

SFB 649 Discussion Paper 2013-047

# Tie the straps: uniform bootstrap confidence bands for bounded influence curve estimators

Wolfgang Karl Härdle\*  
Ya'acov Ritov\*\*  
Weining Wang\*



\*Humboldt-Universität zu Berlin, Germany  
\*\*University of Jerusalem Mount Scopus, Israel

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

# Tie the straps: uniform bootstrap confidence bands for bounded influence curve estimators <sup>‡</sup>

Wolfgang Karl Härdle<sup>§</sup>, Ya'acov Ritov<sup>¶</sup>, Weining Wang<sup>||</sup>

October 28, 2013

## Abstract

We consider theoretical bootstrap “coupling” techniques for non-parametric robust smoothers and quantile regression, and verify the bootstrap improvement. To cope with curse of dimensionality, a variant of “coupling” bootstrap techniques are developed for additive models with both symmetric error distributions and further extension to the quantile regression framework. Our bootstrap method can be used in many situations like constructing confidence intervals and bands. We demonstrate the bootstrap improvement over the asymptotic band theoretically, and also in simulations and in applications to firm expenditures and the interaction of economic sectors and the stock market.

Keywords: Nonparametric Regression, Bootstrap, Quantile Regression, Confidence Bands, Additive Model, Robust Statistics

JEL classification: C00, C14

---

\*The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged.

<sup>†</sup>We acknowledge the friendly environment atmosphere of MFO mathematical research institute of Oberwolfach during an Research in pairs stay.

<sup>‡</sup>The financial support from ISF is greatly acknowledged.

<sup>§</sup>Professor at Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz chair of statistics and Director of C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany and School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899.

<sup>¶</sup>Professor at Department of Statistics Hebrew University of Jerusalem Mount Scopus, Jerusalem 91905, Israel.

<sup>||</sup>Hermann Otto Hirschfeld Junior Professor at the Ladislaus von Bortkiewicz chair of statistics and C.A.S.E., Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de

# 1 Introduction

Confidence bands are important tools for model specifications. However it is hard to construct precise confidence bands for nonparametric curves as usually a supreme norm is involved in the statistics. Traditional methods based on 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.

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

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

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 Bickel and Rosenblatt (1973), which is further extended by Johnston (1982), and has recently been studied by Härdle and Song (2010) for  $L$ -smoothers. However, it has also been shown by Härdle, Ritov and Song (2012) that such an asymptotic confidence band has much lower coverage probability in finite sample than what it is supposed to have. The poor performance of such kind of band in finite sample has been well attributed to its slow convergence, see Hall (1991). To address improvement of finite sample performance, there is a class of literature on the bootstrap confidence band, see Claeskens and Van Keilegom (2003), Härdle and Marron (1991), among others. Figure 1 shows a bootstrap confidence band and an asymptotic band for an  $M$ -smoother with  $\rho(\cdot)$  being the Tukey's bisquare loss. One sees that the asymptotic band is narrower than the bootstrap one. Moreover, the asymptotic band does not cover the true curve while the bootstrap one does.

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 "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 Hall (1992), Mammen (1992), Horowitz (2001a), Härdle, Horowitz and Kreiss (2003), 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

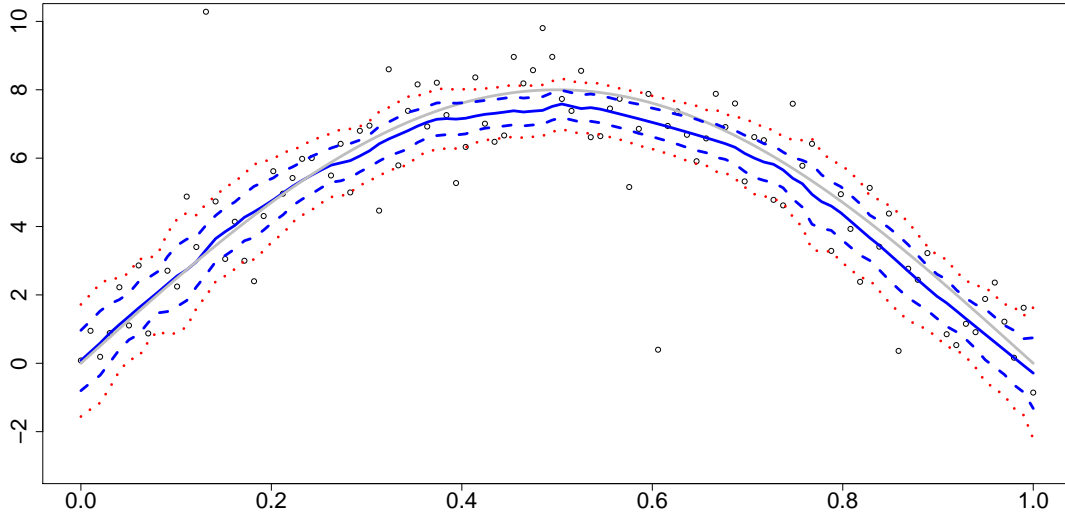


Figure 1: Plot of true curve (grey), robust estimate with Tukey biweight loss and 5% confidence bands (blue dashed), local polynomial estimate (black), bootstrap band (red dotted),  $n = 150$ .

in a nonparametric framework. But very few articles have focused on non-linear statistics (e.g. maximum) in nonparametric regression. Härdle et al. (2012) proposed a bootstrap procedure and showed its improvement properties. This stimulated the current research on finding common properties that loss functions have to share to attain such improvement. Accordingly, we prove a generalized version in the univariate case for the class of loss function with bounded influence. For the high dimensional case, the bootstrap improvement 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, Härdle (1990). 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 well known that (2) solves the dimensionality curse 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 Horowitz (2001b), Horowitz and Lee (2005), Horowitz,

Klemelä and Mammen (2006), and among many others. Horowitz (2001b) focuses on generalized additive models with unknown link functions, Horowitz and Lee (2005) propose a two-stage estimation for quantile regression in additive models, Horowitz et al. (2006) show the equivalence between spline, kernel and other methods in terms of optimal minimax rate in 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 sum, we investigate a coupling technique that allows us to “tie the straps” even a little tighter for a class of estimators. We mean by that, theoretically speaking, confidence band construction is made more precise in a variety of the estimation problems 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 the paper is organized as follows. In Section 2 we explain in details the model setup and the bootstrap method. Section 3 presents the main results. In Section 4 a small simulation study is presented. Finally, we show in Section 5 applications on managerial compensation and impacts on stock markets.

## 2 Models and bootstrap confidence sets

This section describes the estimator and our coupling techniques, motivates 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:

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

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. Other examples for  $\rho(\cdot)$  include the trimmed mean, Huber (1964)

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

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:

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

where  $K_h(u) = K(u/h)/h$  is a kernel function with bandwidth  $h$ . Now generate a bootstrap sample using an 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  $Y_i$  are the same. We will show that for a class of loss functions the following approximation holds:

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

where  $B$  is a closed compact set in  $[0, 1]$ ,  $\Gamma_n$  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  $h$ ,

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

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:

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

where the  $\nu_1(\cdot), \nu_2(\cdot), \dots$  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 define the width  $H$  subintervals  $[lH, (l+1)H]$ ,  $0 \leq l \leq (H^{-1} - 1)$ . The linear  $B$ -spline basis is:

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

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

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

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  $\mathbf{E} B_{j,l}(x_j) = 0$ ,  $\mathbf{E} B_{j,l}(x_j)^2 = 1$ .

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

$$\begin{aligned} \mathbf{a}_j &= (a_{1,j}, \dots, a_{H,j})^\top, \\ \mathbf{g}(x_{i,j})^\top &= \{\nu_1(x_{i,j}), \dots, \nu_H(x_{i,j})\}^\top. \end{aligned}$$

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 (only for bounded influence functions), 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  $(\varepsilon^*, \varepsilon^\sharp)$ , the bootstrap residuals and its associate theoretical couple respectively are:

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

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

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

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

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  $V_i \stackrel{\text{def}}{=} F_{i,\text{sgn}(\varepsilon_i)}(|\varepsilon_i|)$  have a standard uniform distribution,



and if  $Z_i$  is as above, then  $\varepsilon_i$  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  $\varepsilon_i$ . Hence, for  $t > 0$ :

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

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

$$\mathcal{L}(\varepsilon_i^\sharp) = \mathcal{L}(\varepsilon_i). \tag{19}$$

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 trying to construct an empirically feasible bootstrap sample that is strongly approximating a sample from the true distribution. One example of the coupled bootstrap approach was already 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  $\varepsilon_i$  are symmetric. Then the coupling may be performed via a Rademacher randomized variables  $Z_i$  with

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

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

$$\begin{aligned}\varepsilon_i^* &\stackrel{\text{def}}{=} Z_i |\hat{\varepsilon}_i| \\ \varepsilon_i^\sharp &\stackrel{\text{def}}{=} Z_i \eta_i.\end{aligned}\tag{20}$$

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  $(\varepsilon_i^*, \varepsilon_i^\sharp)$  is given by (15) and (16). It was argued that the distributions of  $\varepsilon_i^\sharp$  and  $\varepsilon_i$  are identical and also the conditional distributions given  $\{V_i\}_{i=1}^n$  are the same.

The resampling technique will be applied to nonparametric estimation of an additive quantile regression model. The reanalysis of the data used by Horowitz and Lee (2005) 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

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

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  $M < \infty$ .  $\mathbb{E}\{\psi(\varepsilon)|X\} = 0$  w.p. 1. Also  $\psi(\cdot)$  is Liptchitz 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  $\mathbb{E}\{g_l^2(X_j)\} = 1$  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. A.2 assumes W.L.O.G., the covariates are on bounded support and imposes 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 smaller as we need to reduce the bias, see Härdle and Marron (1991). 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.

We show first convergence results for 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

$$\mathbf{E}_{\hat{F}_{\varepsilon|X=x_i}} \psi(\varepsilon_i^*) = 0 = \mathbf{E}_{F_{\varepsilon|X=x_i}} \psi(\varepsilon_i^\#) \quad (25)$$

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  $\|\hat{F}_{(\varepsilon|X_i=x)}(\cdot) - F_{(\varepsilon|X_i=x)}(\cdot)\|_\infty = \mathcal{O}(h^2\Gamma_n)$ , then a similar coupling argument as in (7) can be used. Recall  $\varepsilon_i^* = \hat{F}_{(\varepsilon|X_i=x)}^{-1}(F_{(\varepsilon|X_i=x)})$ , 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  $\tilde{l}_h^*(\cdot)$  as in (9). Then*

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

where

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

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

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

**Remark 1.** The aforementioned strong approximation results mean that the stochastic behavior of  $\hat{l}_h^\sharp(x) - l(x)((\hat{m}_j^\sharp - m_j)(x))$  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  $\sup_x |\hat{l}_h^\sharp(x) - l(x)|(\sup_{x \in B} |(\hat{m}_j^\sharp - m_j)(x)|)$  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 Kreiss and Neumann (1998). Therefore, the bootstrap confidence band is a direct consequence of the results.

**Remark 2.** Similar result was proved by Härdle et al. (2012) 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 Härdle (1989), we set the joint pdf of  $(X, Y)$  as,

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

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

$n$	95%		90%	
	Rel. Err.	Area	Rel. Err.	Area
100	0.07(0.02)	1.23(2.51)	0.08(0.02)	1.02(2.20)
200	0.08(0.01)	0.89(1.95)	0.10(0.02)	0.74(1.76)
400	0.05(0.01)	0.78(1.32)	0.09(0.01)	0.64(1.15)

Table 1: Relative errors and areas (estimated as averaged width of the confidence bands over the grid points) of asymptotic (bootstrap) with 100 repetitions per sample, and 200 samples. Relative error  $\stackrel{\text{def}}{=} |(\text{true coverage} - \text{nominal coverage})|/\text{nominal coverage}$ .

- 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 (33) with Gaussian kernel

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

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

- 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

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

- 6) Calculate the  $1 - \alpha$  quantile  $d_\alpha^*$  of  $d_1, \dots, d_{n^*}$ .

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

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

Figure 1 shows the theoretical signal curve, the robust estimate using Huber's loss function with corresponding 95% uniform confidence band from the asymptotic theory and the confidence band using the bootstrap method. The real curve is marked as the grey solid line. We then compute the confidence band based on asymptotic theory according to Härdle (1989). 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 robust estimation. To compare which method is more precise, Table 1 presents respectively the relative errors of simulated bands compared with their nominal levels together with the calculated area of the 95% and 90% confidence band, for sample size  $n = 100, 200, 400$ . 100 simulation runs are carried out and 200 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) obtain 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 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  $x_1, x_2, x_3, x_4 \sim U(-2.5, 2.5)$ ,

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

and  $\varepsilon_i$  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  $\varepsilon_i^*$ ,  $i = 1, \dots, n$  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

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

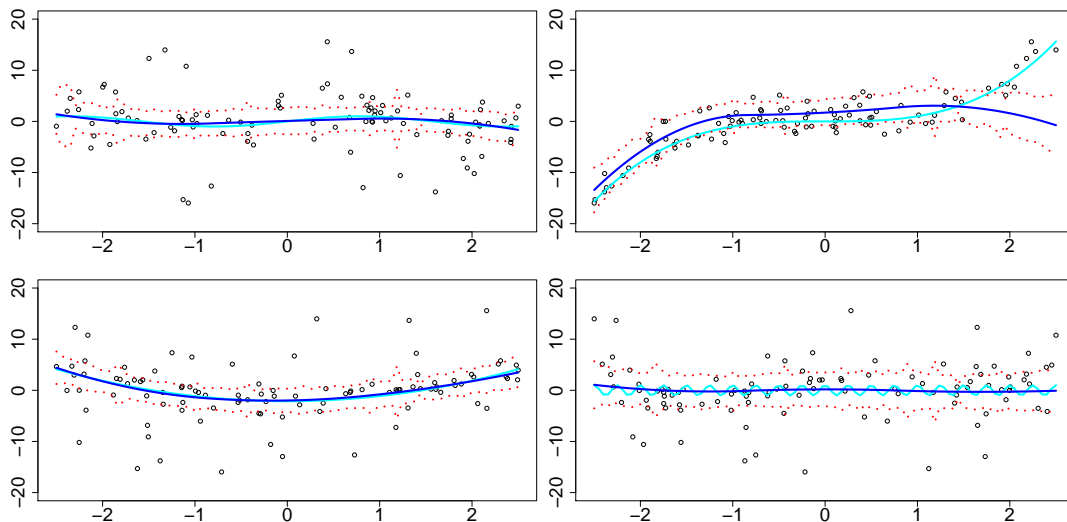


Figure 2: Plot of true curve (dark blue), robust estimates and bands (cyan), bootstrap band (red dotted)

- 5) Calculate the  $1 - \alpha$  quantile  $d_\alpha^*$  of  $d_1, \dots, d_{n^*}$ .
- 6) Construct the bootstrap uniform band centered around  $\hat{m}_j(x_j)$

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

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

The relative errors between the bootstrap bands and their nominal levels are shown in Table 2. The coverage are close to the nominal levels and the widths of band are clearly shrinking w.r.t. the sample sizes.

## 5 Empirical analysis

### 5.1 Firm expenses analysis

Yafeh and Yosha (2003) use 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 Horowitz and Lee (2005). The dependent variable  $Y$  is: general sales and administrative expenses deflated by sales (denoted by MH5), which is one of five measures of



	$n$	Rel. Err.	Area
95%	100	0.07, 0.03, 0.05, 0.01	6.06, 5.37, 5.44, 5.21
	200	0.07, 0.03, 0.02, 0.04	5.50, 4.74, 4.54, 4.65
	400	0.04, 0.02, 0.01, 0.03	4.83, 3.63, 3.76, 3.70
90%	100	0.05, 0.08, 0.06, 0.03	5.88, 5.07, 5.04, 5.30
	200	0.03, 0.02, 0.04, 0.02	4.84, 3.84, 3.85, 4.00
	400	0.06, 0.02, 0.02, 0.01	4.02, 3.25, 3.11, 3.03

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 100 repetitions per sample, and 200 samples.

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), sample size= 185. The regression model we consider here is:

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

The estimated additive components and its bootstrap confidence bands are shown in Figure 3. Similarly, it can be seen that the nonlinear effects are  $\log(\text{asset})$  and *TOPTEN*, and the firm age effects are minor compared to the other three. Differently, the effect of leverage is also a little bit nonlinear, and the shape of curves deviates from what Horowitz and Lee (2005) present, especially for the effect of *TOPTEN*. This may 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 stock market

We analyze how the four markets (oil, currency, bond, real estate) affect the stock market. This study would give implications to the interactions of the economic conditions among different sectors. The data source is ProQuest Statistical Datasets, we focus on the US 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 weekly frequency. We select the data during the period 20080903 – 20111128.

It can be observed that all the four markets have non linear effects on the stock indices values, Figure 4, but only exchange rate EUR-USD and crude

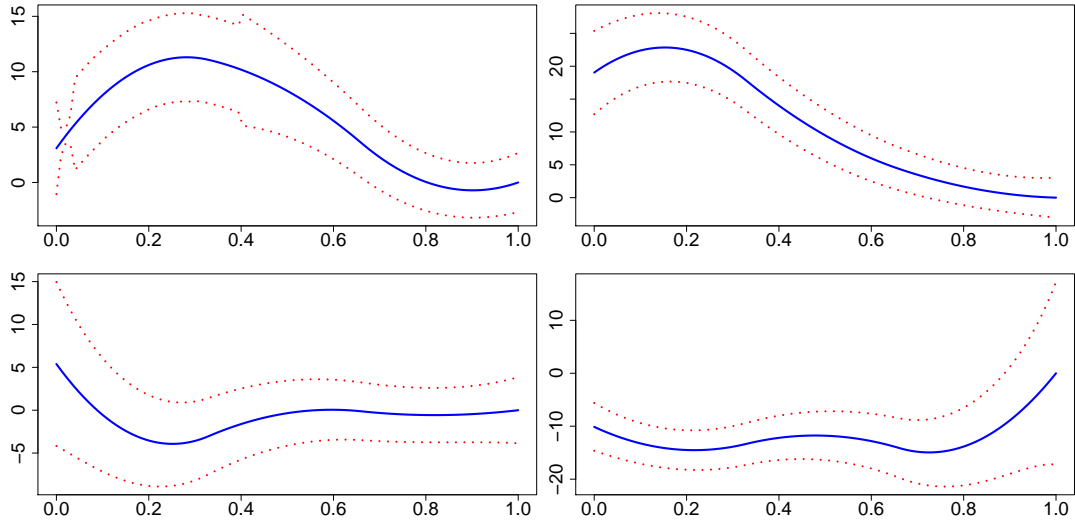


Figure 3: Robust estimates (blue), bootstrap bands (red dotted), left up:  $\text{Log}(\text{Asset})$ , right up: Leverage, left below: Age, right below: TOPTEN.

oil prices affect the the stock indices returns nonlinearly, Figure 5. It is not difficult to interpret the relationships: In Figure 4, for the exchange rate EUR-USD, the weakness of EUR up to a certain level ( $< 1.27$ ) are 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 negative impact on stock indices at every level, but the effects decrease when the prices raise. 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. So one sees a negative correlation there when the inflation index is bigger than (0.7). Finally, increasing real estate index can be a sign of booming economic condition, therefore the stock indices raise when the real estate index gets higher. However, when the real estate index is too high, it is likely that there exist bubble, so one sees a drop in the market indices.

In Figure 5, we see difference effects on S&P log returns, exchange rate EUR- USD are positively correlated with returns until a high lever ( $> 1.40$ ), the crude oil has majorally negative effects on stock returns. More non-linearity is presented in the plots for inflation index and real estate index.

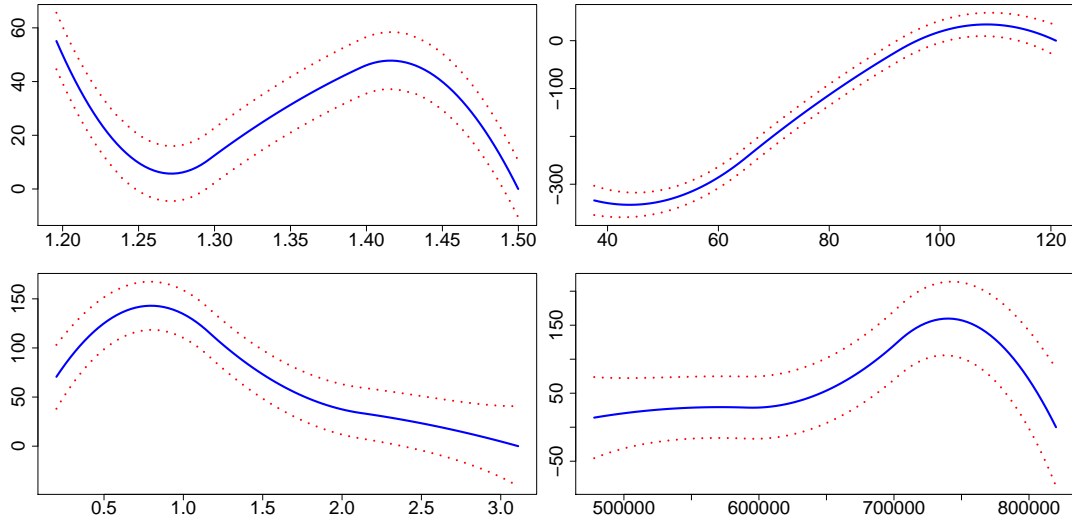


Figure 4: Robust estimate (blue), bootstrap bands (red dotted), Y: S&P index, left up: exchange rates EUR-USD, right up: crude oil price, left below: inflation index, right below: real estate index.

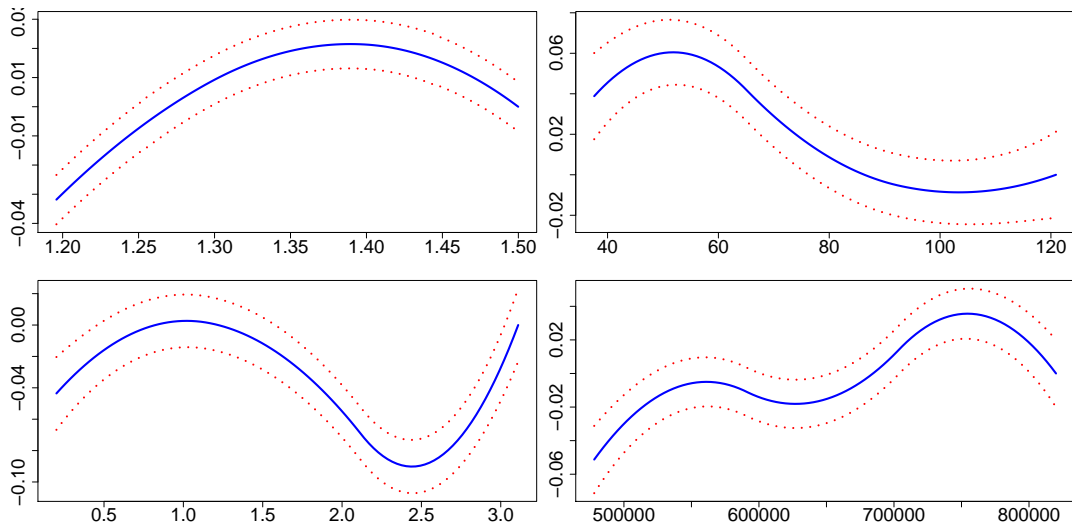


Figure 5: Robust estimate (blue), bootstrap bands (red dotted), Y: S&P index log return, left up: exchange rates EUR-USD, right up: crude oil price, left below: inflation index, right below: real estate index.

## 6 Conclusion

We have developed and proved the bootstrap improvement for a wide class of smoothers with bounded influence functions. Moreover, we extend our results to additive models to cope with curse of dimensionality.

## A Appendix

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

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

$$\max_i [(\hat{l}_h^\sharp - l)(X_i) - \{(\hat{l}_{h,g}^* - \hat{l}_g)(X_i)\}] = \mathcal{O}_p(h^2\Gamma_n), \quad (32)$$

This is extended to any point  $x$  in a compact set  $B$  in (43).

Proving (32) can be done by showing first:

$$\max_i |\psi(\varepsilon_i^*) - \psi(\varepsilon_i^\sharp)| = \mathcal{O}_p(h^2\Gamma_n). \quad (33)$$

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 Franke and Mwita (2011) we have for a small  $b$ :

$$\sup_{|t| \leq b, i=1, \dots, n} |\tilde{F}_{(\varepsilon|X)}(t) - F_{(\varepsilon|X)}(t)| = \mathcal{O}_p(h^2\Gamma_n). \quad (34)$$

Also recall (24)  $\hat{F}_{(\varepsilon|X)}$  is a scaled version of  $\tilde{F}_{(\varepsilon|X)}$ . 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

$$\max_i |\varepsilon_i^\sharp - \varepsilon_i^*| = \max_i |F_{(\varepsilon|X)}^{-1}(U_i) - \hat{F}_{(\varepsilon|X)}^{-1}(U_i)| = \mathcal{O}_p(h^2\Gamma_n) \quad (35)$$

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  $X$ ,

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

by A.2. Therefore, (33) is proved. As  $\psi(\cdot)$  plays a role in the estimation via the zero functions defined below, to prove (32), 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 focused the difference in the second moment,

$$\mathbf{E}_{\hat{F}_{\varepsilon|X_i=x}} \psi(\varepsilon_i^*) = 0 = \mathbf{E}_{F_{\varepsilon|X_i=x}} \psi(\varepsilon_i^\#). \quad (36)$$

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

$$\tilde{l}(x) \stackrel{\text{def}}{=} \operatorname{argmin}_\theta \mathbf{E} \rho(Y - \theta) K_h(X - x) \quad (37)$$

$$\tilde{l}_g(x) \stackrel{\text{def}}{=} \operatorname{argmin}_\theta \mathbf{E}^* \rho(Y^* - \theta) K_h(X - x) \quad (38)$$

$$T_n(X_i) \stackrel{\text{def}}{=} G_n^* \{\tilde{l}_g(X_i), X_i\} - G_n^\# \{\tilde{l}(X_i), X_i\}. \quad (39)$$

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

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

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

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

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 Kong, Linton and Xia (2010).

According to A.1, excluding 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  $\max_i |T_n(X_i)|$  by 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 (34), and the second term  $\max_i T_{n,2}(X_i)$  is concerning the convergence rate of  $\hat{l}_g(\cdot)$ ,

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

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

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

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 Stone (1982). 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 Härdle and Marron (1991). 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 Pollard (1984). Our loss function have the following expansion with the remainder term defined as  $r\{y, \theta(x)\}$

$$\rho\{y - \theta(x)\} = \rho\{y - l(x)\} + \{\theta - l(x)\}\psi\{y - l(x)\} + |\theta - l(x)|r\{y, \theta(x)\}, \quad (40)$$

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  $\eta$  and  $\epsilon$ ,  $\exists \delta$  such that,

$$\limsup_n \mathbb{P} \left( \sup_{l': \|l'_g - \tilde{l}_g\| < \delta} \|R_n^*(l'_g) - R_n^*(\tilde{l}_g)\| > \eta \right) < \epsilon \quad (41)$$

$$\limsup_n \mathbb{P} \left( \sup_{l': \|l' - \tilde{l}\| < \delta} \|R_n(l') - R_n(\tilde{l})\| > \eta \right) < \epsilon. \quad (42)$$

According to the equicontinuity lemma in Pollard (1984), first of all we can prove that  $|r(y, \theta)|$  has an envelope. As the bounded influence condition in A.2,

$$\begin{aligned} |r(y, \theta)| &= \{|\rho(y - \theta) - \rho(y - l(x)) - (\theta - l(x))\psi(y - l(x))|\} / |\theta - l(x)| \\ &\leq \{|\theta - l(x)|\psi(y - \theta_0)\} / |\theta - l(x)| + \{|\theta - l(x)|\psi(y - l(x))\} / |\theta - l(x)| \\ &\leq 2C_1. \end{aligned}$$

The  $\rho\{Y - l(x)\}$  is differentiable in quadratic mean at  $l(x)$ . Moreover, we observe that the condition on covering numbers is satisfied, since the class of function  $r(y, \theta)$  can be expressed as a union or intersection of classes of polynomial discrimination, therefore the covering number is bounded.

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

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

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

This means,

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

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

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

and (32) 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 (32) using the fact that,

$$\sup |A_n(x)| \leq \max_i |A_n(X_{(i)})| + \max_i \sup_{x \in [X_{(i)}, X_{(i+1)}]} |A_n(X_{(i)}) - A(x)| \quad (43)$$

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

$$\max_i \sup_{x \in [X_{(i)}, X_{(i+1)}]} |A_n(X_{(i)}) - A(x)| \leq c_2 \max_i \sup_x |X_i - x|, \quad (44)$$

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 Penrose (1964), Theorem 1.1.

**Remark** For the non-differentiable  $\psi(\cdot)$  cases, in particular 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  $\psi_{i,j} \stackrel{\text{def}}{=} |\psi\{\varepsilon_j^* - \tilde{l}_g(X_i) + \hat{l}_g(X_j)\} - \psi\{\varepsilon_i^\# - \tilde{l}(X_i) + l(X_j)\}| = \mathcal{O}_p(h^2)$  because, so

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

Then the argument follows from the above proof.

Moreover, one can also use the strong consistency of  $G_n^\#, G_n^*$  to  $\mathbb{E}(G_n^\#)$  and  $\mathbb{E}^* G_n^*$  respectively based on A.1 and lemma 2.4 of Newey and McFadden (1986).

## A.2 Proof of second part of Theorem 3.1

The number of regressors in (11) is of order  $p = H^{-1}d + 1$ . Portnoy (1997) 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,

$$n^{-1} n^{1/5*2/3} \log n = \mathcal{o}(1) \quad (45)$$

and therefore the condition is satisfied.

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

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

Recall that

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

Lemma 14 of Stone (1985) 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

$$\sup_{x \in B} |\bar{m}(x) - m(x)| \leq C_\infty H^2. \quad (46)$$

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

1.

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

We know first that,

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

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

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

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

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

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

Finally, we can derive that,

$$n^{-1} \sum_{i=1}^n \xi_{i,j} = \mathcal{O}_{a.s.}(H^2 n^{-1/2} \Gamma_n), j \neq 1 \quad (47)$$

The last term

$$n^{-1} |\{\sum_{i=1}^n \varepsilon_i \Phi(X_i)\}_j| = \mathcal{O}_{a.s.}(n^{-1/2} \Gamma_n) \quad (48)$$

Therefore, one has collective term from (47) and (48),

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

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

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

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

where  $\alpha$  and  $\beta$  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

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

since

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

According to assumption A.7,

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

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  $Y_i^\# = m(X_i) + \varepsilon_i^\#$ .

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

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

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

One can derive using coupling argument,

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

## References

- Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Ann. Statist.* **1**: 1071–1095.
- Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives, *Ann. Statist.* **31**: 1852–1884.
- Franke, J. and Mwita, P. (2011). Nonparametric estimates for conditional quantiles of time series, *Scientific Commons*.
- Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications, *Statistics: A Journal of Theoretical and Applied Statistics* **22**: 215–232.
- Hall, P. (1992). *The Bootstrap and Edgeworth expansion*, Springer Verlag.
- Härdle, W. (1989). Asymptotic maximal deviation of  $M$ -smoothers, *J. Multivariate Anal.* **29**: 163–179.
- Härdle, W., Horowitz, J. and Kreiss, J.-P. (2003). Bootstrap method for time series, *International Statistical Review / Revue Internationale de Statistique* **71**(2): 435–459.
- Härdle, W. K. (1990). *Applied Nonparametric Regression*, Cambridge University Press.
- Härdle, W. K., Ritov, Y. and Song, S. (2012). Partial linear quantile regression and bootstrap confidence bands, *Journal of Multivariate Analysis* **107**: 244–262.
- Härdle, W. K. and Song, S. (2010). Confidence bands in quantile regression, *Econometric Theory* **26**: 1180–1200.
- Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *The Annals of Statistics* **19**(2): 778–796.

- Horowitz, J., Klemelä, J. and Mammen, E. (2006). Optimal estimation in additive regression models, *Bernoulli* **12**(2): 271–298.
- Horowitz, J. L. (2001a). *The Bootstrap*, Handbook of Econometrics.
- Horowitz, J. L. (2001b). Nonparametric estimation of a generalized additive model with an unknown link function, *Econometrica* **69**(2): 499–513.
- Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model, *Journal of the American Statistical Association* **100**(472): 1238–1249.
- Huber, P. J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics* **35**(1): 73–101.
- Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates, *Journal of Multivariate Analysis* **12**(402-414).
- Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model, *Econometric Theory* **26**: 159–166.
- Kreiss, J.-P. and Neumann, M. H. (1998). Regression-type inference in nonparametric autoregression, *Annals of Statistics* **26**: 1570–1613.
- Mammen, E. (1992). *When does bootstrap work?: Asymptotic results and simulations*, Springer Verlag.
- Newey, W. K. and McFadden, D. L. (1986). Large sample estimation and hypothesis testing, *Handbook of Econometrics* **4**: 2111–2245.
- Penrose, M. D. (1964). A strong law for the largest nearest-neighbour link between random points, *Journal London Mathematical Society* **60**(3): 951–960.
- Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer Verlag.
- Portnoy, S. (1997). Local asymptotics for quantile smoothing splines, *The Annals of Statistics* **25**(1).
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *The Annals of Statistics* **10**(4): 1040–1053.
- Stone, C. J. (1985). Additive regression and other nonparametric models, *The Annals of Statistics* **13**(2): 689–705.

Yafeh, Y. and Yosha, O. (2003). Large shareholders and banks: Who monitors and how?, *Economic Journal* **113**: 128–146.

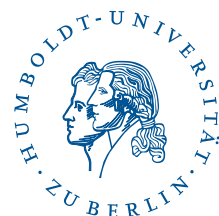
# SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
- 002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
- 003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
- 004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
- 005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
- 006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
- 007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
- 008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
- 009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
- 010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.
- 011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.
- 012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.
- 013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.
- 014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.
- 015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.
- 016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.
- 017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.
- 018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.
- 019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



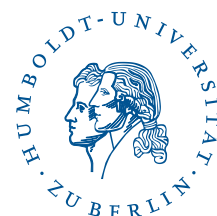
## SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 020 "Disaster Risk in a New Keynesian Model" by Maren Brede, April 2013.
- 021 "Econometrics of co-jumps in high-frequency data with noise" by Markus Bibinger and Lars Winkelmann, May 2013.
- 022 "Decomposing Risk in Dynamic Stochastic General Equilibrium" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 023 "Reference Dependent Preferences and the EPK Puzzle" by Maria Grith, Wolfgang Karl Härdle and Volker Krättschmer, May 2013.
- 024 "Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 025 "The 'Celtic Crisis': Guarantees, transparency, and systemic liquidity risk" by Philipp König, Kartik Anand and Frank Heinemann, May 2013.
- 026 "State Price Densities implied from weather derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Huei-Wen Teng, May 2013.
- 027 "Bank Lending Relationships and the Use of Performance-Sensitive Debt" by Tim R. Adam and Daniel Streitz, May 2013.
- 028 "Analysis of Deviance in Generalized Partial Linear Models" by Wolfgang Karl Härdle and Li-Shan Huang, May 2013.
- 029 "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps" by Markus Bibinger and Mathias Vetter, May 2013.
- 030 "Can expert knowledge compensate for data scarcity in crop insurance pricing?" by Zhiwei Shen, Martin Odening and Ostap Okhrin, May 2013.
- 031 "Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, May 2013.
- 032 "CDO Surfaces Dynamics" by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ostap Okhrin, July 2013.
- 033 "Estimation and Inference for Varying-coefficient Models with Nonstationary Regressors using Penalized Splines" by Haiqiang Chen, Ying Fang and Yingxing Li, July 2013.
- 034 "Robust Estimation and Inference for Threshold Models with Integrated Regressors" by Haiqiang Chen, July 2013.
- 035 "A new perspective on the economic valuation of informal care: The well-being approach revisited" by Konstantin Kehl and Stephan Stahlschmidt, July 2013.
- 036 "Herding in financial markets: Bridging the gap between theory and evidence" by Christopher Boortz, Simon Jurkatis, Stephanie Kremer and Dieter Nautz, July 2013.
- 037 "Default Risk Calculation based on Predictor Selection for the Southeast Asian Industry" by Wolfgang Karl Härdle and Dedi Dwi Prastyo, August 2013.
- 038 "ECB monetary policy surprises: identification through cojumps in interest rates" by Lars Winkelmann, Markus Bibinger and Tobias Linzert, August 2013.
- 039 "Limited higher order beliefs and the welfare effects of public information" by Camille Cornand and Frank Heinemann, August 2013.
- 040 "Privacy Concerns, Voluntary Disclosure of Information, and Unraveling: An Experiment" by Volker Benndorf, Dorothea Kübler and Hans-Theo Normann, September 2013.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".





## SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 041 "Goodness-of-fit Test for Specification of Semiparametric Copula Dependence Models" by Shulin Zhang, Ostap Okhrin, Qian M. Zhou and Peter X.-K. Song, September 2013.
- 042 "Volatility linkages between energy and agricultural commodity prices" by Brenda López Cabrera and Franziska Schulz, September 2013.
- 043 "Testing the Preferred-Habitat Theory: The Role of Time-Varying Risk Aversion" by Till Strohsal, September 2013.
- 044 "Assortative matching through signals" by Friedrich Poeschel, September 2013.
- 045 "Intertemporal Consumption and Debt Aversion: An Experimental Study" by Thomas Meissner, September 2013.
- 046 "Automated Valuation Modelling: A Specification Exercise" by Rainer Schulz, Martin Wersing and Axel Werwatz, October 2013.
- 047 "Tie the straps: uniform bootstrap confidence bands for bounded influence curve estimators" by Wolfgang Karl Härdle, Ya'acov Ritov and Weining Wang, October 2013.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

