Ann Inst Stat Math (2008) 60:519–543
DOI 10.1007/s10463-006-0111-3

A bootstrap approach to model checking for linear models under length-biased data

J. L. Ojeda · J. A. Cristóbal · J. T. Alcalá

Received: 2 May 2005 / Revised: 15 May 2006 / Published online: 31 January 2007

Abstract In this paper, we propose two bootstrap-based model checking tests for a parametric linear model when data are affected by length-bias. These tests are based on the measure of the discrepancy between nonparametric and parametric estimators for the regression function when the data are drawn under a length-biased mechanism. We consider two different discrepancy measures: the supremum and the integral of the quadratic difference between the parametric and nonparametric estimators.

Keywords Bootstrap · Length-biased data · Model checking · Lack-of-fit test · Local linear estimator

1 Introduction

Length-biased data appear naturally in many fields of research where direct observation of the random phenomena of interest is not possible or is difficult. For example, when studying wildlife populations, larger individuals or units are more likely to be sighted, and hence they are more likely to be registered in a sample. In some other situations, as is common in econometric and epide-

J. L. Ojeda (✉) · J. A. Cristóbal · J. T. Alcalá
Department of Statistical Methods, Fac. de Ciencias (Edf. Matemáticas),
University of Zaragoza, Pedro Cerbuna, num. 12., 50009 Zaragoza, Spain
e-mail: jojeda@unizar.es

J. A. Cristóbal
e-mail: cristo@unizar.es

J. T. Alcalá
e-mail: jtalcala@unizar.es

miological contexts, when time is involved in a study, length-biased sampling is related to the underlying renewal process driving the phenomena. We can clearly see that, in those circumstances, observations cannot always be recorded in such a way that they have an equal chance of appearing in the sample; rather, the probability of an individual or an event being selected is proportional to its size or duration. In Patil and Rao (1978), Patil et al. (1988), Rao (1997) and Cristóbal and Alcalá (2001) we can find a wide range of examples illustrating this and other sampling biases. It is worth mentioning that the sampling bias in the response variable not only causes a distortion in the determination of the true regression function, but also affects the variance estimation, see Vardi (1982).

Whereas the problem of linear regression under usual data observation assumptions has been extensively studied in the literature, the development of methods to cope with biased data has mainly been carried out in the last half century, and much of the theory is devoted to censored or truncated data. While censored and truncated data are extreme cases of biased observations (some individuals are unobservable or partially observable), most of the problems they exhibit are shared by other kinds of biased data, see for example Quesenberry and Jewell (1986) and the references therein. Although length-bias is not an extreme situation, it should be pointed out that, as happens with truncation and censoring, standard parametric estimation methods are not suitable, or cannot properly be applied. In this regard, both the parametric and nonparametric methods we propose in order to estimate the regression function (i.e.: least squares and local least squares estimators) are based on compensating the effect that the length-bias produces in the observations.

Smoothing methods have attracted a lot of attention during the last few decades. While computer development has made possible the use of computer-intensive methods and smoothing methods, what really makes these techniques appealing is their ability to highlight data structure almost without any other assumption being made about the data (see, for example, Fan and Gijbels 1996). In particular, the lack of any assumption about the functional form makes local polynomial estimates suitable for overcoming the problem of lack of specification in the parametric model (see Kozek 1990; Härdle and Mammen 1993; Hart 1997) and, therefore, they avoid the drawbacks caused by the poor performance of the power of the F-test.

When observed data are affected by lengthéd-bias, the problem of variance misestimation in a linear model can be even worse. A possible solution to this well known drawback consists of comparing both residuals, that is to say, the parametric regression residuals and the local polynomial regression residuals, in order to keep track of the deviation under different alternative hypotheses. This approach has been suggested in the works of Cox et al. (1988), Kozek (1990), Härdle and Mammen (1993) and Alcalá et al. (1999) amongst others, and is the one we adopt here. As our data are affected by length-bias, we will follow the approaches proposed by Cristóbal and Alcalá (2000) or Wu (2000) to provide the appropriate estimators, while the distributional behavior of our statistics will be based on the wild bootstrap technique, see Wu (1986) and

Härdle and Mammen (1993). As is mentioned in Delgado and González Manteiga (2001), bootstrap methodology may help to overcome difficulties in the complex asymptotic analysis and could lead to an improvement on the asymptotic convergence rates, see Hall (1991).

In this work, the main issue is how to obtain a view of the distributional behavior of both parametric and local polynomial estimators, see Bickel and Freedman (1981) and Freedman (1981). In the particular case of the regression problem, it is crucial that the bootstrap procedure resembles the structure exhibited by the residuals, see Wu (1986). The approach we follow in order to validate the proposed tests consists of proving that the statistics used to carry out these tests and their bootstrap counterparts are based on some gaussian process whose stochastic behavior is the same, and hence leading to statistics with the same asymptotic distribution.

Section 2 of this paper introduces the linear models and the assumptions required. Section 3 is devoted to the properties of the proposed estimators and statistics, in particular to their strong uniform representation in terms of an appropriate gaussian process. In Sect. 4 we derive the bootstrap tests proving their consistency. Finally, in Sect. 5 we present a simulation study of the behavior of these tests.

2 The model

We assume throughout the paper that (X, Y) is a two-dimensional random variable with distribution function F and density function $f_{XY}(x, y)$, such that $Y > C > 0$ and $X \in [0, 1]$ with probability 1. The regression function $m(x)$ is then given by:

$$m(x) = \mathbf{E}[Y|X = x].$$

In some situations, $m(x)$ can be supposed to be a linear combination of given functions g_j:

$$m(x) = \mathbf{g}(x)^{\mathrm{T}}\boldsymbol{\beta} = \sum_{j=1}^{k} \beta_j g_j(x), \tag{1}$$

where $\boldsymbol{\beta}$ is the vector of linear combination coefficients $(\beta_1, \ldots, \beta_k) \in \Omega$, a compact in $\mathbf{R}^k$, and $\mathbf{g}(x) = (g_1(x), \ldots, g_k(x))$, a column vector of functions. In this way, we can define a class of linear models $\mathcal{M}_0$ as

$$\mathcal{M}_0 = \left\{ \sum_{j=1}^{k} \beta_j g_j(x) \ : \ (\beta_1, \ldots, \beta_k) \in \Omega \subset \mathbf{R}^k \right\}. \tag{2}$$

Hence, if for example $g_j(x) = x^{j-1}$, the class of the polynomial regressions of degree $k-1$ is considered. Thus depending on g_j, we can deal with a broad class of different parametric functions. Therefore, provided that g_j for $j = 1, \ldots, k$

are suitable for representing m (i.e. $m \in \mathcal{M}_0$), we have to determine the values $\boldsymbol{\beta}_0$ such that $m(x) = \sum_{j=1}^{k} \beta_{0j} g_j(x)$.

The problem that will be addressed is how to test the adequacy of such a model when the observations are affected by length-bias. More precisely, we will address the following hypothesis test:

$$H_0 : m \in \mathcal{M}_0 \quad \text{vs.} \quad H_1 : m \notin \mathcal{M}_0 \tag{3}$$

for a given $\mathcal{M}_0$, when the data are affected by length-bias. This will be achieved by means of two different discrepancy measures: the integrated squared difference, a kind of weighted L_2 norm, and the supremum norm, a kind of L_∞ distance.

While $m(x)$ depends on the random phenomena driven by (X, Y), as a consequence of the length-bias sampling we cannot observe this variable directly and, therefore, our sample $(x_1, y_1), \ldots, (x_n, y_n)$ is an i.i.d sample from a random variable with distribution F^w whose density is given by:

$$dF^w(x, y) = f^w_{XY}(x, y)\, \mathrm{d}x\mathrm{d}y = \frac{y f_{XY}(x, y)}{\mu_Y}\, \mathrm{d}x\mathrm{d}y, \tag{4}$$

where $\mu_Y = \int y f_{XY}(x, y) \mathrm{d}x\mathrm{d}y$. This is precisely the meaning of length-bias in the response, namely, the probability of the observation (x, y) is proportional to y. Let us denote by $\mathbf{E}^w[\cdot]$ and $\mathbf{Var}^w[\cdot]$ the mean and variance respectively for the observed data, i.e.: computed with density f^w_{XY}, in order to distinguish them from $\mathbf{E}[\cdot]$ and $\mathbf{Var}[\cdot]$, which are defined from the unobserved random variable (X, Y) (i.e.: computed with f_{XY}). Therefore:

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

and as $c(x)$ is the conditional coefficient of variation, a direct application of the standard estimation techniques will lead to inconsistent estimators. Note also that $f^w_X(x)$, the X marginal density for the biased distribution, is precisely $\mu_Y^{-1} m(x) f_X(x)$ because of the length-bias.

One of the most commonly used estimators in the literature to obtain a value for $\boldsymbol{\beta}$ in (1) is the least square method:

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} w_i \left(y_i - \mathbf{g}(x_i)^{\mathrm{T}} \boldsymbol{\beta}\right)^2 \tag{5}$$

where w_i are suitable weights. Under appropriate assumptions for $\mathcal{M}_0$ and (X, Y), there exists a $\hat{\boldsymbol{\beta}}_n$ such that $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_0 + o(1)$ almost surely; thus, $\hat{\boldsymbol{\beta}}_n$ is strongly consistent for $\boldsymbol{\beta}_0$. Therefore:

$$\tilde{m}_n(x) = \mathbf{g}(x)^{\mathrm{T}} \hat{\boldsymbol{\beta}}_n$$

is a strongly consistent estimator for $m(x)$ if $\mathcal{M}_0$ is the appropriate model (i.e. $m \in \mathcal{M}_0$). Note that when the model is misspecified, $\mathbf{g}(x)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n$ does not agree with $m(x)$ in (1), and then $\tilde{m}_n(x)$ converges to $\tilde{m}(x) = \mathbf{g}(x)^{\mathrm{T}}\boldsymbol{\beta} \neq m(x)$, that is to say, the closest function to m within the class $\mathcal{M}_0$ in the least squares sense.

In order to carry out the least squares estimator for the regression function, we require:

B1 The functions g_j, $i = 1, \ldots, k$ are twice continuous differentiable in $(0, 1)$.
B2 The Matrix

$$\mathbf{L} = \mathbf{E}^w\left[\frac{\mu_Y}{Y}\mathbf{g}(X)\mathbf{g}(X)^{\mathrm{T}}\right]$$

is not singular.

The local polynomial estimator of order p for $m(x)$ is given by $\hat{m}_n(x) = \hat{\alpha}_0$, where $\hat{\alpha}_0, \ldots, \hat{\alpha}_p$ are the solutions to the following weighted least square problem:

$$\min_{\alpha_0,\ldots,\alpha_p} \sum_{i=1}^{n} w_i\big(y_i - \alpha_0 - \cdots - \alpha_p(x_i - x)^p\big)^2 K_h(x_i - x), \tag{6}$$

where $K_h(u)$ is $h^{-1}K\left(u h^{-1}\right)$ for a given kernel function K and bandwidth h. Note in this expression the dependence of $\alpha_0, \ldots, \alpha_p$ on x. The local polynomial estimator uses the weight $K_h(x_i - x)$ in every x to estimate the value $m(x)$. Therefore, using a suitable function K we are penalizing observations with x_i distant from x and, thus, considering the local behavior. As a consequence, misspecification in the model does not affect this estimator, which is adapted to the functional form of m because of the local estimation. So, a comparison between $\tilde{m}_n$ and $\hat{m}_n$ can reveal any possible misspecification in the model. More precisely, if the difference between $\tilde{m}_n$ and $\hat{m}_n$ is not statistically significant, then we can accept that the regression function m belongs to $\mathcal{M}_0$; in other words, the class of functions $\mathcal{M}_0$ is suitable for representing m.

As in the parametric case, we will require some additional assumptions in order properly to carry out the local linear estimation ($p = 1$):

A1 $m(x)$, $f_X(x)$ and $v^w(x)$ are twice continuously differentiable in $(0, 1)$ and there exists a constant C such that $0 < C < f_X(x)$, $v^w(x)$ in $[0, 1]$, where

$$v^w(x) = \mathbf{E}^w\left[\left(\frac{Y - m(X)}{Y}\right)^2 \middle| X = x\right]$$

L1 The kernel K is an even function with support $[-a, a]$, which is twice continuously differentiable in the interior of the support, decreasing in $[0, a]$ with $K(a) = 0$, and such that $\int K(u)\,\mathrm{d}u = 1$.
L2 The bandwidth h_n used in the local linear estimation is an $O\left(n^{-1/5}\right)$ quantity.

We also denote $\int u^j K(u)\,\mathrm{d}u$ by μ_j, and $\int K^{(j-1)}(x-u)K(u)\,\mathrm{d}u$ by $K^{(j)}(x)$, where $K^{(1)} = K$, while δ_{nh_n} stands for $\sqrt{\log n/(nh_n)}$.

Although it is common to use a quadratic discrepancy measure to perform model checking, it is also sometimes useful, or even desirable, to consider another measure based on the supremum norm. Note the different qualitative behavior of these measurements: local differences determine the supremum distance behavior, while quadratic distance is a global measurement. In the following sections, we will derive the asymptotic distribution for these two discrepancy measures when observations are length-biased. Therefore, we will be able to use the following statistics

$$K_n^\infty = \sup_{x\in[0,1]} \left|\mu_Y^{-1}\sqrt{nh_n} f_X(x)\big(\hat{m}_n(x) - \tilde{m}_n(x)\big)\right|,$$
$$W_n^2 = \mu_Y^{-2}\int_{[0,1]} nh_n f_X(x)^2 \big(\hat{m}_n(x) - \tilde{m}_n(x)\big)^2\,\mathrm{d}x,$$

to perform the test given in (3), where $nh_n f_X(x)^2 \mu_Y^{-2}$ will be estimated from the remark following Proposition 3.

In the next section, we will derive the statistics we use to handle the length-bias in data when using both the least square estimator over the class $\mathcal{M}_0$ and the local polynomial estimator. We will also study their main properties from the point of view of our purposes: the consistency and the strong uniform representation in terms of appropriate gaussian processes that will characterize their stochastic behaviour. Thereafter, we will derive the bootstrap tests proving their consistency.

3 Supremum and quadratic statistics

As can be seen from Eq. (4), the reciprocal of the responses can be used to compensate the length–bias, see Cristóbal and Alcalá (2000). Hence, we can use the reciprocal of each observation as a weight in (5), obtaining the following optimization problem:

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{1}{y_i}\Big(y_i - \mathbf{g}(x_i)^{\mathrm{T}}\boldsymbol{\beta}\Big)^2, \tag{7}$$

from which we obtain:

$$\tilde{\boldsymbol{\beta}}_n = \left(\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{Y},$$

where $\mathbf{Y}$ is the column vector with observations y_i, $\mathbf{G}$ is a $n \times k$ matrix with entries $g_j(x_i)$, and $\mathbf{B}$ is given by $\mathrm{diag}(y_1^{-1},\ldots,y_n^{-1})$. Besides the fact that the reciprocal of the responses compensates the effect of length–bias present in the data, it is worth mentioning that the nonparametric maximum likelihood estimator of the distribution function F under length–biased observations is proportional to $1/y_i$ at every given observation (x_i, y_i), see Cox (1969) and Vardi (1982). Under the assumptions made in the previous section, it can be proved that this estimator is strongly consistent.

Proposition 1 *If assumptions B1 and B2 are fulfilled and if the regression function m belongs to the class of functions $\mathcal{M}_0$, then $\tilde{\boldsymbol{\beta}}_n$ is a strongly consistent estimator of $\boldsymbol{\beta}_0$, and, therefore:*

$$\tilde{m}_n(x) = \mathbf{g}(x)^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n = m(x) + O\left(\sqrt{\frac{\log\log n}{n}}\right)$$

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

It is interesting to note that this last statement means that there exists a positive constant C, such that:

$$\lim\sup_{n\to\infty} \sqrt{\frac{n}{\log\ \log\ n}} \sup_{x\in[0,1]} \left|\tilde{m}_n(x) - m(x)\right| \leq C \quad \text{a.s.}$$

That is to say, with probability one, the supremum on $[0, 1]$ of the parametric error process $\tilde{m}_n(x) - m(x)$ decreases to zero as fast as $C\sqrt{\log\ \log\ n/n}$. It is also worth mentioning that from the point of view of the test we are considering this is precisely the important point. As we will see, this convergence rate to zero is faster than the rate of the local linear error process $\left(\hat{m}_n(x) - m(x)\right)$. In this way, we are able to use the stochastic behavior of the local linear error process to address the stochastic behavior of the difference between the least square estimator and the local linear estimator.

In the case of the local linear estimator, the use of the reciprocal of the responses to compensate the length–bias in Eq. (6) leads to the following weighted least squares problem:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \frac{1}{y_i}\left(y_i - \alpha_0 - \alpha_1(x_i - x)\right)^2 K_h(x_i - x)$$

The solution for α_0 in this estimation equation for every x can be written in the following form

$$\hat{m}_n(x) = \hat{\alpha}_0 = \sum_{i=1}^{n} \frac{w^w_{ih_n}(x)}{\sum_{i=1}^{n} w^w_{ih_n}(x)} y_i \tag{8}$$

where

$$w^w_{ih}(x) = \frac{1}{y_i}\left(s^w_2(x;h)K\left(\frac{x_i - x}{h}\right) - s^w_1(x;h)K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)\right),$$

and where

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

While in the parametric case we were only concerned with the strong uniform consistency, in the case of the nonparametric estimator we are mainly interested in its stochastic behavior. In this regard, and from our point of view, the strong and uniform representation of the local linear error process by means of a suitable sequence of gaussian processes will be an invaluable tool. In order to achieve the desired strong uniform approximation, we recall from Cristóbal et al. (2004) that under our assumptions:

$$\hat{m}_n(x) = m(x) + \frac{h_n^2}{2}m''(x) + \frac{\mu_Y\, e_0^w(x;h_n)}{f_X(x)} + O\left(h_n\delta_{nh_n}\right) \tag{9}$$

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

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

As we can see, the process $e_j^w(x;h_n)$ comprises the main stochastic features of the nonparametric error, and can be strongly and uniformly represented by a sequence of gaussian processes of known covariance function. This, together with the fact that $e_j^w(x;h_n)$ is an $O\left(\delta_{nh_n}\right)$ quantity uniformly in $[0,1]$ and almost surely, leads not only to the strong uniform consistency of the nonparametric estimator, but also to the strong uniform approximation of the nonparametric error process by means of a suitable sequence of known gaussian processes. Therefore, in this way we have completely characterized the stochastic behavior of the nonparametric estimator. Note also that Eq. (9) shows that the local linear estimator in this framework where data are length–biased shares the main properties of the ordinary local linear estimator with respect to the asymptotic rate of convergence of both the bias and variance terms. In this regard, it should be mentioned that, with the corresponding changes in the variance asymptotic expression, MSE and MISE bandwidth selectors are obtained in the same way, see Cristóbal and Alcalá (2000), while a cross–validation bandwidth selector is also proposed in Wu (2000).

Theorem 1 *If assumptions L1, L2, A1 are satisfied, then*

$$\hat{m}_n(x) = m(x) + O\left(h_n^2 + \delta_{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely. Moreover, we have the following strong uniform approximation:*

$$\hat{m}_n(x) = m(x) + \frac{\mu_Y}{\sqrt{nh_n} f_X(x)} Z_n^w(x) + O\left(h_n^2 + \delta_{nh_n}^2 + h_n \delta_{nh_n}\right) + o\left(\frac{\log^2 n}{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely, where* $Z_n^w(x)$ *is a sequence of second-order gaussian processes with null expectation and covariance function given by:*

$$\mathbf{Cov}\left[Z_n^w(s), Z_n^w(t)\right] = \mathbf{E}^w\left[\frac{1}{h_n} K\left(\frac{s-X}{h_n}\right) K\left(\frac{t-X}{h_n}\right)\left(\frac{Y-m(X)}{Y}\right)^2\right].$$

As $\left(\tilde{m}_n(x) - m(x)\right)$ decreases to zero faster than $\left(\hat{m}_n(x) - m(x)\right)$, because $\sqrt{\log\ \log\ n/n} = o\left(h_n^2 + \delta_{nh_n}\right)$, we can use the stochastic behavior of this last process to obtain that of the difference between the parametric and local linear estimators defining our test statistics. As a consequence, from these expressions, it is now possible to derive the distributions of K_n^∞ and W_n^2.

Theorem 2 *Under the assumptions made in Proposition 1 and Theorem 1, if* H_0 *is true, then:*

$$K_n^\infty = \sup_{x\in[0,1]} \left|Z_n^w(x)\right| + O\left(\sqrt{h_n \log\ \log\ n}\right),$$
$$W_n^2 = \int_{[0,1]} Z_n^w(x)^2\, \mathrm{d}x + O\left(\sqrt{h_n\ \log\ n \log\ \log\ n}\right),$$

almost surely.

Here, it is worth mentioning that although these results are useful for obtaining asymptotics, they suffer from a poor convergence rate. Because of such a slow convergence, and because we have to work with finite samples, the bootstrap can help us to avoid such a poor performance. In the next section, we present a bootstrap scheme that will make it possible to use the bootstrap in this setting, where data are length–biased in the response.

The asymptotic distribution for the supremum statistic K_n^∞ can be addressed in the same way as in Cristóbal et al. (2004), where confidence bands for the regression function in this setting are given. In the case of the quadratic statistic W_n^2, it is possible to use the results in de Jong (1987) in a similar manner to Härdle and Mammen (1993).

4 Bootstrap statistics

In the previous section, we have presented several tools that allow us to perform testing hypotheses using two different criteria. While these tools enable us to obtain asymptotic distributions for both statistics, and to prove the consistency of the testing procedure, they suffer, as has been earlier mentioned, from a very poor convergence rate. In the particular case of the statistic based on the supremum, it is well known that the convergence rate is too slow (see, for example, Leadbetter et al. 1983). In this regard, as is shown in Hall (1991) in the context of local density estimation, the bootstrap may help to overcome these difficulties.

As the subject of our study is data affected by length–bias, where the observations do not come directly from the random phenomena driven by F, but from the random variable with distribution F^w, the bootstrap works by mimicking the way data behaves, and this imposes new difficulties when trying to implement bootstrap ideas. We will show that for the estimators proposed in the previous section we can use a bootstrap scheme that is similar to the one used in the case of ordinary unbiased data.

Note first that our main interest lies in the distributional behavior of the statistics K_n^∞ and W_n^2. Moreover, as can be seen in the previous section, the stochastic aspects of these quantities are determined by the compensated residuals of the nonparametric regression, see (9). Thus, to mimic the random behavior of K_n^∞ and W_n^2, we should be able to model the stochastic behavior of these compensated residuals.

To this end, let us consider the regression function estimators defined in Sect. 3, namely $\hat{m}_n(x)$ and $\tilde{m}_n(x)$, both adapted to length–biased data and, further, let $\hat{\epsilon}_i$ be the local linear residuals:

$$\hat{\epsilon}_i = y_i - \hat{m}_n(x_i) \quad i = 1, \ldots, n. \tag{10}$$

The bootstrap sample is defined in the following manner:

$$x_i^* = x_i;\ y_i^* = \tilde{m}_n\left(x_i^*\right) + \hat{\epsilon}_i^*;\ \hat{\epsilon}_i^* = \hat{\epsilon}_i\gamma_i \quad i = 1, \ldots, n; \tag{11}$$

where γ_i, the wild bootstrap random variable, see Eq. (17) in the appendix, is independent of $\hat{\epsilon}_i$ and can take only two values, having null expectation, and variance and third moment equal to 1; see, for example Härdle and Mammen (1993) and the references therein. It is not difficult to see that γ_i changes the sign of the local linear residuals $\hat{\epsilon}_i$ randomly, making minor changes in their absolute value. However, it does not change its main stochastic properties, namely the first, second and third moments. In this way, and because the nonparametric estimator is a consistent estimator for the regression function, it is clear how this bootstrap sample mimics the stochastic behavior of the real sample.

In addition to all these considerations, and from a theoretical perspective, it is also interesting to point out that the use of the bootstrap random variable γ_i under these conditions makes the bootstrap sample lie in a probability space that enlarges the probability space where the given sample lies. Hence, we can think of both samples as if they were in this enlarged probability space.

The bootstrap counterpart of $\tilde{\boldsymbol{\beta}}_n$ is given by the following expression:

$$\tilde{\boldsymbol{\beta}}_n^* = \left(\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{Y}^*, \tag{12}$$

where $\mathbf{Y}^*$ is a column vector with y_i^* entries and therefore $\tilde{m}_n^*(x) = \mathbf{g}(x)^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n^*$. It is worth noting that, in this case, and as a consequence of the expectation of $\hat{\epsilon}_i^*$ being null, it is sure that $\tilde{m}_n(x)$ belongs to the class of functions $\mathcal{M}_0$. Moreover, the bootstrap counterpart of $\hat{m}_n(x)$ is given by

$$\hat{m}_n^*(x) = \sum_{i=1}^{n}\frac{w_{ih_n}^{w}(x)}{\sum_{i=1}^{n} w_{ih_n}^{w}(x)}y_i^* \tag{13}$$

where, as a consequence of $x_i^* = x_i$, we can use the weights $w_{ih_n}^{w}(x)$ that were defined in (8) to compensate the effect of the length–bias by means of the reciprocal. Once the bootstrap scheme and the bootstrap estimators have been defined, we can plug both into the definition of K_n^{∞} and W_n^2, obtaining, in this way, a bootstrap observation of the test statistics by means of

$$K_n^{\infty^*} = \sup_{x\in[0,1]}\left|\mu_Y^{-1}\sqrt{nh_n}f_X(x)\left(\hat{m}_n^*(x) - \tilde{m}_n^*(x)\right)\right|,$$
$$W_n^{2^*} = \mu_Y^{-2}\int_{[0,1]} nh_n f_X(x)^2\left(\hat{m}_n^*(x) - \tilde{m}_n^*(x)\right)^2 \mathrm{d}x.$$

Note that, because $x_i^* = x_i$, we estimate $nh_n f_X(x)^2\mu_Y^{-2}$ in the same way as in the non-bootstrap case, see the remark following Proposition 3.

Using these statistics, we can perform the proposed tests in the previous sections with the aid of the bootstrap distribution for them. Let us denote by $K_{n(1-\alpha)}^{\infty^*}$ and $W_{n(1-\alpha)}^{2^*}$ the $K_n^{\infty^*}$ and $W_n^{2^*}$ bootstrap $(1-\alpha)$-quantile, respectively. Thus, in order to test the adequacy of a model for the regression function using the supremum norm, we should reject H_0 at a given confidence level $1-\alpha$ if

$$K_n^{\infty} > K_{n(1-\alpha)}^{\infty^*}.$$

If we use the integrated squared error, then we should reject H_0 at a given confidence level $1-\alpha$ when

$$W_n^2 > W_{n(1-\alpha)}^{2^*}.$$

The consistency of this procedure follows from the results presented below, which prove that the assertions made in the previous section for $\hat{m}_n(x)$ and $\tilde{m}_n(x)$ are also valid for the bootstrap estimators $\hat{m}_n^*(x)$ and $\tilde{m}_n^*(x)$. I should be recalled that, in this case, as we have chosen the bootstrap response with mean value $\tilde{m}_n(x)$, the regression function we are estimating is simply $\tilde{m}_n(x)$, the closest function in $\mathcal{M}_0$ to the true regression function of data $m(x)$. Furthermore, note that, in any event, as both bootstrap estimators are estimating something that belongs to the class $\mathcal{M}_0$, this bootstrap procedure works under the null and alternative hypotheses, just as in the case of ordinary samples; see Härdle and Mammen (1993).

Proposition 2 *If assumptions B1 and B2 are fulfilled, then, the estimator $\tilde{\boldsymbol{\beta}}_n^*$ is a strongly consistent estimator of $\tilde{\boldsymbol{\beta}}_n$, therefore:*

$$\tilde{m}_n^*(x) = \mathbf{g}(x)^{\mathrm{T}} \tilde{\boldsymbol{\beta}}_n^* = \tilde{m}_n(x) + O\left(\sqrt{\frac{\log\ \log\ n}{n}}\right)$$

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

The consistency of the bootstrap nonparametric estimator is proved in the next result, showing that, as happened in the asymptotic case, the bootstrap local linear error process $\left(\hat{m}_n^*(x) - \tilde{m}_n(x)\right)$ decreases to 0 at a slower rate than the bootstrap parametric error process $\left(\tilde{m}_n^*(x) - \tilde{m}_n(x)\right)$. However, the most important fact from our point of view is that the bootstrap local linear error process can be uniformly approximated by a suitable sequence of gaussian processes, whose asymptotic stochastic properties are the same as in the sequence of processes $Z_n^w(x)$.

Theorem 3 *If assumptions L1, L2, A1, and B1 are fulfilled, then*

$$\hat{m}_n^*(x) = \tilde{m}_n(x) + O\left(h_n^2 + \delta_{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely. Moreover, we have the following strong uniform approximation:*

$$\hat{m}_n^*(x) = \tilde{m}_n(x) + \frac{\mu_Y}{\sqrt{nh_n} f_X(x)} Z_n^*(x) + O\left(h_n^2 + \delta_{nh_n}^2 + h_n \delta_{nh_n}\right) + o\left(\frac{\log^2\ n}{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely, where $Z_n^*(x)$ is a second–order gaussian process with null expectation and covariance function given by:*

$$\mathbf{Cov}\left[Z_n^*(s), Z_n^*(t)\right] = \mathbf{E}^w\left[\frac{1}{h_n} K\left(\frac{s-X}{h_n}\right) K\left(\frac{t-X}{h_n}\right)\left(\frac{Y-m(X)}{Y}\right)^2\right].$$

As a consequence of Theorem 3, both bootstrap estimators $\hat{m}_n^*(x)$ and $\tilde{m}_n^*(x)$ behave as their non–bootstrap counterparts $\hat{m}_n(x)$ and $\tilde{m}_n(x)$ regarding the rates of convergence. Moreover, both processes $Z_n(x)$ and $Z_n^*(x)$, are second–order gaussian processes with the same mean and covariance function, which means that $Z_n(x)$ and $Z_n^*(x)$ are stochastically equivalent. Furthermore, in the bootstrap case the estimators $\hat{m}_n^*(x)$ and $\tilde{m}_n^*(x)$ tend to the parametric estimator $\tilde{m}_n(x)$ which, as has been mentioned, ensures the consistency of the tests under both the null and the alternative hypotheses. Finally, using the same arguments given in the previous section, we obtain the stochastic behavior of $K_n^{\infty^*}$ and $W_n^{2^*}$.

Theorem 4 *Under the assumptions made in Proposition 2 and Theorem 3, if H_0 is true, then:*

$$K_n^{\infty^*} = \sup_{x\in[0,1]} \left|Z_n^*(x)\right| + O\left(\sqrt{h_n \log\ \log\ n}\right),$$
$$W_n^2 = \int_{[0,1]} Z_n^*(x)^2\, \mathrm{d}x + O\left(\sqrt{h_n\ \log\ n \log\ \log\ n}\right),$$

almost surely.

Remark 1 Note that both second–order gaussian processes $Z_n(x)$ and $Z_n^*(x)$ are stochastically equivalent, because they are defined in the same probability space that depends on the sample data and the bootstrap sample. Hence, the consistency of the bootstrap procedure follows from Theorems 2 and 4, given that the statistics K_n^{∞}, W_n^2 and their bootstrap versions $K_n^{\infty^*}$ and $W_n^{2^*}$ can be asymptotically written as continuous functionals of $Z_n(x)$ and $Z_n^*(x)$, respectively, which are second–order stochastically equivalent gaussian processes. In this way, the statistics $K_n^{\infty^*}$ and $W_n^{2^*}$ behave as K_n^{∞} and W_n^2 given that the distributional behavior of Z_n^w and Z_n^* is the same.

5 Brief simulation study

In order to obtain an idea of the finite sample behavior of the proposed tests, we have carried out a brief simulation study in this section. We have considered a bivariate random variable (X, Y) defined, as in Cristóbal et al. (2004), in such a way that X is distributed uniformly in $[0, 1]$ and $Y = m(X)(1 + 0.1\epsilon)$, where ϵ is a uniform random variable in $[-\sqrt{3}, \sqrt{3}]$ independent of X, but in this case, and following Härdle and Mammen (1993), we are going to consider a polynomial regression function because our main concern is model checking. Therefore $m(x)$ is defined as:

$$m(x) = g(x) + A\Delta(x) = 2x - x^2 + A\Delta(x). \tag{14}$$

When $A = 0$, m belongs to the following class of functions

$$\mathcal{M}_0 = \left\{ax + bx^2\ :\ a, b \in [-5, 5]\right\},$$

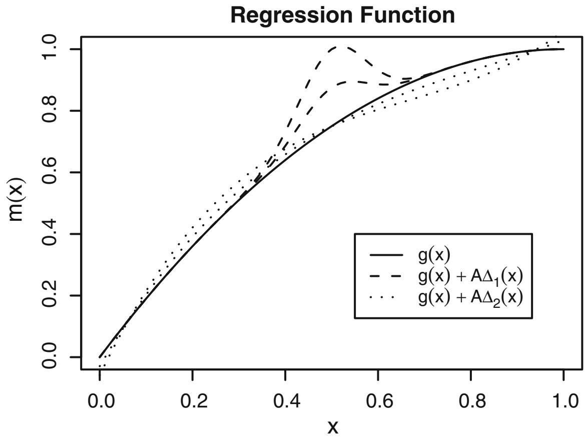

Fig. 1 Regression functions for $A = 0$ (continuous line), and $A = 0.5, 1$ with $\Delta = \Delta_1$ (dashed line), $\Delta = \Delta_2$ (dotted line)

and, allowing $\varDelta$ to be

$$\varDelta_1(x) = \frac{1}{4}\exp\left(-100(x-1/2)^2\right),$$
$$\varDelta_2(x) = 2(x-1/16)(x-1/2)(x-15/16),$$

when $A \neq 0$, we obtain that $m \notin \mathcal{M}_0$. In this way, the term $A\varDelta(x)$ in (14) acts as a perturbation whose intensity depends on A. Using both kinds of function, we can examine the behavior of the proposed tests under two different situations. Note that in the case of $\varDelta_1$, the regression function has an extreme value that is considerably larger than those of $\varDelta_2$; on the other hand $\varDelta_2$ is much flatter, see Fig. 1.

The hypothesis test (3) has been performed using the statistics presented at the end of Sect. 2. In the following tables, the rate of acceptances of the null hypotheses is presented for 500 simulations of each of the different values of the sample size $n = 50, 100, 200$, the confidence level $1 - \alpha = 0.9, 0.95$, and the intensity $A = 0, 0.5, 1$, with both perturbation functions $\varDelta_1$ and $\varDelta_2$. We have considered a bootstrap sample of size $B = 8000$ (Tables 1–3).

The numerical computation of the statistics for every bootstrap sample $\left(x_1^*, y_1^*\right), \ldots, \left(x_n^*, y_n^*\right)$, has been carried out obtaining the values of $\tilde{m}_n^*(x)$ and $\hat{m}_n^*(x)$ for x on the grid $0, 0.01, 0.02, \ldots, 0.99, 1$ using a Cross-Validation bandwidth selector (see Wu, 2000) for h_n. To estimate $nh_nf_X(x)\mu_Y^{-2}$ on that grid we have used the remark following Proposition 3. Next, we have computed the error process $\mu_Y^{-1}\sqrt{nh_n}f_X(x)\left(\hat{m}_n^*(x) - \tilde{m}_n^*(x)\right)$ on that grid, and then, W_n^{2*} by means of the integral of the square of this error process using a simple Riemann

Table 1 Acceptance rate of H_0 for K_n^∞ with Δ_1

α	n	A	K_n^∞ Accep.	α	n	A	K_n^∞ Accep.
0.10	50	0.0	0.898	0.05	50	0.0	0.948
		0.5	0.576			0.5	0.662
		1.0	0.150			1.0	0.186
	100	0.0	0.880		100	0.0	0.950
		0.5	0.302			0.5	0.364
		1.0	0.012			1.0	0.018
	200	0.0	0.854		200	0.0	0.944
		0.5	0.078			0.5	0.098
		1.0	0.002			1.0	0.006

Table 2 Acceptance rate of H_0 for W_n^2 with Δ_1

α	n	A	W_n^2 Accep.	α	n	A	W_n^2 Accep.
0.10	50	0.0	0.876	0.05	50	0.0	0.964
		0.5	0.570			0.5	0.630
		1.0	0.114			1.0	0.132
	100	0.0	0.896		100	0.0	0.942
		0.5	0.220			0.5	0.260
		1.0	0.008			1.0	0.006
	200	0.0	0.856		200	0.0	0.958
		0.5	0.046			0.5	0.056
		1.0	0.002			1.0	0.004

Table 3 Acceptance rate of H_0 for K_n^∞ with Δ_2

α	n	A	K_n^∞ Accep.	α	n	A	K_n^∞ Accep.
0.10	50	0.0	0.884	0.05	50	0.0	0.944
		0.5	0.824			0.5	0.852
		1.0	0.544			1.0	0.690
	100	0.0	0.870		100	0.0	0.940
		0.5	0.622			0.5	0.734
		1.0	0.188			1.0	0.304
	200	0.0	0.870		200	0.0	0.922
		0.5	0.440			0.5	0.572
		1.0	0.048			1.0	0.080

Table 4 Acceptance rate of H_0 for W_n^2 with Δ_2

α	n	A	W_n^2 Accep.	α	n	A	W_n^2 Accep.
0.10	50	0.0	0.894	0.05	50	0.0	0.944
		0.5	0.854			0.5	0.864
		1.0	0.606			1.0	0.704
	100	0.0	0.896		100	0.0	0.936
		0.5	0.710			0.5	0.802
		1.0	0.218			1.0	0.318
	200	0.0	0.882		200	0.0	0.930
		0.5	0.494			0.5	0.588
		1.0	0.068			1.0	0.104

summation formula, and the supremum $K_n^{\infty^*}$ as the maximum of the absolute value of this error process over the grid.

As can be seen in tables 1, 2, 3, 4 in the case $A = 0$, the null hypotheses is accepted in about $100(1-\alpha)\%$ of the simulations in each different situation. When $A \neq 0$, this rate decreases when n increases in a noticeable manner, as was expected. This illustrates, from an empirical perspective, the consistency of the test procedures we have presented.

Acknowledgements The authors would like to express their thanks to the associate editor and the anonymous referee for their helpful comments, observations and suggestions on an earlier version of this paper. This research has been partially supported by Spanish MCyT grant BMF2002-00080 and MEC grant MTM2005-01464.

A Appendix

A.1 The estimators

Proof of Proposition 1 Using a matrix notation, Eq. (7) can be written as

$$\Phi(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{G}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{B}(\mathbf{Y} - \mathbf{G}\boldsymbol{\beta}).$$

The value $\tilde{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ that minimizes this expression is given by

$$\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G}\tilde{\boldsymbol{\beta}}_n = \mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{Y}.$$

Note that $\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G}$ is a matrix whose (j,l)th element is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i}g_j(x_i)g_l(x_i),$$

and that

$$\mathbf{E}^w\left[\frac{1}{Y}g_j(X)g_l(X)\right] = \frac{1}{\mu_Y}\mathbf{E}\left[g_j(X)g_l(X)\right].$$

As all these matrix elements have finite second–order moments, the application of the Law of the Iterated Logarithm gives that

$$\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G} = \frac{1}{\mu_Y}\mathbf{L} + O\left(\sqrt{\frac{\log\ \log\ n}{n}}\right) \tag{15}$$

almost surely, where $\mathbf{L}$ is given in Assumption B2. Hence, for a sufficiently large n, we know $\mathbf{G}^T\mathbf{B}\mathbf{G}$ is a non-singular matrix. Moreover, as a consequence of y_i being $m(x_i) + \epsilon_i$, for $m \in \mathcal{M}_0$ we obtain that $y_i = \mathbf{g}(x_i)^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon_i$, and we can write $\mathbf{Y} = \mathbf{G}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$ for a column vector $\boldsymbol{\epsilon}$ with entries ϵ_i. Therefore:

$$\tilde{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_0 + \left(\mathbf{G}^{\mathrm{T}}\mathbf{B}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{B}\boldsymbol{\epsilon}.$$

Now, as $\mathbf{G}^{\mathrm{T}}\mathbf{B}\boldsymbol{\epsilon}$ is a vector with entries given by

$$\frac{1}{n}\sum_{i=1}^{n} g_j\left(x_i\right)\frac{\epsilon_i}{y_i},$$

the application, once more, of the Law of the Iterated Logarithm to each of these entries means that $\mathbf{G}^{\mathrm{T}}\mathbf{B}\boldsymbol{\epsilon}$ is a matrix whose elements are $O\left(\sqrt{\log\ \log\ n/n}\right)$ quantities almost surely and, hence,

$$\tilde{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_0 + O\left(\sqrt{\frac{\log\ \log\ n}{n}}\right). \qquad \square$$

For the proofs related to the strong uniform consistency of the nonparametric error we will follow those given in Cristóbal et al. (2004). The following result shows the strong uniform convergence of the processes $s_j^w\left(x; h_n\right)$ and $e_j^w\left(x; h_n\right)$, exhibiting their convergence rate.

Proposition 3 *Under assumptions L1, L2, we have that:*

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

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

Proof See Proposition A3 and A4 in Cristóbal et al. (2004). $\square$

Remark 2 As a consequence of the previous result, we obtain that

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

From here we can obtain an estimation of $nh_n f_X\left(x\right)^2 \mu_Y^{-2}$ by means of $\left(nh_n\mu_2\right)^{-1}\sum_{i=1}^{n} w_{ih_n}^w\left(x\right)$.

Now, we tackle the strong uniform approximation of the process $e_0^w\left(x; h_n\right)$ by means of a sequence of second–order gaussian processes, where we will follow part of the proof of Proposition A.5 in Cristóbal et al. (2004).

Proposition 4 *Under assumptions L1 and L2 in Sect. 2:*

$$e_0^w(x;h_n) = \frac{1}{\sqrt{nh_n}} Z_n^w(x) + o\left(\delta_{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely, where* Z_n^w *is a second–order gaussian process with null expectation and covariance function given by:*

$$\mathbf{Cov}\left[Z_n^w(s), Z_n^w(t)\right] = \mathbf{E}^w\left[\frac{1}{h_n}K\left(\frac{s-X}{h_n}\right)K\left(\frac{t-X}{h_n}\right)\left(\frac{Y-m(X)}{Y}\right)^2\right].$$

Proof Let us denote by $Y_n^0(x)$ the following process:

$$\begin{aligned} Y_n^0(x) &= \sqrt{nh_n} e_0^w(x;h_n) \\ &= \sqrt{nh_n}\int\left(\frac{y-m(z)}{y}\right)\frac{1}{h_n}K\left(\frac{z-x}{h_n}\right)\mathrm{d}E_n^w(z,y), \end{aligned}$$

where $E_n^w(\cdot)$ is $\sqrt{n}\big(F_n^w(\cdot) - F^w(\cdot)\big)$, that is to say, the empirical process of the length–biased sample. Using results presented in Tusnády (1977), this empirical process can be approximated uniformly in $\mathbf{R}^2$ and almost surely by means of a suitable sequence of Brownian Motions $B_n(H(z,y))$, in such a way that

$$\left\| E_n^w(\cdot) - B_n(H(\cdot))\right\|_\infty = O\left(\frac{\log^2 n}{\sqrt{n}}\right)$$

almost surely, where H is the so-called Rosenblatt transformation

$$H(z,y) = \left(F_X^w(z), F_{Y|X}^w(y \mid z)\right),$$

with $F_X^w(z)$ and $F_{Y|X}^w(y \mid z)$ being the X^w marginal and $Y^w|X^w$ conditional distributions, respectively. Hence, using integration by parts and as $(y-m(z))y^{-1}K\left(h_n^{-1}(z-x)\right)$ has bounded variation, and vanishes at the boundary, we obtain that

$$\begin{aligned} Y_n^0(x) = &\int\left(\frac{y-m(z)}{y}\right)\frac{1}{h_n}K\left(\frac{z-x}{h_n}\right)\mathrm{d}B_n(H(z,y)) \\ &+ O\left(\frac{\log^2 n}{\sqrt{n}}\right) \end{aligned}$$

uniformly in $[0,1]$ and almost surely. Therefore, and defining $Z_n^w(x)$ as:

$$\int \left(\frac{y-m(z)}{y}\right)\frac{1}{h_n}K\left(\frac{z-x}{h_n}\right)\mathrm{d}B_n(H(z,y))$$

we obtain the first part of the proof.

For the second part, note that if $W_n(u,v)$ is the bivariate Brownian Motion in $[0,1]\times[0,1]$, such that $B_n(u,v) = W_n(u,v) - uvW_n(1,1)$. As a consequence of the identity $(u,v) = H(z,y)$ we obtain that

$$\mathbf{Cov}\left[Z_n^w(s), Z_n^w(t)\right] = \int \frac{1}{h_n}K\left(\frac{s-z}{h_n}\right)K\left(\frac{t-z}{h_n}\right)\left(\frac{y-m(z)}{y}\right)^2 f_{XY}^w(z,y)\,\mathrm{d}z\mathrm{d}y. \qquad \square$$

Proof of Theorem 1 The proof of the strong uniform consistency follows the same reasoning that was given in Theorem 2.1 and its corollary in Cristóbal et al. (2004).

Now, using Proposition 4 jointly with the strong uniform representation given in Eq. (9), the strong uniform approximation by means of a sequence of gaussian processes is proved. $\square$

Proof of Theorem 2 Let us introduce the following notation:

$$\tilde{\zeta}_n(x) = \mu_Y^{-1}\sqrt{nh_n}f_X(x)\left(\tilde{m}_n(x) - m(x)\right)$$
$$\hat{\zeta}_n(x) = \mu_Y^{-1}\sqrt{nh_n}f_X(x)\left(\hat{m}_n(x) - m(x)\right)$$
$$\zeta_n(x) = \mu_Y^{-1}\sqrt{nh_n}f_X(x)\left(\hat{m}_n(x) - \tilde{m}_n(x)\right).$$

As a consequence of Proposition 1, we obtain that

$$\zeta_n(x) = \hat{\zeta}_n(x) - \tilde{\zeta}_n(x) = \hat{\zeta}_n(x) + O\left(\sqrt{h_n \log\ \log\ n}\right)$$

uniformly in $[0,1]$ and almost surely. Furthermore, the second consequence of Theorem 1 gives that

$$\hat{\zeta}_n(x) = Z_n^w(x) + o\left(\frac{\log^2\ n}{\sqrt{nh_n}}\right)$$

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

$$\zeta_n(x) = Z_n^w(x) + O\left(\sqrt{h_n \log\ \log\ n}\right) \tag{16}$$

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

Now, bearing in mind the last equation, for the first test statistic we proposed we obtain that

$$K_n^{\infty} = \|\zeta_n(x)\|_{\infty} = \left\|Z_n^w(x)\right\|_{\infty} + O\left(\sqrt{h_n \log\ \log\ n}\right)$$

almost surely.

In the case of W_n^2, note that because of Theorem 1:

$$\begin{aligned} W_n^2 &= \int_0^1 \zeta_n(s)^2\, \mathrm{d}s = \int_0^1 \left(\hat{\zeta}_n(s) + O\left(\sqrt{h_n \log\ \log\ n}\right)\right)^2 \mathrm{d}s \\ &= \int_0^1 Z_n^w(s)^2\, \mathrm{d}s + O\left(\sqrt{h_n\ \log\ n \log\ \log\ n} + h_n \log\ \log\ n\right), \end{aligned}$$

almost surely. □

A.2 Bootstrap estimators

From the bootstrap scheme given in Sect. 4, and the regression estimators presented in Sect. 3, the proofs in this Appendix follow essentially those argumentations given in Appendix A.1.

In what follows, the main changes are due to the introduction of the wild bootstrap random variable, whose involvement is such that we must deal with the bootstrap distribution $F_n^{w^*}$, where F^{w^*} is defined as

$$\mathrm{d}F^{w^*}(z, y, \gamma) = \mathrm{d}F^w(z, y)\, p_\gamma$$

where p_γ, $\gamma \in \{a, b\}$ is the probability function of the wild bootstrap random variable Ξ, verifying the following equations:

$$\begin{aligned} p_a + p_b &= 1 \\ a p_a + b p_b &= 0 \\ a^l p_a + b^l p_b &= 1 \quad l = 2, 3 \end{aligned} \tag{17}$$

see Wu (1986) or Härdle and Mammen (1993). We will use $\mathbf{E}^{w^*}[\cdot]$ to denote expectations with regard to the distribution F^{w^*}. Note also that both samples, the original observations $(x_1, y_1), \ldots, (x_n, y_n)$, and the bootstrap sample $\left(x_1^*, y_1^*\right), \ldots, \left(x_n^*, y_n^*\right)$ have the same size n as a consequence of the bootstrap scheme employed.

Proof of Proposition 2 As $x_i^* = x_i$, Eq. (15) is also valid in this setting, and because of $y_i^* = \mathbf{g}(x_i)^{\mathrm{T}} \tilde{\boldsymbol{\beta}}_n + \epsilon_i^*$, we have that the vector $\mathbf{Y}^*$ is $\mathbf{G}^{\mathrm{T}} \tilde{\boldsymbol{\beta}}_n + \boldsymbol{\epsilon}^*$ and hence

$$\tilde{\boldsymbol{\beta}}_n^* = \tilde{\boldsymbol{\beta}}_n + \left(\mathbf{G}^{\mathrm{T}} \mathbf{B} \mathbf{G}\right)^{-1} \mathbf{G}^{\mathrm{T}} \mathbf{B} \boldsymbol{\epsilon}^*.$$

The rest of the proof follows the same arguments that were given in the proof of Proposition 1, as the random variables in $\mathbf{B}\epsilon^*$ are $y_i^{-1}\epsilon_i^* = \gamma_i y_i^{-1}\hat{\epsilon}_i$, which have null expectation and finite second–order moment under the assumptions. □

From the local linear estimator perspective, and bearing in mind that $x_i^* = x_i$ we have to consider the same $s_j^w(x;h)$ as before. The main change that the bootstrap scheme introduces with respect to the previous part of the Appendix is that, in this case, we have to deal with the process

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

instead of the process $e_j^w(x;h_n)$.

Proposition 5 *Under assumptions L1, L2 we have that:*

$$e_j^*(x;h_n) = O\left(\delta_{nh_n}\right)$$

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

Proof Note that

$$\mathbf{E}^{w^*}\left[\Xi\left(\frac{Y - m(X)}{Y}\right)\right] = 0$$

and that follow the same argument used in the proof of Proposition 3. □

Proposition 6 *Under assumptions L1 and L2 in Sect. 2:*

$$e_0^*(x;h_n) = \frac{1}{\sqrt{nh_n}}Z_n^*(x) + o\left(\delta_{nh_n}\right)$$

uniformly in $[0,1]$ *and almost surely, where* Z_n^* *is a second–order gaussian process with null expectation and covariance function given by:*

$$\begin{aligned}\mathbf{Cov}\left[Z_n^*(s), Z_n^*(t)\right] &= \mathbf{E}^{w^*}\left[\frac{1}{h_n}K\left(\frac{s-X}{h_n}\right)K\left(\frac{t-X}{h_n}\right)\Xi^2\left(\frac{Y-m(X)}{Y}\right)^2\right]\\ &= \mathbf{E}^{w}\left[\frac{1}{h_n}K\left(\frac{s-X}{h_n}\right)K\left(\frac{t-X}{h_n}\right)\left(\frac{Y-m(X)}{Y}\right)^2\right].\end{aligned}$$

Proof If we define r_i to be $\gamma_i(y_i - m(x_i))y_i^{-1}$, the process can be written in the following way:

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

where $(x_i, r_i), \ldots, (x_i, r_i)$ is an i.i.d. sample from the random variable (X, R) with distribution F^r, and where $R = \Xi Q$ and $Q = Y^{-1}(Y - m(X))$. Note that the distribution of (X, R) can be written in terms of the distribution F^q of (X, Q) as a mixture:

$$F^r(x,r) = F^q\left(x, \frac{r}{a}\right)p_a + F^q\left(x, \frac{r}{b}\right)p_b,$$

and, as a consequence, F^r is a continuous distribution. Moreover, the expectation of R is null, and its variance is given by

$$\begin{aligned}
\mathbf{E}\left[R^2\right] &= \int r^2\,\mathrm{d}F^r(x,r) \\
&= \int \frac{r^2}{a}\,\mathrm{d}F^q\left(x, \frac{r}{a}\right)p_a + \int \frac{r^2}{b}\,\mathrm{d}F^q\left(x, \frac{r}{b}\right)p_b \\
&= \int a^2 q^2\,\mathrm{d}F^q(x,q)p_a + \int b^2 q^2\,\mathrm{d}F^q(x,q)p_b \\
&= \mathbf{E}\left[Q^2\right]\left(a^2 p_a + b^2 p_b\right) = \mathbf{E}^w\left[\left(\frac{Y - m(X)}{Y}\right)^2\right].
\end{aligned}$$

In a similar way, it can also be shown that

$$\mathbf{E}\left[R^2 g(X)\right] = \mathbf{E}^w\left[\left(\frac{Y - m(X)}{Y}\right)^2 g(X)\right],$$

and furthermore, not only is the conditional mean $\mathbf{E}[R|X = x]$ null for every $x \in [0,1]$, but we also have that $\mathbf{E}\left[R^2|X = x\right] = v^w(x)$.

Now, the proof follows the same argumentation that was given to prove Proposition 4, but with the distribution F^r instead of F^w. Let us denote by $Y_n^{0^*}(x)$ the following process

$$\begin{aligned}
Y_n^{0^*}(x) &= \sqrt{nh_n}\, e_0^*(x; h_n) \\
&= \sqrt{nh_n}\int r\frac{1}{h_n}K\left(\frac{z - x}{h_n}\right)\mathrm{d}E_n^r(z,r)
\end{aligned}$$

where $E_n^r(\cdot)$ is the empirical process $\sqrt{n}\left(F_n^r(\cdot) - F^r(\cdot)\right)$. As a consequence, in this case we obtain that

$$\begin{aligned}
Y_n^0(x) &= \int r\frac{1}{h_n}K\left(\frac{z - x}{h_n}\right)\mathrm{d}B_n\left(H^r(r,y)\right) + O\left(\frac{\log^2 n}{\sqrt{n}}\right) \\
&= Z_n^*(x) + O\left(\frac{\log^2 n}{\sqrt{n}}\right).
\end{aligned}$$

where $H^r(r,y) = \left(F_X^r(z), F_{R|X}^r(r \mid z)\right)$, and

$$\mathbf{Cov}\left[Z_n^*(s), Z_n^*(t)\right] = \int \frac{1}{h_n} K\left(\frac{s-z}{h_n}\right) K\left(\frac{t-z}{h_n}\right) r^2 \mathrm{d}F^r(z,y)$$
$$= \mathbf{E}^w\left[\left(\frac{Y-m(X)}{Y}\right)^2 \frac{1}{h_n} K\left(\frac{s-X}{h_n}\right) K\left(\frac{t-X}{h_n}\right)\right]. \qquad \square$$

Proof of Theorem 3 This proof is quite similar to the proof of Theorem 1, but using Propositions 5 and 6 to handle $e_0^*(x;h_n)$.

Notice first that, as a consequence of $y_i^* = \tilde{m}_n(x_i) + \hat{\epsilon}_i^*$, and $\hat{\epsilon}_i^* = \gamma_i\left(\epsilon_i + O\left(h_n^2 + \delta_{nh_n}\right)\right)$ almost surely because of Theorem 1, we have:

$$\hat{m}_n^*(x) = \frac{\sum_{i=1}^n w_{ih_n}^w(x)\tilde{m}_n(x_i)}{\sum_{i=1}^n w_{ih_n}^w(x)} + \frac{\sum_{i=1}^n w_{ih_n}^w(x)\gamma_i\epsilon_i}{\sum_{i=1}^n w_{ih_n}^w(x)}$$
$$+ \frac{\sum_{i=1}^n w_{ih_n}^w(x)\gamma_i}{\sum_{i=1}^n w_{ih_n}^w(x)} O\left(h_n^2 + \delta_{nh_n}\right) = A + B + C.$$

The second–order asymptotic expansion of $\hat{m}_n^*(x_i)$ at x takes the following form:

$$\tilde{m}_n(x) + h_n\left(\frac{x_i - x}{h_n}\right)\tilde{m}_n'(x) + h_n^2\left(\frac{x_i - x}{h_n}\right)^2 \frac{\tilde{m}_n''(x)}{2}.$$

Hence, using Proposition 3 we have that:

$$A = \tilde{m}_n(x) + 0 + h_n^2 \frac{\tilde{m}_n''(x)}{2} \frac{s_2^w(x;h_n)s_2^w(x;h_n) - s_1^w(x;h_n)s_3^w(x;h_n)}{s_2^w(x;h_n)s_0^w(x;h_n) - s_1^w(x;h_n)s_1^w(x;h_n)}$$
$$= \tilde{m}_n(x) + \mu_2 h_n^2 \frac{m''(x)}{2}\left(1 + O\left(h_n^2 + \delta_{nh_n}\right)\right)$$

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

$$B = \frac{s_2^w(x;h_n)e_0^*(x;h_n) - s_1^w(x;h_n)e_1^*(x;h_n)}{s_2^w(x;h_n)s_0^w(x;h_n) - s_1^w(x;h_n)s_1^w(x;h_n)}$$
$$= \frac{\mu_Y\, e_0^*(x;h_n)}{f_X(x)} + O\left(h_n\delta_{nh_n}\right)$$

uniformly in $[0,1]$ and almost surely. For the third term C, arguing as in Proposition 5, but bearing in mind that Ξ and X are independent and Ξ has null expectation, we obtain

$$\frac{\sum_{i=1}^n w_{ih_n}^w(x)\gamma_i}{\sum_{i=1}^n w_{ih_n}^w(x)} = O\left(\delta_{nh_n}\right),$$

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

$$C = O\left(\delta_{nh_n}\right) O\left(h_n^2 + \delta_{nh_n}\right),$$

almost surely.

All these arguments lead to the following uniform and almost sure representation for $\hat{m}_n^*(x)$:

$$\hat{m}_n^*(x) = \tilde{m}_n(x) + \mu_2 h_n^2 \frac{m''(x)}{2}\left(1 + O\left(h_n^2 + \delta_{nh_n}\right)\right) + \frac{\mu_Y\, e_0^*(x; h_n)}{f_X(x)} + O\left(h_n \delta_{nh_n}\right) + O\left(\delta_{nh_n}^2 + h_n^2 \delta_{nh_n}\right). \tag{18}$$

This concludes the proof of the strong uniform consistency since $e_0^*(x; h_n) = O\left(\delta_{nh_n}\right)$ uniformly and almost surely. Moreover, Proposition 4, jointly with the strong uniform representation given in Eq. (18), proves the strong uniform approximation by means of gaussian processes. □

Proof of Theorem 4 In this case, the changes the bootstrap scheme introduces require the following changes in notation:

$$\tilde{\zeta}_n^*(x) = \mu_Y^{-1}\sqrt{nh_n} f_X(x)\left(\tilde{m}_n^*(x) - \tilde{m}_n(x)\right)$$
$$\hat{\zeta}_n^*(x) = \mu_Y^{-1}\sqrt{nh_n} f_X(x)\left(\hat{m}_n^*(x) - \tilde{m}_n(x)\right)$$
$$\zeta_n^*(x) = \mu_Y^{-1}\sqrt{nh_n} f_X(x)\left(\hat{m}_n^*(x) - \tilde{m}_n^*(x)\right).$$

Because of Proposition 2:

$$\zeta_n^*(x) = \hat{\zeta}_n^*(x) - \tilde{\zeta}_n^*(x) = \hat{\zeta}_n^*(x) + O\left(\sqrt{h_n \log\ \log\ n}\right)$$

uniformly in $[0, 1]$ and almost surely, and as a consequence of Theorem 3 we obtain

$$\hat{\zeta}_n^*(x) = Z_n^*(x) + o\left(\frac{\log^2\ n}{\sqrt{nh_n}}\right)$$

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

$$\zeta_n^*(x) = Z_n^*(x) + O\left(\sqrt{h_n \log\ \log\ n}\right) \tag{19}$$

uniformly in $[0, 1]$ and almost surely. The rest of the proof follows the same arguments that were given in Theorem 2. □

References

Alcalá, J. T., Cristóbal, J. A., González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statistics & Probability Letters, 42*(1), 39–46.

Bickel, P. J., Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics, 9*(6), 1196–1217.

Cox, D. R. (1969). Some Sampling Problems in Technology In *New Developments in Survey Sampling*. (pp. 506–527). New York: Wiley.

Cox, D., Koh, E., Wahba, G., Yandell, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *The Annals of Statistics, 16*(1), 113–119.

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

Cristóbal, J. A., Alcalá, J. T. (2001). An overview of nonparametric contributions to the problem of functional estimation from biased data. *Test, 10*(2), 309–332.

Cristóbal, J. A., Ojeda, J. L., Alcalá, J. T. (2004). Confidence bands in nonparametric regression with length biased data. *Annals of the Institute of Statistical Mathematics, 56*(3), 475–496.

Delgado, M. A., González Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *The Annals of Statistics, 29*(5), 1469–1507.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics, 9*(6), 1218–1228.

Hall, P. (1991). On convergence rates of suprema. *Probability Theory and Related Fields, 89*(4), 447–455.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics, 21*(4), 1926–1947.

Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests*. Springer Series in Statistics. Heidelberg: Springer

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields, 75*(2), 261–277.

Kozek, A. S. (1990). A nonparametric test of fit of a linear model. *Communications in Statistics. Theory and Methods, 19*(1), 169–179.

Leadbetter, M. R., Lindgren, G., Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Heidelberg: Springer

Patil, G. P., Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics, 34*(2), 179–189.

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

Quesenberry, C. P. Jr., Jewell, N. P. (1986). Regression analysis based on stratified samples. *Biometrika, 73*(3), 605–614.

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

Tusnády, G. (1977). A remark on the approximation of the sample *DF* in the multidimensional case. *Periodica Mathematica Hungarica., 8*(1), 53–55.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics, 10*(2), 616–620.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics, 14*(4), 1261–1350. With discussion and a rejoinder by the author.

Wu, C. O. (2000). Local polynomial regression with selection biased data. *Statistica Sinica, 10*(3), 789–817.