Robust nonparametric kernel regression estimator

Ge Zhao, Yanyuan Ma *

University of South Carolina, Columbia, SC 29208, United States

ABSTRACT

In robust nonparametric kernel regression context, we prescribe method to select trimming parameter and bandwidth. Through solving estimating equations, we control outlier effect through combining weighting and trimming. We show asymptotic consistency, establish bias, variance properties and derive asymptotics.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Nonparametric kernel regression is a familiar tool to explore the underlying relation between response variable and covariates (Rosenblatt et al., 1956; Parzen, 1962). Like in parametric regression estimation, kernel estimator can be affected by outliers, hence the need of considering robustness (Huber, 1979; Huber and Ronchetti, 2009; Maronna et al., 2006). See early works in Härdle and Gasser (1984), Cheng and Cheng (1987), Hall and Jones (1990), Wang and Scott (1994) and Fan et al. (1994). While these methods largely are based on the idea of truncation, Fan and Truong (1993) investigated the nonparametric estimation with heavy tailed errors in covariates. Christmann and Steinwart (2007) considered the robustness of kernel estimator from a convex risk point of view. Robust kernel regression is also extended to more complex cases, such as generalized regression models, derivative estimation and missing or dependent data treatment (Boente and Fraiman, 1989; Boente and Rodriguez, 2006; Boente et al., 2009; Bianco et al., 2011).

One important tuning parameter in nonparametric estimation is bandwidth. The most familiar practice in bandwidth selection in both non-robust and robust methods is to minimize the asymptotic mean squared error (MSE), see for example, Cleveland (1979), Härdle and Marron (1985), Cheng and Cheng (1987), Hall and Jones (1990) and Cantoni and Ronchetti (2001). Plug-in method is proposed as an important tool of bandwidth selection in robust nonparametric regression (Boente et al., 1997; Boente and Orellana, 2004; Bianco and Boente, 2007; Boente and Rodriguez, 2008). Cross-validation, as another popular bandwidth selection tool, is studied in Rice (1984), Leungi et al. (1993) and Leung (2005). These studies conclude that the asymptotic properties of the various bandwidth selection procedures are identical to the leading order. In addition, the performances of the nonparametric estimators are similar as long as a near optimal bandwidth is used because the criterion curves as functions of the bandwidth are quite flat near the optimal value (Hall and Jones, 1990; Leung, 2005). This result indicates that a near optimal bandwidth is sufficiently good for nonparametric estimation in practice. However, except for the plug-in method (Boente et al., 1997), it is largely overlooked that the bandwidth selection procedure also needs to

---

be robustified in the robust kernel regression. As a remedy, in this paper, we will propose a robust method of bandwidth selection.

In addition to bandwidth, the trimming parameter in the truncation function is also an important tuning parameter. Surprisingly, not much attention has been paid to the selection of trimming parameter in the literature. For example, Härdle and Gasser (1984) did not study this issue carefully and only selected several particular values to illustrate the performance numerically. To fill this gap, we propose a robust trimming parameter selection method in this paper.

To thoroughly investigate robust nonparametric regression, we propose a robust kernel estimator by using truncation function and derive its consistency and asymptotic normality under mild conditions. More general than the existing literatures, our method does not limit to any specific truncation function. We prescribe an automatic data driven procedure to simultaneously and robustly select both the bandwidth and the trimming parameter. We carefully study the theoretical properties of the method and establish the consistency and asymptotic normality of the robust procedure. We also derive the first order bias and variance properties of the estimator. The robust estimator we propose also automatically corrects the boundary bias without the need to employ any boundary correction procedures.

2. Robust estimator and implementation

Consider the familiar problem of nonparametric kernel regression $Y_i = m(X_i) + \epsilon_i$, $i = 1, \ldots, n$, where ϵ_i is a mean zero random variable. Our goal is to estimate the regression mean function $m(x)$ nonparametrically. However, the data we rely on for the estimation, (X_i, Y_i), are contaminated and hence a robust procedure to estimate m is needed. Generally speaking, at fixed h, k, we propose to obtain the robust estimator $\widehat{m}(x_0, h, k)$ through solving

$$\sum_{i=1}^{n} w(x_i, x_0, h) \psi_k \{y_i - m(x_0, h, k)\} = 0,$$

where w is a weighting function, such as $w(x_i, x_0, h) = h^{-1} \int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2} K\{(x_0 - u)/h\} du$ for a kernel function $K(\cdot)$. ψ is truncation function defined as

$$\psi_k(s) = sI(|s| < k) + kI(s \geq k) - kI(s \leq -k) \tag{1}$$

where k is a trimmer and $I(\cdot)$ is the indicator function.

Assume we have observations (x_i, y_i), $i = 1, \ldots, n$, and we would like to estimate the nonparametric regression of Y on x at $x = x'_1, \ldots, x'_r$ against $100q\%$ outliers. If desired, x' can be chosen the same as the $x_{(i)}$ and they are not required to be equally spaced. The bandwidth and the trimming parameter are determined robustly by using the cross-validation to obtain $\widehat{h}(k)$, and selecting k so that $100q\%(0 < q < 1)$ of the data are trimmed. Thus, if we create a width $2k$ band centered at the estimated curve, then the non-outliers are within the band. The robust kernel regression estimator algorithm is the following.

1. Start with an initial trimming parameter $k = k_0$.
2. At each k, select bandwidth h via leave-one-out cross-validation:
   (a) Solve $\sum_{j \neq i} w(x_j, x_i, h) \psi_k(y_j - \widehat{y}_{-1,i}) = 0$ to obtain $\widehat{y}_{-1,i}$.
   (b) Solve $\underset{h}{\text{argmin}} \, \phi_{1-q} \{|\widehat{y}_{-1,1} - y_1|, \ldots, |\widehat{y}_{-1,n} - y_n|\}$ to obtain h, where ϕ_p is defined as $\phi_p(s_1, \ldots, s_n) = \sum_{i=1}^{\lfloor np \rfloor} s_{(i)}, s_{(i)}$ is the ith smallest of $s_1, \ldots, s_n$.
3. Choose a set of points between $x_{(1)}$ and $x_{(n)}$, x'_t, $t = 1, \ldots, r$, update k via:
   (a) Solve $\sum_{i=1}^{n} w(x_i, x'_t, h) \psi_k \{y_i - \widehat{m}(x'_t, h)\} = 0$ to obtain $\widehat{m}(x'_t, h)$. Here x'_t is the point between $x_{(1)}$ and $x_{(n)}$.
   (b) Solve $\sum_{t=1}^{r} \sum_{i=1}^{n} I \{|y_i - \widehat{m}(x'_t, h)| < k\} = r \lceil n(1-q) \rceil$ to obtain k.
4. Repeat step 2 and 3 until convergence of k and h to k^* and h^* respectively.

The final estimation $\widehat{m}_t(x, h^*, k^*)$ is provided as the solution of

$$\sum_{i=1}^{n} w(x_i, x, h^*) \psi_{k^*} \{y_i - \widehat{m}(x, h^*)\} = 0.$$

Here, k_0 is the initial value of the trimming parameter k, and is set at $\max_i y_i - \min_i y_i$ in all our implementations. We have suggested a leave-one-out cross-validation procedure for selecting the bandwidth, which could be computation intensive when sample size is large. An alternative is to use M-fold cross-validation procedures for relatively small M, such as $M = 10$, to ease the computation burden. We can see that the selection of the bandwidth and trimming parameter is a robust type of profiling procedure and has the leave-one-out cross-validation feature.

3. Asymptotic properties

To formally establish the asymptotic properties of the robust kernel regression estimator, we first introduce some notations. Assume $x \in [0, 1]$ and the conditional probability density function of Y on x is $f(y, x)$. Let $m(x) \equiv \int_{-\infty}^{\infty} yf(y, x)dy$

be the regression mean function of Y on x. Let $Y_i = m(x_i) + \epsilon_i$, $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1$ where ϵ_i are independent and have expectation zero and variance σ^2. Now consider the robust kernel regression estimator of $m(x)$, $\widehat{m}(x)$, which solves the equation

$$H_n\{x, \widehat{m}(x)\} \equiv \sum_{i=1}^{n} \frac{1}{h_n} \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h_n}\right) du \cdot \psi\{Y_i - \widehat{m}(x)\} = 0,$$

where K is a kernel function and $\psi(\cdot)$ is a general truncation function. Note that here we allow $\psi(\cdot)$ to have the form defined in (1), but $\psi(\cdot)$ can also be any other function as long as it satisfies the regularity conditions listed below. Let $s_i = (x_{i-1}+x_i)/2$ for $i = 2, \ldots, n$, $s_0 = 0$, $s_1 = x_1/2$ and $s_n = 1$. We assume the following regularity conditions.

C1 $\psi(\cdot)$ is an odd, monotone, bounded function with bounded first and second derivatives. Let $M_\psi \equiv \sup\{|\psi(x)|\}$, $M_{\psi'} \equiv \sup\{|\psi'(x)|\}$ and $M_{\psi''} \equiv \sup\{|\psi''(x)|\}$.
C2 $f(y, x)$ is Lipschitz continuous as a function of x at any y. As a function of y, at any x, $f(y, x)$ belongs to L_1. $\partial f(y, x)/\partial x$ exists and is an integrable of y and $\int_{-\infty}^{\infty} |\partial f(y, x)/\partial x| dy < C$ for all x, where C is a positive constant.
C3 The error term ϵ is symmetrically distributed, i.e., the pdf of ϵ, denoted $f(\epsilon, x)$, is an even function of ϵ at any x.
C4 The bandwidth $h_n > 0$ depends on the sample size n and satisfies $h_n \to 0$, $nh_n^2 \to \infty$.
C5 The interpolating sequence $\{s_i\}_{i=0}^n$ satisfies $\sup_i |s_i - s_{i-1} - D_i/n| = O(n^{-\delta})$, where $\delta > 1$, $d_1 < D_i < d_2$ and d_1, d_2 are two constants. This leads to $\sup_i |s_i - s_{i-1}| = O(n^{-1})$ and $\sup_i |x_i - x_{i-1}| = O(n^{-1})$.
C6 The kernel function $K(u)$ is Lipschitz-continuous and has compact support $[-1, 1]$, satisfying $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} u^{p-1} K(u)du = 0$ and $\int_{-1}^{1} u^p K(u)du \neq 0$ for $p = 2$.
C7 Let $Q(\cdot)$ be a function defined on $[0, 1+1/(2n)]$ and it satisfies $Q\{(2i-1)/(2n)\} = x_i$, $Q\{1+1/(2n)\} = 1$. Q is monotone, smooth and has bounded first and second derivatives. Let $M_{Q'} \equiv \sup_{t \in (0,1+1/(2n))}\{Q'(t)\}$. Let $G(x)$ be the inverse function of Q, i.e. $Q\{G(x)\} = x$ for all x. Let $g(x)$ and $g'(x)$ be the first and second derivatives of $G(x)$ respectively. There exist positive constants C_1, C_2 such that $C_1 < g(x) < C_2$ and $C_1 < g'(x) < C_2$ for all x.
C8 $m(x)$ is p-times differentiable at all x.

Condition C1 imposes some basic conditions on the trimming function. Condition C2 concerns smoothness and boundedness of the relevant function and is very minor. Condition C3 is also routinely required and we further relaxed it in Remark 1. Conditions C4–C8 are basic requirements on bandwidth, data distribution, kernel function and the mean function and are very mild. Let $\text{var}_t\{g(Y_i, x)\} = \text{var}\{g(Y_i, x) \mid x_i = t\}$, $E_t\{g(Y_i, x)\} = E\{g(Y_i, x) \mid x_i = t\}$. We now present the theoretical properties, while leaving all the proofs in the supplementary document (see Appendix A).

Theorem 1. *Under Conditions* C1–C6 *and* C8, $\widehat{m}(x)$ *is consistent, i.e.*

$$\widehat{m}(x) - m(x) \xrightarrow{p} 0. \tag{2}$$

Furthermore the leading order of the difference $\widehat{m}(x) - m(x)$ *is* $O_p(n^{-1/2}h_n^{-1/2} + h_n^p)$.

Theorem 2. *Under Conditions* C1–C6 *and* C8,

$$\text{bias}\{\widehat{m}(x, h_n)\} = \frac{(-h_n)^p \partial^p E_u[\psi\{Y - m(x)\}]}{E_x \psi'\{Y - m(x)\} p! \partial u^p}\Big|_{u=x} \int_{-1}^{1} K(u)u^p du + O\{(nh_n)^{-1} + h_n^{2p}\}, \tag{3}$$

which is of order $O\{h_n^p + (nh_n)^{-1}\}$.

Theorem 2 shows that the bias of the robust kernel regression estimator is of order h_n^p, identical to the case for classical kernel estimator. The effect of the truncation function ψ is only limited to the coefficient h_n^p, not the bias order itself. It is easy to verify that when $\psi(\epsilon) = \epsilon$, the result in Theorem 2 reduces to the familiar asymptotic bias result for the classical regression estimator.

Theorem 3. *Under Conditions* C1–C8, *the variance of robust kernel estimator* $\widehat{m}(x)$ *is*

$$\text{var}\{\widehat{m}(x)\} = \frac{\text{var}[\psi\{Y - m(x)\}]}{nh_n g(x)E_x^2 \psi'\{Y - m(x)\}} \int_{-1}^{1} K^2(u)du + O(n^{-1} + n^{-\delta}h_n^{-1} + n^{-2}h_n^{-3}), \tag{4}$$

which is of order $O\{(nh_n)^{-1}\}$.

Theorem 3 shows that the variance of the robust kernel regression estimator is of order $1/(nh_n)$, identical to the case for classical kernel estimator. Like in the case for bias, the effect of the truncation function ψ is also limited to the coefficient only and when $\psi(\epsilon) = \epsilon$, the variance result in Theorem 3 reduces to the familiar asymptotic variance result for the classical regression estimator.

Theorem 4. *Under Conditions* C1–C7, *the mean squared error of* $\widehat{m}(x)$ *is*

$$MSE\{\widehat{m}(x)\} = \frac{h_n^{2p}}{p!^2 E_x^2 \psi'\{Y - m(x)\}} \left(\frac{\partial^p E_u[\psi\{Y - m(x)\}]}{\partial u^p}\Big|_{u=x} \int_{-1}^{1} K(u)u^p du \right)^2$$

$$+ \frac{\text{var}\left[\psi\{Y - m(x)\}\right]}{nh_n g(x) E_x^2 \psi'\{Y - m(x)\}} \int_{-1}^{1} K^2(u)du + O(n^{-1} + h_n^{3p} + n^{-\delta}h_n^{-1} + n^{-2}h_n^{-3}) \tag{5}$$

which is of order $O\{(nh_n)^{-1} + h_n^{2p}\}$.

Obviously, the optimal bandwidth that minimizes MSE has order $O(n^{-1/5})$, which yields a MSE of order $n^{-4/5}$. To provide a complete picture of the asymptotic properties of the robust kernel regression estimator, we further provide the asymptotic normality of the estimator in Theorem 5.

Theorem 5. *Under Conditions* C1–C8,

$$\frac{\widehat{m}(x) - m(x) - B(x)/E_x \psi'\{Y - m(x)\}}{\sigma /E_x \psi'\{Y - m(x)\}} \to N(0, 1) \tag{6}$$

in distribution, where

$$\sigma^2 = \frac{\text{var}_x\left[\psi\{Y - m(x)\}\right]}{nh_n g(x)} \int_{-1}^{1} K^2(u)du.$$

4. Simulation studies

We perform comprehensive simulation studies to illustrate the performance of the robust kernel regression estimator in comparison with the classical one. To take into account the possible boundary issues, we further incorporated boundary kernels for both the robust and the classical estimators. Across all experiments, we generated 1000 data sets with sample size 100 each, and we estimate the regression function at 50 points. One simulation takes 1.8 s on average. The quartic kernel function $K(u) = 15/16(1 - u^2)^2 I(-1 < u < 1)$ is used for all models, and the bandwidth h is selected via a leave-one-out cross-validation targeting at minimizing the overall MSE for the classical kernel estimators. We implemented our proposed procedure to obtain bandwidth for the robust estimators. For comparison, we also implemented the robust plug-in bandwidth selection procedure of Boente et al. (1997), with its pilot bandwidth obtained from the optimal bandwidth of the classical kernel estimator. To demonstrate the performance of the methods in various models, we computed the bias, variance and MSE.

We considered two different regression functions. The first regression function (M1) is the density of the Student t distribution with 11 degrees of freedom on $[-6, 6]$. In the second model (M2), the true mean function is $m(x) = 1/(1 + e^{-x})$, $x \in [-10, 10]$. This imitates the real data pattern.

We generated the regression errors ϵ_i from three different settings. (a) We generated n errors from a normal distribution with mean zero and standard deviation 0.1. (b) We generated half of the errors from the same normal distribution as in (a) and half of the errors from a Cauchy distribution scaled by 0.03, then added the n errors randomly to all $m(x)$ values. (c) We generated the errors in the same way as in (b), and we assigned the $n/2$ Cauchy errors to the same $n/2$ $m(x_i)$'s throughout the 1000 simulations. (a) can be viewed as the standard case without any outliers, while (b) and (c) can be viewed as containing 50% outliers, either randomly positioned (b) or positioned at fixed observations (c). Note that q is the percentage of data that are expected to deviate away from the regression relation that we aim to recover, and our method can handle any q value in (0, 1). Thus, unlike the traditional definition, "outlier" here indicates observations that deviate from the regression relation that we are interested in recovering.

The estimation biases, variances and MSEs of the five estimators are given in Table 1. We first examine the performance of the various estimators when no outliers exist. In this case, both the classical kernel estimator and the robust estimator are consistent estimators, while the classical estimator should have superior performance because the robust estimator has to take more precaution against model violation which is not needed here. This is indeed the case for M1. However, the opposite is true in M2. There, although the robust estimator is not making the most use of all the observations, it out-performs the classical estimator. A closer inspection reveals that this is caused by the boundary effect, in that the classical kernel estimator is severely distorted near the boundaries. Thus, we applied the boundary correction to both the classical and robust estimators, resulting in the five estimators described at the beginning of this section. Indeed, with the boundary correction, the classical estimator greatly improved its performance and exceeds that of the robust ones. It is interesting to see that although the original robust estimator was not particularly designed to handle the boundary effect, it does the correction automatically because it detects large discrepancies between the response and the estimate and then down weights the observations. In other words, boundary correction is a by-product of the robust estimation procedure and further correction is not needed. This is evidenced by the identical performance of the two robust estimators, with and without boundary correction respectively.

Table 1
Simulation results of four estimators. Classical, Robust, Classical B and Robust B indicate the classical kernel estimator, the robust kernel estimator and the two estimators with boundary kernel employed to correct the boundary effect with bandwidth selected by cross-validation. Robust plug-in indicates the robust kernel estimator that selects bandwidth using plug-in method.

Mean model	Criterion	Method	No outliers	Random outliers	Fixed outliers
M1	Bias	Classical	0.0170	0.0344	0.0603
		Classical B	0.0091	0.0302	0.0498
		Robust	0.0174	0.0246	0.0256
		Robust B	0.0167	0.0222	0.0227
		Robust plug-in	0.0298	0.0471	0.0273
	Variance	Classical	0.0013	0.8306	2.6166
		Classical B	0.0010	0.5055	3.6064
		Robust	0.0014	0.0017	0.0018
		Robust B	0.0014	0.0019	0.0021
		Robust plug-in	0.0012	0.0017	0.4510
	MSE	Classical	0.0019	0.8329	2.6220
		Classical B	0.0012	0.5072	3.6116
		Robust	0.0020	0.0031	0.0033
		Robust B	0.0021	0.0032	0.0034
		Robust plug-in	0.0030	0.0061	0.4534
M2	Bias	Classical	0.0484	0.0643	0.0929
		Classical B	0.0133	0.0338	0.0535
		Robust	0.0187	0.0194	0.0204
		Robust B	0.0182	0.0191	0.0203
		Robust plug-in	0.0404	0.0525	0.0439
	Variance	Classical	0.0025	1.2187	3.2792
		Classical B	0.0008	0.5143	0.9880
		Robust	0.0013	0.0015	0.0014
		Robust B	0.0015	0.0020	0.0019
		Robust plug-in	0.0023	0.0020	0.0019
	MSE	Classical	0.0124	1.2307	3.3018
		Classical B	0.0012	0.5161	0.9915
		Robust	0.0021	0.0023	0.0023
		Robust B	0.0022	0.0027	0.0027
		Robust plug-in	0.0051	0.0064	0.0052

When outliers occur, the advantages of the robust estimators are more prominent. Whether the outliers are randomly positioned or positioned at fixed locations, the robust estimators perform consistently well and much better than the classical counter parts. Compared to robust plug-in bandwidth selection method, our procedure is also superior because of its stability. In fact, the plug-in method is highly dependent on the pilot bandwidth (Boente and Rodriguez, 2008). For example, for fixed outliers, in M2, the plug-in method has its MSE twice that of the cross-validation method, while in M1, its MSE deteriorates further.

To further showcase the improved performance of the robustestimator, we illustrate in Figs. 1 and 2 the performance of these estimators under the various situations considered above. In all the situations, because of the similarity of the robust estimators with or without the boundary kernel correction, we only present the result without the correction in the curve estimation plots. We also omit the confident band of the plug-in method because its overlap with the cross-validation method makes the plots difficult to visualize.

First of all, for model M1, when there is no outliers in the data (first column of Fig. 1), the classical kernel estimator (red) has overall performance better than the robust estimator in terms of both bias and confidence band. However, for model M2 (Fig. 2, first column), the classical kernel estimator encounters serious difficulties on the boundary. In comparison, for data without any contaminations, the robust estimator (blue) has relatively large bias and variability, especially at the peak for M1 and at the places where the curve flattens in M2. On the boundary, however, the robust estimator outperforms the classical one due to its automatic bias correction property.

When outliers occur, regardless they are positioned randomly (column 2 of both figures) or fixed (column 3 of both figures), the superiority of our robust estimator is very clear. In both Figs. 1 and 2, the 95% confidence bands of the robust estimator cover the true curve, and are much narrower than the classical estimator, with or without boundary correction. In fact, the performance of the robust estimator is very close to the case of no outliers. In contrast, the performance of the classical estimator, with or without boundary correction, deteriorates dramatically, with increased bias and much increased variability. Furthermore, the robust method with cross-validation based bandwidth also out-performs the one with plug-in method when outliers occur.

In summary, the classical kernel estimator has good performance for data without outliers in the interior of the estimation range, but its performance is not reliable at the boundary. The boundary kernel estimation resolves the boundary issue. However, both estimators suffer from contaminations when outliers occur and can have dramatically worse performance. In contrast, the robust estimator, as promised, is indeed very robust against the outliers and continues to perform well.
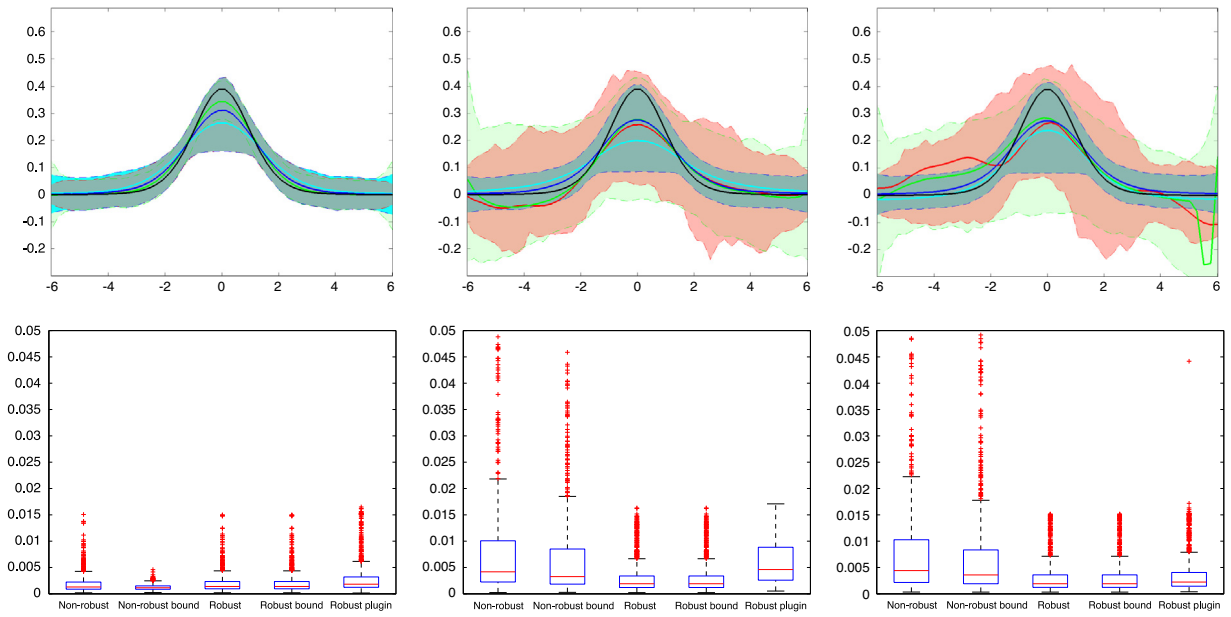
Fig. 1. Estimation of M1. Column 1: No outliers; Column 2: Randomly positioned outliers; Column 3: Fix-positioned outliers. Row 1: mean estimated curve and 95% confidence band. Black: true function. Red: classical estimator. Green: classical estimator with boundary correction. Blue: robust estimator with cross-validation bandwidth. Cyan: robust estimator with plug-in bandwidth. Row 2: Boxplot of MSEs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
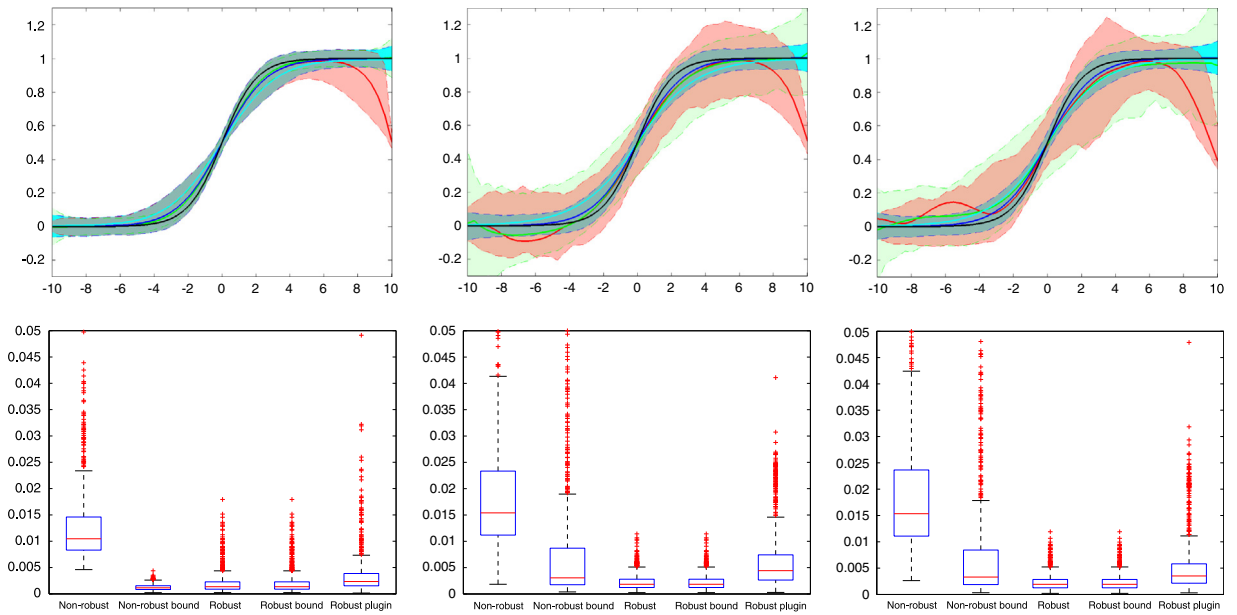

Fig. 2. Estimation of M2. Column 1: No outliers; Column 2: Randomly positioned outliers; Column 3: Fix-positioned outliers. Row 1: mean estimated curve and 95% confidence band. Black: true function. Red: classical estimator. Green: classical estimator with boundary correction. Blue: robust estimator with cross-validation bandwidth. Cyan: robust estimator with plug-in bandwidth. Row 2: Boxplot of MSEs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In addition, the robust estimator automatically corrects the boundary bias. Meanwhile, based on the numerical performance, the robust model has competitive performance compared with the classical estimators even when the data are not contaminated. Lastly, the proposed model with cross-validation bandwidth selection has a more stable performance.

Table 2
Performance on wind data analysis from three estimators, Classical estimator (Classical), Classical estimator with boundary correction (Classical B) and Robust estimator (Robust) in non-outliers and outliers. Performance measured with root mean squared error (RMSE), mean truncated absolute error (TMAE), median absolute error (MDAE), integrated mean squared error (IMSE), integrated mean truncated absolute error (ITMAE) and rate of monotonicity violation (RMV). The relative difference of Robust estimator to Classical B estimator is in the brackets.

	Criteria	Models		
		Classical	Classical B	Robust
Nonoutliers	RMSE	5.29	5.26	4.90 (−6.76%)
	TMAE	4.35	4.34	4.03 (−7.21%)
	MDAE	4.03	4.02	3.53 (−12.11%)
	IMSE	339.41	297.79	269.97 (−9.34%)
	IMTAE	47.23	45.79	44.09 (−3.70%)
	RMV	0.83	0.64	0.05 (−91.37%)
Outliers	RMSE	18.43	18.44	19.18 (4.02%)
	TMAE	9.68	9.68	9.71 (0.31%)
	MDAE	15.70	15.71	16.05 (2.14%)
	IMSE	3893.51	3894.66	4236.27 (8.77%)
	IMTAE	99.38	99.38	99.74 (0.36%)
	RMV	0.35	0.34	0.006 (−98.31%)

5. Wind power generation

We further illustrate the performance of the robust kernel regression estimator on a data set regarding wind speed and wind power from a wind farm in China. The study at the wind farm aims at understanding the relation between wind speed and the power the wind generates so that better human intervention can be applied to minimize the potential hazard to equipment and to improve efficiency. The response variable Y is the total power in kilowatts per hour generated in the farm, and the covariate x is the wind speed in meter per second measured at 70 m above ground. Normally, the wind power generated increases as the wind speed increases, hence we expect the mean function $m(x) = E(Y \mid x)$ to be a monotonically increasing function. Meanwhile, the total power has an upper limit due to the equipment constraints therefore it resembles the model M2 in this aspect. We analyze the data set with 3387 unequally spaced observations.

We implemented three estimators, the classical regression estimator without and with boundary correction and the robust kernel regression estimator without boundary correction on the data set, where the crossvalidation procedure for bandwidth selection takes about 0.9 s per iteration.

The results are summarized in Table 2, where we show the square root of MSE (RMSE), mean truncated absolute error (TMAE), median absolute error (MDAE), integrated mean squared error (IMSE), integrated mean truncated absolute error (ITMAE) and rate of monotonicity violation (RMV). Here, TMAE $= n^{-1} \sum_{i=1}^{n} |\psi\{Y_i - \widehat{m}(x_i)\}|$, RMV $= n^{-1} \sum_{i=1}^{n} I[\text{rank}\{\widehat{m}(x_{(i)})\} = i]$, where $x_{(i)}$ is the ith smallest x value. We compared the performance at outliers and non-outliers, where the 20% observations with the largest residuals are considered as outliers. Smaller values of all these measures indicate a better fit. Clearly, for the non-outlying observations, the robust estimator has the best performance compared with both non-robust estimators in terms of all six measures indicating a good fit of the regression curve to the "good" data. In contrast, for the outliers, the robust estimator yields the "worst fit" in terms of five measures, which is exactly the performance to be expected of a robust estimator, in that the estimator should not be distorted by the outlying observations, hence should not provide a good fit to these observations. In terms of RMV, the robust estimator still outperforms the two nonrobust estimators in the outliers, indicating that the outlying observations likely did not violate the relative order.

We also show the plots of estimated curves with their pointwise 95% confidence band in Fig. 3. First of all, the robust estimated curve appears much smoother than both classical estimated curves. In fact, the two non-robust curves fluctuate at several locations in the middle of the x range, clearly an outlier effect and do not agree with the general monotonicity of the wind speed-power relation. When the wind speed becomes very large (near upper limit in the plot), the power generating equipment can experience more efficiency loss due to engineering and management reasons, resulting in a potential decreasing relation between wind speed and power. This is reflected in the extreme right region of the plots. Overall, the robust estimated curve demonstrates an increasing trend, while the classical curves experience spurious maxima and minima. Next we estimated the pointwise 95% confidence band, based on $\widehat{m}(x) - \text{bias}\{\widehat{m}(x)\} \pm 1.96\text{sd}\{\widehat{m}(x)\}$, where bias$\{\widehat{m}(x)\}$ and sd$\{\widehat{m}(x)\}$ are respectively calculated using (3) and (4), with $m(x)$ replaced by $\widehat{m}(x)$. We can see clearly that the two non-robust estimators have much wider confidence bands than the robust one. At the boundary, there are relatively few observations hence all the confidence bands experience increased width. However, the robust estimator still has relatively narrower band in this region.

6. Discussion

Despite of the wide and long lasting interest in robustness in parametric models, robust issue in nonparametric estimation received relatively less attention. Part of the reason may be philosophical, in that resistance to data contamination
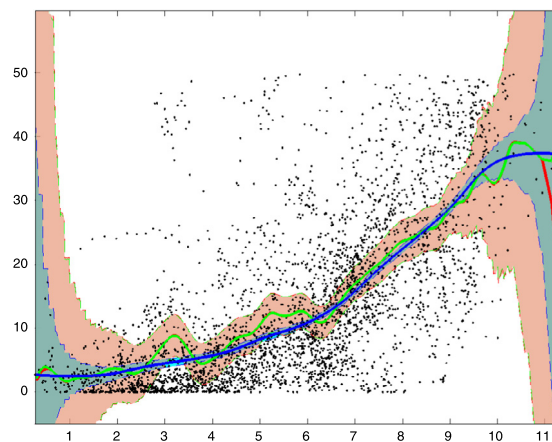
Fig. 3. Estimated curve and 95% confidence band for wind power data. Black: raw data. Red: classical estimator (Classical). Green: classical estimator with boundary correction (Classical B). Blue: robust estimator (Robust). Brown: confidence band of Classical and Classical B. light blue: confidence band of Robust estimator. The confidence band of Classical and that of Classical B almost overlap. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

seems contradictory to the ideas of letting data determine the regression relation. Benefiting from several pioneering works in this area and the truncating and weighting idea fully developed in the robust statistic literature, we propose a fully automatic way of performing robust nonparametric regression using kernel approach. The procedure is able to produce estimate that has the usual bias, variance and asymptotic distribution properties as the classical estimator, while at the same time able to reach resistance to a targeted percentage of outliers steadily.

Throughout the article, we have assumed that the error distribution is symmetric. To generalize the method for non-symmetric errors, one can modify the truncation function accordingly so that truncation applies at different trimmers on the positive and negative sides. Although the work is prompted by a practical problem in wind power generation, the method is general and can be used in any nonparametric regression problems regardless outliers are of concern or not. It will be of great interest to further investigate how to handle leverage points, multivariate covariates and even high dimensional covariates in a robust fashion. We expect that the situation there will be much more complex, due to the various ways outliers and leverage points can occur as well as various further assumptions needed to avoid curse of dimensionality. Research along this line is much needed and can be fruitful.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spl.2016.04.010.

References

Bianco, A., Boente, G., 2007. Robust estimators under semi-parametric partly linear autoregression: Asymptotic behaviour and bandwidth selection. J. Time Series Anal. 28 (2), 274–306.
Bianco, A.M., Boente, G., Sombielle, S., 2011. Robust estimation for nonparametric generalized regression. Statist. Probab. Lett. 81 (12), 1986–1994.
Boente, G., Fraiman, R., 1989. Robust nonparametric regression estimation for dependent observations. Ann. Statist. 1242–1256.
Boente, G., Fraiman, R., Meloche, J., 1997. Robust plug-in bandwidth estimators in nonparametric regression. J. Statist. Plann. Inference 57 (1), 109–142.
Boente, G., González-Manteiga, W., Pérez-González, A., 2009. Robust nonparametric estimation with missing data. J. Statist. Plann. Inference 139 (2), 571–592.
Boente, G., Orellana, L., 2004. Robust plug-in estimators in proportional scatter models. J. Statist. Plann. Inference 122 (1), 95–110.
Boente, G., Rodriguez, D., 2006. Robust estimators of high order derivatives of regression functions. Statist. Probab. Lett. 76 (13), 1335–1344.
Boente, G., Rodriguez, D., 2008. Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. Comput. Statist. Data Anal. 52 (5), 2808–2828.
Cantoni, E., Ronchetti, E., 2001. Resistant selection of the smoothing parameter for smoothing splines. Stat. Comput. 11 (2), 141–146.
Cheng, K., Cheng, P., 1987. Robust nonparametric estimation of a regression function. Sankhyā: The Indian Journal of Statistics. Series 9–22.
Christmann, A., Steinwart, I., 2007. Consistency and robustness of kernel-based regression in convex risk minimization. Bernoulli 799–819.
Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc. 74 (368), 829–836.
Fan, J., Hu, T.-C., Truong, Y.K., 1994. Robust non-parametric function estimation. Scand. J. Statist. 433–446.
Fan, J., Truong, Y.K., 1993. Nonparametric regression with errors in variables. Ann. Statist. 1900–1925.
Hall, P., Jones, M., 1990. Adaptive m-estimation in nonparametric regression. Ann. Statist. 1712–1728.
Härdle, W., Gasser, T., 1984. Robust non-parametric function fitting. J. R. Stat. Soc. Ser. B Methodol. 42–51.
Härdle, W., Marron, J.S., 1985. Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist. 1465–1481.
Huber, P.J., 1979. Robust Smoothing. Academic Press, New York.
Huber, P.J., Ronchetti, M.E., 2009. Robust Statistics, second ed. Wiley, New York.
Leung, D.H.-Y., 2005. Cross-validation in nonparametric regression with outliers. Ann. Statist. 2291–2310.
Leungi, D.H., Marriott, F.H., Wu, E.K., 1993. Bandwidth selection in robust smoothing. J. Nonparametr. Stat. 2 (4), 333–339.
Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics. John Wiley & Sons, Chichester, ISBN.
Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 1065–1076.
Rice, J., 1984. Bandwidth choice for nonparametric regression. Ann. Statist. 1215–1230.
Rosenblatt, M., et al., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27 (3), 832–837.
Wang, F.T., Scott, D.W., 1994. The l 1 method for robust nonparametric regression. J. Amer. Statist. Assoc. 89 (425), 65–76.