

Tie the straps: Uniform bootstrap confidence bands for semiparametric additive models^{☆,☆☆}



Wolfgang Karl Härdle^{a,b}, Ya'acov Ritov^c, Weining Wang^{a,*}

^a C.A.S.E. – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany

^b Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore

^c Department of Statistics Hebrew University of Jerusalem Mount Scopus, Jerusalem 91905, Israel

ARTICLE INFO

Article history:

Received 9 November 2013

Available online 20 November 2014

AMS subject classifications:

62G08

62G09

62G20

62H12

62P20

Keywords:

Nonparametric regression

Bootstrap

Quantile regression

Confidence bands

Additive model

Robust statistics

ABSTRACT

This study considers the theoretical bootstrap “coupling” techniques for nonparametric robust smoothers and quantile regression, and we verify the bootstrap improvement. To handle the curse of dimensionality, a variant of “coupling” bootstrap techniques is developed for additive models both in a robust mean regression and in a quantile regression framework. Our bootstrap method can be used in many situations such as constructing confidence intervals and bands. We demonstrate the bootstrap improvement over the theoretical asymptotic band in simulations and in applications to firm expenditures and the interaction of economic sectors and the stock market.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Confidence bands are important tools for model specifications. However, it is difficult to construct precise confidence bands for nonparametric curves since a supreme norm is usually involved in the statistics. Traditional methods based on the asymptotic theory have natural drawbacks in their finite sample performance, and this motivates bootstrap methods to attain more precise bands. In this article, we deal with bootstrap bands construction for a general class of nonparametric M - and L -estimates; moreover, we adopt additive models to handle the multivariate covariates case. We believe that the developed technique is essential for componentwise shape inspection of additive models in empirical economics. Applications can be found in the work of [6] or [4].

Consider $Y, X \in \mathbb{R}^{d+1}$ with variable Y and $X \in \mathbb{R}^d$.

$$l(x) = \arg \min_{\theta} E_{Y|X=x} \rho(Y - \theta), \quad (1)$$

[☆] We thank the Editors and two referees for helpful comments. The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged.

^{☆☆} We appreciate the friendly working environment and atmosphere of the MFO Mathematical Research Institute of Oberwolfach during our Research in pairs stay.

* Corresponding author.

E-mail addresses: haerdle@wiwi.hu-berlin.de (W.K. Härdle), yaacov.ritov@gmail.com (Y. Ritov), wangwein@cms.hu-berlin.de (W. Wang).

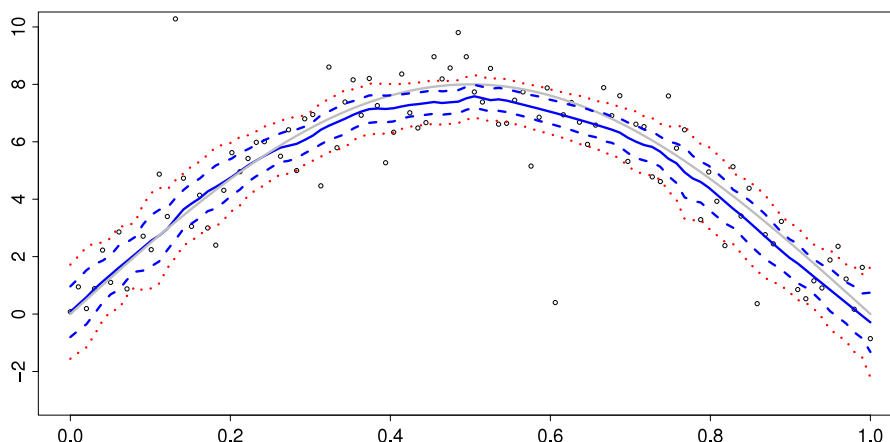


Fig. 1. Plot of true curve (gray), robust estimate with Tukey biweight loss and 5% confidence bands (blue dashed), local polynomial estimate (black), bootstrap band (red dotted), $n = 150$.

with $\rho(\cdot)$ as a loss function described in detail in Section 2.1. For the confidence bands construction, one stream of literature using empirical process theory follows the asymptotic results of [2], which is further extended by [19], and has recently been studied by [13] for L -smoothers. However, it has also been shown by [31] that such an asymptotic confidence band has a much lower coverage probability in a finite sample than what it is supposed to have. The poor performance of such a band in a finite sample has been well attributed to its slow convergence, see [7]. To address improvement of a finite sample performance, there is a class of literature on the bootstrap confidence band, see [3,12], among others. Fig. 1 shows a bootstrap confidence band and an asymptotic band for an M -smoother with $\rho(\cdot)$ being Tukey's bisquare loss. One can see that the asymptotic band is narrower than the bootstrap one. Moreover, the asymptotic band does not cover the true curve while the bootstrap band does.

Many different resampling techniques have been developed, among which the permutation tests are very popular and can be adapted to many application scenarios, see [28,1,22,23]. The bootstrap is a class of data driven resampling techniques that provide non-asymptotic approximations of finite sample distributions of different statistics. In a location model (more generally a regression model), resampling is done from the estimated residuals, and a typical theoretical analysis leads to the conclusion that “bootstrap works” in the sense that a suitably centered bootstrap estimator converges to the same asymptotic normal distribution as the original estimator under consideration. A large literature body has focused on showing bootstrap improvements and refinements of approximations via bootstrap resampling, see [8,21,15,11], which discuss the conditions for bootstrap consistency, and also prove the bootstrap accuracy as an approximation to the exact finite sample distribution for special types of statistics in a nonparametric framework. However, very few articles have focused on nonlinear statistics (e.g. maximum) in nonparametric regression. [31] proposes a bootstrap procedure and shows its improvement properties.

This stimulates the current research on finding common properties that loss functions have to share in order to attain such an improvement. Accordingly, we prove a generalized version in the univariate case for a class of loss function with bounded influence. The bootstrap becomes difficult when the dimension d of the regressors gets large. One way to avoid this problem is to impose a structure, such as an additive model, on the multivariate nonparametric function. The additive structure assumes that the covariates' effects are separable, and this effect is presented in many economic applications, [10]. Specifically, we consider the regression function

$$m(x_1, \dots, x_d) = \sum_{j=0}^d m_j(x_j), \quad (2)$$

with $m_0(x_0)$ a constant. It is worth noting that the additive structure implicitly assumes that the covariates effects are separable, and of course this assumption needs to be tested in advance. [32] develops a test on testing the interaction effects; correspondingly, our method can also be extended to implement similar tests.

It is well known that (2) avoids the “curse the dimensionality” in the sense that one-dimensional convergence rates are achieved for the estimation of $m(x_1, \dots, x_d)$, but keeps enough flexibility of the marginal influence of the different variables. See [14,17,16] among many others. [14] focuses on generalized additive models with unknown link functions, [17] proposes a two-stage estimation for quantile regression in additive models, [16] shows the equivalence between spline, kernel and other methods in terms of optimal minimax rate in the additive model estimation. The resulting estimate $\hat{m}_j(x_j)$ in (2) though needs to be screened for closeness to $m_j(x_j)$. This requires construction of confidence intervals and bands as a function of x_j . For such screening tests, our tightened bootstrap techniques will be verified. Namely, the bootstrap-based confidence bands are shown to be very close to the true finite sample distribution-based ones.

In summary, we investigate a coupling technique that allows us to “tie the straps” even a little tighter for a class of bounded influence estimators. Theoretically speaking, confidence band construction is made more precise in a variety of the estimation problems that we consider. The coupling idea is based on mimicking the distribution of the original data via a controllable random mechanism. Similar results like (9) will be derived for additive models.

The remainder of this paper is organized as follows. In Section 2, we explain the model setup and the bootstrap method in more detail. Section 3 presents the main results. In Section 4, a small simulation study is presented. Moreover, we show in Section 5 applications on managerial compensation and the impact on stock markets. Finally, Section 6 concludes with some comments and directions for future research.

2. Models and bootstrap confidence sets

This section describes the estimator and our coupling techniques, which motivate the obtainable theoretical results, and discusses some of the assumptions.

2.1. Univariate case

Let us describe the coupled bootstrap in the simple case of nonparametric minimum contrast curve estimation. Here $(X, Y)^\top \in \mathbb{R}^2$. The object of estimation is identified via:

$$l(x) = \arg \min_{\theta} E_{(Y|X=x)} \rho(Y - \theta), \quad (3)$$

where $\rho(\cdot)$ is a loss function of e.g. Hampel/Huber type or more generally (up to a constant) a negative (pseudo) log likelihood. In the quantile regression case, $\rho(x) = |x| \{\tau - \mathbf{1}(x \leq 0)\}$ is the check function. Another example for $\rho(\cdot)$ includes the trimmed mean, [18]

$$\rho(x) = \begin{cases} x^2/2, & |x| \leq k, \\ -k^2/2 + k|x|, & |x| > k \end{cases} \quad (4)$$

or a form of Winsorized mean:

$$\rho(x) = \begin{cases} x^2, & |x| \leq k, \\ k^2, & |x| > k. \end{cases} \quad (5)$$

A sample based version of (3) is:

$$\hat{l}_h(x) = \arg \min_{\theta} n^{-1} \sum_{i=1}^n \rho(Y_i - \theta) K_h(x - X_i), \quad (6)$$

where $K_h(u) = K(u/h)/h$ is a kernel function with bandwidth h . Now we generate a bootstrap sample using i.i.d. uniform random variables $U_1, \dots, U_n \in U(0, 1)$, and then generate:

$$Y_i^* = \hat{l}_g(X_i) + \varepsilon_i^*, \quad i = 1, \dots, n, \quad (7)$$

where $\varepsilon_i^* \sim \hat{F}_{(\varepsilon|X=X_i)}^{-1}(U_i)$ (discussed in detail in (24)) and $g = \mathcal{O}(n^{-1/9})$ a slightly larger bandwidth than h . The basic idea of coupling is based on comparing this sample to the pseudo observations:

$$Y_i^\# = l(X_i) + \varepsilon_i^\#, \quad i = 1, \dots, n, \quad (8)$$

where $\varepsilon_i^\# = F_{(Y|X=X_i)}^{-1}(U_i)$. Note that given $\{X_i\}_{i=1}^n$, the distribution of $Y_i^\#$ and Y_i is the same. We will show that for a class of loss functions the following approximation holds:

$$\sup_{x \in B} [\hat{l}_h^\#(x) - l(x) - \{\hat{l}_{h,g}^*(x) - \hat{l}_g(x)\}] = \mathcal{O}_p(h^2 \Gamma_n), \quad (9)$$

where B is a closed compact set in $[0, 1]$, Γ_n a slowly increasing sequence (a sequence a_n is slowly increasing if $n^{-\alpha} a_n \rightarrow 0$ for any $\alpha > 0$), $\hat{l}_h^\#(\cdot)$ is the nonparametric estimate calculated from $\{(X_i, Y_i^\#)\}$, $\hat{l}_{h,g}^*(X_i)$ is an estimate calculated from the bootstrap sample $\{(X_i, Y_i^*)\}$ with bandwidth h ,

$$\hat{l}_{h,g}^* \stackrel{\text{def}}{=} \arg \min_{\theta} \sum_{i=1}^n \rho(Y_i^* - \theta) K_h(x - X_i) \quad (10)$$

and $\hat{l}_g(X_i)$ is calculated as in (6) from the original sample with bandwidth g . The basic elements in proving (9) are smoothness of $F_{\varepsilon|X=x}(\cdot)$ and bounded influence of $\rho(\cdot)$ in (3).

2.2. Additive models and bootstrap confidence sets

For any $x \in \mathbb{R}^d$ ($d > 1$), the nonparametric approach in (6) is not appropriate when d is large, as the standard nonparametric optimal convergence rate, $\mathcal{O}_p(n^{-4/(4+d)})$, would be too slow when d is large. Additive models were suggested to remedy the problems posed by the dimensionality. Recall (2), and impose the additive structure:

$$Y_i = \sum_{j=0}^d m_j(x_{i,j}) + \varepsilon_i. \quad (11)$$

Further, approximate the additive model via a basis function approach:

$$m_j(x_{i,j}) \approx \sum_{l=1}^{L_j+1} a_{l,j} v_l(x_{i,j}),$$

where the $v_1(\cdot), v_2(\cdot), \dots$ could be any sequence of functions spanning the L^2 space. Our implementation uses B -splines, for example, linear B -splines: consider a sequence of H^{-1} equally spaced knots on the interval $[0, 1]$, which defines the width H subintervals $[lH, (l+1)H]$, $0 \leq l \leq (H^{-1} - 1)$. The linear B -spline basis is:

$$v_l(x) = \begin{cases} H^{-1}x - l + 1 & (l-1)H \leq x \leq lH, \\ l + 1 - H^{-1}x & lH \leq x \leq (l+1)H, \\ 0 & \text{otherwise.} \end{cases}$$

Denote the theoretical standardized B spline basis $\phi_l(\cdot)$, for the j th variable, $j = 1, \dots, d$,

$$\begin{aligned} \phi_{l,j}(x_j) &\stackrel{\text{def}}{=} v_l(x_j) - v_{l-1}(x_j) c_{l,j} / c_{l-1,j}, \\ B_{j,l}(x_j) &= \phi_{l,j}(x_j) / \|\phi_{l,j}(x_j)\|_2, \end{aligned}$$

where $l = 0, \dots, H^{-1}$, $c_{l,j} \stackrel{\text{def}}{=} \int \phi_{l,j}(u) f_j(u) du$ with $f_j(u)$ is the density for x_j , so that $EB_{j,l}(x_j) = 0$, $EB_{j,l}(x_j)^2 = 1$.

The additive estimate can then be obtained as follows. Define the vectors in $(\mathbb{R}^{H^{-1}d+1})$,

$$\begin{aligned} A &= (a_0, \mathbf{a}_1^\top, \dots, \mathbf{a}_d^\top)^\top, \\ \Phi(X_i) &= \{1, \mathbf{g}(x_{i,1})^\top, \dots, \mathbf{g}(x_{i,d})^\top\}^\top, \end{aligned}$$

where

$$\begin{aligned} \mathbf{a}_j &= (a_{1,j}, \dots, a_{H,j})^\top, \\ \mathbf{g}(x_{i,j})^\top &= \{v_1(x_{i,j}), \dots, v_H(x_{i,j})\}^\top. \end{aligned}$$

Finally, let $\hat{A}$ be the estimation of A :

$$\hat{A} = \arg \min_A \sum_{i=1}^n \rho\{Y_i - A^\top \Phi(X_i)\}, \quad (12)$$

and

$$\hat{m}_j(x_{i,j}) = \hat{\mathbf{a}}_j^\top \mathbf{g}(x_{i,j}), \quad (13)$$

with $j = 1, \dots, d$.

2.3. Coupled bootstrap for cases of high dimensional covariates

The additive structure in (11) is one solution to the curse of dimensionality, however, the bootstrap approach in (7) does not work for this modeling scenario as nonparametric estimation of $F_{\varepsilon|X}(\cdot)$ would again run into the “curse of dimensionality” problem. We suggest another bootstrap technique, and prove that it strongly approximates a model with the same asymptotic properties as the original model.

Define

$$Z_i = \begin{cases} 1 & \text{with prob } \tau \\ -1 & \text{with prob } 1 - \tau, \end{cases} \quad i = 1, \dots, n. \quad (14)$$

This includes the special case for symmetric error distributions with $\tau = 1/2$, which is the usual assumption for mean or robust M -smoothers. Moreover, it generally adapts to the case of quantile regression for asymmetric error distributions. The bootstrap couple $(\varepsilon^*, \varepsilon^\#)$, the bootstrap residuals and its associate theoretical couple respectively are:

$$\varepsilon_i^* \stackrel{\text{def}}{=} Z_i |\hat{\varepsilon}_i| \quad (15)$$

$$\varepsilon_i^\# \stackrel{\text{def}}{=} Z_i \eta_i, \quad i = 1, \dots, n, \quad (16)$$

where

$$\eta_i \stackrel{\text{def}}{=} F_{i,Z_i}^{-1}\{F_{i,\text{sgn}(\varepsilon_i)}(|\varepsilon_i|)\}, \quad i = 1, \dots, n, \quad (17)$$

and

$$F_{i,s}(t) \stackrel{\text{def}}{=} P(|\varepsilon_i| \leq t | s\varepsilon_i > 0), \quad i = 1, \dots, n, \quad s \in \{1, -1\}. \quad (18)$$

Note that the same Z_i appears both in (15) and in (16).

Recall that $F_{Y|X=X_i}\{l(X_i)\} = \tau$ and, hence, $F_{\varepsilon|X=X_i}(0) = \tau$. Now, it is easy to see that $V_i \stackrel{\text{def}}{=} F_{i,\text{sgn}(\varepsilon_i)}(|\varepsilon_i|)$ has a standard uniform distribution, and if Z_i is as above, then ε_i and $Z_i F_{i,Z_i}^{-1}(V_i)$ have the same distribution. Formally, note that

$$F_{i,+1}(t) = \frac{F_i(t) - 1 + \tau}{\tau},$$

$$F_{i,-1}(t) = \frac{1 - \tau - F_i(-t)}{1 - \tau},$$

where $F_i(\cdot)$ is the cdf of ε_i . Hence, for $t > 0$:

$$\begin{aligned} P(\varepsilon_i^\# < t) &= \tau P[F_{i,+1}^{-1}\{F_{i,\text{sgn}(\varepsilon_i)}(|\varepsilon_i|)\} < t] + 1 - \tau \\ &= \tau P\{F_{i,\text{sgn}(\varepsilon_i)}(|\varepsilon_i|) < F_{i,+1}(t)\} + 1 - \tau \\ &= \tau P\{\varepsilon_i < 0, F_{i,-1}(-\varepsilon_i) < F_{i,+1}(t)\} + \tau P\{\varepsilon_i > 0, F_{i,+1}(\varepsilon_i) < F_{i,+1}(t)\} + 1 - \tau \\ &= \tau P\left\{\varepsilon_i < 0, \frac{1 - \tau - F_i(\varepsilon_i)}{1 - \tau} < \frac{F_i(t) - 1 + \tau}{\tau}\right\} + \tau P(0 < \varepsilon_i < t) + 1 - \tau \\ &= \tau P\left[1 - \tau > F_i(\varepsilon_i) > \frac{1 - \tau}{\tau}\{1 - F_i(t)\}\right] + \tau P(0 < \varepsilon_i < t) + 1 - \tau \\ &= \tau \left[1 - \frac{1 - \tau}{\tau}\{1 - F_i(t)\} - \tau\right] + \tau\{F_i(t) - 1 + \tau\} + 1 - \tau \\ &= F_i(t). \end{aligned}$$

The case $t < 0$ is dealt similarly. It follows

$$\mathcal{L}(\varepsilon_i^\#) = \mathcal{L}(\varepsilon_i). \quad (19)$$

Our confidence “ideal” interval is conditional on $\{V_i\}_{i=1}^n$ which has a direct link to the absolute value of errors $\{|\varepsilon_i|\}_{i=1}^n$. Note, however, that the estimator is asymptotically consistent and its bias does not depend on these absolute values. Moreover, by the law of large numbers, the pointwise width of the conditional confidence interval is within a factor of $1 + \mathcal{O}_p(1)$ of the unconditional one.

2.4. How does the coupling work?

The basic idea of our approach is to construct an empirically feasible bootstrap sample that strongly approximates a sample from the true distribution. One example of the coupled bootstrap approach has already been explained in (7) and (8). It, however, relies on estimators of the conditional distribution $F_{Y|X=x}(\cdot)$, which become very imprecise when d is large.

Another approach proposed in Section 2.3 motivated as the wild bootstrap is based on randomizing the obtained residuals and using the same random source to mimic the stochastic of the unobservable errors. To get the basic idea, let us assume for a moment that the distributions of ε_i are symmetric. Then the coupling may be performed via a Rademacher randomized variable Z_i with

$$P(Z_i = 1) = P(Z_i = -1) = 1/2$$

and generation of the couple ε_i^* (the bootstrapped residuals), $\varepsilon_i^\#$ (the theoretical coupling), where $\{\varepsilon_i^*, \varepsilon_i^\#\}$ is:

$$\varepsilon_i^* \stackrel{\text{def}}{=} Z_i |\hat{\varepsilon}_i|$$

$$\varepsilon_i^\# \stackrel{\text{def}}{=} Z_i \eta_i. \quad (20)$$

With this construction, we are able to establish a result similar to (9).

In a non symmetric distribution (required for quantile regression), one defines Z_i with $P(Z_i = 1) = \tau$ and $P(Z_i = -1) = 1 - \tau$ assuming the centering $F_{Y|X_i}\{l(X_i)\} = \tau$, and the couple $(\varepsilon_i^*, \varepsilon_i^\#)$ is given by (15) and (16). It was argued that the distributions of $\varepsilon_i^\#$ and ε_i are identical and also the conditional distributions given $\{V_i\}_{i=1}^n$ are the same.

The resampling technique will be applied to a nonparametric estimation of an additive quantile regression model. The reanalysis of the data used by [17] provides us with sharper bands that have not been calculated in that paper.

2.5. Extension to generalized linear models

The model in Section 2 can be extended to generalized linear models, with the relation $g\{E(Y|X)\} = l(X)$, where $g(\cdot)$ is the known prespecified link function. For continuous random variable Y , the extension to the above bootstrap procedure is trivial. Moreover, it can be also generalized to the models with discrete Y_i . For example, for the binary logistic model with Y_i as binary data, define

$$\hat{\varepsilon}_i = Y_i - g\{\hat{l}_h(X_i)\}, \quad (21)$$

then $\hat{\varepsilon}_i$ will be bounded in $[-1, 1]$, and one can apply the same bootstrap procedure as in (20).

3. Main results

The section gives asymptotic results for the estimators described in Section 2. To establish the asymptotic property, some assumptions are needed:

Assumptions

A.1 The solution $l(\cdot)$ of (1) is two-times differentiable and bounded. We abbreviate $\psi(Y, \cdot)$ to $\psi(\cdot)$, and $\psi(\cdot) = \rho'(\cdot)$ (or subgradient in the quantile regression case) is a.s. bounded by $M < \infty$. $E\{\psi(\varepsilon)|X\} = 0$ w.p. 1. Also $\psi(\cdot)$ is Lipschitz continuous except for a finite number of points in the compact set B .

A.2 Assume the support of X is $[0, 1]^d$, the conditional density of $\varepsilon f_{(\varepsilon|X=\cdot)}(\cdot|\cdot)$, and is bounded by C_1 and is twice differentiable in the interior point of $[0, 1]^d$, and bounded away from 0 by $c_1 > 0$.

A.3 The kernel function $K_h(\cdot)$ is a product kernel composed from one dimensional kernels with bandwidth h :

$$K_h(s) = \prod_{j=1}^d K(s_j/h)/h, \quad s = (s_1, \dots, s_d)^\top \in \mathbb{R}^d. \quad (22)$$

A.4 The kernel bandwidth satisfies $h \sim n^{-1/5}$. Let g be another bandwidth sequence $g \gg h$, e.g., $g = \mathcal{O}(n^{-1/9})$.

A.5 For each j , the true regression function $m_j(\cdot)$, $j \in 1, \dots, d$, is at least one time continuous differentiable function on $[0, 1]$.

A.6 $E\{g_j^2(X_j)\} = 1$ for $j \in 1, \dots, d$, $\|\Phi_l(X_j)\|_\infty \leq C_3 H$, a.s., where $\Phi_l(X_j) \stackrel{\text{def}}{=} \{\phi_l^2(x_{1,j}), \dots, \phi_l^2(x_{n,j})\}^\top$, with $j \in 1, \dots, d$.

A.7 The number of regressors in (12) is of $\mathcal{O}(p)$, where $p = dH^{-1} + 1$ with $H^{-1} \sim n^{1/5}$, and $d = \mathcal{O}(n^{2/3})$.

A.1 is about the continuity and the bounded influence structure of the loss function, it is quite essential for proving the bootstrap improvement. Also, A.1 includes a very general class of regression loss functions although it does not include the usual mean regression loss. A.2 assumes W.L.O.G., the covariates are on bounded support and impose assumption on the conditional density of the error term. A.3 is a standard assumption on the kernel function. A.4 is about theoretical rates of the bandwidths h and g . h is of the standard optimal rate in nonparametric regression, while g is required to be larger as we need to mimic the bias of estimation in the original nonparametric regression, see [12]. A.5–A.7 are assumptions on additive models. A.5 assumes that additive components behave properly. A.6 imposes conditions on the basis functions, and the linear B -spline satisfies A.6. The above assumptions are all very mild or standard assumptions according to the literature.

We show first convergence results for the bootstrap methods in (7) and (8). The resampling step has been defined in (7), where the smooth estimate of the conditional distribution is:

$$\tilde{F}_{(\varepsilon|X=x)}(t) \stackrel{\text{def}}{=} \sum_{i=1}^n W_{h,i}(x) \mathbf{1}[\{Y_i - \hat{l}_h(X_i)\} \leq t], \quad (23)$$

with $W_{h,i}(x) = n^{-1} K_h(x - X_i) / \hat{f}_h(x)$ and $\hat{f}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i)$ the kernel density estimator. To have a correctly centered estimation for $F_{(\varepsilon|X=x)}(t)$, we define

$$d\hat{F}_{(\varepsilon|X=x)}(t) = \begin{cases} \frac{d\tilde{F}_{(\varepsilon|X=x)}(t)}{\tilde{F}_{(\varepsilon|X=x)}(0) + C_0(1 - \tilde{F}_{(\varepsilon|X=x)}(0))} & \text{if } t < 0, \\ \frac{C_0 d\tilde{F}_{(\varepsilon|X=x)}(t)}{\tilde{F}_{(\varepsilon|X=x)}(0) + C_0(1 - \tilde{F}_{(\varepsilon|X=x)}(0))} & \text{otherwise} \end{cases} \quad (24)$$

where C_0 is a constant defined as

$$C_0 \stackrel{\text{def}}{=} \frac{\int_{-\infty}^0 \psi(u) d\tilde{F}_{(\varepsilon|X=x)}(u)}{\int_0^\infty \psi(u) d\tilde{F}_{(\varepsilon|X=x)}(u)}.$$

Note that the estimator in (24) is centered so that

$$E_{\tilde{F}_{(\varepsilon|X=x_i)}} \psi(\varepsilon_i^*) = 0 = E_{F_{(\varepsilon|X=x_i)}} \psi(\varepsilon_i^*). \quad (25)$$

The influence function of the estimator is proportional to $\psi(\cdot) = \rho'(\cdot)$. If it is bounded with bounded derivatives a.e. and $\mathcal{L}(\varepsilon|X)$ is such that $\|\hat{F}_{(\varepsilon|X_i=X)}(\cdot) - F_{(\varepsilon|X=X_i)}(\cdot)\|_\infty = \mathcal{O}(h^2 \Gamma_n)$, then a similar coupling argument as in (7) can be used. Recall $\varepsilon_i^* = \hat{F}_{(\varepsilon|X_i=X)}^{-1}(F_{(\varepsilon|X_i=X)}(\varepsilon_i^*))$, then

$$|\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\#)| = \mathcal{O}_p(h^2 \Gamma_n). \quad (26)$$

This ensures that

$$n^{-1} \sum_{i=1}^n \{\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\#)\} K_h(x - X_i) = \mathcal{O}_p(h^2 \Gamma_n). \quad (27)$$

The argument is based on two facts. First from (25) the means are zero and second that (26) holds.

Once the Y_i^* s are generated, one applies (10) to the bootstrap data $\{(X_i, Y_i^*)\}_{i=1}^n$ to obtain $\hat{l}_{h,g}^*(x)$. Summarizing, we have:

Theorem 3.1. *Let assumptions A.1–A.4 be fulfilled and define $\hat{l}_h^\#(\cdot)$ as in (9). Then*

$$\sup_{x \in B} |A_n(x)| = \mathcal{O}_p(h^2 \Gamma_n), \quad (28)$$

where

$$A_n(x) \stackrel{\text{def}}{=} (\hat{l}_h^\# - l)(x) - \{(\hat{l}_{h,g}^* - \hat{l}_g)(x)\}. \quad (29)$$

Let assumptions A.1–A.7 be fulfilled and consider the additive model of (11) with the estimator (12), and the resampling scheme is considered as in (15) and (16), then

$$\sup_{x \in B} |(\hat{m}_j^\# - m_j)(x) - \{(\hat{m}_j^* - \hat{m}_j)(x)\}| = \mathcal{O}_{a.s.}(H^2 \Gamma_n).$$

Remark 1. The aforementioned strong approximation results mean that the stochastic behavior of $\hat{l}_h^\#(x) - l(x) ((\hat{m}_j^\# - m_j)(x))$ is well approximated by its bootstrap counterpart $\hat{l}_{h,g}^*(x) - \hat{l}_g(x) ((\hat{m}_j^* - \hat{m}_j)(x))$. This implies in particular that the distribution of $\sup_x |\hat{l}_h^\#(x) - l(x)| (\sup_{x \in B} |(\hat{m}_j^\# - m_j)(x)|)$ is consistently approximated by that of $\sup_x |\hat{l}_{h,g}^*(x) - \hat{l}_g(x)| (\sup_{x \in B} |(\hat{m}_j^* - \hat{m}_j)(x)|)$. Also the rate $H^2 \Gamma_n$ is sufficient for the validity of the bootstrap for supremum-functionals, see [24]. Therefore, the bootstrap confidence band is a direct consequence of the results.

Remark 2. A similar result was proved by [31] for quantile regression. There the centering ensures that the bootstrap error distribution has the proper quantile. Here it is generalized to a wider class of centering, and to additive models.

4. Simulation

This section is divided into two parts. First, we concentrate on the univariate $x \in [0, 1]$ case and the bootstrap procedure (7), (8), check the validity of the bootstrap procedure, and compare it with asymptotic uniform confidence bands. Second, we adopt the bootstrap procedure for the additive model as in (20), and check the validity of the bootstrap band in the same fashion.

4.1. Univariate case

The simulation setup in the univariate case is:

- (1) Simulate (X_i, Y_i) , $i = 1, \dots, n$ according to the predefined joint probability density function (pdf) $f(x, y)$. In order to compare with [9], we set the joint pdf of (X, Y) as,

$$f(x, y) = g\{y - \sin(\pi x)\} \mathbf{1}(x \in [0, 1]), \quad (30)$$

$$g(u) = 9\varphi(u)/10 + \varphi(u/9)/90. \quad (31)$$

- (2) Compute $\hat{l}_h(x)$ as in (6), with $\rho(\cdot)$ as a biweight loss, $\hat{\varepsilon}_i \stackrel{\text{def}}{=} Y_i - \hat{l}_h(X_i)$.

- (3) Compute the estimated conditional edf as in (A.2) with a Gaussian kernel

$$K_h(u) = (\sqrt{2\pi})^{-1} \exp\{-u^2/2h\}/h,$$

and $h = 0.06$ is selected by cross validation.

Table 1

Relative errors and areas (estimated as averaged width of the confidence bands over the grid points) of asymptotic and bootstrap with 1000 repetitions per sample, and 5000 samples. Relative error $\stackrel{\text{def}}{=} |(\text{true coverage} - \text{nominal coverage})|/\text{nominal coverage}$. Standard deviations measuring variability over samples are shown in brackets.

	n	95%		90%	
		Rel. Err.	Area	Rel. Err.	Area
Boot.	100	0.02(0.015)	2.52(0.547)	0.02(0.014)	2.21(0.746)
	200	0.01(0.010)	2.01(0.387)	0.02(0.010)	1.82(0.679)
	400	0.01(0.004)	1.33(0.201)	0.01(0.003)	1.13(0.301)
Asym.	100	0.09(0.034)	1.22(0.596)	0.11(0.070)	1.02(0.711)
	200	0.06(0.026)	0.91(0.602)	0.10(0.051)	0.67(0.518)
	400	0.05(0.018)	0.72(0.314)	0.09(0.043)	0.68(0.325)

(4) For each $i = 1, \dots, n$, generate random variable $\varepsilon_i^* \sim \hat{F}_{(\varepsilon|X)}(t)$, $i = 1, \dots, n$:

$$Y_i^* = \hat{l}_g(X_i) + \varepsilon_i^*,$$

with $g = 0.2$.

(5) For each sample $\{X_i, Y_i^*\}_{i=1}^n$, compute $\hat{l}_{h,g}^*(\cdot)$ and the random variable

$$d_i^* \stackrel{\text{def}}{=} \sup_{x \in B} [|\hat{l}_{h,g}^*(x) - \hat{l}_g(x)| \sqrt{\hat{f}_X(x) \hat{f}_{(\varepsilon|X=x_i)}(\varepsilon_i^*)} / \sqrt{\hat{E}_{Y|X}\{\psi^2(\varepsilon_i^*)\}}].$$

(6) Calculate the $1 - \alpha$ quantile d_α^* of $d_1, \dots, d_n^*$.

(7) Construct the bootstrap uniform band centered around $\hat{l}_h(x)$

$$\hat{l}_h(x) \pm [\sqrt{\hat{f}_X(x) \hat{f}_{(\varepsilon|X=x_i)}(\varepsilon_i^*)} / \sqrt{\hat{E}_{Y|X}\{\psi^2(\varepsilon_i^*)\}}]^{-1} d_\alpha^*.$$

Fig. 1 shows the theoretical signal curve, the robust estimate using Huber loss function with a corresponding 95% uniform confidence band from the asymptotic theory and the confidence band using the bootstrap method. The real curve is marked as the gray solid line. We then compute the confidence band based on asymptotic theory according to [9]. Here, we notice that the asymptotic band is narrower than the bootstrap band. The width of the bands has not been affected by outliers since we adopt a robust estimation. To compare which method is more precise, Table 1 presents respectively the simulated coverage probabilities together with the calculated area of the 95% and 90% confidence band, for sample sizes $n = 100, 200, 400$. (Practically we would like to keep the sample size larger than 100 to achieve a reasonably precise estimation.) 5000 simulation runs are carried out and 1000 bootstrap samples are generated for each simulation. From Table 1, we observe that, for the asymptotic method, coverage probabilities improve with increased sample size and the bootstrap method (shown in brackets) obtains a larger coverage probability than the asymptotic one, as it has smaller relative errors. It is also observed that the sizes of the bands decrease with increased sample sizes. Overall, the bootstrap method displays a better convergence rate, while not sacrificing much on the width of the bands.

4.2. Additive model

We now extend the study to the case of multivariate covariates, where we use an additive model for the estimation. The bootstrap procedure is as follows:

(1) Simulate (X_i, Y_i) , $i = 1, \dots, n$ following model (11). The variables $x_1, x_2, x_3, x_4 \sim U(-2.5, 2.5)$,

$$m_1(x_1) = \sin(\pi x_1), \quad m_2(x_2) = \Phi(3x_2), \quad m_3(x_3) = x_3^3, \quad m_4(x_4) = x_4^4,$$

and ε_i is simulated from a mixture normal density function with density $\varphi(u/9)/90 + \varphi(u)/10$.

(2) Compute the estimation $\hat{m}_1(x_1), \hat{m}_2(x_2), \hat{m}_3(x_3), \hat{m}_4(x_4)$ via (13) and $\hat{\varepsilon}_i = Y_i - \sum_{j=1}^4 \hat{m}_j(x_{i,j})$.

(3) For each $i = 1, \dots, n$, generate random variable ε_i^* , $i = 1, \dots, n$ as in (15):

$$Y_{i,i^*} = \sum_{j=1}^4 \hat{m}_j(x_{i,j}) + \varepsilon_i^*.$$

(4) For each sample $\{x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}, y_i^*\}$, compute $\hat{m}_j^*(\cdot)$ and the random variable

$$d_i^* \stackrel{\text{def}}{=} \sup_{x \in [-2.5, 2.5]} \{ \sqrt{\hat{f}_{X_j}(x) \hat{f}_{(\varepsilon|X_j=x_{i,j})}(\varepsilon_i^*)} / \sqrt{\hat{E}_{Y|X_j=x_{i,j}}\{\psi^2(\varepsilon_i^*)\}} | \hat{m}_j^*(x) - \hat{m}_j(x) | \}.$$

(5) Calculate the $1 - \alpha$ quantile d_α^* of $d_1, \dots, d_n^*$.

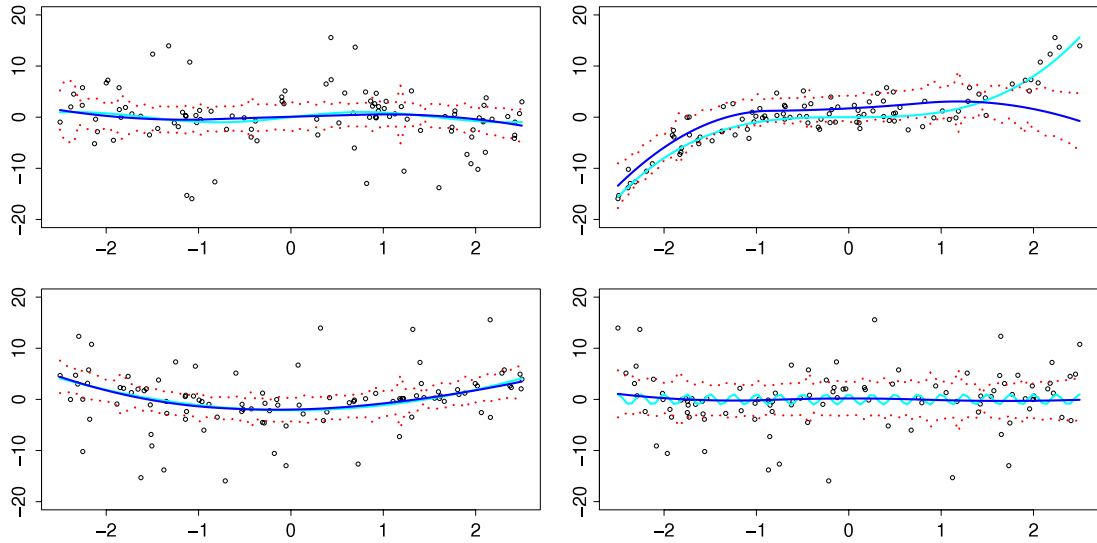


Fig. 2. Plot of true curve (dark blue), robust estimates and bands (cyan), bootstrap band (red dotted). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Relative errors and areas of bootstrap bands for $\hat{m}_1(\cdot)$, $\hat{m}_2(\cdot)$, $\hat{m}_3(\cdot)$, $\hat{m}_4(\cdot)$ with 1000 repetitions per sample, and 5000 samples. Standard deviations measuring variability over samples are shown in brackets.

		n	$\hat{m}_1(\cdot)$	$\hat{m}_2(\cdot)$	$\hat{m}_3(\cdot)$	$\hat{m}_4(\cdot)$
Rel. Err.	95%	100	0.08(0.051)	0.05(0.027)	0.08(0.037)	0.09(0.020)
		200	0.07(0.042)	0.03(0.015)	0.02(0.018)	0.04(0.019)
		400	0.04(0.023)	0.02(0.012)	0.01(0.010)	0.03(0.012)
	90%	100	0.06(0.050)	0.09(0.079)	0.06(0.066)	0.04(0.022)
		200	0.04(0.038)	0.05(0.026)	0.03(0.048)	0.02(0.014)
		400	0.03(0.028)	0.02(0.013)	0.02(0.021)	0.01(0.010)
Area	95%	100	6.76(1.291)	6.99(1.213)	6.87(1.241)	6.69(1.546)
		200	5.54(1.112)	4.84(1.007)	4.98(1.115)	4.79(1.134)
		400	4.56(0.988)	3.78(1.001)	3.21(0.943)	3.76(1.019)
	90%	100	5.99(1.667)	5.45(1.472)	5.75(1.987)	6.01(1.654)
		200	4.84(1.331)	4.38(1.112)	4.13(1.198)	4.11(1.219)
		400	3.51(0.969)	3.23(0.989)	2.98(0.823)	3.09(1.012)

(6) Construct the bootstrap uniform band centered around $\hat{m}_j(x_j)$

$$\hat{m}_j(x_{i,j}) \pm [\sqrt{\hat{f}_{x_{i,j}}(x_j) \{ \hat{f}_{(e|X_j=x_{i,j})}(e_i^*) \}} / \sqrt{\hat{E}_{Y|X_j=x_{i,j}} \{ \psi^2(e_i^*) \}}] d_\alpha^*.$$

The estimation of $\hat{m}_j(x_j)$ s ($j = 1, \dots, 4$) and their bootstrap confidence bands is shown in Fig. 2.

The simulated coverage probabilities are shown in Table 2. The coverage probabilities are close to the nominal levels and the widths of the bands are clearly shrinking w.r.t. the sample sizes.

5. Empirical analysis

5.1. Firm expenses analysis

[35] uses a sample of Japanese firms in the chemical industry to examine whether a concentrated shareholding is associated with lower expenditure on activities with scope for managerial private benefits. We focus on the same regression problem as proposed in [17]. The dependent variable Y is: general sales and administrative expenses deflated by sales (denoted by MH5), which is one of five measures of expenditures on activities with scope for managerial private benefits considered. The covariates are: ownership concentration (denoted by TOPTEN, cumulative shareholding by the largest ten shareholders), and firm characteristics: the log of assets, firm age, and leverage (the ratio of debt to debt plus equity), the sample size is 185. The regression model we consider here is:

$$MH5 = m_0 + m_1(TOPTEN) + m_2\{\log(Assets)\} + m_3(Age) + m_4(Leverage) + error.$$

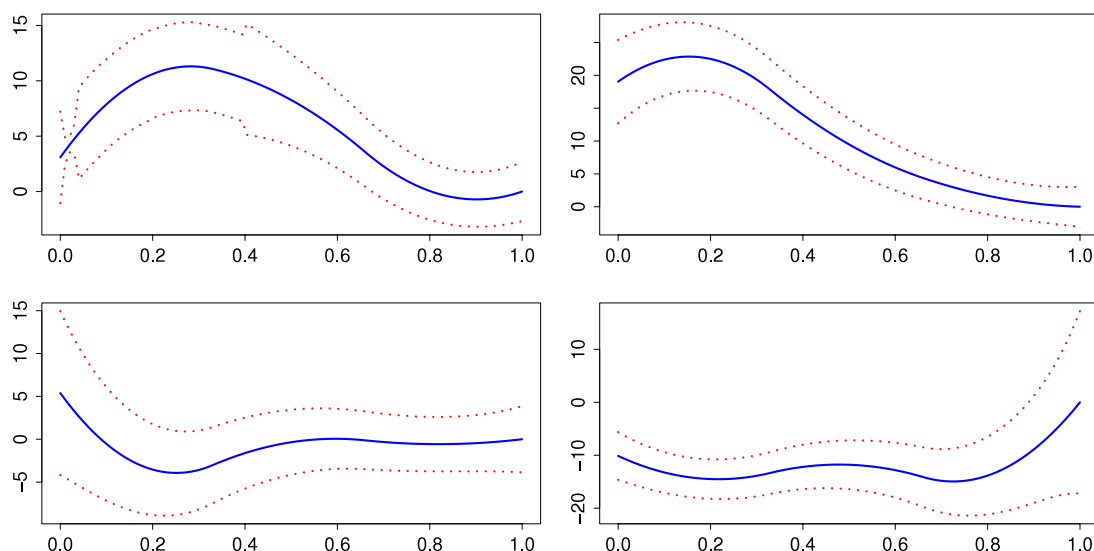


Fig. 3. Robust estimates (blue), bootstrap bands (red dotted), left up: Log(Asset), right up: Leverage, left below: Age, right below: TOPTEN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The estimated additive components and its bootstrap confidence bands are shown in Fig. 3. Similarly, it can be seen that the nonlinear effects are log(asset) and TOPTEN, and the firm age effects are minor compared to the other three. The effect of leverage is also a little bit nonlinear, and the shape of the curves deviates from what [17] presents, especially for the effect of TOPTEN. This may be due to the different subjects studied: in our case, robust estimation with Tukey biweight loss, while in their case the conditional median curve.

5.2. The impact on the stock market

We analyze how four assets (oil, currency, bonds, real estate) affect the stock market. This study would give implications on the interactions of the economic conditions among different sectors. The data source is ProQuest Statistical Datasets, and we focus on the United States market. Therefore, the covariates are taken as: the crude oil index, EUR–USD exchange rate, the 10-year treasury constant maturity inflation index %, the real estate index, and the Y variable is S&P 500 index returns. The data are synchronized to a weekly frequency. We selected the data during the period of 09/03/2008–11/28/2011.

It can be observed that all of the four markets have nonlinear effects on the stock indices values, Fig. 4, but only the exchange rate EUR–USD and crude oil prices affect the stock indices returns nonlinearly, Fig. 5. It is not difficult to interpret the relationships: In Fig. 4, for the exchange rate EUR–USD, the weakness of EUR up to a certain level (< 1.27) is negatively correlated with the stock indices, and then a positive correlation follows, but this relationship is again reversed when the EUR is too high (> 1.43). Oil prices have a negative impact on stock indices at every level, but the effects decrease when the prices rise. As for the inflation index, when the inflation rate is high, interest rates are typically high, this may reduce the consumption and investments in the stock market. One can see a negative correlation there when the inflation index is larger than 0.7. Finally, an increasing real estate index can be a sign of booming economic conditions, therefore the stock indices rise when the real estate index gets higher. However, when the real estate index is too high, it is likely that there exists a bubble, so a drop in the market indices is seen.

In Fig. 5, we see a difference in effects on S&P log returns, exchange rate EUR–USD is positively correlated with returns until a high level (> 1.40), the crude oil has major negative effects on stock returns. More nonlinearity is presented in the plots for the inflation index and the real estate index.

6. Conclusion

We have developed and proved the bootstrap improvement for a wide class of smoothers with bounded influence functions. Moreover, we extended our results to additive models in order to cope with the curse of dimensionality. Our bootstrap method can be further extended to serve various purposes in additive models. It may be used for componentwise hypothesis testing, testing the separability or interaction and bias correction, see [32]. Furthermore, similar bootstrap improvement results can be extended to other semiparametric models, e.g. partial linear models, and single index models. Moreover, different types of resampling procedures can be adopted to directly resample (Y_i, X_i) , such as in [27,22]. And, according to [1], we are able to deal with cases with discrete covariates.

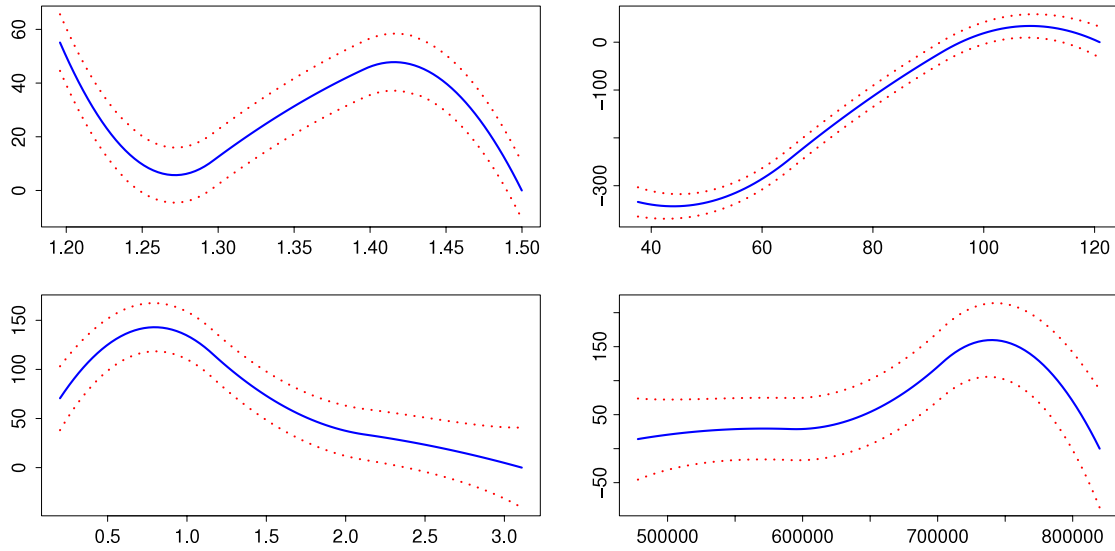


Fig. 4. Robust estimate (blue), bootstrap bands (red dotted), Y : S&P index, left up: exchange rates EUR–USD, right up: crude oil price, left below: inflation index, right below: real estate index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

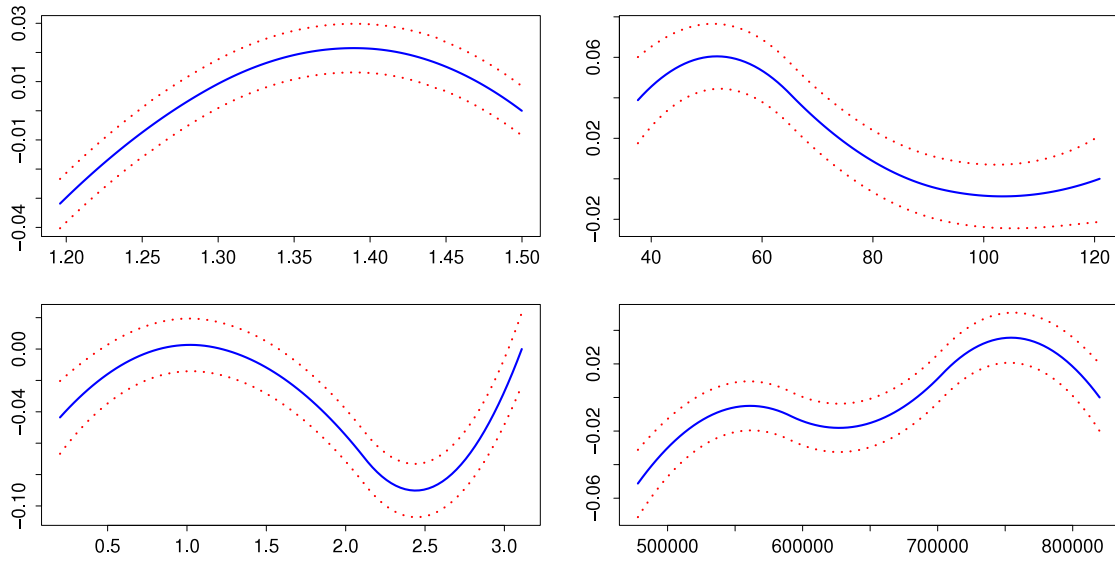


Fig. 5. Robust estimate (blue), bootstrap bands (red dotted), Y : S&P index log return, left up: exchange rates EUR–USD, right up: crude oil price, left below: inflation index, right below: real estate index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgment

The financial support from ISF (Complex statistical models and time (ISF, 2010–2014)) is greatly acknowledged.

Appendix. Proof

A.1. Proof of Theorem 3.1 in the $d = 1$ case

To prove Theorem 3.1, we show finally, with (optimal rate) $h = \mathcal{O}(n^{-1/5})$:

$$\max_i [(\hat{l}_h^* - l)(X_i) - \{(\hat{l}_{h,g}^* - \hat{l}_g)(X_i)\}] = \mathcal{O}_p(h^2 \Gamma_n). \quad (\text{A.1})$$

This is extended to any point x in a compact set B in (A.12).

Proving (A.1) can be done by showing first:

$$\max_i |\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\#)| = \mathcal{O}_p(h^2 \Gamma_n). \quad (\text{A.2})$$

Recall the definition of $\tilde{F}_{\psi|X}(\cdot)$ in (24), it is the induced conditional cdf of $\{\psi(\varepsilon_i^*)\}_{i=1}^n$. By [5] we have for a small b :

$$\sup_{|t| \leq b, i=1, \dots, n} |\tilde{F}_{\psi|X}(t) - F_{(\varepsilon|X)}(t)| = \mathcal{O}_p(h^2 \Gamma_n). \quad (\text{A.3})$$

Also recall (24) $\hat{F}_{(\varepsilon|X)}$ is a scaled version of $\tilde{F}_{\psi|X}$. Then according to A.2, the inverse function of $\hat{F}_{(\varepsilon|X)}(\cdot)$ and $F_{(\varepsilon|X)}(\cdot)$ will also be close in the sense that

$$\max_i |\varepsilon_i^\# - \varepsilon_i^*| = \max_i |F_{(\varepsilon|X)}^{-1}(U_i) - \hat{F}_{(\varepsilon|X)}^{-1}(U_i)| = \mathcal{O}_p(h^2 \Gamma_n). \quad (\text{A.4})$$

Moreover, for a $\delta_i = \mathcal{O}_p(h^2 \Gamma_n)$ uniformly, define $F_{\psi|X}^{-1}(\cdot)$ as the conditional c.d.f. of $\psi(\varepsilon)$ on X ,

$$\begin{aligned} |\psi(\varepsilon_i^\#) - \psi(\varepsilon_i^*)| &= |F_{\psi|X}^{-1}[\hat{F}_{\psi|X}\{\psi(\varepsilon_i^*)|X_i\}|X_i] - \psi(\varepsilon_i^*)|, \\ &= |F_{\psi|X}^{-1}[F_{\psi|X}\{\psi(\varepsilon^*)|X_i\} + \delta_i|X_i] - \psi(\varepsilon_i^*)|, \\ &= |F_{\psi|X}^{-1}[F_{\psi|X}\{\psi(\varepsilon^*)|X_i\} + \delta_i|X_i] - F_{\psi|X}^{-1}[F_{\psi|X}\{\psi(\varepsilon^*)|X_i\}|X_i]|, \\ &\leq \frac{1}{c_1} \delta_i = \mathcal{O}_p(h^2 \Gamma_n), \end{aligned}$$

by A.2. Therefore, (A.2) is proved. As $\psi(\cdot)$ plays a role in the estimation via the zero functions defined below, to prove (A.1), we first want to write the estimation difference $(\hat{l}_h - l)(X_i)$ and $(\hat{l}_{h,g}^* - \hat{l}_g)(X_i)$ written w.r.t. to their zero functions defined as follows.

$$\begin{aligned} G_n^*(\theta, X_i) &\stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) [\psi\{\varepsilon_j^* - \theta + \hat{l}_g(X_j)\}], \\ &= \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) \{\psi(Y_j^* - \theta)\}, \\ G_n^\#(\theta, X_i) &\stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) [\psi\{\varepsilon_j^\# - \theta + l(X_j)\}], \\ &= \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) \{\psi(Y_j^\# - \theta)\}. \end{aligned}$$

Note that, for the first moment, we have the natural equality as follows, so we can focus on the difference in the second moment,

$$\mathbb{E}_{\hat{F}_{\varepsilon|X_i=X}} \psi(\varepsilon_i^*) = 0 = \mathbb{E}_{F_{\varepsilon|X_i=X}} \psi(\varepsilon_i^\#). \quad (\text{A.5})$$

We abbreviate $\mathbb{E}_{F_{\varepsilon|X_i=X_i}}$ as $\mathbb{E}$ and $\mathbb{E}_{\hat{F}_{\varepsilon|X_i=X_i}}$ as $\mathbb{E}^*$, define

$$\tilde{l}(x) \stackrel{\text{def}}{=} \underset{\theta}{\operatorname{argmin}} \mathbb{E} \rho(Y - \theta) K_h(X - x), \quad (\text{A.6})$$

$$\tilde{l}_g(x) \stackrel{\text{def}}{=} \underset{\theta}{\operatorname{argmin}} \mathbb{E}^* \rho(Y^* - \theta) K_h(X - x), \quad (\text{A.7})$$

$$T_n(X_i) \stackrel{\text{def}}{=} G_n^*(\tilde{l}_g(X_i), X_i) - G_n^\#(\tilde{l}(X_i), X_i) \quad (\text{A.8})$$

as the unbiased versions of the true function. It is not hard to derive that the bias has the order of $\mathcal{O}_p(h^2)$.

$$\begin{aligned} \tilde{l}(X_i) - l(X_i) &= \frac{\mathbb{E} G_n^\#(l(X_i), X_i)}{\partial \mathbb{E} G_n^\#(l(X_i), X_i)} + \mathcal{O}(\tilde{l}(X_i) - l(X_i)), \\ &= \{h^2 l''(X_i)/2 + f_X'(X_i) l'(X_i) h^2 / f_X(X_i)\} \|K\|_s^2 + \mathcal{O}_p(h^2). \end{aligned}$$

Similarly for $\tilde{l}_g(X_i) - \hat{l}_g(X_i)$, we have

$$\begin{aligned} \tilde{l}_g(X_i) - \hat{l}_g(X_i) &= \frac{\mathbb{E}^* G_n^*(l(X_i), X_i)}{\partial \mathbb{E}^* G_n^*(l(X_i), X_i)} + \mathcal{O}(\tilde{l}_g(X_i) - \hat{l}_g(X_i)), \\ &= \{h^2 \hat{l}_g''(X_i)/2 + f_X'(X_i) \hat{l}_g'(X_i) h^2 / f_X(X_i)\} \|K\|_s^2 + \mathcal{O}_p(h^2), \end{aligned}$$

where $\|K\|_s^2 \stackrel{\text{def}}{=} \int s^2 K^2(s) ds$. A balance between bias and variance term would lead to the choice of the rate h as $\mathcal{O}(n^{-1/5} \Gamma_n)$, and we have $|\tilde{l}(X_i) - l(X_i) - \{\tilde{l}_g(X_i) - \hat{l}_g(X_i)\}| = \mathcal{O}_p(h^2)$, so we can write the difference of the bias term as $\mathcal{O}_p(h^2)$ in the following derivation.

Remark. In the case when $\psi(\cdot)$ is not differentiable, in particular for the quantile case, the above stochastic expansion is still valid, see the stochastic expansion in [20].

According to A.1, excluding the non-differentiable $\psi(\cdot)$ case, one can use the Lipschitz condition of the function $\psi(\cdot)$, and $\psi(\cdot)$ is bounded, we have, $\exists$ a constant C , such that,

$$\begin{aligned} \max_i |T_n(X_i)| &= \max_i \left| \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) \{ \psi(\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)) - \psi(\varepsilon_j^\# - \tilde{l}(X_i) + l(X_j)) \} \right|, \\ &\leq \max_i \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) |C[\{\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)\} - \{\varepsilon_j^\# - \tilde{l}(X_i) + l(X_j)\}]|. \end{aligned}$$

So we can break the upper bound of $\max_i |T_n(X_i)|$ into two terms, the first term $\max_i T_{n,1}(X_i)$ involves the bootstrapped error and its theoretical couple, which has been proved by (A.3), and the second term $\max_i T_{n,2}(X_i)$ is concerning the convergence rate of $\hat{l}_g(\cdot)$,

$$\begin{aligned} \max_i |T_n(X_i)| &\leq \max_i \frac{1}{n} C \sum_{j=1}^n W_{h,j}(X_i) |\varepsilon_j^* - \varepsilon_j^\#| + \max_i \frac{1}{n} C \sum_{j=1}^n W_{h,j}(X_i) |[\{\hat{l}_g(X_j) - \tilde{l}_g(X_i)\} - \{l(X_j) - \tilde{l}(X_i)\}]|, \\ &\leq \max_i T_{n,1}(X_i) + \max_i T_{n,2}(X_i). \end{aligned}$$

$\max_i |T_{n,1}(X_i)|$ is known to have the rate $\mathcal{O}_p(n^{-1/2} h^{3/2} \Gamma_n)$, and

$$\max_i T_{n,2}(X_i) = \max_i \frac{1}{n} C \sum_{j=1}^n W_{h,j}(X_i) |\{\hat{l}_g(X_{i,j,0}) - l(X_{i,j,0})\}(X_i - X_j)| + \mathcal{O}_p(h^2),$$

where $X_{i,j,0}$ is a point between X_i and X_j , and C is a constant, according to the mean value theorem. $\sup_{x \in B} |\hat{l}_g(x) - l(x)|$ is of the rate $\mathcal{O}_p(g^{-1}(ng)^{-1/2} \Gamma_n + g^3)$, see [33]. Therefore the optimal rate for g would be $\mathcal{O}(n^{-1/9})$ in our case (as in A.4), this rate is slower than the choice of h , which confirms the results in [12]. Then we can achieve

$$\max_i T_{n,2}(X_i) = \mathcal{O}_p(h^2 \Gamma_n).$$

As the second derivative of the loss function $\psi(\cdot)$ does not exist at zero, we use the local version of equicontinuity lemma from chapter VII.1 in [29]. Our loss function has the following expansion with the remainder term defined as $r\{y, \theta(x)\}$

$$\rho\{y - \theta(x)\} = \rho\{y - l(x)\} + \{\theta - l(x)\} \psi\{y - l(x)\} + |\theta - l(x)| r\{y, \theta(x)\}, \quad (\text{A.9})$$

define $R_n^*(Y^*, \tilde{l}_g(X_i)) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n r(Y_j^*, \tilde{l}_g(X_i))$, $R_n(Y, \tilde{l}(X_i)) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n r(Y_j, \tilde{l}(X_i))$ are the high order function, we need to prove it satisfies the equicontinuity condition around the point $\tilde{l}_g(X_i)$ and $\tilde{l}(X_i)$, namely, for each η and ϵ , $\exists \delta$ such that,

$$\limsup_n P\left(\sup_{l': \|\tilde{l}_g' - \tilde{l}_g\| < \delta} \|R_n^*(l'_g) - R_n^*(\tilde{l}_g)\| > \eta\right) < \epsilon, \quad (\text{A.10})$$

$$\limsup_n P\left(\sup_{l': \|\tilde{l}' - \tilde{l}\| < \delta} \|R_n(l') - R_n(\tilde{l})\| > \eta\right) < \epsilon. \quad (\text{A.11})$$

According to the equicontinuity lemma in [29], first of all we can prove that $|r(y, \theta)|$ has an envelope, by having the function $r(y, \theta)$ satisfying certain conditions.

So we can achieve the estimations are linked to the zero functions around the point $\tilde{l}_g(X_i)$ and $\tilde{l}(X_i)$ respectively, as follows:

$$\begin{aligned} \hat{l}_{h,g}^*(X_i) - \tilde{l}_g(X_i) &= -\frac{G_n^*\{\tilde{l}(X_i), X_i\}}{\partial \text{EE}^* G_n^*\{\tilde{l}_g(X_i), X_i\}} + \mathcal{O}_p((nh)^{-1/2}), \\ \tilde{l}_h^*(X_i) - \tilde{l}(X_i) &= -\frac{G_n^*\{\tilde{l}(X_i), X_i\}}{\partial \text{EC}_n^*\{\tilde{l}(X_i), X_i\}} + \mathcal{O}_p((nh)^{-1/2}), \end{aligned}$$

where $\partial \text{EE}^* G_n^*\{\tilde{l}_g(X_i), X_i\}$ denotes the partial derivative of $\text{EE}^* G_n^*\{\theta, X_i\}$ w.r.t. θ taking value at the point $\tilde{l}_g(X_i)$.

This means,

$$\begin{aligned} |\hat{l}_{h,g}^*(X_i) - \hat{l}_g(X_i) - \tilde{l}_h^*(X_i) + l(X_i)| &= -\frac{G_n^* \{\tilde{l}_g(X_i), X_i\}}{\partial E E^* G_n^* \{\tilde{l}_g(X_i), X_i\}} + \frac{G_n^\sharp \{\tilde{l}(X_i), X_i\}}{\partial E G_n^\sharp \{\tilde{l}(X_i), X_i\}} + \mathcal{O}_p(h^2), \\ &= -\frac{\partial E G_n^\sharp \{\tilde{l}(X_i), X_i\} [G_n^* \{\tilde{l}_g(X_i), X_i\} - G_n^\sharp \{\tilde{l}(X_i), X_i\}]}{\partial E E^* G_n^* \{\tilde{l}_g(X_i), X_i\} \partial E G_n^\sharp \{\tilde{l}(X_i), X_i\}} \\ &\quad + \frac{G_n^\sharp \{\tilde{l}(X_i), X_i\} [\partial E E^* G_n^* \{\tilde{l}_g(X_i), X_i\} - \partial E G_n^\sharp \{\tilde{l}(X_i), X_i\}]}{\partial E E^* G_n^* \{\tilde{l}_g(X_i), X_i\} \partial E G_n^\sharp \{\tilde{l}(X_i), X_i\}} + \mathcal{O}_p(h^2). \end{aligned}$$

Therefore, since $G_n^* \{\tilde{l}_g(X_i), X_i\}$ and $G_n^\sharp \{\tilde{l}(X_i), X_i\}$ are known by S.L.L.N. to have strong consistency to $E^* G_n^* \{\tilde{l}_g(X_i), X_i\}$ and $E G_n^\sharp \{\tilde{l}(X_i), X_i\}$, we have,

$$\max_i |\hat{l}_{h,g}^*(X_i) - \hat{l}_g(X_i) - \tilde{l}_h^*(X_i) + l(X_i)| = \mathcal{O}(\max_i T_n(X_i)) + \mathcal{O}_p(h^2 \Gamma_n) + \mathcal{O}_p(h^2 \Gamma_n),$$

and (A.1) is proved.

Define the order statistics $X_{(1)}, \dots, X_{(n)}$ of $X_1, \dots, X_n$, so the claim (28) can be proved from (A.1) using the fact that

$$\sup |A_n(x)| \leq \max_i |A_n(X_{(i)})| + \max_i \sup_{x \in [X_{(i)}, X_{(i+1)}]} |A_n(X_{(i)}) - A(x)|, \quad (\text{A.12})$$

it suffices to consider the speed of the last term. With Lipschitz continuity of $A_n(\cdot)$:

$$\max_i \sup_{x \in [X_{(i)}, X_{(i+1)}]} |A_n(X_{(i)}) - A(x)| \leq c_2 \max_i \sup_x |X_i - x|, \quad (\text{A.13})$$

where $c_2 > 0$ is a constant, this upper random bound is of order $\mathcal{O}_p(n^{-1/d} \log n) = \mathcal{O}_p(h^2 \Gamma_n)$. The uniform bound for $\|X_i - x\|$ results from the uniform law of large numbers over a ball of size $n^{-1/d}$, see [26, Theorem 1.1].

Remark. For the non-differentiable $\psi(\cdot)$ cases, in particular the quantile regression case, one can still establish similar inequality, as

$$\begin{aligned} \max_i \left| \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) \{ \psi(\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)) - \psi(\varepsilon_i^\sharp - \tilde{l}(X_i) + l(X_j)) \} \right| \\ \leq \max_i \frac{1}{n} \sum_{j=1}^n W_{h,j}(X_i) | \psi(\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)) - \psi(\varepsilon_i^\sharp - \tilde{l}(X_i) + l(X_j)) |. \end{aligned}$$

Define $\psi_{i,j} \stackrel{\text{def}}{=} | \psi(\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)) - \psi(\varepsilon_i^\sharp - \tilde{l}(X_i) + l(X_j)) | = \mathcal{O}_p(h^2)$ because, so

$$\begin{aligned} P(\psi_{i,j} > c_3 h^2) &\leq E(|\psi_{i,j}|) / (c_3 h^2) \\ &= \mathcal{O}_p((\tilde{l}_g(X_i) + \hat{l}_g(X_j) - \tilde{l}(X_i) + l(X_j)) / h^2). \end{aligned}$$

Then the argument follows from the above proof.

Moreover, one can also use the strong consistency of $G_n^\sharp, G_n^*$ to $E(G_n^\sharp)$ and $E^* G_n^*$ respectively based on A.1 and Lemma 2.4 of [25].

A.2. Proof of second part of Theorem 3.1

The number of regressors in (11) is of order $p = H^{-1}d + 1$. [30] shows that as long as $n^{-1}(p \log n)^{3/2} \rightarrow 0$ then the estimators of the regression parameters are consistent and have the standard variance. In our situation,

$$n^{-1} n^{1/5 \cdot 2/3} \log n = \mathcal{O}(1), \quad (\text{A.14})$$

and therefore the condition is satisfied.

Now we have a look at the behavior of the design matrix in (12).

$$\begin{aligned} \hat{\mathcal{L}}(A) &\stackrel{\text{def}}{=} -n^{-1} \sum_{i=1}^n \rho\{Y_i - A^\top \Phi(X_i)\}, \\ \nabla \hat{\mathcal{L}}(A) &\stackrel{\text{def}}{=} -n^{-1} \sum_{i=1}^n \psi\{Y_i - A^\top \Phi(X_i)\} \Phi(X_i), \\ \nabla^2 E \hat{\mathcal{L}}(A) &\stackrel{\text{def}}{=} -\nabla E \psi\{Y_i - A^\top \Phi(X_i)\} \Phi(X_i) \Phi(X_i)^\top. \end{aligned}$$

Recall that

$$\hat{A} = \operatorname{argmin}_A \hat{\mathcal{L}}(A).$$

Lemma 14 of [34] ensures that with probability approaching 1, $\hat{A}$ exists uniquely and that $\nabla \hat{\mathcal{L}}(\hat{A}) = 0$. In addition, there exists $\bar{m}(x) = \bar{A}^\top \Phi(x)$, such that

$$\sup_{x \in B} |\bar{m}(x) - m(x)| \leq C_\infty H^2. \quad (\text{A.15})$$

According to the bounded influence condition A.1, the j th term, $j \in 1, 2, \dots, H^{-1}d + 1$.

$$\begin{aligned} |\{\nabla \hat{\mathcal{L}}(\bar{A})\}_j| &= \left| \left\{ -n^{-1} \sum_{i=1}^n [\psi\{m(X_i) + \varepsilon_i - \bar{A}^\top \Phi(X_i)\} - \psi(0)] \Phi(X_i) \right\}_j \right|, \\ &\leq \left[n^{-1} \sum_{i=1}^n C_3 |\{A^\top \Phi(X_i) + \varepsilon_i - \bar{A}^\top \Phi(X_i)\} \Phi(X_i)| \right]_j, \\ &\leq \left[n^{-1} \sum_{i=1}^n C_3 |\{A^\top \Phi(X_i) + \varepsilon_i - \bar{A}^\top \Phi(X_i)\} \Phi(X_i)| \right]_j. \end{aligned}$$

We know first that,

$$\mathbb{E}[\{m(X_i) - \bar{m}(X_i)\} \Phi(X_i)]_j = \mathcal{O}(H^2).$$

Let $\xi_{i,j} \stackrel{\text{def}}{=} \{|m(X_i) - \bar{m}(X_i)\} \Phi(X_i)| - \mathbb{E}[m(X_i) - \bar{m}(X_i)\} \Phi(X_i)]_j$, by Bernstein's Lemma:

Lemma A.1. Let $Z_1, \dots, Z_n$ be independent r.v.s.

$$\log \mathbb{E} \exp(tZ_i) \leq \mathbb{E}(Z_i^2)t^2/2$$

for all $t \in [0, \infty]$. Then

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_i \right| \geq t \sqrt{2 \sum_{i=1}^n \mathbb{E} Z_i^2} \right) \leq 2 \exp(-t^2).$$

Finally, we can derive that,

$$n^{-1} \sum_{i=1}^n \xi_{i,j} = \mathcal{O}_{a.s.}(H^2 n^{-1/2} \Gamma_n), \quad j \neq 1. \quad (\text{A.16})$$

The last term

$$n^{-1} \left| \left\{ \sum_{i=1}^n \varepsilon_i \Phi(X_i) \right\}_j \right| = \mathcal{O}_{a.s.}(n^{-1/2} \Gamma_n). \quad (\text{A.17})$$

Therefore, one has a collective term from (A.16) and (A.17),

$$\|\nabla \hat{\mathcal{L}}(\bar{A})\| = \mathcal{O}_{a.s.}(H^{3/2} + H^{-1/2} n^{-1/2} \Gamma_n),$$

where $\|\cdot\|$ denotes the L_2 norm.

By assumptions A.5, A.7, $\forall l = 1, \dots, H^{-1}$, the d dimensional vector $\Phi_l^\top(X_i)$ satisfies,

$$\beta \|b\|^2/d \geq \mathbb{E} b^\top \Phi_l^\top(X_i) \Phi_l(X_i) b \geq \alpha \|b\|^2/d,$$

where α and β are two constants.

Lemma A.2. Assume A.1 and A.5, as $n \rightarrow \infty$,

$$\begin{aligned} \|\hat{A} - A\| &= \mathcal{O}_{a.s.}(H^{3/2} + H^{-1/2} n^{-1/2} \Gamma_n), \\ \max_{i \in 1, \dots, n} \|\hat{m}(X_i) - \bar{m}(X_i)\| &= \mathcal{O}_{a.s.}(H + H^{-1} n^{-1/2} \Gamma_n). \end{aligned}$$

Proof. According to similar equicontinuity arguments, exists an $(H^{-1}d + 1) \times (H^{-1}d + 1)$ matrix $\bar{A}$, such that

$$\|\hat{A} - \bar{A}\| = \mathcal{O}(\|\nabla^2 \mathbb{E} \hat{\mathcal{L}}(\bar{A}_0)\|^{-1} \{-\nabla \hat{\mathcal{L}}(\bar{A})\|),$$

since

$$\nabla^2 \mathbb{E} \hat{\mathcal{L}}(X_i) = \Phi(X_i) \Phi(X_i)^\top \nabla \mathbb{E} \psi(Y_i - \hat{A}_0^\top \Phi(X_i)).$$

According to assumption A.7,

$$c_3 I_{H^{-1}d+1} \geq \nabla^2 \mathbb{E} \hat{\mathcal{L}}(X_i) \geq c_4 I_{H^{-1}d+1},$$

therefore

$$\|\hat{A} - \bar{A}\| = \mathcal{O}_{a.s.}(H^{3/2} + H^{-1/2} n^{-1/2} \Gamma_n).$$

Moreover, by Cauchy–Schwarz inequality

$$\begin{aligned} \max_{i \in 1, \dots, n} |\hat{m}(X_i) - \bar{m}(X_i)| &\leq \max_i \|\hat{A} - \bar{A}\| \|\Phi(X_i)\|, \\ &= \mathcal{O}_{a.s.}(H^{3/2} + H^{-1/2} n^{-1/2} \Gamma_n) \mathcal{O}_{a.s.}(H^{-1/2}), \\ &= \mathcal{O}_{a.s.}(H + H^{-1} n^{-1/2} \Gamma_n). \end{aligned}$$

We would like to check for the pseudo observations $Y_i^\# = m(X_i) + \varepsilon_i^\#$.

$$\begin{aligned} &|\{\hat{m}_k^*(X_{i,k}) - \hat{m}_k(X_{i,k})\} - \{\hat{m}_k^\#(X_{i,k}) - m_k(X_{i,k})\}| \\ &\leq |(\hat{A}^{*\top} - \hat{A}^\top - \hat{A}^{\# \top} + A^\top) \Phi(X_{i,k})|, \\ &= \mathcal{O}(|[\nabla^2 \mathbb{E} \hat{\mathcal{L}}^*(\hat{A})^{-1} \{-\nabla \hat{\mathcal{L}}^*(\hat{A})\} - \nabla^2 \mathbb{E} \hat{\mathcal{L}}(\bar{A})^{-1} \{-\nabla \hat{\mathcal{L}}(\bar{A})\}]^\top \Phi(X_i)| + H^2), \end{aligned}$$

as $\nabla^2 \mathbb{E} \hat{\mathcal{L}}^*(\hat{A})^{-1}$ and $\nabla^2 \mathbb{E} \hat{\mathcal{L}}(\bar{A})^{-1}$ are both bounded,

$$\begin{aligned} &|\{\hat{m}_k^*(X_{i,k}) - \hat{m}_k(X_{i,k})\} - \{\hat{m}_k^\#(X_{i,k}) - m_k(X_{i,k})\}| \\ &= \mathcal{O} \left(\left| n^{-1} \sum_{i=1}^n \{\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\#) + \bar{A}^\top \Phi(X_i) - A^\top \Phi(X_i)\} \Phi(X_i)^\top \Phi(X_i) \right| + H^2 \right), \\ &= \mathcal{O} \left(\left| n^{-1} \sum_{i=1}^n \{\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\#)\} \Phi(X_i)^\top \Phi(X_i) \right| + H^2 \right), \\ &= \mathcal{O} \left(\left| n^{-1} \sum_{i=1}^n \{\varepsilon_i^* - \varepsilon_i^\#\} \Phi(X_i)^\top \Phi(X_i) \right| + H^2 \right), \\ &= \mathcal{O} \left(\left| n^{-1} \sum_{i=1}^n \{Z_i |\hat{\varepsilon}_i| - Z_i \eta_i\} \Phi(X_i)^\top \Phi(X_i) \right| + H^2 \right). \end{aligned}$$

One can derive using a coupling argument,

$$\begin{aligned} |\{\hat{m}_k^*(X_{i,k}) - \hat{m}_k(X_{i,k})\} - \{\hat{m}_k^\#(X_{i,k}) - m_k(X_{i,k})\}| &= \mathcal{O}_{a.s.}(n^{-1/2}(H^2 + n^{-1/2} \Gamma_n) H^{-1/2} + H^2) \\ &= \mathcal{O}_{a.s.}(H^2). \end{aligned}$$

References

- [1] D. Basso, M. Chiarandini, L. Salmaso, Synchronized permutation tests in replicated $i \times j$ designs, *J. Statist. Plann. Inference* 137 (8) (2007) 2564–2578.
- [2] P.J. Bickel, M. Rosenblatt, On some global measures of the deviations of density function estimates, *Ann. Statist.* 1 (6) (1973) 1071–1095.
- [3] G. Claeskens, I. Van Keilegom, Bootstrap confidence bands for regression curves and their derivatives, *Ann. Statist.* 31 (6) (2003) 1852–1884.
- [4] A. Deaton, J. Muellbauer, An almost ideal demand system, *Amer. Econ. Rev.* 40 (3) (1980) 312–326.
- [5] J. Franke, P. Mwita, W. Wang, Nonparametric estimates for conditional quantiles of time series, SFB 649 Discussion Paper.
- [6] M. Fuss, D. McFadden, Y. Mundlak, A survey of functional forms in the economic analysis of production, Vol. 1, McMaster University Archive for the History of Economic Thought, 1978.
- [7] P. Hall, Edgeworth expansions for nonparametric density estimators, with applications, *Statistics* 22 (2) (1991) 215–232.
- [8] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer Verlag, 1992.
- [9] W.K. Härdle, Asymptotic maximal deviation of M -smoothers, *J. Multivariate Anal.* 29 (2) (1989) 163–179.
- [10] W.K. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, 1990.
- [11] W.K. Härdle, J. Horowitz, J.-P. Kreiss, Bootstrap method for time series, *Int. Statist. Rev.* 71 (2) (2003) 435–459.
- [12] W.K. Härdle, S.J. Marron, Bootstrap simultaneous error bars for nonparametric regression, *Ann. Statist.* 19 (2) (1991) 778–796.
- [13] W.K. Härdle, S. Song, Confidence bands in quantile regression, *Econometric Theory* 26 (4) (2010) 1180–1200.
- [14] J.L. Horowitz, Nonparametric estimation of a generalized additive model with an unknown link function, *Econometrica* 69 (2) (2001) 499–513.
- [15] J.L. Horowitz, The bootstrap, in: *Handbook of Econometrics*, Vol. 5, Elsevier, 2001.

- [16] J.L. Horowitz, J. Klemelä, E. Mammen, Optimal estimation in additive regression models, *Bernoulli* 12 (2) (2006) 271–298.
- [17] J.L. Horowitz, S. Lee, Nonparametric estimation of an additive quantile regression model, *J. Amer. Statist. Assoc.* 100 (472) (2005) 1238–1249.
- [18] P.J. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.* 35 (1) (1964) 73–101.
- [19] G.J. Johnston, Probabilities of maximal deviations for nonparametric regression function estimates, *J. Multivariate Anal.* 12 (3) (1982) 402–414.
- [20] E. Kong, O. Linton, Y. Xia, Uniform Bahadur representation for local polynomial estimates of M -regression and its application to the additive model, *Econometric Theory* 26 (05) (2010) 1529–1564.
- [21] E. Mammen, *When Does Bootstrap Work?: Asymptotic Results and Simulations*, Springer Verlag, 1992.
- [22] M. Marozzi, Applications in business, medical and industrial statistics of bi-aspect nonparametric tests for location problems, *Stat. Methods Appl.* 12 (2) (2003) 187–194.
- [23] M. Marozzi, Multivariate tri-aspect non-parametric testing, *J. Nonparametr. Stat.* 19 (6–8) (2007) 269–282.
- [24] M.H. Neumann, J.-P. Kreiss, Regression-type inference in nonparametric autoregression, *Ann. Statist.* 26 (4) (1998) 1570–1613.
- [25] W.K. Newey, D.L. McFadden, Large sample estimation and hypothesis testing, in: *Handbook of Econometrics*, Vol. 4, 1986, pp. 2111–2245.
- [26] M.D. Penrose, A strong law for the largest nearest-neighbour link between random points, *J. Lond. Math. Soc.* 60 (3) (1999) 951–960.
- [27] F. Pesarin, L. Salmaso, Finite-sample consistency of combination-based permutation tests with application to repeated measures designs, *J. Nonparametr. Stat.* 22 (5) (2010) 669–684.
- [28] F. Pesarin, L. Salmaso, *Permutation Tests for Complex Data: Theory, Applications and Software*, John Wiley & Sons, 2010.
- [29] D. Pollard, *Convergence of Stochastic Processes*, Springer Verlag, 1984.
- [30] S. Portnoy, Local asymptotics for quantile smoothing splines, *Ann. Statist.* 25 (1) (1997) 414–434.
- [31] S. Song, Y. Ritov, W.K. Härdle, Bootstrap confidence bands and partial linear quantile regression, *J. Multivariate Anal.* 107 (2012) 244–262.
- [32] S. Sperlich, D. Tjøstheim, L. Yang, Nonparametric estimation and testing of interaction in additive models, *Econometric Theory* 18 (02) (2002) 197–251.
- [33] C.J. Stone, Optimal global rates of convergence for nonparametric regression, *Ann. Statist.* 10 (4) (1982) 1040–1053.
- [34] C.J. Stone, Additive regression and other nonparametric models, *Ann. Statist.* 13 (2) (1985) 689–705.
- [35] Y. Yafeh, O. Yosha, Large shareholders and banks: who monitors and how? *Econ. J.* 113 (2003) 128–146.