Robust estimators of high order derivatives of regression functions *

Graciela Boente
Universidad de Buenos Aires and CONICET, Argentina
and
Daniela Rodriguez
Universidad de Buenos Aires and Fundación Antorchas, Argentina

Abstract

In this paper, robust estimates for the derivatives of order ν of the regression function are considered. This estimator extend the proposals given when $\nu = 1, 2$. Uniform consistency, which allows to construct a robust data–driven bandwidth, is established. Besides, the robust estimates introduced are asymptotically normally distributed and their asymptotic efficiency is that of the related $M-$location estimators.

AMS Subject Classification 1990: Primary 62F35, Secondary 62H25.
Key words and phrases: Robust Estimation, Smoothing Techniques, Bandwidth Selectors, Kernel Weights, Asymptotic Properties.

Corresponding Author
Graciela Boente
Moldes 1855, 3^o A
Buenos Aires, C1428CRA, Argentina
email: gboente@mate.dm.uba.ar

*This research was partially supported by Grant PICT # 03-00000-006277 from ANPCYT, Grant X-094 from the Universidad de Buenos Aires and Grant 13900-6 of the Fundación Antorchas at Buenos Aires, Argentina. The research of Daniela Rodriguez was supported by a scholarship of the Fundación Antorchas.

1 Introduction

The estimation of derivatives from noisy data has been studied by several authors. For random carriers, Schuster and Yakowitz (1979) showed the uniform convergence of the estimates derived from kernel weights while Sarda and Vieu (1988) considered the case of dependent observations. For fixed design, Gasser and Müller (1984) and Gasser et al. (1985) studied estimators based on kernel weights which are linear in the response variables.

As it is well known, nonparametric regression estimators depend on a smoothing parameter that should be choosen by the practitioner. Large bandwidths produce regression estimators with small variance but high bias, while small values produce more wiggly curves. This trade–off between bias and variance lead to several proposals to select the smoothing parameter, such as cross-validation procedures and plug–in methods. The optimum bandwidth is defined as the value that minimizes the integrated mean square error. Obviously, this optimum bandwidth depends on both the regression function and the kernel K. More precisely, if the regression function φ has continuous derivatives up to order ν and a kernel of at least the same order is considered in the estimation procedure, the optimal bandwidth depends on $\varphi^{(\nu)}$. Therefore, the popular plug–in procedure to select a bandwidth estimates the unknown quantities in the expression for the optimal bandwidth, in particular, it substitutes $\varphi^{(\nu)}$ by its estimate. Therefore, estimates of the derivatives can also be used to provide adaptive selectors for the smoothing parameter.

It is well known that in nonparametric regression least squares estimators can be seriously affected by anomalous data. As was noted by Härdle and Gasser (1985) linear estimates of the derivatives are much more sensitive to single outliers than the estimates of the regression function. Härdle and Gasser (1985) considered an homoscedastic regression model and proposed kernel M–estimators to estimate nonparametrically the first derivative of the regression function. They heuristically extend their proposal to higher order derivatives. However, if $\nu > 2$, their extension introduces a bias in the estimation that should be corrected. Our proposal solves this problem.

The sensitivity of the classical bandwidth selectors to anomalous data was discussed by several authors, such as, Leung, Marrot and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997) and Cantoni and Ronchetti (2001). Härdle and Gasser (1984) (for fixed designs) and Hall and Jones (1990) (for random designs) studied approximations to the bias and variance of local M–estimates and considered the mean square error of a local M–estimate, obtaining the optimal bandwidth for it. Therefore, our procedure also allows to construct robust data–driven bandwidths for the regression function that will converge to the optimal one. This entails that the final robust regression estimator will have the same asymptotic distribution as the one defined using the optimal smoothing parameter.

This paper is organized as follows. In Section 2, we propose robust estimators for the derivatives of the regression function under a nonparametric regression model with fixed designs. Uniform consitency is derived in Section 3. All proofs are given in the Appendix.

2 Robust estimation of the derivative of order ν

In this section, we will introduce a robust estimator of the derivative of order ν which generalizes the proposal considered, when $\nu = 2$, by Boente, Fraiman and Meloche (1997). On the other hand, our proposal corrects the bias of the estimates considered by Härdle and Gasser (1985), when $\nu > 2$.

Let $y_i \in I\!R$ be independent observations such that

$$y_i = \varphi(t_i) + u_i \qquad 1 \le i \le n\,, \tag{1}$$

where the errors u_i are independent and identically distributed with symmetric common distribution $F(\cdot/\sigma_u)$ and $0 \le t_1 \le \ldots \le t_n \le 1$ are fixed design points.

As mentioned in the Introduction, robust estimates for the first derivative of the regression function have been studied by Härdle and Gasser (1985), when the scale is known. These authors also discussed the estimation of higher derivatives by formally differentiating the equation defining the local $M-$estimator and then, ignoring residual terms. However, even if these arguments lead to asymptotically unbiased estimators of the second order derivatives of φ, when $\nu > 2$ an asymptotic bias is present and should be corrected. Our proposal corrects this bias by introducing a term involving estimates of the derivatives up to the order $\nu - 2$. As mentioned above, when $\nu = 2$, Boente, Fraiman and Meloche (1997) considered kernel–based estimates for the second derivative of the regression function to provide a robust bandwidth selector in the nonparametric regression setting.

In order to define the estimates, let us denote by $\psi^{(j)}$ the $j-$th derivatives of the score function ψ while $w_{ni}(t,h)$ and $w_{ni}^{(\nu)}(t,h)$ stand for the kernel weights used to estimate the regression function and its $\nu-$th derivative, respectively. More precisely, let $w_{ni}(t,h)$ and $w_{ni}^{(\nu)}(t,h)$ be defined as

$$w_{ni}(t,h) = \frac{1}{nh} K_0\left(\frac{t-t_i}{h}\right), \tag{2}$$

$$w_{ni}^{(\nu)}(t,h) = \frac{1}{nh^{\nu+1}} K^{(\nu)}\left(\frac{t-t_i}{h}\right), \tag{3}$$

with h the bandwidth parameter, $K_0 : I\!R \to I\!R$ a continuous integrable function with compact support and $\int K_0(t)dt = 1$ and $K : I\!R \to I\!R$ is an integrable function differentiable up to order ν with $\nu-$th derivative $K^{(\nu)}$ that satisfies the conditions to be stated below. Note that two different kernels can be used in (2) and (3). Also, Gasser and Müller weights, as in Härdle and Gasser (1985) can be considered and a kernel K_ν of order (ν, k) with $k \ge \nu + 2$ either both even or both odd, as defined by Gasser, Müller and Mammitzsch (1984) can be used in (3). However, according to Lemma 1 in Gasser and Müller (1984), a kernel of order (ν, k) equals the derivative of order ν of a function K with bounded support $[-\tau, \tau]$ satisfying $K^{(j)}(-\tau) = K^{(j)}(\tau) = 0$ for $0 \le j \le \nu - 1$, $\int K(t)dt = 1$, $\int t^j K(t)du = 0$, for $1 \le j \le k - \nu - 1$ and $\int t^{k-\nu} K(t)dt = (-1)^\nu (k-\nu)!\beta/k!$ with $\beta \ne 0$. Assuming that the scale σ_u is known, Härdle and Gasser (1985) suggested to use

$$\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i - \widehat{\varphi}(t)}{\sigma_u}\right)\left[\sum_{i=1}^n w_{ni}(t,h)\psi^{(1)}\left(\frac{y_i - \widehat{\varphi}(t)}{\sigma_u}\right)\right]^{-1}$$

as an estimate of $\varphi^{(\nu)}(t)$ where $\widehat{\varphi}(t)$ is a preliminary robust estimate of the regression function. However, as mentioned above this estimate will be biased if $\nu > 2$. In order to correct the bias, define

$$\widehat{B}_\nu\left(t,\sigma,\varphi\right) = \sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i-\varphi(t)}{\sigma}\right) \tag{4}$$

$$\widehat{C}_\nu\left(t,\sigma,\varphi\right) = \sum_{\substack{3\le j\le \nu\\ j:odd}} \frac{1}{j!\;\sigma^j}\;\lambda_{j,n}(t,\sigma,\varphi)\;\widehat{H}_j(\nu,t) \tag{5}$$

where

$$\lambda_{j,n}(t,\sigma,\varphi) = \sum_{i=1}^n w_{ni}(t,h)\psi^{(j)}\left(\frac{y_i-\varphi(t)}{\sigma}\right) \tag{6}$$

$$H_j(\nu,t) = \left.\{[\varphi(u)-\varphi(t)]^j\}^{(\nu)}\right|_{u=t} \tag{7}$$

and $\widehat{H}_j(\nu,t)$ is an estimate of $H_j(\nu,t)$. The term $\widehat{B}_\nu$ is the numerator of the estimate introduced by Härdle and Gasser (1985) while $\widehat{C}_\nu$ is related to the correction term needed to deal with the bias appearing in their estimation procedure.

Lemma P.1 shows that for $j\ge 3$, $H_j(\nu,t)$ depends continuously only on the derivatives of $\varphi(t)$ of order lower or equal to $\nu-2$ and thus, it can be estimated plugging–in estimators of the derivatives of φ up to order $\nu-2$ in the expression of H_j. It also allows to calculate easily H_j. For instance, using that $H_2(1,t)=0$, $H_2(2,t)=2\;[\varphi^{(1)}(t)]^2$, $H_3(3,t)=6\;[\varphi^{(1)}(t)]^3$ and $H_2(3,t)=6\;\varphi^{(1)}(t)\;\varphi^{(2)}(t)$, we easily obtain $H_3(4,t)=\;36\;[\varphi^{(1)}(t)]^2\varphi^{(2)}(t)$. Therefore, in order to estimate the $4-$th derivative of φ, we only need preliminary estimates of the regression function and of its first and second derivatives. In general, if $H_j(\nu,t)=\Phi_j(\varphi^{(1)}(t),\ldots,\varphi^{(\nu-2)}(t))$, we can define $\widehat{H}_j(\nu,t)=\Phi_j(\widehat{\varphi}^{(1)}(t),\ldots,\widehat{\varphi}^{(\nu-2)}(t))$, where $\widehat{\varphi}^{(j)}$ denote preliminary estimates of $\varphi^{(j)}$.

Remark 2.1. Note also that for any $\mathcal{I}\subset[0,1]$ such that $\sup_{t\in\mathcal{I}}\max_{1\le k\le\nu-2}|\varphi_{k,n}(t)-\varphi_k(t)|\to 0$, we have that $\sup_{t\in\mathcal{I}}|\Phi_j(\varphi_{1,n},\ldots,\varphi_{\nu-2,n})-\Phi_j(\varphi_1,\ldots,\varphi_{\nu-2})|\to 0$. In particular, if we have uniform strongly consistent estimates of $\varphi^{(k)}$, for $1\le k\le \nu-2$ then $\widehat{H}_j(\nu,t)\to H_j(\nu,t)$ uniformly on $\mathcal{I}$.

Let $\widehat{\sigma}_u$ be a robust estimate of the residuals scale such as the robust Rice–type estimator, i.e., $\widehat{\sigma}_u=\frac{1}{2}\operatorname*{median}_{1\le i\le n}|y_i-y_{i-1}|$, and $\widehat{\varphi}_{\mathrm{R}}(\cdot)=\widehat{\varphi}_{\mathrm{R}}(\cdot,h_0)$ a kernel–based M–estimate of the regression function with initial bandwidth h_0, i.e., a solution of

$$\sum_{i=1}^n w_{ni}(t,h_0)\psi\left(\frac{y_i-\widehat{\varphi}_{\mathrm{R}}(t,h_0)}{\widehat{\sigma}_u}\right)=0\;.$$

The robust estimator, $\widehat{\varphi}_{\rm R}^{(\nu)}(t,h)$, of the derivative of order ν of the regression function φ is, then, defined as

$$\widehat{\varphi}_{\rm R}^{(\nu)}(t,h) \quad = \quad \frac{\widehat{B}_\nu\left(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}\right)-\widehat{C}_\nu\left(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}\right)}{\lambda_{1,n}\left(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}\right)}\,\widehat{\sigma}_u\,. \tag{8}$$

This procedure depends on the pilot bandwidth h_0 used to estimate $\widehat{\varphi}_{\rm R}$ and on the preliminary estimates of the derivatives of $\varphi(t)$ up to order $\nu-2$, which obviously also involve a choice for the smoothing parameter.

For instance, when $\nu=4$ we have that a robust estimator of the fourth derivative of the regression function, $\widehat{\varphi}_{\rm R}^{(4)}(t,h)$, is given by

$$\frac{\widehat{\sigma}_u\left\{\sum_{i=1}^n w_{ni}^{(4)}(t,h)\psi\left(\frac{y_i-\widehat{\varphi}_{\rm R}(t)}{\widehat{\sigma}_u}\right)-\frac{6}{\widehat{\sigma}_u^3}\left[\widehat{\varphi}_{\rm R}^{(1)}(t)\right]^2\widehat{\varphi}_{\rm R}^{(2)}(t)\sum_{i=1}^n w_{ni}(t,h)\psi^{(3)}\left(\frac{y_i-\widehat{\varphi}_{\rm R}(t)}{\widehat{\sigma}_u}\right)\right\}}{\sum_{i=1}^n w_{ni}(t,h)\psi^{(1)}\left(\frac{y_i-\widehat{\varphi}_{\rm R}(t)}{\widehat{\sigma}_u}\right)}\,,$$

where $\widehat{\varphi}_{\rm R}(t)=\widehat{\varphi}_{\rm R}(t,h_0)$, $\widehat{\varphi}_{\rm R}^{(1)}(t)$ and $\widehat{\varphi}_{\rm R}^{(2)}(t)$ denote preliminary estimates of $\varphi^{(1)}(t)$ and $\varphi^{(2)}(t)$, respectively. It is worthwhile noticing, that the estimate considered by Härdle and Gasser (1985), involves only the first term of the numerator, while the second one is the one which corrects its asymptotic bias.

3 Consistency

Under regularity conditions, Boente, Fraiman and Meloche (1997) showed that the robust estimates of the second derivatives of regression function converges to the second derivatives of regression function. In this section, we will consider the following assumptions to derive the weak consistency of the estimates $\widehat{\varphi}_{\rm R}^{(\nu)}(t,h)$ defined in (8).

C.1. $\{t_i\}_{i=1}^n$ are fixed design points in $[0,1]$, $0\le t_1\le\ldots\le t_n\le 1$, such that $t_0=0$ and $t_{n+1}=1$ and $\max\limits_{1\le i\le n+1}\left|(t_i-t_{i-1})-\frac{1}{n}\right|=O(n^{-\delta})$ for some $\delta>1$.

C.2. $\{u_i: 1\le i\le n\}$ are i.i.d. random variables such that $u_i\sim F(./\sigma_u)$, F symmetric.

C.3. The function $\varphi(\cdot)$ has ν continuous derivatives on [0,1].

C.4. **a)** $K(\cdot)$ is a function with support $[-1,1]$, $\nu-$th differentiable. $K_0(\cdot)$ is a continuous function with support $[-1,1]$.

b) $\int K_0(u)du=1$, $\int K(u)du=1$, $K^{(j)}(1)=K^{(j)}(-1)=0$ for $0\le j\le\nu-1$.

c) $K^{(\nu)}$ and K_0 are Lipschitz function of order one.

C.5. For $1\le k\le\nu-2$, $\sup\limits_{t\in[h,1-2h]}|\widehat{\varphi}_{\rm R}^{(k)}(t)-\varphi^{(k)}(t)|\xrightarrow{a.s.}0$, as $n\to\infty$.

C.6. The initial estimators $\widehat{\varphi}_{\mathrm{R}}(t)$ satisfy
a) $\sup_{t\in[0,1]} |\widehat{\varphi}_{\mathrm{R}}(t) - \varphi(t)| \xrightarrow{a.s.} 0$ as $n \to \infty$.
b) There exists $\epsilon_0 > 0$ such that

$$\sum_{n=1}^{\infty} P\left(\sup_{t\in[h,1-2h]} |\widehat{\varphi}_{\mathrm{R}}(t) - \varphi(t)| > \epsilon_0\, h \right) < \infty \,.$$

C.7. $\lim_{n\to\infty} \dfrac{nh^{2\nu+1}}{\log n} = +\infty$, $\lim_{n\to\infty} h = 0$.

C.8. $\psi : R \to R$ is an odd, strictly increasing, bounded and continuous function such that $\lim_{t\to\infty} \psi(t) = a > 0$.

C.9. ψ is $(\nu+1)$-th differentiable. Moreover, for any $1 \le \ell \le \nu+1$, , the ℓ-th derivative $\psi^{(\ell)}$ and $\upsilon_\ell(t) = t\,\psi^{(\ell)}(t)$ are bounded.

C.10. The estimator $\widehat{\sigma}_u$ de σ_u satisfy $\widehat{\sigma}_u \xrightarrow{a.s.} \sigma_u$ if $n \to \infty$.

Theorem 3.1. *Under* **C.1** *to* **C.10***, if* $n^{\delta-1}h^{\nu+1} \to \infty$, $nh^{\nu+2} \to \infty$ *and* $E\psi^{(1)}\left(\dfrac{u_1}{\sigma_u}\right) \neq 0$, *we have that*

$$\sup_{t\in[h,1-2h]} |\widehat{\varphi}_{\mathrm{R}}^{(\nu)}(t,h) - \varphi^{(\nu)}(t)| \xrightarrow{a.s.} 0 \quad as \quad n \to \infty.$$

Remark 3.1. Note that if $\nu \ge 2$ and **C.7** holds then $n\,h^{\nu+2} \to \infty$. Moreover, using similar arguments as those considered in Theorem 2 from Härdle and Luckhaus (1984), it is easy to show that **C.6** holds for the kernel M–smoother.

4 Asymptotic Normality

Under regularity conditions, Härdle and Gasser (1985) showed that the robust estimates of the first derivative of regression function is asymptotically normally distributed. In this section, we will consider the following assumptions to derive the asmptotic distribution of the estimates $\widehat{\varphi}_{\mathrm{R}}^{(\nu)}(t,h)$ defined in (8).

N.1. **a)** $K(\cdot)$ is a function with support $[-1,1]$, $k-$th differentiable with $k \ge \nu+2$.

b) $\int K(u)du = 1$, $K^{(j)}(1) = K^{(j)}(-1) = 0$ for $0 \le j \le \nu-1$, $\int u^j K(u)du = 0$ for $1 \le j \le k-\nu-1$, $\int u^{k-\nu} K(u)du = (-1)^\nu\, \frac{(k-\nu)!}{k!} \beta \neq 0$.

c) $K_0(\cdot)$ is a continuous function with support $[-1,1]$. $\int K_0(u)du = 1$ and for some $\ell \ge 2$ it holds $\int u^j K_0(u)du = 0$ for $1 \le j \le \ell$, $\int u^\ell K_0(u)du \neq 0$.

c) $K^{(\nu)}$ and K_0 are Lipschitz function of order one.

N.2. The function φ is $r-$th differentiable with $r = \max(k, \ell)$, $k \geq \nu + 2$ and $\varphi^{(r)}$ is continuous at t.

N.3. For $1 \leq s \leq \nu - 2$, $|\widehat{\varphi}_{\rm R}^{(s)}(t) - \varphi^{(s)}(t)| \xrightarrow{p} 0$ and $\sqrt{nh^{2\nu+1}}\left(\widehat{H}_j(\nu, t) - H_j(\nu, t)\right) \xrightarrow{p} 0$.

N.4. The initial estimators $\widehat{\varphi}_{\rm R}(t)$ satisfy $\sqrt{nh}\,(\widehat{\varphi}_{\rm R}(t) - \varphi(t)) = O_p(1)$.

N.5. The estimator $\widehat{\sigma}_u$ de σ_u satisfy $\widehat{\sigma}_u \xrightarrow{p} \sigma_u$.

N.6. $\lim_{n\to\infty} nh^{2\nu+1} = +\infty$, $\lim_{n\to\infty} nh^{2k+1} = 0$, $\lim_{n\to\infty} nh^{2\ell+2\nu+1} = 0$.

Remark 4.1. Lemma 1 in Gasser and Müller (1984) shows that assumption **N.1** entails that $\int u^k K^{(\nu)}(u)du = \beta$ and $\int u^\nu K^{(\nu)}(u)du = (-1)^\nu \nu!$. Assumption **N.5** is fulfilled if $\widehat{\sigma}_u$ is the robust Rice–type estimator while **N.4** is satisfied when $\widehat{\varphi}_{\rm R}(\cdot) = \widehat{\varphi}_{\rm R}(\cdot, h)$.

Theorem 4.1. *Under **C.1**, **C.2**, **C.8**, **C.9** and **N.1** to **N.6** if in addition ψ is continuously differentiable up to order $\nu + \ell$ and $E\psi^{(1)}\left(\frac{u_1}{\sigma_u}\right) \neq 0$, we have that*

$$\sqrt{nh^{2\nu+1}}\left(\widehat{\varphi}_{\rm R}^{(\nu)}(t,h) - \varphi^{(\nu)}(t)\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma_u^2\left[\int_{-1}^{1}\left(K^{(\nu)}(u)\right)^2 du\right] \frac{E\psi^2\left(\frac{u_1}{\sigma_u}\right)}{E^2\psi^{(1)}\left(\frac{u_1}{\sigma_u}\right)}\right).$$

It is worthnoticing that Theorem 4.1 entails that the robust estimators introduced will have the same asymptotic efficiency as the location $M-$estimators defined through the score function ψ.

5 Concluding Remarks

Selection of the smoothing parameter is an important step in any nonparametric analysis, even when robust estimates are used. The classical procedures based on least squares cross–validation or on a plug–in rule turn out to be non–robust since they lead to over or undersmoothing as noted for nonparametric regression by Leung, Marrot and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997) and Cantoni and Ronchetti (2001). For any robust smoother $\widehat{\varphi}_{\rm R}(t,h)$, the optimal bandwidth $h_{\rm opt}^{\rm R}$, is defined as the value of h minimizing the integrated mean square error curve

$$\text{MISER}(h) = E\int_0^1 \left(\widehat{\varphi}_{\rm R}(t,h) - \varphi(t)\right)^2 dt.$$

According to Härdle and Gasser (1984), if the function has $\nu \geq 2$ continuous derivatives and a kernel, K_ν, of order ν is used to estimate $\varphi(t)$, $h_{\rm opt}^{\rm R}$ is related to the optimal bandwidth for the classical estimates, $h_{\rm opt}$, through the relation $h_{\rm opt}^{\rm R} = h_{\rm opt} V^{\frac{1}{2\nu+1}}$, where

$V = E\psi^2\left(\frac{u_1}{\sigma_u}\right)\left[E^2\psi^{(1)}\left(\frac{u_1}{\sigma_u}\right)\right]^{-1}$ gives the efficiency of the local M–estimate and

$$h_{\text{opt}} = n^{\frac{1}{2\nu+1}}\left(\frac{(\nu!)^2\, C_1\left(K_\nu\right)\sigma_u^2}{2\nu C_2^2\left(K_\nu\right) T\left(\varphi^{(\nu)}\right)}\right)^{\frac{1}{2\nu+1}}$$

where $C_1(K_\nu) = \int K_\nu^2(t)\,dt$, $C_2(K_\nu) = \int u^\nu\, K_\nu(u)\,du$ and $T(f) = \int_0^1 f^2(u)\,du$. Therefore, a robust bandwidth selector, $\widehat{h}$, based on the estimates defined in Section 2 can be constructed. This robust bandwidth selector tends to overcome the well known sensitivity of the classical selectors.

Theorem 3.1 entails that $\frac{\widehat{h}}{h_{\text{opt}}^{\text{R}}} \stackrel{a.s.}{\longrightarrow} 1$ and so the the data–driven is asymptotic equivalent to the optimal bandwidth. Theorems 3.1 and 3.2 from Boente, Fraiman and Meloche (1997) allow us to conclude that the kernel based M–estimates of φ obtained using a plug–in bandwidth selector, $\widehat{h}$, based on $\widehat{\varphi}_{\text{R}}^{(\nu)}$ will be consistent. Moreover, it will be asymptotically equivalent to the kernel based M–estimates obtained using the optimal bandwidth, $h_{\text{opt}}^{\text{R}}$, based on a $\nu-$th degree of smoothness, in the sense, that $(n\, h_{\text{opt}}^{\text{R}})^{1/2}\left[\widehat{\varphi}_{\text{R}}(t, h_{\text{opt}}^{\text{R}}) - \widehat{\varphi}_{\text{R}}(t, \widehat{h})\right] \stackrel{p}{\longrightarrow} 0$. This asymptotically equivalence, entails that, $(n\, h_{\text{opt}}^{\text{R}})^{1/2}\left[\widehat{\varphi}_{\text{R}}(t, \widehat{h}) - \varphi(t)\right]$ is asymptotically normally distributed.

Our proposal for the derivatives of the regression function can therefore be used to provide a data–driven bandwidth selector. It also corrects the bias of the estimates considered by Härdle and Gasser (1985), when $\nu > 2$. It is worth noticing that a ready–to–use robust plug–in bandwidth selector can be defined by using a robust version of the iterative schemes proposed discussed in Ruppert, Sheater and Wand (1995).

P Appendix: Proofs.

Lemma P.1. *For any fixed t and $j \geq 2$, let $v_j(u) = [\varphi(u) - \varphi(t)]^j$. Then we have that for $k \geq j$, $v_j^{(k)}(t) = H_j(k,t)$ depends only on $\varphi^{(\ell)}(t)$ with $\ell \leq k-1$. Moreover, if $j \geq 3$, and $k \geq j$, $v_j^{(k)}(t)$ depends only on $\varphi^{(\ell)}(t)$ with $\ell \leq k-2$.*

Proof. Since $v_1^{(k)}(u) = \varphi^{(k)}(u)$, we have that

$$v_j^{(k)}(u) \;=\; (v_{j-1}(u)v_1(u))^{(k)} = \sum_{r=0}^{k}\binom{k}{r} v_{j-1}^{(r)}(u)\, v_1^{(k-r)}(u)\,.$$

Using that $v_1(t) = 0$ and $v_{j-1}(t) = 0$, and that $v_{j-1}^{(1)}(t) = 0$, for $j \geq 3$, we get

$$v_j^{(k)}(t) \;=\; \begin{cases} \displaystyle\sum_{r=1}^{k-1}\binom{k}{r} v_{j-1}^{(r)}(t)\varphi^{(k-r)}(t) & \text{for} \quad j \geq 2 \\ \displaystyle\sum_{r=2}^{k-1}\binom{k}{r} v_{j-1}^{(r)}(t)\varphi^{(k-r)}(t) & \text{for} \quad j \geq 3 \end{cases}$$

and thus, the result follows. □

In order to prove Theorem 3.1, we will need the following lemmas. From now on, denote by $\mathcal{I} = [h, 1-2h]$.

Lemma P.2. *If* **C.1**, **C.3** *and* **C.4** *hold and, in addition,* $\lim_{n\to\infty} nh^{\nu+2} = +\infty$ *and* $\lim_{n\to\infty} n^{\delta-1}h^{\nu+1} = +\infty$, *we have that*

(a) $\lim_{n\to\infty} \sup_{t\in\mathcal{I}} \left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right) \right| = 0.$

(b) $\lim_{n\to\infty} \sup_{t\in\mathcal{I}} \left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right)(\varphi(t_i)-\varphi(t)) - \frac{1}{h^{\nu}}\int_{-1}^{1} K^{(\nu)}(u)\varphi(t-uh)du \right| = 0.$

Proof. The proof follows the same steps as in Lemma A.1 from Boente, Fraiman and Meloche (1997). (a) Note first that **C.4** entails that

$$\int_0^1 K^{(\nu)}\left(\frac{t-u}{h}\right) du = 0 \qquad \text{for} \quad t \in [h, 1-h]. \tag{P.1}$$

Using **C.1** we have that, for some $C > 0$ there exists n_0 such that for $n \geq n_0$, $nh > C$ and $|t_n - 1| \leq \frac{C}{n}$, and so, for $t \in [h, 1-2h]$,

$$\int_{t_n}^1 K^{(\nu)}\left(\frac{t-u}{h}\right) du = 0 \quad \forall n \geq n_0 \ . \tag{P.2}$$

Denote by M_ν the Lipschitz constant for $K^{(\nu)}$ and by ξ_i an intermediate point $t_{i-1} \leq \xi_i \leq t_i$. Then, from (P.1) and (P.2), we get

$$\begin{aligned}
\left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right) \right| &= \left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right) - \frac{1}{h^{\nu+1}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} K^{(\nu)}\left(\frac{t-u}{h}\right) du \right| \\
&= \frac{1}{h^{\nu+1}} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ K^{(\nu)}\left(\frac{t-t_i}{h}\right) - n(t_i - t_{i-1}) K^{(\nu)}\left(\frac{t-\xi_i}{h}\right) \right\} \right| \\
&\leq \frac{M_\nu}{h^{\nu+1}} \quad \frac{1}{n} \sum_{i=1}^{n} \left| \frac{t_i - \xi_i}{h} \right| + \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} \left| K^{(\nu)}\left(\frac{t-\xi_i}{h}\right) \right| |1 - n(t_i - t_{i-1})| \\
&\leq \frac{M_\nu}{h^{\nu+2}} \quad \sup_i |t_i - t_{i-1}| + ||K^{(\nu)}||_\infty \frac{1}{h^{\nu+1}} \sum_{i=1}^{n} \left| \frac{1}{n} - (t_i - t_{i-1}) \right| \\
&\leq \frac{M_\nu C}{nh^{\nu+2}} \quad + \|K^{(\nu)}\|_\infty \frac{O(1)}{h^{\nu+1} n^{\delta-1}}
\end{aligned}$$

and the result follows.
(b) As in (a), using that $K^{(\nu)}$ has compact support on $[-1, 1]$, we have that

$$\frac{1}{h^{\nu+1}} \int_0^1 K^{(\nu)}\left(\frac{t-u}{h}\right) \varphi(u) du = \frac{1}{h^{\nu}} \int_{-1}^1 K^{(\nu)}(u) \varphi(t - uh) du \ .$$

On the other hand, the boundness of φ and (a) imply that,

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)\varphi(t)\right|\to 0\ .$$

Thus, it remains to show that

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)\varphi(t_i)-\frac{1}{h^{\nu+1}}\int_0^1 K^{(\nu)}\left(\frac{t-u}{h}\right)\varphi(u)du\right|\to 0\ .$$

As in (a), let n_0 be such that for $n\geq n_0$, $nh>C$ and $1-t_n\leq\dfrac{C}{n^{\delta}}$. Then, using analogous arguments as those considered in (a), we get that for $n\geq n_0$

$$\begin{aligned}
&\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)\varphi(t_i)-\frac{1}{h^{\nu+1}}\int_0^1 K^{(\nu)}\left(\frac{t-u}{h}\right)\varphi(u)du\right|\\
=&\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}\left\{K^{(\nu)}\left(\frac{t-t_i}{h}\right)\varphi(t_i)-n\int_{t_{i-1}}^{t_i}K^{(\nu)}\left(\frac{t-u}{h}\right)\varphi(u)du\right\}-\right.\\
&\left.-\frac{1}{h^{\nu+1}}\int_{t_n}^{1}K^{(\nu)}\left(\frac{t-u}{h}\right)\varphi(u)du\right|\\
\leq&\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}\left\{K^{(\nu)}\left(\frac{t-t_i}{h}\right)\varphi(t_i)-n(t_i-t_{i-1})K^{(\nu)}\left(\frac{t-\xi_i}{h}\right)\varphi(\xi_i)\right\}\right|+\\
&+\|K^{(\nu)}\|_\infty\|\varphi\|_\infty\frac{C}{n^{\delta}\,h^{\nu+1}}\ .
\end{aligned}$$

which concludes the proof using that $K^{(\nu)}$ and φ are bounded Lipschitz functions. $\square$

Using similar arguments, we obtain the following Lemma.

Lemma P.3. *If **C.1**, **C.3** and **C.4** hold and if, in addition,* $\lim_{n\to\infty}nh^2=+\infty$ *and* $\lim_{n\to\infty}n^{\delta-1}h^1=+\infty$, *we have that*

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh}\sum_{i=1}^{n}\left|K_0\left(\frac{t-t_i}{h}\right)^p\right|-\int_{-1}^{1}|K_0(u)|^p du\right|\to 0,\quad p=1,2$$

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h}\right)^p\right|-\int_{-1}^{1}|K^{(\nu)}(u)|^p du\right|\to 0,\quad p=1,2.$$

Lemma P.4. *Let $\{y_i : i\geq 1\}$ be a sequence of i.i.d random variables. Assume that η is a continuously differentiable and bounded function such that $\upsilon(t)=t\,\eta'(t)$ is bounded. Then, under **C.1**, **C.4**, **C.7**, for any bounded continuous function $m:[0,1]\to I\!R$, we have that for any compact set $\mathcal{C}\subset I\!R^+$*

$$a)\ \sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)m(t_i)\left[\eta\left(\frac{y_i}{\sigma}\right)-E\eta\left(\frac{y_i}{\sigma}\right)\right]\right|\xrightarrow{a.s.}0\ .$$

b) $\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} \left| \frac{1}{nh} \sum_{i=1}^{n} K_0\left(\frac{t-t_i}{h}\right) m(t_i) \left[\eta\left(\frac{y_i}{\sigma}\right) - E\eta\left(\frac{y_i}{\sigma}\right)\right]\right| \xrightarrow{a.s.} 0$.

PROOF. We will only prove a), since the proof of b) follows similarly. Denote $S_n(t,\sigma) = \frac{1}{nh^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right) m(t_i) V_{i,\sigma}$ with $V_{i,\sigma} = \sigma \left[\eta\left(\frac{y_i}{\sigma}\right) - E\eta\left(\frac{y_i}{\sigma}\right)\right]$. It is enough to show that $\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} S_n(t,\sigma) \xrightarrow{a.s.} 0$.

Since η is bounded and $\mathcal{C}$ is a compact set, we have that for some fixed constant M, $|V_{i,\sigma}| < M$, for all $\sigma \in \mathcal{C}$. Thus, using that m is bounded and

$$\sup_{t\in\mathcal{I}} \left| \frac{1}{nh} \sum_{i=1}^{n} \left[K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right]^2 - \int_{-1}^{1} [K^{(\nu)}(u)]^2 du \right| \to 0$$

Bernstein's inequality implies that, for some positive constant α, we have that for $n \ge n_0$

$$\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} P(|S_n(t,\sigma)| > \epsilon) \le 2e^{-\alpha n h^{2\nu+1}} . \tag{P.3}$$

Denote $\mathcal{T} = [0,1] \times \mathcal{C}$ and $B_r(a,b)$ the ball of radius r and center $(a,b) \in \mathcal{T}$. Then, for any $\gamma > \nu + 2$ there exist $(a_1, b_1), \ldots, (a_\ell, b_\ell) \in \mathcal{T}$ with $\ell = \ell_n = O(h^{-\gamma})$ such that $\mathcal{T} \subset \bigcup_{i=1}^{\ell} B_{h^\gamma}(a_i, b_i)$. For the sake of notation simplicity, let $\mathcal{B}_i$ stand for $B_{h^\gamma}(a_i, b_i)$, then,we have that

$$\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} |S_n(t,\sigma)| \le \max_{1\le j\le \ell} \sup_{(t,\sigma)\in\mathcal{B}_j} |S_n(t,\sigma) - S_n(a_j,b_j)| + \max_{1\le j\le l} |S_n(a_j,b_j)| . \tag{P.4}$$

Using that m, η, υ and $K^{(\nu)}$ are bounded functions and that $K^{(\nu)}$ is Lipschitz, straighforward calculations lead to $|S_n(t,\sigma) - S_n(a_j,b_j)| \le A_1 h^{\gamma-(\nu+2)}$ for $(t,\sigma) \in \mathcal{B}_j$, for some positive constant A_1, which entails that for $n \ge n_1$

$$\max_{1\le j\le \ell} \sup_{(t,\sigma)\in\mathcal{B}_j} |S_n(t,\sigma) - S_n(a_j,b_j)| \le \epsilon . \tag{P.5}$$

Finally, (P.3), (P.4) and (P.5) entail that, for $n \ge \max\{n_1, n_0\}$,

$$\begin{aligned} P(\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} |S_n(t,\sigma)| > 2\epsilon) &\le P\left(\max_{1\le j\le \ell_n} |S_n(a_j,b_j)| \ge \epsilon\right) \le 2\ell_n \, e^{-\alpha n h^{2\nu+1}} \\ &\le C h^{-\gamma} n^{-\alpha\delta_n} \le C\,(nh)^{-\gamma}\, n^{\gamma-\alpha\delta_n} , \end{aligned}$$

where, from **C.7**, $\delta_n = nh^{2\nu+1}/\log n \to \infty$. Taking $\gamma = \nu + 3$, and since $\delta_n \to \infty$ and $nh \to \infty$ we have that for $n \ge n_3$, $(nh)^{-\gamma} < 1$ and $\gamma - \alpha\delta_n < -2$. Hence, for $n \ge \max_{1\le i\le 3}(n_i)$, $P(\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} |S_n(t,\sigma)| > 2\epsilon) \le Cn^{-2}$, which shows that $\sum_{i=1}^{n} P(\sup_{t\in\mathcal{I}} \sup_{\sigma\in\mathcal{C}} |S_n(t,\sigma)| > 2\epsilon) < \infty$ concluding the proof. $\square$

Remark P.1. It is worthwhile noticing that Lemma P.2 and Lemma P.4 entail that under the conditions stated therein,

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)(\varphi(t_i)-\varphi(t))^j-\frac{1}{h^\nu}\int_{-1}^{1}K^{(\nu)}(u)(\varphi(t-uh)-\varphi(t))^j du\right| \quad\to\quad 0$$

$$\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)(\varphi(t_i)-\varphi(t))^j\left[\eta\left(\frac{y_i}{\sigma}\right)-E\eta\left(\frac{y_i}{\sigma}\right)\right]\right| \quad\xrightarrow{a.s.}\quad 0\,.$$

Moreover, if we denote $v_j(u)=(\varphi(u)-\varphi(t))^j$ using that $\frac{1}{h^\nu}\int_{-1}^{1}K^{(\nu)}(u)\left[\varphi(t-uh)-\varphi(t)\right]^j du=\int_{-1}^{1}K(u)v_j^{(\nu)}(t-uh)$ and $\int_{-1}^{1}K(u)v_j^{(\nu)}(t-uh)\to v_j^{(\nu)}(t)=H_j(\nu,t)$, we get

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)(\varphi(t_i)-\varphi(t))^j-H_j(\nu,t)\right| \quad\to\quad 0$$

$$\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}\left|\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{t-t_i}{h}\right)(\varphi(t_i)-\varphi(t))^j\eta\left(\frac{y_i}{\sigma}\right)-E\eta\left(\frac{Z_1}{\sigma}\right)H_j(\nu,t)\right| \quad\xrightarrow{a.s.}\quad 0\,.$$

Lemma P.5. *Under* ***C.1*** *to* ***C.10****, if $n^{\delta-1}h^{\nu+1}\to\infty$ as $n\to\infty$, we have that for any compact set $\mathcal{C}\subset I\!R^+$*

(a) $\displaystyle\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}\left|\sum_{i=1}^{n}w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i-\varphi(t)}{\sigma}\right)-\varphi^{(\nu)}(t)\lambda_1(\sigma)-\sum_{\substack{3\le j\le\nu\\ j:odd}}\frac{\lambda_j(\sigma)}{\sigma^j\,j!}H_j(\nu,t)\right|\xrightarrow{a.s.}0$*, with* $w_{ni}^{(\nu)}(t,h)$ *and H_j defined in (3) and (7), respectively and* $\lambda_j(\sigma)=E\psi^{(j)}\left(\frac{u_1}{\sigma}\right)$.

(b) $\displaystyle\sup_{t\in\mathcal{I}}|\lambda_{j,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R})-\lambda_j(\sigma_u)|\xrightarrow{a.s.}0$*, where $\lambda_{j,n}(t,\sigma,\varphi)$ is defined in (6).*

Proof: (a) A Taylor expansion of order ν gives

$$\sum_{i=1}^{n}w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i-\varphi(t)}{\sigma}\right)=\sum_{j=0}^{\nu}S_{jn}(t,\sigma)+\frac{1}{\sigma^\nu\,\nu!}R_n(t,\sigma)$$

where

$$\begin{aligned}S_{jn}(t,\sigma)&=\frac{1}{\sigma^j\,j!}\sum_{i=1}^{n}w_{ni}^{(\nu)}(t,h)\psi^{(j)}\left(\frac{u_i}{\sigma}\right)\left[\varphi(t_i)-\varphi(t)\right]^j\\ R_n(t,\sigma)&=\sum_{i=1}^{n}w_{ni}^{(\nu)}(t,h)\left[\psi^{(\nu)}\left(\frac{u_i+\xi_i}{\sigma}\right)-\psi^{(\nu)}\left(\frac{u_i}{\sigma}\right)\right]\left[\varphi(t_i)-\varphi(t)\right]^\nu\end{aligned}$$

with $|\xi_i|\le|\varphi(t_i)-\varphi(t)|$.
Using the oddness of ψ and **C.2**, we get that $E(\psi^{(j)}(u_1))=0$ for j even. Therefore, Lemma P.4 entail that $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}|S_{jn}(t,\sigma)|\xrightarrow{a.s.}0$ for j even.

Let C_1 and C_2 be the Lipschitz constants of $\psi^{(\nu)}$ and φ, respectively. Then, since $|\xi_i| \leq |\varphi(t_i) - \varphi(t)| \leq C_2|t_i - t|$, and K has compact support on $[-1,1]$, we get

$$\begin{aligned} |R_n(t,\sigma)| &\leq C_2^\nu \frac{1}{nh^{\nu+1}} \sum_{i=1}^n \left| K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right| |\psi^{(\nu)}\left(\frac{u_i+\xi_i}{\sigma}\right) - \psi^{(\nu)}\left(\frac{u_i}{\sigma}\right)| |t-t_i|^\nu \\ &\leq \frac{C_1\, C_2^{\nu+1}}{\sigma} \frac{1}{nh^{\nu+1}} \sum_{i=1}^n \left| K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right| |t-t_i|^{\nu+1} \leq \frac{C_1\, C_2^{\nu+1}}{\sigma}\, h\, \frac{1}{nh} \sum_{i=1}^n \left| K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right| . \end{aligned}$$

Lemma P.3 entails that $\sup_{t\in\mathcal{I}} \left| \frac{1}{nh}\sum_{i=1}^n \left|K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right| - \int_{-1}^1 |K^{(\nu)}(u)|du \right| \to 0$, which together with the compactness of $\mathcal{C}$, allow to conclude that $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}} \frac{1}{\sigma^\nu}|R_n(t,\sigma)| \xrightarrow{a.s.} 0$ using that $\lim_{n\to\infty} h = 0$.

To conclude the proof of (a), it remains to show that

i) $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}} \left| S_{1n}(t,\sigma) - \frac{1}{\sigma} E\psi^{(1)}\left(\frac{u_1}{\sigma}\right)\varphi^{(\nu)}(t)\right| \xrightarrow{a.s.} 0$

ii) $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}} \left| S_{jn}(t,\sigma) - \frac{1}{\sigma^j j!} E\psi^{(j)}\left(\frac{u_1}{\sigma}\right)\ H_j(\nu,t)\right| \xrightarrow{a.s.} 0$, when j is odd and $j \geq 3$.

i) Lemma P.4 implies that

$$\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}} \left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h}\right) \left[\psi^{(1)}\left(\frac{u_i}{\sigma}\right) - E\psi^{(1)}\left(\frac{u_1}{\sigma}\right)\right] [\varphi(t_i) - \varphi(t)] \right| \xrightarrow{a.s.} 0\ ,$$

thus, it will be enough to show that

$$\sup_{t\in\mathcal{I}} \left| \frac{1}{nh^{\nu+1}} \sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h}\right) [\varphi(t_i) - \varphi(t)] - \varphi^{(\nu)}(t) \right| \to 0\ , \tag{P.6}$$

which follows from Lemma P.2, the fact that $\varphi^{(\nu)}$ is Lipschitz and the equality

$$\frac{1}{h^\nu}\int_{-1}^1 K^{(\nu)}(u)\varphi(t-uh)du = \int_{-1}^1 K(u)\varphi^{(\nu)}(t-uh)du\ .$$

ii) Again, using Lemma P.4 we have that

$$\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}} \left| \sum_{i=1}^n w_{ni}^{(\nu)}(t,h) \left[\psi^{(1)}\left(\frac{u_i}{\sigma}\right) - E\psi^{(1)}\left(\frac{u_1}{\sigma}\right)\right] [\varphi(t_i) - \varphi(t)]^j \right| \xrightarrow{a.s.} 0\ ,$$

and so, as in i), it will be enough to show that,

$$\sup_{t\in\mathcal{I}} \left| \sum_{i=1}^n w_{ni}^{(\nu)}(t,h)[\varphi(t_i) - \varphi(t)]^j - H_j(\nu,t) \right| \to 0\ ,$$

which follows from Remark P.1.

(b) From Lemma P.4, we get $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}|\lambda_{j,n}(t,\sigma,\varphi(t))-\lambda_j(\sigma)|\xrightarrow{a.s.}0$. Thus, in order to obtain (b), from **C.10** and the continuity of $\lambda_j(\sigma)$, it remains to show that $\sup_{t\in\mathcal{I}}|\lambda_{j,n}(t,\widehat{\sigma}_u,\varphi)-\lambda_{j,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})|\xrightarrow{a.s.}0$.

Using that $\psi^{(j)}$ is a Lipschitz function, we obtain

$$|\lambda_{j,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\lambda_{j,n}(t,\widehat{\sigma}_u,\varphi)| \le \frac{1}{\widehat{\sigma}_u}\|\psi^{(j+1)}\|_\infty|\widehat{\varphi}_{\mathrm{R}}(t)-\varphi(t)|\frac{1}{nh}\sum_{i=1}^n\left|K_0\left(\frac{t-t_i}{h}\right)\right|,$$

which implies the desired result using **C.6**, **C.10** and the fact that

$$\sup_{t\in\mathcal{I}}\left|\frac{1}{nh}\sum_{i=1}^n\left|K_0\left(\frac{t-t_i}{h}\right)\right|-\int_{-1}^1|K_0(u)|du\right|\to 0.\ \square$$

Proof of Theorem 3.1 Let $\widehat{N}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})=\widehat{B}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\widehat{C}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\varphi^{(\nu)}(t)\lambda_{1,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})$ and note that

$$\begin{aligned}\sup_{t\in\mathcal{I}}\left|\widehat{\varphi}_{\mathrm{R}}^{(\nu)}(t,h)-\varphi^{(\nu)}(t)\right| &= \sup_{t\in\mathcal{I}}|\lambda_{1,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})|^{-1}\left|\widehat{N}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})\right|\\ &\le \{\inf_{t\in\mathcal{I}}|\lambda_{1,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})|\}^{-1}\sup_{t\in\mathcal{I}}\left|\widehat{N}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})\right|.\end{aligned}$$

Since $\inf_{t\in\mathcal{I}}|\lambda_{1,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})|\ge\left|E\psi^{(1)}\left(\frac{u_1}{\widehat{\sigma}_u}\right)\right|-\sup_{t\in\mathcal{I}}\left|\lambda_{1,n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-E\psi^{(1)}\left(\frac{u_1}{\widehat{\sigma}_u}\right)\right|$, from Lemma P.5(b), **C.10** and the fact that $E\psi^{(1)}\left(\frac{u_1}{\sigma_u}\right)\neq 0$, it will be enough to show that $\sup_{t\in\mathcal{I}}|\widehat{N}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})|\xrightarrow{a.s.}0$. According to Lemma P.5, if $\mathcal{C}$ denotes the closure of a neighborhood of σ_u, it suffices to show that

$$\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}\left|\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\left[\psi\left(\frac{y_i-\widehat{\varphi}_{\mathrm{R}}(t)}{\sigma}\right)-\psi\left(\frac{y_i-\varphi(t)}{\sigma}\right)\right]\right| \xrightarrow{a.s.} 0 \quad \text{(P.7)}$$

$$\sup_{t\in\mathcal{I}}|\widehat{H}_j(\nu,t)-H_j(\nu,t)| \xrightarrow{a.s.} 0 \quad \text{(P.8)}$$

Since $H_j(\nu,t)=\Phi_j(\varphi^{(1)}(t),\ldots,\varphi^{(\nu-2)}(t))$ and $\widehat{H}_j(\nu,t)=\Phi_j(\widehat{\varphi}^{(1)}(t),\ldots,\widehat{\varphi}^{(\nu-2)}(t))$, using **C.5** and the uniform continuity of Φ_j, we get (P.8). In order to prove (P.7) using a Taylor's expansion of order ν, as in Lemma P.5 (a), we obtain that

$$\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i-\widehat{\varphi}_{\mathrm{R}}(t)}{\sigma}\right)=\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\psi\left(\frac{y_i-\varphi(t)}{\sigma}\right)+\sum_{j=1}^\nu\frac{1}{\sigma^j\, j!}S_{jn}^*(t)+\frac{1}{\sigma^\nu\,\nu!}R_n^*(t)$$

where $|\xi_{i,n}^*|\le|\varphi(t)-\widehat{\varphi}_{\mathrm{R}}(t)|$ and

$$\begin{aligned}S_{jn}^*(t,\sigma) &= [\varphi(t)-\widehat{\varphi}_{\mathrm{R}}(t)]^j\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\psi^{(j)}\left(\frac{y_i-\varphi(t)}{\sigma}\right)\\ R_n^*(t,\sigma) &= [\varphi(t)-\widehat{\varphi}_{\mathrm{R}}(t)]^\nu\sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\left[\psi^{(\nu)}\left(\frac{y_i+\xi_{i,n}^*}{\sigma}\right)-\psi^{(\nu)}\left(\frac{y_i-\varphi(t)}{\sigma}\right)\right].\end{aligned}$$

Let $m_j(u,\sigma) = E\psi^{(j)}\left(\dfrac{u_1+\varphi(u)-\varphi(t)}{\sigma}\right)$. Lemmas P.2(a) and P.4 together with **C6** and the equality $\dfrac{1}{h^{\nu-j}}\displaystyle\int_{-1}^{1} K^{(\nu)}(u)\, m_j(t-uh,\sigma) = \int_{-1}^{1} K^{(j)}(u)\, m_j^{(\nu-j)}(t-uh,\sigma)$ imply that $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}|S^*_{jn}(t,,\sigma)| \xrightarrow{a.s.} 0$. Therefore, the proof will be concluded if we show that $\sup_{t\in\mathcal{I}}\sup_{\sigma\in\mathcal{C}}|R^*_n(t,\sigma)| \xrightarrow{a.s.} 0$. Let C_1 and C_2 denote the Lipschitz constants for $\psi^{(\nu)}(t)$ and $\varphi(t)$, respectively. Then, we have

$$\sigma|R^*_n(t,\sigma)| \;\leq\; C_1|\varphi(t)-\widehat{\varphi}_{\mathrm{R}}(t)|\,\frac{|\varphi(t)-\widehat{\varphi}_{\mathrm{R}}(t)|^{\nu}}{h^{\nu}}\,\frac{1}{n\,h}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right|$$

which together with Lemma P.3 and **C.6** conclude the proof.□

Proof of Theorem 4.1. Denote

$$\begin{aligned}
T_n(\sigma,m) &= \sqrt{nh^{2\nu+1}}\left\{\sigma\left[\widehat{B}_\nu(t,\sigma,m)-C_\nu(t,\sigma,m)\right]-\varphi^{(\nu)}(t)\lambda_1(\sigma)\right\}\\
\widehat{T}_n(\sigma,m) &= \sqrt{nh^{2\nu+1}}\left\{\sigma\left[\widehat{B}_\nu(t,\sigma,m)-\widehat{C}_\nu(t,\sigma,m)\right]-\varphi^{(\nu)}(t)\lambda_{1,n}(t,\sigma,m)\right\}
\end{aligned}$$

with $C_\nu(t,\sigma,m) = \displaystyle\sum_{\substack{3\leq j\leq\nu\\ j:odd}}\frac{1}{j!\,\sigma^j}\,\lambda_j(\sigma)\,H_j(\nu,t)$ and $\lambda_j(\sigma) = E\psi^{(j)}\left(\dfrac{u_1}{\sigma}\right)$. As in Lemma P.5., it is easy to derive that $\lambda_{1n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}}) \xrightarrow{p} \lambda_1(\sigma_u)$. Thus, it will be enough to show that,

$$\widehat{T}(\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}}) \xrightarrow{\mathcal{D}} N\left(0,\sigma_u^2\int_{-1}^{1}\left[K^{(\nu)}(u)\right]^2 du E\psi^2\left(\frac{u_1}{\sigma_u}\right)\right).$$

The Central Limit Theorem and the expansion for the bias given in Gasser and Müller (1984) together with the fact that $nh^{2k+1}\to 0$ entail that

$$T_n(\sigma_u,\varphi) \xrightarrow{\mathcal{D}} N\left(0,\sigma_u^2\int_{-1}^{1}\left[K^{(\nu)}(u)\right]^2 du E\psi^2\left(\frac{u_1}{\sigma_u}\right)\right).$$

Therefore, the proof will be completed if we show that $\widehat{T}_n(\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-T_n(\sigma_u,\varphi) \xrightarrow{p} 0$. Note that $\widehat{T}_n(\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-T_n(\sigma_u,\varphi) = \sqrt{nh^{2\nu+1}}\{S_{1n}+S_{2n}+S_{3n}+S_{4n}+S_{5n}\}$, where

$$\begin{aligned}
S_{1n} &= \widehat{\sigma}_u\widehat{B}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\sigma_u\widehat{B}_\nu(t,\sigma_u,\varphi)\\
S_{2n} &= \sum_{\substack{3\leq j\leq\nu\\ j:odd}}\frac{1}{j!}\left[\frac{1}{\widehat{\sigma}_u^{j-1}}\lambda_{jn}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\frac{1}{\sigma_u^{j-1}}\lambda_{jn}(t,\sigma_u,\varphi)\right]\widehat{H}_j(\nu,t)\\
S_{3n} &= \sum_{\substack{3\leq j\leq\nu\\ j:odd}}\frac{1}{j!\sigma_u^{j-1}}\left[\lambda_{jn}(t,\sigma_u,\varphi)-\lambda_j(\sigma_u)\right]\widehat{H}_j(\nu,t)\\
S_{4n} &= \sum_{\substack{3\leq j\leq\nu\\ j:odd}}\frac{1}{j!\sigma_u^{j-1}}\lambda_j(\sigma_u)[\widehat{H}_j(\nu,t)-H_j(\nu,t)]\\
S_{5n} &= \widehat{\varphi}_{\mathrm{R}}^{(\nu)}(t)(\lambda_{1n}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\lambda_1(\sigma_u))\,.
\end{aligned}$$

It follows inmediately that **N.3** entails that $\sqrt{nh^{2\nu+1}}S_{4n} \xrightarrow{p} 0$

Therefore to obtain the desired result, it remains to show that

a) $\sqrt{nh^{2\nu+1}}(\lambda_{jn}(t,\sigma_u,\varphi)-\lambda_j(\sigma_u)) \xrightarrow{p} 0$ for $1 \leq j \leq \nu$, j odd,

b) $\sqrt{nh^{2\nu+1}}\left[\frac{1}{\widehat{\sigma}_u^{j-1}}\lambda_{jn}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\frac{1}{\sigma_u^{j-1}}\lambda_{jn}(t,\sigma_u,\varphi)\right] \xrightarrow{p} 0$ for $3 \leq j \leq \nu$, j odd,

and

c) $\sqrt{nh^{2\nu+1}}\left[\widehat{\sigma}_u\widehat{B}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\mathrm{R}})-\sigma_u\widehat{B}_\nu(t,\sigma_u,\varphi)\right] \xrightarrow{p} 0.$

a) We have that

$$\begin{aligned}
\lambda_{jn}(t,\sigma_u,\varphi)-\lambda_j(\sigma_u) &= \sum_{i=1}^n w_{ni}(t,h)\left[\psi^{(j)}\left(\frac{y_i-\varphi(t)}{\sigma_u}\right)-E\psi^{(j)}\left(\frac{y_i-\varphi(t)}{\sigma_u}\right)\right] \\
&+ \sum_{i=1}^n w_{ni}(t,h)E\psi^{(j)}\left(\frac{y_i-\varphi(t)}{\sigma_u}\right)-E\psi^{(j)}\left(\frac{u_1}{\sigma_u}\right) \\
&= \sum_{i=1}^n w_{ni}(t,h)z_{j,i}+\sum_{i=1}^n w_{ni}(t,h)\left[E\psi^{(j)}\left(\frac{u_1+\varphi(t_i)-\varphi(t)}{\sigma_u}\right)-E\psi^{(j)}\left(\frac{u_1}{\sigma_u}\right)\right] \\
&+ \left[\sum_{i=1}^n w_{ni}(t,h)-1\right]E\psi^{(j)}\left(\frac{u_1}{\sigma_u}\right) \\
&= A_{1n}+A_{2n}+A_{3n}
\end{aligned}$$

The Central Limit Theorem entail that,

$$\sqrt{nh}\sum_{i=1}^n w_{ni}(t,h)z_{j,i} \xrightarrow{\mathcal{D}} N(0,V_j^2)$$

which entails that $\sqrt{nh^{2\nu+1}}A_{1n} \xrightarrow{p} 0$, since $h \to 0$. On the other hand, denoting

$$m_j(v)=E\left[\psi^{(j)}\left(\frac{u_1+\varphi(v)-\varphi(t)}{\sigma_u}\right)-\psi^{(j)}\left(\frac{u_1}{\sigma_u}\right)\right]$$

straightforward calculations similar to those given in Lemma P.2 show that,

$$\sqrt{nh^{2\nu+1}}\left|\sum_{i=1}^n w_{ni}(t,h)m_j(t_i)-\frac{1}{h}\int_0^1 K\left(\frac{t-u}{h}\right)m_j(u)du\right| \to 0$$

In particular, if $m_j(v)=1$ we obtain that $\sqrt{nh^{2\nu+1}}A_{3n} \to 0$. Using that

$$\sqrt{nh^{2\nu+1}}\frac{1}{h}\int_0^1 K\left(\frac{t-u}{h}\right)m_j(u)du=\sqrt{nh^{2\nu+1}}\int_0^1 K(u)m_j(t-uh)du$$
$$=(-1)^l\sqrt{nh^{2\nu+1}}h^l\int_0^1 u^lK(u)m_j^{(l)}(\xi)du=(-1)^l\sqrt{nh^{2\nu+1+2l}}\int_0^1 u^lK(u)m_j^{(l)}(\xi)du$$

and **N.3** we obtain that $\sqrt{nh^{2\nu+1+2l}}A_{2n} \to 0$, which concludes the proof of a). Using that $\widehat{H}_j(\nu,t) \xrightarrow{p} H_j(\nu,t)$ we get that $\sqrt{nh^{2\nu+1}}S_{3n} \xrightarrow{p} 0$ and $\sqrt{nh^{2\nu+1}}S_{5n} \xrightarrow{p} 0$.

b) The Mean Value Theorem, **N.4** and straighforward calculations allow to derive that

$$\sqrt{nh^{2\nu+1}}\left[\lambda_{jn}(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}) - \lambda_{jn}(t,\widehat{\sigma}_u,\varphi)\right] \xrightarrow{p} 0$$

Thus, to show b) and therefore that $\sqrt{nh^{2\nu+1}}S_{2n} \xrightarrow{p} 0$ it will be enough to

$$\sqrt{nh^{2\nu+1}}\left[\frac{1}{\widehat{\sigma}_u^{j-1}}\lambda_{jn}(t,\widehat{\sigma}_u,\varphi) - \frac{1}{\sigma_u^{j-1}}\lambda_{jn}(t,\sigma_u,\varphi)\right] \xrightarrow{p} 0$$

Denote

$$\widehat{\lambda}_{jn}(t,\sigma) = \sum_{i=1}^{n} w_{ni}(t,h)\psi^{(j)}\left(\frac{y_i-\varphi(t_i)}{\sigma}\right) = \sum_{i=1}^{n} w_{ni}(t,h)\psi^{(j)}\left(\frac{u_i}{\sigma}\right).$$

Then, using that $E\psi^{(j)}\left(\dfrac{u_1}{\sigma}\right) = 0$ for all σ, a tightness argument similar to that used in Boente and Fraiman (1990) allow to show that

$$\sqrt{nh^{2\nu+1}}\left[\frac{1}{\widehat{\sigma}_u^{j-1}}\widehat{\lambda}_{jn}(t,\widehat{\sigma}_u) - \frac{1}{\sigma_u^{j-1}}\widehat{\lambda}_{jn}(t,\sigma_u)\right] \xrightarrow{p} 0.$$

On the other hand, using a Taylor's expansion of order two and **N.3** straightforward calculations entail that,

$$\sqrt{nh^{2\nu+1}}\left[\widehat{\lambda}_{jn}(t,\widehat{\sigma}_u) - \lambda_{jn}(t,\widehat{\sigma}_u,\varphi)\right] \xrightarrow{p} 0.$$
$$\sqrt{nh^{2\nu+1}}\left[\widehat{\lambda}_{jn}(t,\sigma_u) - \lambda_{jn}(t,\sigma_u,\varphi)\right] \xrightarrow{p} 0.$$

which concludes the proof of b).

c) It only remanins to show c) which entails that $\sqrt{nh^{2\nu+1}}S_{1n} \xrightarrow{p} 0$. First note that

$$\begin{aligned}
\widehat{B}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}) - \widehat{B}_\nu(t,\widehat{\sigma}_u,\varphi) &= \frac{1}{nh^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right)\psi'\left(\frac{y_i-\varphi(t)}{\widehat{\sigma}_u}\right)[\widehat{\varphi}_{\rm R}(t)-\varphi(t)] \\
&+ \frac{1}{nh^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h}\right)\psi''(\xi_i)[\widehat{\varphi}_{\rm R}(t)-\varphi(t)]^2 \\
&= A_{1n} + A_{2n}
\end{aligned}$$

with

$$|A_{2n}| \le \|\psi''\|_\infty[\widehat{\varphi}_{\rm R}(t)-\varphi(t)]^2\frac{1}{nh^{\nu+1}}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h}\right)\right|$$

Hence, Lemma P.3 and P.4 toghether with **N.4** entail that $\sqrt{nh^{2\nu+1}}\left[\widehat{B}_\nu(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}) - \widehat{B}_\nu(t,\widehat{\sigma}_u,\varphi)\right] \xrightarrow{p} 0$.

Thus, it wil be enough to show that,

$$\sqrt{nh^{2\nu+1}}\left[\widehat{B}_\nu(t,\widehat{\sigma}_u,\varphi)\widehat{\sigma}_u - \widehat{B}_\nu(t,\sigma_u,\varphi)\sigma_u\right] \xrightarrow{p} 0$$

which follows similarly to b) using that $nh^{2\nu+1} \to 0$ to deal with the bias term. □

References

[1] Boente, G. and Fraiman, R. (1990). Asymptotic distribution of robust estimates for nonparametric models from mixing observations. *Annals of Statistics*, **18**, 891-906.

[2] Boente, G., Fraiman, R. and Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *J. Statist. Planning and Inference*, **57**, 109-142.

[3] Boente, G. and Rodriguez, D. (2005). Robust estimators of high order derivatives of regression functions. Preprints from the Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Available at http://www.ic.fcen.uba.ar/preprints/boenterodriguez2005.pdf

[4] Cantoni, and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing* , **11**, 141-146.

[5] Gasser Th., Kneip A., Kohler W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Stat. Assoc.* **86**, 643-652.

[6] Gasser, T. and Härdle, W. (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.*, **12**, 233-240.

[7] Gasser, T. and Müller, H. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171-185.

[8] Gasser, T., Müller, H., Köhler, W., Molinari, L. and Prader, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12**, 210-229.

[9] Gasser, T., Müller, H. and Mammitzsch (1984). Kernel for nonparametric curve estimation. *J. Royal Statist. Soc.*, Ser. B, **11**, 171-185.

[10] Härdle, W. and Luckhaus, S. (1988). Uniform consistency of a class of regression function estimator. *Ann. Statist.*, **12**,612-623.

[11] Hall, P. and Jones, M. (1990)Hall. Adaptive M-estimation in nonparametric regression.*Ann. Statist.*, **18**, no. 4, 1712-1728.

[12] Leung, D., Marrot, F. and Wu, E. (1993). Bandwidth selection in robust smoothing. *J. Nonparametr. Statist.,* **4**, 333–339.

[13] Ruppert, D., Sheater, S. and Wand, M. (1995) An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257-1270.

[14] Sarda, P. and Vieu, P. (1988). Vitesses de convergence déstimateurs nonparamétriques dúne regression et de ses dérivées. *C. R. Acad. Sci. Paris*, Serie I, 83-86.

[15] Schuster, E. and Yakowitz, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Statist.*, **7**, no. 1, 139-149.

[16] Wang F. and Scott D. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Stat. Assoc.*, **89**, 65-76.