Bootstrap confidence intervals in nonparametric regression without an additive model

Dimitris N. Politis

Abstract The problem of confidence interval construction in nonparametric regression via the bootstrap is revisited. When an additive model holds true, the usual residual bootstrap is available but it often leads to confidence interval under-coverage; the case is made that this under-coverage can be partially corrected using predictive—as opposed to fitted—residuals for resampling. Furthermore, it has been unclear to date if a bootstrap approach is feasible in the absence of an additive model. The main thrust of this paper is to show how the transformation approach put forth by Politis (2010, 2013) in the related setting of prediction intervals can be found useful in order to construct bootstrap confidence intervals *without* an additive model.

1 Introduction

Consider regression data of the type $\{(Y_t, x_t),\ t = 1, \ldots, n\}$. For simplicity of presentation, the regressor x_t is assumed univariate and deterministic; the case of a multivariate regressor is handled similarly. As usual, it will be assumed that $Y_1, \ldots, Y_n$ are independent but not identically distributed. Attention focuses primarily on the first two moments of the response Y_t, namely

$$\mu(x_t) = E(Y_t|x_t) \ \text{ and } \ \sigma^2(x_t) = Var(Y_t|x_t). \tag{1}$$

In the nonparametric setting, the functions $\mu(\cdot)$ and $\sigma(\cdot)$ are considered unknown but assumed to possess some degree of smoothness (differentiability, etc.). There are many approaches towards nonparametric estimation of the functions μ and σ, e.g., wavelets and orthogonal series, smoothing splines, local polynomials, and kernel smoothers. For concreteness, this paper will focus on one of the oldest methods, namely the Nadaraya-Watson (N-W) kernel estimators; see Li and Racine (2007) and the references therein.

Beyond point estimates of the functions μ and σ, it is important to be able to additionally provide interval estimates in order to have a measure of their statistical accuracy. Suppose, for example, that a practitioner is interested in the expected response to be observed at a future point $x_{\rm f}$. A confidence interval for $\mu(x_{\rm f})$ is then desirable. Under regularity conditions, such a confidence interval can be given either via a large-sample

Dimitris N. Politis
University of California at San Diego, La Jolla, CA 92093-0112, USA , e-mail: dpolitis@ucsd.edu

normal approximation, or via a resampling approach; see e.g. Freedman (1981), Härdle and Bowman (1988), Härdle and Marron (1991), Hall (1993), or Neumann and Polzehl (1998).

Typical regularity conditions for the above bootstrap approaches involve the assumption of an additive model with respect to independent and identically distributed (i.i.d.) errors. In Section 2, we revisit the usual model-based bootstrap for regression adding the dimension of employing predictive as opposed to fitted residuals as advocated by Politis (2010, 2013) in a related context. More importantly, in Section 3 we address the problem of constructing a bootstrap confidence interval for $\mu(x_{\rm f})$ *without* an underlying additive model.

The *model-free* approach developed in this paper is totally automatic, relieving the practitioner from the need to find an optimal transformation towards additivity and variance stabilization; this is a significant practical advantage because of the multitude of such proposed transformations, e.g. the Box/Cox power family, ACE, AVAS, etc.—see Linton et al. (1997) and the references therein. The finite-sample simulations provided in Section 4 confirm the viability and good performance of the model-free confidence intervals.

2 Model-based nonparametric regression

2.1 Nonparametric regression with an additive model

An additive model for nonparametric regression is given by the equation

$$Y_t = \mu(x_t) + \sigma(x_t)\,\varepsilon_t\ ,\ \ t = 1,\ldots,n, \tag{2}$$

with $\varepsilon_t \sim$ i.i.d. (0,1) from an (unknown) distribution F. The N-W estimator of $\mu(x)$ is defined as

$$m_x = \sum_{i=1}^{n} Y_i \tilde{K}\left(\frac{x-x_i}{h}\right)\ \text{ with }\ \tilde{K}\left(\frac{x-x_i}{h}\right) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{x-x_k}{h}\right)} \tag{3}$$

where h is the bandwidth, and $K(x)$ is a symmetric kernel function with $\int K(x)dx = 1$. Similarly, the N-W estimator of $\sigma^2(x)$ is given by $s_x^2 = M_x - m_x^2$ where $M_x = \sum_{i=1}^{n} Y_i^2 \tilde{K}\left(\frac{x-x_i}{h}\right)$.

For $t = 1,\ldots,n$, let $e_t = (Y_t - m_{x_t})/s_{x_t}$ denote the *fitted* residuals, and $\tilde{e}_t = (Y_t - m_{x_t}^{(t)})/s_{x_t}^{(t)}$ the *predictive* residuals. Here, $m_x^{(t)}$ and $M_x^{(t)}$ denote the estimators m_x and M_x respectively computed from the *delete-Y_t* dataset: $\{(Y_i, x_i),\ i = 1,\ldots,t-1$ and $i = t+1,\ldots,n\}$. As before, define $s_{x_t}^{(t)} = \sqrt{M_{x_t}^{(t)} - (m_{x_t}^{(t)})^2}$. Choosing the bandwidth h is often done by *cross-validation*, i.e., picking h to minimize $\sum_{t=1}^{n} \tilde{e}_t^2$, or its L_1 analog: $\sum_{t=1}^{n} |\tilde{e}_t|$.

2.2 Model-based confidence intervals

Consider the problem of constructing a confidence interval for the regression function $\mu(x_{\rm f})$ at a point of interest $x_{\rm f}$. A normal approximation to the distribution of the estimator $m_{x_{\rm f}}$ implies an approximate $(1-\alpha)100\%$ equal-tailed, confidence interval for $\mu(x_{\rm f})$ given by:

$$[m_{x_{\rm f}} + v_{x_{\rm f}} \cdot z(\alpha/2),\ m_{x_{\rm f}} + v_{x_{\rm f}} \cdot z(1-\alpha/2)] \tag{4}$$

where $v_{x_f}^2 = s_{x_f}^2 \sum_{i=1}^n \tilde{K}^2(\frac{x_f - x_i}{h})$ with $\tilde{K}$ defined in eq. (3), and $z(\alpha)$ being the α–quantile of the standard normal. If the 'density' (e.g. histogram) of the design points $x_1, ..., x_n$ can be thought to approximate a given functional shape (say, $f(\cdot)$) for large n, then the large-sample approximation

$$\sum_{i=1}^{n} \tilde{K}^2(\frac{x_f - x_i}{h}) \sim \frac{\int K^2(x)dx}{nh\, f(x_f)} \tag{5}$$

can be used which relies on the assumption that $\int K(x)dx = 1$; see e.g. Li and Racine (2007).

Interval (4) may be problematic in two respects: (a) it ignores the bias of m_x, so it must be either explicitly bias-corrected, or a suboptimal bandwidth must be used to ensure undersmoothing; and (b) it is based on a Central Limit Theorem which may not be a good finite-sample approximation if the errors are skewed and/or leptokurtic, or when the sample size is not large enough. For both above reasons, practitioners often prefer bootstrap methods over the normal approximation interval (4). When using fitted residuals, the following algorithm is the well-known residual bootstrap pioneered by Freedman (1981) in a linear regression setting, and extended to nonparametric regression by Härdle and Bowman (1988), and other authors. As an alternative, we also propose the use of predictive residuals for resampling as advocated by Politis (2010, 2013) in a related context. The predictive residuals have an empirical distribution that has similar shape as that of the fitted residuals but it has larger scale. This is a finite-sample phenomenon only but it may help alleviate the well-known phenomenon of under-coverage of bootstrap confidence intervals.

Our goal is to approximate the distribution of the confidence *root*: $\mu(x_f) - m_{x_f}$ by that of its bootstrap counterpart.

RESAMPLING ALGORITHM FOR MODEL-BASED CONFIDENCE INTERVALS FOR $\mu(x_f)$

1. Based on the $\{(Y_t, x_t), t = 1, ..., n\}$ data, construct the estimates m_x and s_x from which the fitted residuals e_i, and predictive residuals $\tilde{e}_i$ are computed for $i = 1, ..., n$.
2. For the traditional model-based bootstrap approach (MB), let $r_i = e_i - n^{-1}\sum_j e_j$, for $i = 1, ..., n$. For the predictive residual approach (PRMB) as in Politis (2010), let $r_i = \tilde{e}_i - n^{-1}\sum_j \tilde{e}_j$, for $i = 1, ..., n$.
   a. Sample randomly (with replacement) the residuals $r_1, ..., r_n$ to create the bootstrap pseudo-residuals $r_1^*, ..., r_n^*$ whose empirical distribution is denoted by $\hat{F}_n^*$.
   b. Create pseudo-data in the Y domain by letting $Y_i^* = m_{x_i} + s_{x_i} r_i^*$, for $i = 1, ..., n$.
   c. Based on the pseudo-data $\{(Y_t^*, x_t), t = 1, ..., n\}$, re-estimate the functions $\mu(x)$ and $\sigma(x)$ by the kernel estimators m_x^* and s_x^* (with same kernel and bandwidths as the original estimators m_x and s_x).
   d. Calculate a replicate of the *bootstrap confidence root*: $m_{x_f} - m_{x_f}^*$.
3. Steps (a)—(d) in the above are repeated B times, and the B bootstrap root replicates are collected in the form of an empirical distribution with α–quantile denoted by $q(\alpha)$.
4. Then, a $(1-\alpha)100\%$ equal-tailed confidence interval for $\mu(x_f)$ is given by:

$$[m_{x_f} + q(\alpha/2), m_{x_f} + q(1-\alpha/2)]. \tag{6}$$

Remark 2.1 As in all nonparametric smoothing problems, choosing the bandwidth is often a key issue due to the ever-looming problem of bias; the addition of a bootstrap algorithm as above further complicates things. Different authors have used various tricks to account for the bias. For example, Härdle and Bowman (1988)

construct a kernel estimate for the second derivative $\mu''(x)$, and use this estimate to explicitly correct for the bias; the estimate of the second derivative is known to be consistent but it is difficult to choose its bandwidth. Härdle and Marron (1991) estimate the (fitted) residuals using the optimal bandwidth but the resampled residuals are then added to an oversmoothed estimate of μ; the bootstrapped data are then smoothed using the optimal bandwidth. Neumann and Polzehl (1998) use only one bandwidth but it is of smaller order than the mean square error optimal rate; this *undersmoothing* of curve estimates was first proposed by Hall (1993) and is perhaps the easiest theoretical solution towards confidence band construction although the recommended degree of undersmoothing for practical purposes is not obvious.

Remark 2.2 An important feature of all bootstrap procedures is that they can handle *joint* confidence intervals, i.e., confidence *regions*, with the same ease as the univariate ones. This is especially true in regression where simultaneous confidence intervals are typically constructed in the form of confidence *bands*; the details are well-known in the literature and are omitted due to lack of space.

3 Model-free nonparametric regression

3.1 Nonparametric regression without an additive model

We now revisit the nonparametric regression setup but in a situation where a model such as eq. (2) can not be considered to hold true (not even approximately). As an example of model (2) not being valid, consider the setup where the skewness and/or kurtosis of Y_t depends on x_t, and thus centering and studentization will not result in 'i.i.d.–ness'. The dataset is still $\{(Y_t, x_t),\ t = 1,\ldots,n\}$ where the regressor x_t is univariate and deterministic, and the variables $Y_1, Y_2, \ldots$ are *independent* but not identically distributed. Define the conditional distribution $D_x(y) = P\{Y_{\rm f} \le y | x_{\rm f} = x\}$ where $(Y_{\rm f}, x_{\rm f})$ represents the random response $Y_{\rm f}$ associated with regressor $x_{\rm f}$. Attention still focuses on constructing an interval estimate of $\mu(x_{\rm f}) = E(Y_{\rm f}|x_{\rm f}) = \int y\, D_{x_{\rm f}}(dy)$.

Throughout this section, we will assume that the function $D_x(y)$ is continuous in both x and y. Consequently, we can estimate $D_x(y)$ by the local (weighted) empirical distribution

$$\hat{D}_x(y) = \sum_{i=1}^{n} \mathbf{1}\{Y_i \le y\} \tilde{K}\left(\frac{x - x_i}{h}\right); \tag{7}$$

this is just a N-W smoother of the variables $\mathbf{1}\{Y_t \le y\},\ t = 1,\ldots,n$. Estimator $\hat{D}_x(y)$ enjoys many desirable properties, including asymptotic consistency, but is discontinuous as a function of y. To construct a continuous (and differentiable) estimator, let b be a positive bandwidth parameter and $\Lambda(y)$ be a (differentiable) distribution function that is strictly increasing, and define

$$\bar{D}_x(y) = \sum_{i=1}^{n} \Lambda\left(\frac{y - Y_i}{b}\right) \tilde{K}\left(\frac{x - x_i}{h}\right). \tag{8}$$

Under regularity conditions, Li and Racine (2007, Theorem 6.2) show that

$$Var(\bar{D}_x(y)) = O(\frac{1}{hn}) \ \text{ and } \ Bias(\bar{D}_x(y)) = O(h^2 + b^2) \tag{9}$$

assuming that $h \to 0$, $b \to 0$, $hn \to \infty$ and $\sqrt{hn}(h^3 + b^3) = o(1)$; to minimize the asymptotic Mean Squared Error of $\bar{D}_x(y)$, the optimal bandwidths are $h \sim c_h n^{-1/5}$ and $b \sim c_b n^{-2/5}$ for some positive constants c_h, c_b.

Recall that the Y_ts are non-i.i.d. only because they do not have identical distributions. Since they are continuous random variables, the *probability integral transform* is applicable. If we let $\eta_i = D_{x_i}(Y_i)$ for $i = 1, \ldots, n$, then $\eta_1, \ldots, \eta_n$ are i.i.d. Uniform(0,1). Of course, $D_x(\cdot)$ is not known but we can define

$$u_i = \bar{D}_{x_i}(Y_i) \quad \text{for } \; i = 1, \ldots, n; \tag{10}$$

by the consistency of $\bar{D}_x(\cdot)$, we can now claim that $u_1, \ldots, u_n$ are approximately i.i.d. Uniform(0,1).

Using eq. (10) and following the *Model-free Prediction Principle* of Politis (2010), the quantity

$$\Pi_{x_\mathrm{f}} = n^{-1} \sum_{i=1}^{n} \hat{D}^{-1}_{x_\mathrm{f}}(u_i) \tag{11}$$

was proposed as an L_2–optimal predictor of Y_f, i.e., an approximation to the conditional expectation $\mu(x_\mathrm{f}) = E(Y_\mathrm{f}|x_\mathrm{f})$. Note that $\hat{D}_{x_\mathrm{f}}(y)$ is a step function in y, and thus not invertible; the notation $\hat{D}^{-1}_{x_\mathrm{f}}$ denotes the quantile inverse. Alternatively, one could propose the quantity $n^{-1}\sum_{i=1}^{n} \bar{D}^{-1}_{x_\mathrm{f}}(u_i)$ where a true inverse is used; the difference between the two is negligible, and definition (11) is straightforward.

Note that Π_{x_f} is defined as a function of the approximately i.i.d. variables $u_1, \ldots, u_n$; as such, it may be amenable to the original i.i.d. bootstrap of Efron (1979). Two questions arise: (a) is the estimator Π_{x_f} quite different from the standard N-W estimator m_{x_f}? and (b) could m_{x_f} itself be bootstrapped using i.i.d. resampling? The answers to these questions are NO and YES respectively, due to the following fact. To motivate it, recall that the N-W estimator m_x can be expressed alternatively as

$$m_x = \sum_{i=1}^{n} Y_i \tilde{K}\left(\frac{x - x_i}{h}\right) = \int y \, \hat{D}_x(dy) = \int_0^1 \hat{D}_x^{-1}(u) du. \tag{12}$$

The last equality in (12) is the identity $\int y \, F(dy) = \int_0^1 F^{-1}(u) du$ that holds true for any distribution F.

Fact 3.1 *Assume that $D_x(y)$ is continuous in x, and differentiable in y with derivative that is everywhere positive on its support. Then, Π_{x_f} and m_{x_f} are* asymptotically equivalent*, i.e., $\sqrt{nh}\,(\Pi_{x_\mathrm{f}} - m_{x_\mathrm{f}}) = o_p(1)$ for any x_f that is not a boundary point.*

One way to prove the above is to show that the average appearing in (11) is close to a Riemann sum approximation to the integral at the RHS of (12) based on a grid of n points. The law of the iterated logarithm for order statistics of uniform spacings can be useful here; see Devroye (1981) and the references therein.

Remark 3.1 The above line of arguments indicates that there is a variety of estimators that are asymptotically equivalent to m_{x_f} in the sense of Fact 3.1. For example, the Riemann sum $M^{-1}\sum_{k=1}^{M} \hat{D}^{-1}_{x_\mathrm{f}}(k/M)$ is such an approximation as long as $M \geq n$. A stochastic approximation can also be concocted as $M^{-1}\sum_{i=1}^{M} \hat{D}^{-1}_{x_\mathrm{f}}(W_i)$ where $W_1, \ldots, W_M$ are i.i.d. generated from a Uniform(0,1) distribution and $M \geq n$.

3.2 Bootstrap algorithm for model-free confidence intervals

Let $\hat{\mu}(x_{\rm f})$ denote our chosen estimator of $\mu(x_{\rm f}) = E(Y_{\rm f}|x_{\rm f})$, i.e., either $m_{x_{\rm f}}$ or $\Pi_{x_{\rm f}}$, or even one of the other asymptotically equivalent estimators discussed in Remark 3.1. Our goal is to approximate the distribution of the confidence *root*: $\mu(x_{\rm f}) - \hat{\mu}(x_{\rm f})$ by that of its bootstrap counterpart. The algorithm reads as follows.

RESAMPLING ALGORITHM FOR MODEL-FREE CONFIDENCE INTERVALS FOR $\mu(x_{\rm f})$

1. Based on the $\{(Y_t, x_t), t = 1, ..., n\}$ data, construct the estimates $\hat{D}_x(\cdot)$ and $\bar{D}_x(\cdot)$, and use eq. (10) to obtain the transformed data $u_1, ..., u_n$ that are approximately i.i.d. Uniform (0,1).

   a. Sample randomly (with replacement) the transformed data $u_1, ..., u_n$ to create bootstrap pseudo-data $u_1^*, ..., u_n^*$.
   b. Use the quantile inverse transformation $\hat{D}_x^{-1}$ to create bootstrap pseudo-data in the Y domain, i.e., let $\underline{Y}_n^* = (Y_1^*, ..., Y_n^*)$ where $Y_t^* = \hat{D}_{x_t}^{-1}(u_t^*)$. Note that Y_t^* is paired with the original x_t design point; hence, the bootstrap dataset is $\{(Y_t^*, x_t), t = 1, ..., n\}$.
   c. Based on the pseudo-data $\{(Y_t^*, x_t), t = 1, ..., n\}$, re-estimate the conditional distribution $D_x(\cdot)$; denote the bootstrap estimates by $\hat{D}_x^*(\cdot)$ and $\bar{D}_x^*(\cdot)$.
   d. Calculate a replicate of the *bootstrap confidence root*: $\hat{\mu}(x_{\rm f}) - \hat{\mu}^*(x_{\rm f})$ where $\hat{\mu}^*(x_{\rm f})$ equals either $\int y\, \hat{D}_{x_{\rm f}}^*(dy) = \int_0^1 \hat{D}_{x_{\rm f}}^{*^{-1}}(u)du$ or $n^{-1}\sum_{i=1}^n \hat{D}_{x_{\rm f}}^{*^{-1}}(u_i^*)$ according to whether $\hat{\mu}(x_{\rm f})$ was chosen as $m_{x_{\rm f}}$ or $\Pi_{x_{\rm f}}$.

2. Steps (a)—(d) in the above are repeated B times, and the B bootstrap root replicates are collected in the form of an empirical distribution with α–quantile denoted by $q(\alpha)$.
3. Then, the *Model-Free* (MF) $(1-\alpha)100\%$ equal-tailed, confidence interval for $\mu(x_{\rm f})$ is

$$[\hat{\mu}(x_{\rm f}) + q(\alpha/2), \hat{\mu}(x_{\rm f}) + q(1-\alpha/2)]. \tag{13}$$

Remark 3.2 An alternative way to implement step 1(a) of the above algorithm is:

a′. Generate bootstrap pseudo-data $u_1^*, ..., u_n^*$ i.i.d. from an exact Uniform $(0, 1)$ distribution.

If the above choice is made, then there is no need to use eq. (10) to obtain the transformed data $u_1, ..., u_n$; in this sense, the smooth estimator $\bar{D}_x(\cdot)$ is not needed, and the step function $\hat{D}_x(\cdot)$ suffices for the algorithm.

The downside to the above proposal is that the option to use 'predictive' u–data is unavailable. To elaborate, recall that Politis (2010) defined the model-free 'predictive' u–data as follows. Let $\bar{D}_{x_t}^{(t)}$ denote the estimator $\bar{D}_{x_t}$ as computed from the delete-Y_t dataset, i.e., $\{(Y_i, x_i),\ i = 1, \ldots, t-1$ and $i = t+1, \ldots, n\}$. Now let

$$u_t^{(t)} = \bar{D}_{x_t}^{(t)}(Y_t) \quad \text{for } t = 1, \ldots, n. \tag{14}$$

The $u_t^{(t)}$ variables are the model-free analogs of the predictive residuals $\tilde{e}_t$ of Section 2.

Remark 3.3 We can now define *Predictive Model-Free* (PMF) confidence intervals for $\mu(x_{\rm f})$. The PMF Resampling Algorithm is identical to the above with one exception; replace step 1(a) with the following:

a″. Sample randomly (with replacement) the predictive u–data $u_1^{(1)}, ..., u_n^{(n)}$ to create bootstrap pseudo-data $u_1^*, ..., u_n^*$.

Remark 3.4 Recall that the model-free L_1–optimal predictor of Y_f is given by the median$\{\bar{D}^{-1}_{x_f}(u_i)\}$; see Politis (2010, 2013). Therefore, by analogy to Fact 3.1, we have: median$\{\bar{D}^{-1}_{x_f}(u_i)\} = \bar{D}^{-1}_{x_f}(\text{median}\{u_i\}) \simeq \bar{D}^{-1}_{x_f}(1/2)$ since the u_is are approximately Uniform (0,1). Hence, if the practitioner wanted to estimate the median (as opposed to the mean) of the conditional distribution of Y_f given x_f, then the local median $\bar{D}^{-1}_{x_f}(1/2)$, could be bootstrapped using i.i.d. resampling in the same manner that median$\{\bar{D}^{-1}_{x_f}(u_i)\}$ can be bootstrapped.

4 Simulations

4.1 When a nonparametric regression model is true

The building block for the simulation in Section 4.1 is model (2) with $\mu(x) = \sin(x)$, $\sigma(x) = 1/2$, and errors ε_t i.i.d. N(0,1) or two-sided exponential (Laplace) rescaled to unit variance. Knowledge that the variance $\sigma(x)$ is constant was not used in the estimation, i.e., $\sigma(x)$ was estimated from the data. For each distribution, 500 datasets each of size $n = 100$ were created with the design points $x_1, \ldots, x_n$ being equi-spaced on $(0, 2\pi)$, and N-W estimates of $\mu(x) = E(Y|x)$ and $\sigma^2(x) = Var(Y|x)$ were computed using a normal kernel in R.

Confidence intervals with nominal level $\alpha = 0.90$ were constructed using the two methods presented in Section 2.2: Traditional Model-Based (MB) and Predictive Residual Model-Based (PRMB); the two methods presented in Section 3.2: Model-Free (MF) of eq. (13), and Predictive Model-Free (PMF) from Remark 3.3; and the NORMAL approximation interval (4). The smoothing kernel Λ in eq. (8) was taken to be the standard normal density. All required bandwidths were computed by L_1 cross-validation. For each type of interval, the corresponding empirical coverage level (CVR) and average length (LEN) were recorded together with the (empirical) standard error associated with each average length.

Tables 1, 2, 3, and 4, summarize our findings, and contain a number of important features.

- The standard error of the reported coverage levels over the 500 replications is 0.013.
- By construction, this simulation problem has some symmetry that helps us further appreciate the variability of the CVRs. To elaborate, note that for any $x \in [0, \pi]$ we have $|\mu(x)| = |\mu(2\pi - x)|$ and the same symmetry holds for the derivatives of $\mu(x)$ as well due to the sinusoidal structure. Hence, the expected CVRs should be the same for $x_f = 0.15\pi$ and 1.85π in all methods. So for the NORMAL case of Table 1, the CVR would be better estimated by the average of 0.829 and 0.845, i.e., closer to 0.837; similarly, the PMF CVR for the same points could be better estimated by the average of 0.925 and 0.878, i.e., 0.902.
- The NORMAL intervals are characterized by under-coverage even when the true distribution is Normal. This under-coverage is more pronounced when $x_f = \pi/2$ or $3\pi/2$ due to the high bias of the kernel estimator at the points of a 'peak' or 'valley' that the normal interval (4) 'sweeps under the carpet'.
- The length of the NORMAL intervals is quite less variable than those based on bootstrap; this is not surprising since the extra randomization from the bootstrap is expected to inflate the overall variances.
- Although regression model (2) holds true here, the MB intervals show pronounced under-coverage; this is a phenomenon well-known in the bootstrap literature. As previously mentioned, the predictive residuals have generally larger scale than the fitted ones. Consequently, the PRMB intervals are wider, and manage to partially correct the under-coverage of the MB intervals.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.802	0.735	0.725	0.756	0.811	0.760	0.736	0.738	0.765
PRMB	0.860	0.821	0.813	0.826	0.878	0.840	0.806	0.812	0.837
MF	0.843	0.796	0.780	0.798	0.853	0.821	0.815	0.811	0.831
PMF	0.925	0.856	0.851	0.859	0.891	0.875	0.853	0.858	0.878
Normal	0.829	0.836	0.773	0.805	0.860	0.827	0.774	0.820	0.845

Table 1 Empirical coverage levels (CVR) of confidence intervals according to different methods at several x_f points spanning the interval $(0, 2\pi)$. Nominal coverage was 0.90, and sample size $n = 100$; error distribution: i.i.d. Normal.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.380	0.346	0.332	0.358	0.380	0.359	0.334	0.345	0.377
	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.004
PRMB	0.466	0.432	0.418	0.441	0.473	0.451	0.418	0.427	0.466
	0.008	0.009	0.010	0.007	0.010	0.011	0.009	0.008	0.009
MF	0.448	0.418	0.398	0.424	0.455	0.428	0.399	0.420	0.455
	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.005
PMF	0.518	0.487	0.468	0.490	0.513	0.492	0.470	0.487	0.517
	0.005	0.005	0.005	0.004	0.005	0.005	0.006	0.005	0.005
Normal	0.382	0.368	0.368	0.369	0.368	0.371	0.367	0.367	0.378
	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Table 2 (Average) lengths (LEN)—with standard errors below them—of the confidence intervals reported in Table 1.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.789	0.739	0.752	0.782	0.816	0.770	0.741	0.730	0.801
PRMB	0.860	0.858	0.852	0.878	0.905	0.865	0.852	0.857	0.886
MF	0.853	0.835	0.837	0.877	0.870	0.843	0.831	0.835	0.872
PMF	0.924	0.906	0.915	0.930	0.930	0.897	0.913	0.908	0.929
Normal	0.810	0.836	0.820	0.849	0.877	0.843	0.817	0.844	0.852

Table 3 As in Table 1 but with error distribution: i.i.d. Laplace.

- The performance of MF intervals is better than that of MB intervals despite the fact that the former are constructed without making use of eq. (2). However, as with the MB intervals, the MF intervals also show a tendency towards under-coverage.
- The PMF intervals appear to nicely correct the MF under-coverage in the Normal case although in the Laplace case they yield an over-correction. However, even with this over-correction, the PMF coverages are closer to the nominal in most entries of Tables 1 and 3 with only a few exceptions in Table 3 where the PRMB intervals are more accurate.

4.2 When a nonparametric regression model is not true

In this subsection, we investigate the performance of different confidence intervals in the absence of model (2). For easy comparison with Section 4.1, we will keep the same (conditional) mean and variance, i.e.,

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.362	0.332	0.316	0.348	0.370	0.340	0.304	0.323	0.354
	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.004
PRMB	0.455	0.417	0.398	0.436	0.465	0.424	0.382	0.407	0.444
	0.007	0.006	0.007	0.006	0.006	0.005	0.005	0.005	0.006
MF	0.424	0.394	0.372	0.411	0.441	0.405	0.359	0.388	0.426
	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005
PMF	0.494	0.461	0.440	0.473	0.500	0.468	0.431	0.455	0.489
	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.004
Normal	0.368	0.362	0.362	0.358	0.362	0.361	0.362	0.362	0.366
	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Table 4 As in Table 2 but with error distribution: i.i.d. Laplace.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.741	0.743	0.731	0.766	0.820	0.775	0.732	0.750	0.778
PRMB	0.813	0.817	0.815	0.844	0.892	0.865	0.825	0.845	0.862
MF	0.799	0.775	0.753	0.820	0.877	0.839	0.783	0.804	0.839
PMF	0.881	0.853	0.851	0.891	0.907	0.884	0.837	0.858	0.882
Normal	0.811	0.826	0.758	0.793	0.879	0.813	0.764	0.817	0.814

Table 5 As in Table 1 but with error distribution (15): non-i.i.d. skewed.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.361	0.332	0.314	0.343	0.370	0.349	0.322	0.335	0.366
	0.005	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003
PRMB	0.456	0.417	0.395	0.430	0.463	0.442	0.408	0.420	0.460
	0.009	0.008	0.007	0.006	0.006	0.007	0.008	0.006	0.007
MF	0.422	0.390	0.369	0.405	0.441	0.419	0.388	0.410	0.446
	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.005
PMF	0.492	0.460	0.437	0.467	0.499	0.479	0.452	0.472	0.506
	0.006	0.005	0.004	0.004	0.004	0.005	0.004	0.004	0.004
Normal	0.372	0.358	0.358	0.357	0.358	0.359	0.358	0.358	0.367
	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Table 6 As in Table 2 but with error distribution (15): non-i.i.d. skewed.

we will generate independent Y data such that $E(Y|x) = \sin(x)$, $Var(Y|x) = 1/2$, and design points $x_1, \ldots, x_{100}$ equi-spaced on $(0, 2\pi)$. However, the error structure $\varepsilon_x = (Y - E(Y|x))/\sqrt{Var(Y|x)}$ has skewness and/or kurtosis that depends on x, thereby violating the i.i.d. assumption. For our simulation, we considered

$$\varepsilon_x = \frac{c_x Z + (1 - c_x)W}{\sqrt{c_x^2 + (1 - c_x)^2}} \tag{15}$$

where $c_x = x/(2\pi)$ for $x \in [0, 2\pi]$, and $Z \sim N(0,1)$ independent of W that will either be distributed as $\frac{1}{2}\chi_2^2 - 1$ to capture a changing *skewness*, or as $\sqrt{\frac{3}{5}}\, t_5$, to capture a changing *kurtosis*; note that $EW = 0$ and $EW^2 = 1$.

Our results are summarized in Tables 5, 6, 7, and 8. The findings are qualitatively similar to those in Section 4.1. The PMF intervals are the undisputed winners here in terms of coverage accuracy. By contrast,

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.795	0.754	0.730	0.747	0.792	0.782	0.750	0.756	0.770
PRMB	0.868	0.840	0.809	0.843	0.870	0.852	0.834	0.837	0.845
MF	0.863	0.820	0.804	0.811	0.838	0.833	0.814	0.808	0.834
PMF	0.918	0.896	0.889	0.880	0.890	0.886	0.868	0.869	0.877
Normal	0.808	0.815	0.788	0.800	0.861	0.821	0.781	0.828	0.812

Table 7 As in Table 1 but with error distribution (15): non-i.i.d. kurtotic.

$x_f/\pi =$	0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
MB	0.363	0.339	0.327	0.353	0.380	0.360	0.332	0.341	0.372
	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003
PRMB	0.446	0.416	0.402	0.433	0.465	0.443	0.408	0.420	0.455
	0.007	0.007	0.007	0.007	0.006	0.006	0.007	0.005	0.006
MF	0.423	0.402	0.382	0.416	0.449	0.429	0.397	0.410	0.451
	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
PMF	0.496	0.469	0.452	0.480	0.508	0.489	0.461	0.477	0.507
	0.004	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Normal	0.377	0.365	0.365	0.365	0.365	0.367	0.365	0.365	0.373
	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Table 8 As in Table 2 but with error distribution (15): non-i.i.d. kurtotic.

the NORMAL and the MB bootstrap intervals show pronounced under-covarage; interestingly, these are the two methods that most practitioners use at the moment.

References

1. Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings, *Ann. Probab.*, vol. 9, no. 6, pp. 860-867.
2. Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, 7, 1-26.
3. Freedman, D.A. (1981). Bootstrapping regression models, *Ann. Statist.,* 9, 1218-1228.
4. Hall, P. (1993), On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation, *J. Roy. Statist. Soc., Ser. B*, 55, 291-304.
5. Härdle, W. and Bowman, A.W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands, *J. Amer. Statist. Assoc.*, 83, 102-110.
6. Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *Ann. Statist.*, 19, 778-796.
7. Li, Q. and Racine, J.S. (2007). *Nonparametric Econometrics*, Princeton Univ. Press, Princeton NJ.
8. Linton, O.B., Chen, R., Wang, N. and Härdle, W. (1997). An analysis of transformations for additive nonparametric regression, *J. Amer. Statist. Assoc.*, 92, 1512-1521.
9. Neumann, M. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression, *J. Nonparam. Statist.*, 9, 307-333.
10. Politis, D.N. (2010). Model-free Model-fitting and Predictive Distributions, Discussion Paper, Department of Economics, Univ. of California—San Diego. Retrievable from: `http://escholarship.org/uc/item/67j6s174`
11. Politis, D.N. (2013). Model-free Model-fitting and Predictive Distributions, to appear as a Discussion Paper in the journal *Test* in 2013.