and h_n'' that minimize usual Cross Validation for observations and squared compensated residuals respectively, evaluating the curves in $(h_n, 1 - h_n)$ in order to avoid the so-called boundary effect.

In Figs. 1 and 2 we have also drawn the regression function m(solid line), with 95% level confidence bands(dot-dashed lines) for 20 independent samples of (X, Y) affected by length bias in the response. The construction has been carried out using (3.2), and we have also considered two different values for A in (4.1).

As can be seen, the confidence bands appear to be symmetrically located around the regression function, covering it in most cases. We can also observe that the roughness the confidence bands exhibits is due to the use of a Cross Validation bandwidth selector which, in most cases, gives under-smoothed estimations. It should also be noted that the behavior at the central part is not bad, although it is precisely in this central part where the local linear estimator may be less precise due to of the estimation bias produced by the peaks and valleys the regression function we have chosen exhibits in that zone. In Fig. 3 we have also added a plot of two such length biased samples, with the regression function and the 99%, 95% and 90% level confidence bands similarly being plotted. Note the length bias effect, as well as how the estimator compensates for it.

Acknowledgements

The authors would like to express their thanks to two anonymous referees for their helpful comments and observations on an earlier version of this paper. This work was supported by DGES grants PB98-1587, BMF-2002-00080.

Appendix

A.1 *Previous results about power series and expectations*

The results presented here allow us to obtain the expansions we need in order to perform the analysis of the convergence rates. They will be used for local constant or local linear estimators, i.e.: $p = 0$ or $p = 1$, although the result will be set for a general p.

PROPOSITION A.1. *Let X be a random variable on $[0,1]$ whose distribution function is F and with density function f. Let g also be a real valued function on $[0,1]$, with K verifying assumption A1. Let us suppose that f and g have bounded derivatives up to order p for every x in $(0,1)$. Then:*

$$\boldsymbol{E}\left(g(X)\left(\frac{X-x}{h}\right)^j \frac{1}{h}K\left(\frac{X-x}{h}\right)\right) = \sum_{i=0}^{p-1} h^i \mu_{i+j}\frac{gf^{(i)}(x)}{i!} + O(h^p) \tag{A.1}$$

and

$$\boldsymbol{E}\left(g(X)\left(\frac{X-x}{h}\right)^j \frac{1}{h^2}K^2\left(\frac{X-x}{h}\right)\right) = \sum_{i=0}^{p-1} h^{i-1} \nu_{i+j}\frac{gf^{(i)}(x)}{i!} + O(h^{p-1}) \tag{A.2}$$

where $gf(x)$ is $g(x)f(x)$.

PROOF OF PROPOSITION A.1. Change the variable X by $x+uh$, and integrate on u the p order Taylor expansion of $gf(x+uh)$ in $h=0$. □

As $\boldsymbol{E}^w[\cdot]$ agrees with $\frac{1}{\mu_Y}\boldsymbol{E}[Y\cdot]$, these results hold for biased data when using $\frac{1}{y_i}K(\frac{x_i-x}{h})$ instead of $K(\frac{x_i-x}{h})$. Such results help us to obtain the expectation and variance for the quantities s_j^w, e_j^w, s_j, and e_j that will appear later in this Appendix.

A.2 *Strong convergence results*

The results presented here enable us to obtain strong consistency for the estimators proposed in Sections 2 and 3, and also allow us to achieve the strong consistency rates. Let us assume from now on that $\|\cdot\|$ denotes the supremum norm.

PROPOSITION A.2. *Let h_n be a sequence of positive real numbers, such that $h_n = O(n^{-\eta})$, $\eta > 0$ and let $Y_j(x)$, $j = 1, \dots, n$ be independent real valued processes on $[0,1]$ verifying the following conditions:*

(i) *There exists a constant $A > 0$, such that $|Y_j(x)| < A/h_n$ a.e. and $\mathbf{Var}[Y_j(x)] \le A/h_n$.*

(ii) *$|Y_j(x) - Y_j(y)| < B_j|x-y|$ and $\frac{1}{n}\sum_{j=1}^n B_j$ is $O(\frac{1}{h_n})$ almost surely.*

Then:

$$\exists C > 0 \text{ s.t. } \boldsymbol{P}\left\{\left\|\frac{1}{n}\sum_{j=1}^n (Y_j(x) - \boldsymbol{E}[Y_j(x)])\right\| \le C\delta_n \text{ a.e.}\right\} = 1 \tag{A.3}$$

where $\delta_n = \sqrt{\frac{\log n}{nh_n}}$ and a.e. means that there is an n_0 s.t. the property holds for every $n > n_0$.

PROOF OF PROPOSITION A.2. Without loss of generality, let us assume that $\boldsymbol{E}[Y_j(x)]$ is null.

Let $\{I_k^n\}_{k=0}^{L(n)}$ be disjoint subintervals of $[0,1]$ with the same length, such that $[0,1] = \bigcup_{k=1}^{L(n)} I_k^n$, where the number of intervals $L(n)$ is an $O(\sqrt{\frac{n}{h_n \log n}})$ quantity, and let $\hat{x}_k$ be the middle point of I_k^n. Then:

$$\left\|\frac{1}{n}\sum_{j=1}^n Y_j(x)\right\| \le \max_{1\le k\le L(n)}\left|\frac{1}{n}\sum_{j=1}^n Y_j(\hat{x}_k)\right| + \max_{1\le k\le L(n)}\sup_{x\in I_k^n}\left|\frac{1}{n}\sum_{j=1}^n (Y_j(x) - Y_j(\hat{x}_k))\right|.$$

Note that the second term can be bounded by

$$\max_{1\le k\le L(n)} \text{Length}(I_k^n)\frac{1}{n}\sum_{j=1}^n B_j,$$

and, as the length of I_k^n is $O(L(n)^{-1})$ for every k, this term is $O(\delta_n)$ given (ii). It only remains for us to prove that there is a $C > 0$, such that:

$$\boldsymbol{P}\left\{\max_{1\le k\le L(n)}\left|\frac{1}{n}\sum_{j=1}^n Y_j(\hat{x}_k)\right| < C\delta_n \text{ a.e.}\right\} = 1.$$

This is equivalent to $\boldsymbol{P}\{\max_{1\le k\le L(n)} |\frac{1}{n}\sum_{j=1}^n Y_j(\hat{x}_k)| > C\delta_n \text{ i.o.}\} = 0$, where i.o. means that the property holds for infinitely many integers, and it will be proved to be true by way of the Borel-Cantelli lemma if the following condition holds:

$$\sum_{n\ge 0}\boldsymbol{P}\left\{\max_{1\le k\le L(n)}\left|\frac{1}{n}\sum_{j=1}^n Y_j(\hat{x}_k)\right| > C\delta_n\right\} < \infty. \tag{A.4}$$

To check that this condition is indeed verified, note that we can bound these probabilities for every k by means of Markov inequality as follows:

$$\begin{aligned}(A.5)\qquad &\boldsymbol{P}\left\{\left|\frac{1}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right|>C\delta_n\right\}\\ &\leq\frac{\boldsymbol{E}\left[\exp\left(\frac{\lambda_n}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right)\right]+\boldsymbol{E}\left[\exp\left(-\frac{\lambda_n}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right)\right]}{e^{C\lambda_n\delta_n}},\end{aligned}$$

where λ_n is defined as $\sqrt{nh_n\log n}/A$ and so $|\lambda_n Y_j(\hat{x}_k)/n|<1$ a.e. by assumption (i). Therefore, and as a consequence of the following inequality for the exponential function $\exp(x)\leq 1+x+x^2$ for $|x|<1$, we have that:

$$\boldsymbol{E}\left[\exp\left(\frac{\lambda_n}{n}Y_j(\hat{x}_k)\right)\right]\leq 1+\frac{\lambda_n^2}{n^2}\boldsymbol{E}[Y_j^2(\hat{x}_k)]\leq\exp\left(\frac{\lambda_n^2}{n^2}\mathbf{Var}[Y_j(\hat{x}_k)]\right).$$

Hence, and because of the independence of $Y_j(\hat{x}_k)$:

$$\boldsymbol{E}\left[\exp\left(\frac{\lambda_n}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right)\right]\leq\exp\left(\frac{\lambda_n^2}{n^2}\sum_{j=1}^{n}\mathbf{Var}[Y_j(\hat{x}_k)]\right).$$

The same holds for the second summand in (A.5).

As a consequence of assumptions (i), and the fact that $\lambda_n\delta_n$ is $A^{-1}\log n$, we have the following bound for every k:

$$\boldsymbol{P}\left\{\left|\frac{1}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right|>C\delta_n\right\}\leq 2\frac{e^{\log n}}{e^{(C/A)\log n}},$$

and therefore:

$$\begin{aligned}&\boldsymbol{P}\left\{\max_{1\leq k\leq L(n)}\left|\frac{1}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right|>C\delta_n\right\}\\ &\leq\sum_{k=1}^{L(n)}\boldsymbol{P}\left\{\left|\frac{1}{n}\sum_{j=1}^{n}Y_j(\hat{x}_k)\right|>C\delta_n\right\}\leq\frac{2L(n)}{n^{(C/A)-1}}.\end{aligned}$$

Hence, (A.4) holds for a large enough C. □

Let us now apply this general result to the local linear estimator for length biased data given in equation (2.1).

PROPOSITION A.3. *Assuming* A *conditions, we have that*:

$$s_i^w(x;h_n)=\frac{1}{\mu_Y}\left(\mu_i f_X(x)+\mu_{i+1}f_X'(x)h_n\right)+O(h_n^2+\delta_n)$$

uniformly in $[0,1]$ *and almost surely.*

PROOF OF PROPOSITION A.3. This follows from Propositions A.1 and A.2. Note that $s_j^w(x;h_n) = \frac{1}{n}\sum_{i=1}^n Y_i(x)$, where $Y_i(x) = \frac{1}{y_i}\frac{1}{h_n}K(\frac{x_i-x}{h_n})(\frac{x_i-x}{h_n})^j$, and then:

$$|Y_i(x) - Y_i(x')| \leq \frac{1}{h_n^2}\frac{1}{y_i}\left|g'\left(\frac{x_i - \theta(x-x')}{h_n}\right)\right| |x - x'|$$

where $g(u) = u^j K(u)$, whose absolute value is integrable. Hence, as every $\frac{1}{y_i}$ is bounded a.e. because of the boundedness assumption on Y, condition (ii) in Proposition A.2 is guaranteed, while condition (i) is verified given the boundedness and the finite support of K and the previous observation about $\frac{1}{y_i}$. □

We will now achieve the error process representation up to h_n^2 terms and the strong consistency for $\hat{m}_{h_n}$. This result will enable us to obtain the bias term and a suitable expression to develop the confidence bands via the supremum distribution of a Gaussian process. The key points here are that $y_i = m(x_i) + \epsilon_i$ and the use of the power expansion of $m(x_i)$ on x:

PROPOSITION A.4. *If conditions* A1, A2, A4, *and* A5 *are satisfied and*:

$$e_j^w(x;h_n) = \frac{1}{nh_n}\sum_{i=1}^n \left(\frac{y_i - m(x_i)}{y_i}\right) K\left(\frac{x_i - x}{h_n}\right)\left(\frac{x_i - x}{h_n}\right)^j$$

then:

$$e_j^w(x;h_n) = O(\delta_n)$$

uniformly in $[0,1]$ *and almost surely.*

PROOF OF PROPOSITION A.4. Let $Y_i(x) = \frac{\epsilon_i}{y_i}\frac{1}{h_n}K(\frac{x_i-x}{h_n})(\frac{x_i-x}{h_n})^j$, with null expectation as a consequence of $\boldsymbol{E}^w[\cdot]$ being equal to $\frac{1}{\mu_Y}\boldsymbol{E}[Y\cdot]$. Also note that, as far as $\frac{1}{y_i}$ is bounded a.s., then so is $\frac{\epsilon_i}{y_i}$. As a consequence, conditions (i) and (ii) in Proposition A.2 follow by arguing the same line as in the Proposition A.3. □

PROOF OF THEOREM 2.1. As we have seen in Section 2:

$$\hat{m}_{h_n}(x) = \frac{\sum_{i=1}^n w_{ih_n}^w(x)y_i}{\sum_{i=1}^n w_{ih_n}^w(x)} = \frac{s_2^w(x;h_n)t_0^w(x;h_n) - s_1^w(x;h_n)t_1^w(x;h_n)}{s_2^w(x;h_n)s_0^w(x;h_n) - s_1^w(x;h_n)s_1^w(x;h_n)}.$$

Now, using the asymptotic power expansion for m:

$$m(x_i) = m(x) + h_n\left(\frac{x_i - x}{h_n}\right)m'(x) + h_n^2\left(\frac{x_i - x}{h_n}\right)^2\frac{m''(x)}{2},$$

and the fact that $y_i = m(x_i) + \epsilon_i$, we have that:

$$\begin{aligned}\hat{m}_{h_n}(x) &= \frac{\sum_{i=1}^n w_{ih_n}(x)y_i}{\sum_{i=1}^n w_{ih_n}(x)} = m(x) + 0\\ &\quad + h_n^2\frac{m''(x)}{2}\frac{s_2^w(x;h_n)s_2^w(x;h_n) - s_1^w(x;h_n)s_3^w(x;h_n)}{s_2^w(x;h_n)s_0^w(x;h_n) - s_1^w(x;h_n)s_1^w(x;h_n)}\\ &\quad + \frac{s_2^w(x;h_n)e_0^w(x;h_n) - s_1^w(x;h_n)e_1^w(x;h_n)}{s_2^w(x;h_n)s_0^w(x;h_n) - s_1^w(x;h_n)s_1^w(x;h_n)}.\end{aligned}$$

As a consequence of Propositions A.3 and A.4, we obtain the following uniform and almost sure power series expansions:

$$\begin{aligned}
\mu_Y s_0^w(x;h_n) &= f_X(x) + O(h_n^2 + \delta_n)\\
\mu_Y s_1^w(x;h_n) &= h_n f_X'(x) + O(h_n^2 + \delta_n)\\
\mu_Y s_2^w(x;h_n) &= \mu_2 f_X(x) + O(h_n^2 + \delta_n),
\end{aligned}$$

and also that $s_3^w(x;h_n) = O(h_n + \delta_n)$ and $e_j^w(x;h_n) = O(\delta_n)$ uniformly and almost surely on $[0,1]$. Therefore, we have that:

$$\begin{aligned}
\hat{m}_{h_n}(x) &= m(x) + \mu_2 h_n^2 \frac{m''(x)}{2}(1 + O(h_n^2 + \delta_n))\\
&\quad + \frac{\mu_Y e_0^w(x;h_n)}{f_X(x)} + O(\delta_n^2 + h_n\delta_n).
\end{aligned} \tag{A.6}$$

This concludes the Proof of Theorem 2.1. □

The following results are developed in order to carry out the estimation of $G(x)$ and $C_n^w(x)^2$ defined in Section 2. We will show that under the assumptions proposed in Section 1, the $\hat{G}(x)$, defined in (3.1), is a strongly consistent estimator for $G(x)$. We will follow the same steps as in the case of $\hat{m}_{h_n}$.

PROOF OF THEOREM 3.1. First, note that the weights w_{ih} are built in the same manner as w_{ih}^w, but without the reciproque of the response of the observations. Hence the same argument applied in the Proof of Propositions A.3 leads, in this case, to

$$s_j(x;h_n'') = (\mu_j f_X^w(x) + \mu_{j+1} f_X^{w\prime}(x) h_n'') + O(h_n''^2 + \delta_n'') \tag{A.7}$$

uniformly in $[0,1]$ and almost surely for a given bandwidth h_n'' fulfilling A2, and $\delta_n'' = \sqrt{\frac{\log n}{n h_n''}}$.

It follows from Theorem 2.1 that:

$$y - \hat{m}_n^{(i)}(x) = (y - m(x)) + O(h_n^{p+1} + \delta_n)$$

uniformly in $[0,1]$ and almost surely. Hence:

$$\begin{aligned}
\hat{G}(x) &= \sum_{i=1}^n \left(\frac{y_i - \hat{m}_n^{(i)}(x_i)}{y_i}\right)^2 \frac{w_{ih_n''}(x)}{\sum_{i=1}^n w_{ih_n''}(x)}\\
&= \sum_{i=1}^n \left(\frac{y_i - m(x_i)}{y_i}\right)^2 \frac{w_{ih_n''}(x)}{\sum_{i=1}^n w_{ih_n''}(x)}\\
&\quad + 2\sum_{i=1}^n \left(\frac{y_i - m(x_i)}{y_i}\right) \frac{w_{ih_n''}(x)}{\sum_{i=1}^n w_{ih_n''}(x)} O(h_n^{p+1} + \delta_n)\\
&\quad + O(h_n^{2(p+1)} + 2h_n^{p+1}\delta_n + \delta_n^2)
\end{aligned}$$

uniformly in $[0,1]$ and almost surely. Using Propositions A.3 and A.4 in the second and third terms of the previous equality, we obtain

$$\begin{aligned}
\hat{G}(x) &= \sum_{i=1}^n \left(\frac{y_i - m(x_i)}{y_i}\right)^2 \frac{w_{ih_n''}(x)}{\sum_{i=1}^n w_{ih_n''}(x)}\\
&\quad + O((h_n^{p+1} + \delta_n)\delta_n'') + O(h_n^{2(p+1)} + 2h_n^{p+1}\delta_n + \delta_n^2)
\end{aligned}$$

uniformly in $[0,1]$ and almost surely. The first summand in the second term of this previous equality is simply the local linear smoother of the squared compensated residuals. Thus, if we denote $(\epsilon_i/y_i)^2$ by ϵ_i', then ϵ_i' is a bounded random variable such that $\boldsymbol{E}^w[\epsilon_i' \mid x_i]$ is $G(x_i)$, and arguing as in Theorem 2.1 with weigths $w_{ih_n''}$ instead of $w_{ih_n}^w$, and data (x_i, ϵ_i'), $i = 1, \dots, n$, in place of the original observations, we achieve the result. □

Finally, we establish the strong consistency of $C_n^w(x)^2$.

PROOF OF THEOREM 3.2. Note that, following the Proof of Theorem 2.1, we obtain that

$$\sum_{i=1}^{n} w_{ih_n}^w(x) = (nh_n)^2 \left(\frac{1}{\mu_Y^2} \mu_2 f_X(x)^2 + O(h_n^2 + \delta_{nh_n}) \right)$$

uniformly in $[0,1]$ and almost surely.

A similar argument in the case of w_{ih} shows that, under response length biased data, we have that:

$$\sum_{i=1}^{n} w_{ih_n}(x) = (nh_n)^2 (\mu_2 f_X^w(x)^2 + O(h_n^2 + \delta_{nh_n}))$$

uniformly in $[0,1]$ and almost surely. Now, it simply remains for us to note that

$$C_n^w(x)^2 = \sqrt{\mu_2} \frac{\sqrt{\sum_{i=1}^n w_{ih_n}(x)}}{\sum_{i=1}^n w_{ih_n}^w(x)} \sum_{i=1}^{n} \frac{w_{ih_n}(x)}{\sum_{i=1}^n w_{ih_n}(x)} \hat{G}(x_i)$$

and the result follows from Theorem 3.1. □

A.3 *The distribution of the error process*

We now address the problem of the computation of the supreme distribution of the error process. Such a distribution will be obtained via theorem A1 in Bickel and Rosenblatt ((1973), p. 1084), which will provide us with the distribution of the supremum of certain Gaussian processes.

THEOREM A.1. *Let $Z_n^*(x)$ be the process given in* (2.6):

$$\lim_{n\to\infty} \boldsymbol{P} \left\{ \sqrt{-2\log h_n} \left(\max_{0\le x\le 1} \left| \frac{Z_n^*(x)}{\sqrt{R(K)}} \right| - B\left(\frac{1}{h_n}\right) \right) < \chi_\alpha \right\} = 1 - \alpha$$

where $\chi_\alpha = \log(\frac{-\log(1-\alpha)}{2})$ *and where* $B(\cdot)$ *is given by*:

$$B\left(\frac{1}{h}\right) = \sqrt{-2\log h} + \frac{1}{\sqrt{-2\log h}} \left\{ \log \frac{1}{2\pi} \left(\frac{R(K')}{R(K)} \right)^{1/2} \right\}.$$

PROOF OF THEOREM A.1. By changing z for $z' = \frac{z}{h}$ in the process Z_n^*, the integration domain will be $[0, \frac{1}{h}]$. If we also change x for $x' = \frac{x}{h}$, the maximization interval becomes $[0, \frac{1}{h}]$. These changes lead to the following process:

$$U_n(x') = \int_{[0,1/h]} K(z' - x') dW_n(z'),$$

being $W_n(\cdot) = \frac{1}{\sqrt{n}} W(\cdot)$ and W is a standard univariate Wiener process. The result in Bickel and Rosenblatt (1973) can now be applied to the process $\sqrt{R(K)^{-1}} U_n(x)$, and the thesis follows because $U_n(x)$ and $Z_n^*(x)$ have the same distribution: i.e. their expectation and covariance functions agree. □

We will now follow a modification of the ideas used in Johnston (1982), and Härdle (1989) in order to obtain the desired approximation results for a bounded variation function and under general assumptions about data distribution. This procedure leads to a chain of processes that ends in a final Gaussian process whose maximum distribution can be computed using the following result:

PROPOSITION A.5. *Let h_n be a sequence of positive real numbers, such that h_n is $O(n^{-\eta})$ where $\eta > 0$, and let $p(x, y)$ be a bounded variation real function from $[0,1] \times R$ on R. Assume that:*

(i) *$\{x_i, y_i\}_{i=1}^n$ is an independent and identically distributed sample from the random variable (X, Y), whose distribution and density functions are F and f_{XY}, respectively. We also denote its empirical distribution by F_n.*

(ii) *Let $G(x) = \boldsymbol{E}[p^2(X, Y) \mid X = x]$ and let $K(x)$ be twice continuous differentiable functions on $(0, 1)$ and $(-a, a)$, respectively.*

(iii) *There are positive constants C_1, and C_2, such that $C_1 < f_X(x)$ and $\boldsymbol{E}[|p(X, Y)|] < C_2$.*

Also let:

$$Y_n^*(x) = \frac{1}{n} \sum_{i=1}^n p(x_i, y_i) \frac{1}{h_n} K\left(\frac{x_i - x}{h_n}\right) = \int p(z, y) \frac{1}{h_n} K\left(\frac{z - x}{h_n}\right) dF_n(z, y),$$

and $S(x)$ be $G(x) f_X(x)$. If

$$Y_n(x) = \frac{Y_n^*(x)}{\sqrt{\frac{S(x)}{nh_n}}}$$

then

$$\lim_{n \to \infty} \boldsymbol{P}\{\|Y_n\| \le \theta\} = \lim_{n \to \infty} \boldsymbol{P}\{\|Z_n^*\| \le \theta + o(1)\},$$

where Z_n^* is the process defined in (2.6).

PROOF OF PROPOSITION A.5. Given that

$$\boldsymbol{E}[Y_n(x)] = \sqrt{\frac{nh_n}{S(x)}} \int p(z, y) \frac{1}{h_n} K\left(\frac{z - x}{h_n}\right) dF(z, y)$$

we can write:

$$Y_n^0(x) = Y_n(x) - \boldsymbol{E}[Y_n(x)] = \frac{1}{\sqrt{h_n S(x)}} \int p(z, y) K\left(\frac{z - x}{h_n}\right) dZ_n(z, y)$$

where $Z_n(\cdot)$ is the empirical process: $\sqrt{n}(F_n(\cdot) - F(\cdot))$. Let us now define the following processes, which play important roles in the proof:

$$Y_n^1(x) = \frac{1}{\sqrt{h_n S(x)}} \int p(z, y) K\left(\frac{z - x}{h_n}\right) dB_n(H(z, y)),$$

$$Y_n^2(x) = \frac{1}{\sqrt{h_n S(x)}} \int p(z,y) K\left(\frac{z-x}{h_n}\right) dW_n(H(z,y)),$$
$$Y_n^3(x) = \frac{1}{\sqrt{h_n S(x)}} \int \sqrt{S(z)} K\left(\frac{z-x}{h_n}\right) dW(z),$$

where the integration domain is $[0,1] \times R$ for the first two and $[0,1]$ for the third. Also note that:

- $H(z,y) = (F_X(z), F_{Y|X}(y \mid z))$, is the so-called Rosenblatt transformation.

- $B_n(H(z,y))$ is an appropriate succession of Brownian Bridges, defined in a suitable probability space, that approximates uniformly and almost surely to $Z_n(z,y)$; see Tusnady (1997), or Härdle (1989).

- $W_n(z,y)$ is a bivariate Brownian Motion, such that $B_n(z,y) = W_n(z,y) - yzW_n(1,1)$, where $W_n(\cdot,\cdot) = \frac{1}{\sqrt{n}} W(n\cdot, n\cdot)$, with W denoting a standard bidimensional Wiener process. The same notation has been used for the univariate Wiener process $W(z)$ that appears in Y_n^3.

Now $\|Y_n^0 - Y_n^1\|$ can be bound by a quantity of $O(\frac{1}{\sqrt{h_n}} \|Z_n - B_n\|)$ by means of the integration by parts formula, together with the fact that $p(z,y)K(\frac{z-x}{h_n})$ has bounded variation, and that it vanishes at the boundary. Therefore, and because of the way we have chosen the process $B_n(\cdot)$, we have that:

$$\|Y_n^0 - Y_n^1\| = O\left(\sqrt{\frac{\log^2 n}{n h_n}}\right) \quad \text{a.s.} \tag{A.8}$$

Given the relationship between $B_n(\cdot)$ and $W_n(\cdot)$:

$$Y_n^1(x) - Y_n^2(x) = \frac{W_n(1,1)}{\sqrt{h_n S(x)}} \int p(z,y) K\left(\frac{z-x}{h_n}\right) f_{XY}(z,y) dz dy \tag{A.9}$$

because $f_{XY}(z,y)$ is the jacobian of H. As a consequence of $|W_n(1,1)| = O(\sqrt{\log n})$ a.s. and p having finite expectation, we have that:

$$\|Y_n^1 - Y_n^2\| = O(\sqrt{h_n \log n}) \quad \text{a.s.} \tag{A.10}$$

Y_n^2 and Y_n^3 are related given that they are Gaussian processes that share mean value and the following covariance function:

$$\begin{aligned}&\mathbf{Cov}[Y_n^2(x+t), Y_n^2(x)] \\ &= \frac{1}{\sqrt{h_n^2 S(x+t) S(x)}} \int p^2(z,y) K\left(\frac{z-(x+t)}{h_n}\right) K\left(\frac{z-x}{h_n}\right) f_{XY}(z,y) dz dy\end{aligned}$$

which is exactly the covariance of $Y_n^3(x)$. Consequently:

$$Y_n^2(x) \stackrel{\mathcal{D}}{=} Y_n^3(x). \tag{A.11}$$

It remains to show that $Y_n^3(x)$ is not far away, in the supremum norm, from $Z_n^*(x)$. However, as a consequence of:

$$Y_n^3(x) - Z_n^*(x) = \frac{1}{\sqrt{h_n}} \int \left(\sqrt{\frac{S(z)}{S(x)}} - 1\right) K\left(\frac{z-x}{h_n}\right) dW(z)$$

$$= \sqrt{h_n} \int \left(\sqrt{\frac{S(x+uh_n)}{S(x)}} - 1 \right) K(u) dW(x+uh_n),$$

using integration by parts, together with the finite support of K, condition (ii) and that $|W(x+uh_n)| = O(\sqrt{\log n})$ a.s., we obtain:

$$\|Y_n^3 - Z_n^*\| = O(\sqrt{h_n \log n}) \quad \text{a.s.} \tag{A.12}$$

In summary, what we have obtained is that, in a suitable probability space, the paths of $Y_n^1(x)$, $Y_n^2(x)$, and $Y_n^3(x)$ are uniformly and almost surely close, with the same holding for $Y_n^3(x)$ and $Z_n^*(x)$. Thus:

$$\begin{aligned} \boldsymbol{P}\{\|Y_n^0\| \leq A\} &= \boldsymbol{P}\{\|Y_n^2\| \leq A + O(\gamma_{1n})\} \\ &= \boldsymbol{P}\{\|Y_n^3\| \leq A + O(\gamma_{1n})\} = \boldsymbol{P}\{\|Z_n^*\| \leq A + O(\gamma_{1n} + \gamma_{2n})\} \end{aligned}$$

where $\gamma_{1n} = \sqrt{\frac{\log^2 n}{nh_n}} + \sqrt{h_n \log n}$ and $\gamma_{2n} = \sqrt{h_n \log n}$. Therefore, the result is proved. □

The application of this result to length biased data leads to the Proof of Theorem 2.2.

Proof of Theorem 2.2. As a consequence of equation (2.5), and in order to use the previous proposition, the process we have to analyze is:

$$\frac{1}{nh_n} \sum_{i=1}^{n} \left(\frac{y_i - m(x_i)}{y_i} \right) K\left(\frac{x_i - x}{h_n} \right),$$

where the function $p(x,y)$ is now $(\frac{y-m(x)}{y})$.

As required in Proposition A.5, h_n is a positive quantity of order $O(n^{-\eta})$ with $\eta > 0$ because of assumptions A2. Function p is also a bounded variation function such that $\boldsymbol{E}^w[p^2(X,Y) \mid X = x]$ is twice continuous differentiable given of A4, and $\boldsymbol{E}^w[|p(X,Y)|]$ is bounded as a consequence of $1/Y$ and $m(x)$ being bounded (assumptions A5). K is twice continuous differentiable because of A1. Finally, note that the distribution of the random variable, whose sample appears in assumptions (i) in Proposition A.5 is, in this case, F^w, with f_X^w being the marginal density for X, bounded from below because of assumptions A5.

Hence, all the assumptions in Proposition A.5 are satisfied and the proof is complete. □

References

Ahmad, I. A. (1995). On multivariate kernel estimation for samples from weighted distributions, *Statistics and Probability Letters*, **22**, 121–129.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *The Annals of Statistics*, **1**, 1071–1095.

Cristóbal, J. A. and Alcalá, J. T. (2000). Nonparametric regression estimators for length biased data, *Journal of Statistical Planning and Inference*, **89**, 145–168.

Cristóbal, J. A. and Alcalá, J. T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data, *Test*, **10**, 309–332.

Diebolt, J. (1995). A nonparametric test for the regression function: Asymptotic theory, *Journal of Statistical Planning and Inference*, **44**, 1–17.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression, *Journal of the American Statistical Association*, **88**, 1287–1301.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*, Chapman and Hall, New York.

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

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression, *Journal of Multivariate Analysis*, **27**, 228–254.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers, *Journal of Multivariate Analysis*, **29**, 163–179.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *The Annals of Statistics*, **19**, 778–796.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression, *Journal of Econometrics*, **81**, 223–242.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates, *Journal of Multivariate Analysis*, **12**, 402–414.

Knafl, G., Sacks, J. and Ylvisaker, D. (1985). Confidence bands for regression functions, *Journal of the American Statistical Association*, **80**, 682–692.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates, *Journal of the Royal Statistical Society. Series B. Methodological*, **60**, 797–811.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression, *Journal of Nonparametric Statistics*, **9**, 307–333.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families, *Biometrics*, **34**, 179–189.

Patil, G. P., Rao, C. R. and Zelen, M. (1988). Weighted distributions, *Encyclopedia of Statistical Sciences* (eds. S. Kotz and N. L. Johnson), **9**, 565–571, Wiley, New York.

Rao, C. R. (1997). *Statistic and True. Putting Chance to Work*, 2nd ed., World Scientific Publishing, Singapore.

Sköld, M. (1999). Kernel regression in the presence of size-bias, *Journal of Nonparametric Statistics*, **12**, 41–51.

Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing, *The Annals of Statistics*, **22**, 1328–1345.

Tusnady, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case, *Periodica Mathematica Hungarica. Journal of the János Bolyai Mathematical Society*, **8**, 53–55.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression, *Journal of the Royal Statistical Society. Series B. Methodological*, **60**, 797–811.

Wu, C. O. (2000). Local polynomial regression with selection biased data, *Statistica Sinica*, **10**, 789–817.