Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis

Graciela Boente
Departamento de Matemáticas, Instituto de Cálculo
FCEyN, Universidad de Buenos Aires and CONICET, Argentina
and
Daniela Rodriguez
Departamento de Matemáticas, Instituto de Cálculo
FCEyN, Universidad de Buenos Aires and CONICET, Argentina

Abstract

In this paper, under a semiparametric partly linear regression model with fixed design, we introduce a family of robust procedures to select the bandwidth parameter. The robust plug–in proposal is based on nonparametric robust estimates of the $\nu-$th derivatives and under mild conditions, it converges to the optimal bandwidth. A robust cross–validation bandwidth is also considered and the performance of the different proposals is compared through a Monte Carlo study. We define an empirical influence measure for data–driven bandwidth selectors and, through it, we study the sensitivity of the data–driven bandwidth selectors. It appears that the robust selector compares favorably to its classical competitor, despite the need to select a pilot bandwidth when considering plug–in bandwidths. Moreover, the plug–in procedure seems to be less sensitive than the cross–validation in particular, when introducing several outliers. When combined with the three-step procedure proposed by Bianco and Boente (2004), the robust selectors lead to robust data–driven estimates of both the regression function and the regression parameter.

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

AMS Subject Classification 2000: Primary 62F35, Secondary 62G05.
Key words and phrases: Asymptotic Properties; Bandwidth Selectors; Kernel Weights; Partly Linear Models; Robust Estimation; Smoothing Techniques.

1 Introduction

Partly linear models have become an important tool when modelling biometric data, since they combine the flexibility of nonparametric models and the simple interpretation of the linear ones. These models assume that we have a response $y_i \in I\!R$ and covariates or design points $(\mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}} \in I\!R^{p+1}$ satisfying

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + g(t_i) + \epsilon_i \qquad 1 \leq i \leq n \; , \tag{1}$$

with the errors ϵ_i independent and independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$. The semiparametric nature of model (1) offers more flexibility than the standard linear model, when modelling a complicated relationship between the reponse variable with one of the covariates. At the same time, they keep a simple functional form with the other covariates avoiding the "curse of dimensionality" existing in nonparametric regression.

In many situations, it seems reasonable to suppose that a relationship between the covariates $\mathbf{x}$ and t exists, so as in Speckman (1988), Linton (1995) and Aneiros–Pérez and Quintela del Río (2002) we will assume that for $1 \leq j \leq p$

$$x_{ij} = \phi_j(t_i) + \eta_{ij} \qquad 1 \leq i \leq n \tag{2}$$

where the errors η_{ij} are independent. Moreover, the design points t_i will be assumed to be fixed.

Several authors have considered the semiparametric model (1). See, for instance, Denby (1986), Rice (1986), Robinson (1988), Speckman (1988) and Härdle, Liang and Gao (2000) among others.

All these estimators, as nonparametric estimators, depend on a smoothing parameter that should be choosen by the practitioner. As it is well known, large bandwidths produce 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. Linton (1995), using local polynomial regression estimators, obtained an asymptotic expression for the optimal bandwidth in the sense that it minimizes a second order approximation of the mean square error of the least squares estimate, $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}(h)$, of $\boldsymbol{\beta}$. This expression depends on the regression function we are estimating and on parameters which are unknown, such as the standard deviation of the errors. More precisely, for any $\mathbf{c} \in I\!R^p$, let $\sigma^2 = \sigma_\epsilon^2 \mathbf{c}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \mathbf{c}$ be the asymptotic variance of $U = \mathbf{c}^{\mathrm{T}} n^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}(h) - \boldsymbol{\beta})$, and $n\, MSE(h) = EU^2/\sigma^2$ its standardized mean square error. Then, when the smoothing procedure corresponds to local means, under general conditions, that include that the design points are almost uniform design points, i.e., $\{t_i\}_{i=1}^n$ are fixed design points in $[0,1]$, $0 \leq t_1 \leq \ldots \leq t_n \leq 1$, such that $t_0 = 0$ and $t_{n+1} = 1$ and $\max_{1 \leq i \leq n+1} |(t_i - t_{i-1}) - 1/n| = O(n^{-\delta})$ for some $\delta > 1$, we have that, for $\nu \geq 2$,

$$MSE(h) \;\; = \;\; n^{-1} \{ 1 + (nh)^{-1} A_2 + o(n^{-2\mu}) + (n^{\frac{1}{2}} h^{2\nu} A_1 + o(n^{-\mu}))^2 \} \; ,$$

where $\mu = (4\nu-1)/(2(4\nu+1))$, $\boldsymbol{\phi}^{(\nu)}(t) = (\phi_1^{(\nu)}(t), \dots, \phi_p^{(\nu)}(t))^{\mathrm{T}}$, $\alpha_\nu(K) = \int u^\nu K(u)du$, $K_*(u) = K * K(u) - 2K(u)$ and

$$A_1 \;=\; \alpha_\nu^2(K)(\nu!)^{-2}\, \sigma^{-1}\, \mathbf{c}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1} \int_0^1 g^{(\nu)}(t)\boldsymbol{\phi}^{(\nu)}(t)dt \qquad\qquad A_2 = \int K_*^2(u)du\,.$$

Therefore, the optimal bandwidth in the sense of minimizing the asymptotic $MSE(h)$, is given by $h_{opt} = A_0 n^{-\pi}$, with $\pi = 2/(4\nu+1)$ and

$$A_0 = \left(A_2/(4\nu {A_1}^2)\right)^{\pi/2} = \left\{\int K_*^2(u)du\Big/\left[4\nu\left(\sigma^{-1}\mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\alpha_\nu^2(K)(\nu!)^{-2}\int_0^1 g^{(\nu)}(t)\boldsymbol{\phi}^{(\nu)}(t)dt\right)^2\right]\right\}^{\pi/2}. \tag{3}$$

Linton (1995) considered a plug–in approach to estimate the optimal bandwidth and showed that it converges to the optimal one, while Aneiros–Pérez and Quintela del Río (2002) studied the case of dependent errors.

It is well known that, both in linear regression and in nonparametric regression, least squares estimators can be seriously affected by anomalous data. The same statement holds for partly linear models. To avoid that problem, Bianco and Boente (2004) considered a three–step robust estimate for the regression parameter and the regression function. Besides, for the nonparametric regression setting, i.e., when $\boldsymbol{\beta} = 0$, the sensitivity of the classical bandwidth selectors to anomalous data was discussed by several authors, such as, Leung, Marriott and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997), Cantoni and Ronchetti (2001) and Leung (2005).

In this paper, we consider a robust plug–in selector for the bandwidth, under the partly linear model (1) which converges to the optimal one and leads to robust data–driven estimates of the regression function g and the regression parameter $\boldsymbol{\beta}$. We derive an expression analoguous to (3) for the optimal bandwidth of the three–step estimator introduced in Bianco and Boente (2004). As for its linear relative, this expression will depend on the derivatives of the functions g and $\boldsymbol{\phi}$. In Section 2, we review some of the proposals given to estimate robustly the derivatives of the regression function under a nonparametric regression model. The robust bandwidth selector for the partial linear model is introduced in Section 3, where under mild conditions, consistency to the optimal bandwidth is established. In Section 4, for small samples, the behavior of the classical approach and of the resistant selectors is compared through a Monte Carlo study under normality and contamination. Also, a robust cross–validation procedure is introduced and compared with the plug–in one. Finally, in Section 5 an empirical influence measure for the plug–in bandwidth selector is introduced. We use this measure to study the sensitivity of the plug–in selector on some generated examples. All proofs are given in the Appendix.

2 Robust estimation of the derivative of order ν

In this section, we review some of the approaches given to provide robust estimator of the $\nu-$th derivative of the regression function under a fully nonparametric regression model.

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

$$z_i = \varphi(t_i) + u_i \qquad 1 \le i \le n\,, \tag{4}$$

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.

Robust estimates for the first derivative of the regression function have been introduced by Härdle and Gasser(1985), when the scale is known. Boente and Rodriguez (2006) discussed the estimation of higher derivatives. Their approach is analoguous to that given by Boente, Fraiman and Meloche (1997) when $\nu = 2$. On the other hand, a robust local polynomial approach was introduced by Welsh (1996) and extended to the dependent setting by Jiang and Mack (2001).

In order to define both classes of 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

$$\begin{aligned} w_{ni}(t,h) &= (nh)^{-1} K_0\left((t-t_i)/h\right)\,, & (5)\\ w_{ni}^{(\nu)}(t,h) &= (nh^{\nu+1})^{-1} K^{(\nu)}\left((t-t_i)/h\right)\,, & (6)\end{aligned}$$

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 $K^{(\nu)}$.

2.1 The robust differentiation approach

When scale σ_u is known, Härdle and Gasser (1985) suggested to use as an estimate of $\varphi^{(\nu)}(t)$ the ratio $\sigma_u \widehat{B}_\nu\left(t,\sigma_u,\widehat{\varphi}\right)\left[\lambda_{1,n}(t,\sigma_u,\widehat{\varphi})\right]^{-1}$, with $\widehat{\varphi}(t)$ a preliminary robust estimate of the regression function and

$$\begin{aligned} \widehat{B}_\nu\left(t,\sigma,\varphi\right) &= \sum_{i=1}^n w_{ni}^{(\nu)}(t,h)\Psi\left((z_i-\varphi(t))/\sigma\right) & (7)\\ \lambda_{j,n}(t,\sigma,\varphi) &= \sum_{i=1}^n w_{ni}(t,h)\Psi^{(j)}\left((z_i-\varphi(t))/\sigma\right)\,. & (8)\end{aligned}$$

However, this estimate will be biased if $\nu > 2$, since $E\left[\Psi^{(j)}\left(u_i/\sigma_u\right)\right]$ are not equal to 0 for odd values of j (see Boente and Rodriguez (2006) for a discussion). More precisely, the estimate of $\varphi^{(\nu)}(t)$ introduced by Härdle and Gasser (1985) will converge to

$$\varphi^{(\nu)}(t) + (\lambda_1(\sigma_u))^{-1}\sigma_u \sum_{\substack{3\le j\le \nu\\ j:odd}} (\sigma_u^j\, j!)^{-1}\,\lambda_j(\sigma_u)\, H_j(\nu,t) = \varphi^{(\nu)}(t) + (\lambda_1(\sigma_u))^{-1}\sigma_u C_\nu\left(t,\sigma_u,\varphi\right)$$

instead of $\varphi^{(\nu)}(t)$, where $\lambda_j(\sigma) = E\Psi^{(j)}\left(u_1/\sigma\right)$ and $H_j(\nu,t) = \left.\{[\varphi(u)-\varphi(t)]^j\}^{(\nu)}\right|_{u=t}$. To correct the bias, Boente and Rodriguez (2006) introduced an estimator for $C_\nu\left(t,\sigma_u,\varphi\right)$ as follows

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

with $\widehat{H}_j(\nu,t)$ an estimate of $H_j(\nu,t)$. 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

$$\widehat{\varphi}_{\rm R}^{(\nu)}(t,h) \;=\; \widehat{\sigma}_u\,[\widehat{B}_\nu\,(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R}) - \widehat{C}_\nu\,(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R})]/\lambda_{1,n}\,(t,\widehat{\sigma}_u,\widehat{\varphi}_{\rm R})\;. \tag{9}$$

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

$$\sum_{i=1}^n w_{ni}(t,h_0)\Psi\left((z_i - \widehat{\varphi}_{\rm R}(t,h_0))/\widehat{\sigma}_u\right) = 0\;.$$

As mentioned in Boente and Rodriguez (2006), 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 bandwidth choice, leading to $\nu-1$ choices of pilot bandwidths to estimate the $\nu-$th derivative of the regression function, denoted h_j, $0 \le j \le \nu-2$. In order to guarantee the convergence of the preliminary estimates, these bandwidths must satisfy $h_j \to 0$ and $nh_j^{2j+1}/\log n \to +\infty$. One possible choice for them is to define data–driven bandwidths by robustifying and adapting the iterative scheme proposed by Gasser, Kneip and Kohler (1991).

Under mild conditions, in Theorem 3.1 in Boente and Rodriguez (2006), it is shown that if $nh^{\nu+2}\to\infty$ and $E(\Psi'(u_1/\sigma_u)) \ne 0$, $\sup_{t\in[h,1-2h]} |\widehat{\varphi}_{\rm R}^{(\nu)}(t,h) - \varphi^{(\nu)}(t)| \xrightarrow{a.s.} 0$. The asymptotic distribution of the estimates is also derived.

2.2 The robust polynomial approach

To estimate the derivatives of a regression function a different approach was considered by Welsch (1996) who studied local quantile regression and local heteroscedastic $M-$regression estimators. On the other hand, under an homoscesdatic regression model as in (4), Jiang and Mack (2001) introduced a family of estimators for the regression function and their derivatives based on a local $M-$regression approach that leads to pointwise consistent and asymptotically normally distributed estimates even when the observations satisfy an $\alpha-$mixing condition. These estimates are defined as follows. Let ρ be an outlier resistant function with bounded derivative Ψ and $\widehat{\sigma}_u$ a preliminary robust scale estimator. Jiang and Mack (2001) propose to find a_j to minimize

$$\sum_{i=1}^n w_{ni}(t,h)\rho\left((z_i - \sum_{j=0}^{\nu} a_j(t_i-t)^j)/\widehat{\sigma}_u\right)$$

Equivalently, the solution $\widehat{\mathbf{a}}(t) = (\widehat{a}_0(t), ..., \widehat{a}_\nu(t))$ satisfy the local $M-$estimation equations:

$$\sum_{i=1}^{n} w_{ni}(t,h)\Psi\left((z_i - \sum_{j=0}^{\nu} \widehat{a}_j(t)(t_i-t)^j)/\widehat{\sigma}_u\right)(t_i-t)^k = 0 \qquad k = 0,\dots,\nu\ . \tag{10}$$

The local M-type estimator of $\mathbf{a}(t) = (\varphi(t), ..., \varphi^{(\nu)}(t)/\nu!)$ is the solution, $\widehat{\mathbf{a}}(t)$, to (10). Therefore, the estimate of $\varphi^{(\nu)}(t)$ can be defined as $\widehat{\varphi}^{(\nu)}(t,h) = \nu!\, \widehat{a}_\nu(t)$. Under mild conditions, if $h \to 0$ and $nh^{2\nu+1} \to \infty$ the estimates are pointwise consistent and asymptotically normally distributed.

3 Resistant choice of the smoothing parameter

As is well known an important issue in any smoothing procedure is the choice of the smoothing parameter. As mentioned in the Introduction, under a nonparametric regression model, two commonly used approaches are cross–validation and plug–in. However, these procedures may not be robust and their sensitivity to anomalous data was discussed by several authors, including Leung, Marriott and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997), Cantoni and Ronchetti (2001) and Leung (2005). Wang and Scott (1994) note that, in the presence of outliers, the least squares cross–validation function is nearly constant on its whole domain and thus, essentially worthless for the purpose of choosing a bandwidth. The robustness issue remains valid for the partly linear considered in this paper. With a small bandwidth, a small number of outliers with similar values of t_i could easily drive the estimators of $\boldsymbol{\phi}$ and ϕ_0 and so, the final estimator of g, to dangerous levels.

In the following sections we will describe two data–driven bandwidth selectors. In Section **??**, we introduce a robust plug–in bandwidth that relies on an expansion for the mean square error of the robust estimator of $\boldsymbol{\beta}$. Besides, in Section 3.4, a robust bandwidth based on the cross–validation principles is considered. We begin by reviewing the definition of the three–step estimator introduced in Bianco and Boente (2004).

3.1 Preliminaries: Estimation of the regression parameter

Let $\{(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}\}_{i=1}^n$ be independent observations satisfying (1). We will assume that $\epsilon_i \sim F(\cdot/\sigma_\epsilon)$ where F is symmetric and that $\mathbf{x}_i$ and t_i are nonparametrically related through (2), so that the model can be written as

$$\left\{ \begin{array}{rcll} y_i &=& \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \epsilon_i & \quad 1 \le i \le n, \\ x_{ij} &=& \phi_j(t_i) + \eta_{ij} & \quad 1 \le j \le p, \end{array} \right. \tag{11}$$

with η_{ij} independent and such that $\eta_{ij} \sim G_j(\cdot/\sigma_{\eta,j})$ with G_j symmetric. From now on, we will denote by $\phi_0(t) = \boldsymbol{\phi}(t)^{\mathrm{T}}\boldsymbol{\beta} + g(t)$ and so, $y_i = \phi_0(t_i) + u_i$ with $u_i = \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_i$.

Without loss of generality, we will assume from now on that the fixed design points $t_i \in [0,1]$ are such that $0 \le t_1 \le \ldots \le t_n \le 1$. Moreover, we will assume that assumption **B.4** below holds, i.e, the fixed design points are "almost"uniform, $\max_{1\le i\le n+1} |(t_i - t_{i-1}) - 1/n| = O(n^{-\delta})$ for some $\delta > 1$.

It is well known that, both in linear regression and in nonparametric regression, least square estimators can be seriously affected by anomalous data. In partly linear models, the least squares estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ can be obtained by minimizing

$$\sum_{i=1}^{n}\left[y_i-\widehat{\phi}_{0,\mathrm{LS}}(t_i)-\left(\mathbf{x}_i-\widehat{\boldsymbol{\phi}}_{\mathrm{LS}}(t_i)\right)^{\mathrm{T}}\boldsymbol{\beta}\right]^2, \qquad (12)$$

with $\widehat{\phi}_{0,\mathrm{LS}}$ and $\widehat{\phi}_{j,\mathrm{LS}}$ the linear kernel estimators of $\phi_0(t)$ and $\phi_j(t)$, $1\le j\le p$, respectively. As expected, these estimators are highly sensitive to outliers. To avoid this problem, Bianco and Boente (2004) proposed a class of estimates based on a three step procedure with a more resistant behavior under the partly linear model which can be described as follows:

Step 1: Estimate $\phi_0(t)$ and $\phi_j(t)$, $\quad 1\le j\le p$ through a robust smoothing, as local $M-$type estimates. Let $\widehat{\phi}_{0,\mathrm{R}}$ and $\widehat{\phi}_{j,\mathrm{R}}$ denote the obtained estimates and $\widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t)=(\widehat{\phi}_{1,\mathrm{R}}(t),\ldots,\widehat{\phi}_{p,\mathrm{R}}(t))^{\mathrm{T}}$.

Step 2: Estimate the regression parameter by applying a robust regression estimate to the residuals $\widehat{r}_i=y_i-\widehat{\phi}_{0,\mathrm{R}}$ and $\widehat{\boldsymbol{\eta}}_i=\mathbf{x}_i-\widehat{\boldsymbol{\phi}}_{\mathrm{R}}$. Let $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ denote an estimate of $\boldsymbol{\beta}$.

Step 3: Define the estimate of the regression function g as $\widehat{g}_{\mathrm{R}}(t)=\widehat{\phi}_{0,\mathrm{R}}(t)-\widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t)$.

To make explicit the dependence on the bandwidth h, we will denote these estimates as $\widehat{\phi}_{0,\mathrm{R}}(t,h)$, $\widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t,h)$, $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)$ and $\widehat{g}_{\mathrm{R}}(t,h)$.

Theorem 2 in Bianco and Boente (2004) entails that, under mild conditions, when the estimates of the regression parameter are defined through

$$\sum_{i=1}^{n}\psi_1\left((\widehat{r}_i-\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)^{\mathrm{T}}\widehat{\boldsymbol{\eta}}_i)/s_n\right)w_2\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i=0, \qquad (13)$$

with s_n a robust consistent estimate of σ_ϵ, then $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)-\boldsymbol{\beta}\right)\xrightarrow{\mathcal{D}}N(0,\sigma_\epsilon^2\mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1})$ with

$$\begin{aligned}\mathbf{A} &= E\left(\psi_1'\left(\epsilon/\sigma_\epsilon\right)\right)E\left(w_2\left(\|\boldsymbol{\eta}\|\right)\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\right)=E\left(\psi_1'\left(\epsilon/\sigma_\epsilon\right)\right)\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}\\ \boldsymbol{\Sigma} &= E\left(\psi_1^2\left(\epsilon/\sigma_\epsilon\right)\right)E\left(w_2^2\left(\|\boldsymbol{\eta}\|\right)\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\right)=E\left(\psi_1^2\left(\epsilon/\sigma_\epsilon\right)\right)\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}\,.\end{aligned}$$

Denote $\boldsymbol{\Sigma}_{\mathrm{R},\boldsymbol{\eta}}=\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}^{-1}\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}^{-1}$ and $V(\psi_1)=\left[E\left(\psi_1'\left(\epsilon/\sigma_\epsilon\right)\right)\right]^{-2}E\left(\psi_1^2\left(\epsilon/\sigma_\epsilon\right)\right)$. Thus, for any $\mathbf{c}\in I\!R^p$, the asymptotic variance of $U_{\mathrm{R}}=\mathbf{c}^{\mathrm{T}}n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)-\boldsymbol{\beta})$, is given by $\sigma_{\mathrm{R}}^2=\sigma_\epsilon^2\,V(\psi_1)\,\mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{R},\boldsymbol{\eta}}\,\mathbf{c}$.

3.2 Robust plug–in bandwidth selector

An important step to define a robust plug–in bandwidth is to obtain an asymptotic expansion for $MSE_{\mathrm{R}}(h)=n^{-1}\,EU_{\mathrm{R}}^2/\sigma_{\mathrm{R}}{}^2$. For the sake of simplicity, we will begin by fixing some notation. Let $\eta^*_{ij} = \sigma_{\eta,j}\Psi\left(\eta_{ij}/\sigma_{\eta,j}\right)/E\left(\Psi'\left(\eta_{1j}/\sigma_{\eta,j}\right)\right)$ and $u^*_i = \sigma_0\Psi\left(u_i/\sigma_0\right)/E\Psi'\left(u_1/\sigma_0\right)$ be the bounded modified residuals. Besides, denote $v_i = \boldsymbol{\eta}^{*\,\mathrm{T}}_i\boldsymbol{\beta} - u^*_i$, $\mathbf{D} = ED_{\psi_2}(\boldsymbol{\eta})$ with $D_{\psi_2}(\mathbf{u})$, the Jacobian matrix, with (i,j) element $\frac{\partial}{\partial u_j}\psi_2(\mathbf{u})_i$ and σ_0 the scale of u. When using local $M-$smoothers with score function Ψ, in the Appendix we derive an expression for the $MSE_{\mathrm{R}}(h)$ that will allow to obtain the optimal bandwidth for the robust estimator of $\boldsymbol{\beta}$ solution of (13). Effectively, therein it is shown that, under mild conditions, for $\nu \geq 2$,

$$MSE_{\mathrm{R}}(h) = n^{-1}\left\{1 + (nh)^{-1}A_{\mathrm{R},2} + o(n^{-2\mu}) + (n^{\frac{1}{2}}h^{2\nu}A_{\mathrm{R},1} + o(n^{-\mu}))^2\right\}, \quad (14)$$

where

$$\begin{aligned}
A_{\mathrm{R},1} &= \alpha^2_\nu(K)/(\nu!)^2\ \sigma^{-1}_{\mathrm{R}}\ \mathbf{c}'\boldsymbol{\Sigma}^{-1}_{1,\boldsymbol{\eta}}E\left(D_{\psi_2}(\boldsymbol{\eta})\right)\int_0^1 g^{(\nu)}(t)\boldsymbol{\phi}^{(\nu)}(t)dt \\
A_{\mathrm{R},2} &= \sigma^2_\epsilon/\sigma^2_{\mathrm{R}}\left\{\kappa_1\int K^2(u)du + \kappa_2\int(K*K)^2(u)du - 2\kappa_3\int K(u)K*K(u)du\right\}, \\
\kappa_1 &= \sigma^{-2}_\epsilon E\left(\psi'_1\left(\epsilon/\sigma_\epsilon\right)\right)^2\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}\mathbf{A}^{-1}\mathbf{c}\ E\left(v^2_2\right) + E\left(\psi_1\left(\epsilon/\sigma_\epsilon\right)\right)^2\ E\left(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}D_{\psi_2}\left(\boldsymbol{\eta}_1\right)\boldsymbol{\eta}^*_2\right)^2 \\
&\quad +\sigma^{-2}_\epsilon\mathrm{COV}\left(\psi'_1\left(\epsilon_1/\sigma_\epsilon\right)\ v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\psi_2\left(\boldsymbol{\eta}_1\right), \psi'_1\left(\epsilon_2/\sigma_\epsilon\right)\ v_1\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\psi_2\left(\boldsymbol{\eta}_2\right)\right) \\
&\quad + 2\sigma_\epsilon{}^{-2}E\left(\psi'_1\left(\epsilon_1/\sigma_\epsilon\right)\psi_1\left(\epsilon_2/\sigma_\epsilon\right)u^*_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\psi_2\left(\boldsymbol{\eta}_1\right)\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}D_{\psi_2}\left(\boldsymbol{\eta}_2\right)\boldsymbol{\eta}^*_1\right) \\
\kappa_2 &= \sigma^{-2}_\epsilon\left[E\psi'_1\left(\epsilon/\sigma_\epsilon\right)\right]^2\left\{\mathrm{COV}\left(v_1\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_2, v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_1\right) + \mathrm{VAR}\left(v_1\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_2\right)\right\} \\
\kappa_3 &= \sigma^{-2}_\epsilon E\left[\psi'_1\left(\epsilon/\sigma_\epsilon\right)\right]\left\{\mathrm{COV}\left(\psi'_1\left(\epsilon_1/\sigma_\epsilon\right)v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\psi_2\left(\boldsymbol{\eta}_1\right), v_1\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_2\right)\right. \\
&\quad + \mathrm{COV}\left(\psi'_1\left(\epsilon_1/\sigma_\epsilon\right)v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\psi_2\left(\boldsymbol{\eta}_1\right), v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_1\right) \\
&\quad - \left.\sigma_\epsilon\ \mathrm{COV}\left(\psi_1\left(\epsilon_1/\sigma_\epsilon\right)\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}D_{\psi_2}\left(\boldsymbol{\eta}_1\right)\boldsymbol{\eta}^*_2, v_2\ \mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{D}\boldsymbol{\eta}^*_2\right)\right\} \\
\eta^*_{ij} &= \sigma_{\eta,j}\Psi\left(\eta_{ij}/\sigma_{\eta,j}\right)/E\left(\Psi'\left(\eta_{1j}/\sigma_{\eta,j}\right)\right) \qquad \epsilon^*_i = \sigma_\epsilon\Psi\left(\epsilon_i/\sigma_\epsilon\right)/E\left(\Psi'\left(\epsilon_1/\sigma_\epsilon\right)\right) \\
u^*_i &= \sigma_0\Psi\left(u_i/\sigma_0\right)/E\Psi'\left(u_1/\sigma_0\right),
\end{aligned}$$

with $\mathbf{D} = ED_{\psi_2}(\boldsymbol{\eta})$, $v_i = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}^*_i - u^*_i$, σ_0 the scale of u, $\alpha_\nu(K)$ defined in the Introduction and $D_{\psi_2}(\mathbf{u})$, the Jacobian matrix, with (i,j) element $\frac{\partial}{\partial u_j}\psi_2(\mathbf{u})_i$. Therefore, the optimal bandwidth in the sense of minimizing the asymptotic $MSE_{\mathrm{R}}(h)$, is given by $h_{\mathrm{R},opt} = A_{\mathrm{R},0}n^{-\pi}$, with $\pi = 2/4\nu + 1$ and

$$A_{\mathrm{R},0} = \left(A_{\mathrm{R},2}/4\nu A^2_{\mathrm{R},1}\right)^{\pi/2}. \quad (15)$$

Note that when using the least squares estimates defined in (12), we recover (3) from the formula above.

Since the optimal bandwidth $h_{\mathrm{R},opt}$ depends on the unknown quantities σ^2_ϵ, $V(\psi_1)$, $\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}$, $\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}$, κ_1 to κ_3, $g^{(\nu)}(t)$ and $\boldsymbol{\phi}^{(\nu)}(t)$ robust estimates of them must be considered to define a plug–in selector. To define the robust plug-in bandwidth selection method, we propose to plug–in robust estimators of the derivative of orden ν, as defined in Section 2, into (15). Therefore, a robust plug–in selector for the regression parameter under the partly linear regression model (11), can be obtained as follows

- Let s_0 and s_j be robust consistent estimates of the scales σ_0 of $u = \epsilon + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}$ and $\sigma_{\eta,j}$, respectively. Denote by $\widehat{\phi}_{j,\mathrm{R}}(t)$ and $\widehat{\phi}_{j,\mathrm{R}}^{(\nu)}(t)$ preliminary robust consistent estimates of the regression functions $\phi_j(t)$ and of its derivative $\phi_j^{(\nu)}(t)$, $0 \le j \le p$, computed with a pilot bandwidth h. As robust estimators of the derivatives $\phi_j^{(\nu)}(t)$, $j = 0, \ldots, p$, one can use either the robust differentiation or the robust polynomial approach, described in Section 2.1 and 2.2, respectively.

  Moreover, let $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\widehat{g}_{\mathrm{R}}(t)$ be initial robust consistent estimators of $\boldsymbol{\beta}$ and $g(t)$, respectively. For instance, we can define $\widehat{g}_{\mathrm{R}}(t) = \widehat{\phi}_{0,\mathrm{R}}(t) - \widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t)$, as in Step 3.

- Define a robust estimator of $g^{(\nu)}(t)$ as

$$\widehat{g}_{\mathrm{R}}^{(\nu)}(t) = \widehat{\phi}_{0,\mathrm{R}}^{(\nu)}(t) - \widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_{\mathrm{R}}^{(\nu)}(t) \,, \tag{16}$$

  where $\widehat{\boldsymbol{\phi}}_{\mathrm{R}}^{(\nu)}(t) = \left(\widehat{\phi}_{1,\mathrm{R}}^{(\nu)}(t), \ldots, \widehat{\phi}_{p,\mathrm{R}}^{(\nu)}(t)\right)^{\mathrm{T}}$.

- Denote by $\widehat{\sigma}_\epsilon^2$, $\widehat{V}(\psi_1)$, $\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}$, $\widehat{\boldsymbol{\Sigma}}_{2,\boldsymbol{\eta}}$, $\widehat{\mathbf{D}}$, $\widehat{\kappa}_\ell$, $1 \le \ell \le 3$ robust consistent estimates of σ_ϵ^2, $V(\psi_1)$, $\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}$, $\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}$, $\mathbf{D} = ED_{\psi_2}(\boldsymbol{\eta})$ and κ_ℓ, respectively, obtained using the empirical distribution of the residuals $\widehat{\epsilon}_i = y_i - \widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{\mathrm{T}}\mathbf{x}_i - \widehat{g}_{\mathrm{R}}(t_i)$ and $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t_i)$. Define an estimate of σ_{R}^2 as $\widehat{\sigma}_{\mathrm{R}}^2 = \widehat{\sigma}_\epsilon^2\widehat{V}(\psi_1)\mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathrm{R},\boldsymbol{\eta}}\,\mathbf{c}$.

- The robust bandwidth selector $\widehat{h}_{\mathrm{R}}$ is defined as

$$\widehat{h}_{\mathrm{R}} = \widehat{A}_{\mathrm{R},0}n^{-\pi} \qquad \text{with} \qquad \widehat{A}_{\mathrm{R},0} = \left(\widehat{A}_{\mathrm{R},2}/4\nu\widehat{A}_{\mathrm{R},1}^2\right)^{\pi/2} \tag{17}$$

$$\widehat{A}_{\mathrm{R},1} = \alpha_\nu^2(K)(\nu!)^{-2}\,\widehat{\sigma}_{\mathrm{R}}^{-1}\,\mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}^{-1}\,\widehat{\mathbf{D}}\int_h^{1-h}\widehat{g}_{\mathrm{R}}^{(\nu)}(t)\widehat{\boldsymbol{\phi}}_{\mathrm{R}}^{(\nu)}(t)dt \tag{18}$$

$$\widehat{A}_{\mathrm{R},2} = \widehat{\kappa}_1\int K^2(u)du + \widehat{\kappa}_2\int(K*K)^2(u)du - 2\widehat{\kappa}_3\int K(u)(K*K)(u)du \tag{19}$$

In order to avoid numerical integrations, we can consider $n^{-1}\sum_{i=1}^{n}\widehat{g}_{\mathrm{R}}^{(\nu)}(t_i,h)\widehat{\boldsymbol{\phi}}_{\mathrm{R}}^{(\nu)}(t_i,h)I_{[h,1-h]}(t_i)$ instead of $\int_h^{1-h}\widehat{g}_{\mathrm{R}}^{(\nu)}(t,h)\widehat{\boldsymbol{\phi}}_{\mathrm{R}}^{(\nu)}(t,h)dt$.

As estimates of the scale σ_0 of u_1 and $\sigma_{\eta,j}$ of η_{1j}, we can use M-estimates or the robust Rice–type estimators defined as

$$s_0 = \underset{1\le i\le n-1}{\mathrm{median}}\,|y_{i+1} - y_i|/(0.6754\sqrt{2}) \qquad s_j = \underset{1\le i\le n-1}{\mathrm{median}}\,|x_{i+1\,j} - x_{ij}|/(0.6754\sqrt{2})\,, \tag{20}$$

since, under model (11), we are dealing with homoscedastic errors.

As mentioned in Section 2, this procedure depends on the pilot bandwidth h_0 used to compute $\widehat{\phi}_{j,\mathrm{R}}(t,h_0)$ and, when using the differentiation approach described in Section 2.1, on the preliminary estimates of the derivatives of $\phi_j(t)$ up to order $\nu - 2$, which obviously also involve a choice for the smoothing parameter. As mentioned by Aneiros–Pérez and

Quintela del Río (2002), whatever method is used to estimate $A_{\mathrm{R},0}$ an additional smoothing parameter has to be selected, and in this sense the plug–in method is not fully automatic. A robust version of the iterative scheme proposed by Gasser, Kneip and Kohler (1991) may also be considered. Three strategies for choosing the smoothing parameter using the plug–in approach were discussed in Ruppert, Sheater and Wand (1995). These rules provide ready–to–use plug–in bandwidth selectors for the local linear kernel estimate of the regression function in a fully nonparametric regression model. A robust version of the three iterative schemes proposed therein can be also adapted to the partly linear model (11) by using the robust version of Mallow's C_p introduced by Ronchetti and Staudte (1994) and the robust estimates defined in Section 2. However, our simulation study suggests that, for partly linear models, the final estimates of $\boldsymbol{\beta}$ may not be too sensitive to the choice of the pilot bandwidth.

3.3 Consistency of the plug–in bandwidth selector

The purpose of this section is to show that under regularity conditions, the adaptive bandwidth satisfies

$$\widehat{h}_{\mathrm{R}}/h_{opt} \xrightarrow{a.s.} 1 \quad \text{as} \quad n \to \infty \, .$$

The asymptotic equivalence between the data–driven and the optimal bandwidth implies that the robust estimates of the regression parameter $\boldsymbol{\beta}$ using $\widehat{h}_{\mathrm{R}}$ are asymptotically equivalent to those obtained using $h_{\mathrm{R},opt}$. The proof of the asymptotic normality of the data–driven robust regression estimates can be derived using similar arguments to those considered in Boente, Fraiman and Meloche (1997) combined with the techniques used in Bianco and Boente (2004).

In order to guarantee the convergence of adaptive bandwidth, we will need of the following assumptions:

A.1. The functions $g(\cdot), \phi_1(\cdot), \dots, \phi_p(\cdot)$ have ν continuous derivatives on [0,1].

A.2. The initial estimators $\widehat{g}(t) = \widehat{g}_{\mathrm{R}}(t)$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ satisfy $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}$ and $\sup_{t\in[0,1]} |\widehat{g}(t) - g(t)| \xrightarrow{a.s.} 0$.

A.3. $\sup_{t\in[h,1-2h]} |\widehat{\phi}^{(\nu)}_{j,\mathrm{R}}(t) - \phi^{(\nu)}_j(t)| \xrightarrow{a.s.} 0$ for $j = 0, 1, \dots, p$.

A.4. $\sup_{t\in[0,1]} |\widehat{\phi}_{j,\mathrm{R}}(t) - \phi_j(t)| \xrightarrow{a.s.} 0$ for $j = 0, 1, \dots, p$.

A.5. s_0 and s_j are strong consistent estimates of σ_0 and $\sigma_{\eta,j}$, respectively.

Note that, under mild conditions, Theorem 3.1 in Boente and Rodriguez (2006) show that **A.3** holds when using the estimates defined through the differentiation approach given by (9).

Theorem 3.1. *Let $\nu \geq 2$. Assume that $\widehat{\sigma}_\epsilon^2$, $\widehat{V}(\psi_1)$, $\widehat{\mathbf{\Sigma}}_{1,\boldsymbol{\eta}}$, $\widehat{\mathbf{\Sigma}}_{2,\boldsymbol{\eta}}$, $\widehat{\mathbf{D}}$, $\widehat{\kappa}_\ell$, $1 \leq \ell \leq 3$ are consistent estimates of σ_ϵ^2, $V(\psi_1)$, $\mathbf{\Sigma}_{1,\boldsymbol{\eta}}$, $\mathbf{\Sigma}_{2,\boldsymbol{\eta}}$, $\mathbf{D}$ and κ_ℓ respectively. Under **A.1** to **A.5**, if in addition $E(\Psi'(u_1/\sigma_0)) \neq 0$ and $E(\Psi'(\eta_{1j}/\sigma_{\eta,j})) \neq 0$, we have that,*

$$\widehat{h}_{\mathrm{R}}/h_{\mathrm{R},opt} \xrightarrow{a.s.} 1 \quad as \quad n \to \infty ,$$

where $\widehat{h}_{\mathrm{R}}$ is defined through (17).

Remark 3.1. A similar result is obtained if we estimate $\int_h^{1-h} g^{(\nu)}(u)\phi_1^{(\nu)}(u)du$ through

$$n^{-1}\sum_{i=1}^n \widehat{g}^{(\nu)}(t_i)\widehat{\phi}_1^{(\nu)}(t_i)I_{(h,1-h)}(t_i)$$

which was the procedure used in the simulation study to avoid the calculation of the numeric integral.

The next Proposition provide conditions to obtain strongly consistent estimates of σ_ϵ^2 and $\mathbf{\Sigma}_{\mathrm{R},\boldsymbol{\eta}}$. The following additional assumptions are needed.

B.1. $\{\epsilon_i : 1 \leq i \leq n\}$ is a sequence of i.i.d. random variables $\epsilon_i \sim F(./\sigma_\epsilon)$. Moreover, $u_1 = \epsilon_1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}_1 \sim G_0(./\sigma_0)$ with G_0 a symmetric distribution function.

B.2. For each $1 \leq j \leq p$, $\{\eta_{ij} : 1 \leq i \leq n\}$ is a sequence of i.i.d. random variables such that $\eta_{1j} \sim G_j(./\sigma_{\eta,j})$ with G_j a symmetric distribution function.

B.3. $\{\epsilon_i\}$ is independent of $\{\boldsymbol{\eta}_i\}$.

B.4. $\{t_i\}_{i=1}^n$ are fixed design points in $[0,1]$, $0 \leq t_1 \leq \ldots \leq t_n \leq 1$, such that $t_0 = 0$ and $t_{n+1} = 1$ and $\max_{1\leq i\leq n+1} |(t_i - t_{i-1}) - 1/n| = O(n^{-\delta})$ for some $\delta > 1$.

B.5.
a) ψ_1 is an odd, bounded and twice continuously differentiable function with bounded derivatives ψ_1' and ψ_1'', such that $\varphi_1(t) = t\psi_1'(t)$ and $\varphi_2(t) = t\psi_1''(t)$ are bounded,
b) $E\left(w_2(\|\boldsymbol{\eta}\|)\|\boldsymbol{\eta}\|^2\right) < \infty$ and $\mathbf{\Sigma}_{1,\boldsymbol{\eta}}$ is non–singular,
c) $w_2(u) = \psi_2(u)\,u^{-1} > 0$ is a bounded function, Lipschitz of order 1. Moreover, ψ_2 is also a bounded and continuously differentiable function with bounded derivative ψ_2' such that $\lambda_2(t) = t\psi_2'(t)$ is bounded.

Proposition 3.1. *Let $\widehat{g}_{\mathrm{R}}(t_i)$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ be initial estimators of $g(t)$ and $\boldsymbol{\beta}$ satisfying **A.2**. Denote for $1 \leq i \leq n$, $\widehat{\epsilon}_i = y_i - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i - \widehat{g}_{\mathrm{R}}(t_i)$. Let $\widehat{P}_n$ be the empirical measure of $\widehat{\epsilon}_i$ and P the probability measure related to the distribution of ϵ_1.*
*Let $\sigma^2(\cdot)$ be a continuous scale functional such that $\sigma^2(P) = \sigma_\epsilon^2$. Then, under **B.1**, **B.2** and **B.4**, the estimate defined as $\widehat{\sigma}_\epsilon^2 = \sigma^2(\widehat{P}_n)$ is a strongly consistent estimate of σ_ϵ^2.*

Moreover, if $\widehat{\boldsymbol{\eta}}_i = \mathbf{x}_i - \widehat{\phi}_{\mathrm{R}}(t_i)$ *with* $\widehat{\phi}_{\mathrm{R}}$ *satisfying* ***A.4***, *and* ***B.5*** *holds then, the estimates defined through*

$$\widehat{V}(\psi_1) = \left[n^{-1}\sum_{i=1}^{n}\psi_1'\left(\widehat{\epsilon}_i/\widehat{\sigma}_\epsilon\right)\right]^{-2} n^{-1}\sum_{i=1}^{n}\psi_1^2\left(\widehat{\epsilon}_i/\widehat{\sigma}_\epsilon\right)$$

$$\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}} = n^{-1}\sum_{i=1}^{n} w_2\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i\,\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}} \qquad \widehat{\boldsymbol{\Sigma}}_{2,\boldsymbol{\eta}} = n^{-1}\sum_{i=1}^{n} w_2^2\left(\|\widehat{\boldsymbol{\eta}}_i\|\right)\widehat{\boldsymbol{\eta}}_i\,\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{R},\boldsymbol{\eta}} = \widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}^{-1}\,\widehat{\boldsymbol{\Sigma}}_{2,\boldsymbol{\eta}}\,\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}^{-1} \qquad \widehat{\mathbf{D}} = n^{-1}\sum_{i=1}^{n} D_{\psi_2}\left(\widehat{\boldsymbol{\eta}}_i\right)$$

are strongly consistent estimates of $V(\psi_1)$, $\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}$, $\boldsymbol{\Sigma}_{2,\boldsymbol{\eta}}$, $\boldsymbol{\Sigma}_{\mathrm{R},\boldsymbol{\eta}}$ *and* $\mathbf{D}$, *respectively.*

As a consequence of Proposition 3.1., we have that the estimate defined through $\widehat{\sigma}_\epsilon = \tau_1\,\mathrm{mad}_{1\le i\le n}(y_i - \widehat{\boldsymbol{\beta}}(h_0)^{\mathrm{T}}\mathbf{x}_i - \widehat{g}_{\mathrm{R}}(t_i, h_0))$ is a strongly consistent estimate of σ_ϵ^2. The constant τ_1 is a standarizing constant choosen to ensure Fisher–consistency.

Using analogous arguments, it can be seen that **A.4** entails that the estimates s_0 and s_j defined in (20) satisfy **A.5**. A similar result can be obtained for the estimates of κ_ℓ defined through the residuals, in the iterative process.

3.4 Robust cross–validation selector

For spline–based estimators, Cantoni and Ronchetti (2001) introduced a cross-validation criterion to select the bandwidth parameter while robust cross–validation selectors for kernel $M-$smoothers were considered by Leung, Marriott and Wu (1993), Wang and Scott (1994) and Leung (2005), under a fully nonparametric regression model.

A robust cross-validation criterion similar to that considered by Bianco and Boente (2007) for partly linear autoregression models can be defined. Let $\widehat{\phi}_{j,i}(t,h)$ and $\widehat{\phi}_{0,i}(t,h)$ be the smoothers computed with bandwidth h using all the data except $(y_i, \mathbf{x}_i, t_i)$. Denote by $\widehat{g}_i(t,h) = \widehat{\phi}_{0,i}(t,h) - \widehat{\boldsymbol{\phi}}_i(t,h)^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{\mathrm{R}}(h)$, by $\widetilde{\boldsymbol{\beta}}_{\mathrm{R}}(h)$ the regression estimator obtained considering the residuals $y_i - \widehat{\phi}_{0,i}(t_i,h)$ and $\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_i(t_i,h)$ and by $\widehat{\epsilon}_i(h) = y_i - \left(\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{\mathrm{R}}(h) + \widehat{g}_i\left(t_i,h\right)\right)$. Then, the classical least squares cross–validation method constructs an asymptotically optimal data–driven bandwidth and thus, adaptive data–driven estimators, by minimizing

$$\Upsilon_1(h) = n^{-1}\sum_{i=1}^{n}\left(y_i - \left\{\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{\mathrm{R}}(h) + \widehat{g}_i\left(t_i,h\right)\right\}\right)^2 w^2\left(t_i\right) = n^{-1}\sum_{i=1}^{n}\widehat{\epsilon}_i^{\,2}(h)w^2\left(t_i\right)\ ,$$

where the weight function w protects against boundary effects. In the classical setting, linear smoothers and least squares regression estimators are used, while if one tries to obtain resistant procedures, local $M-$smoothers and robust regression estimators, as described in Section 3.1 should be considered. However, as mentioned above, it is well known that when there are outliers in the data, the least squares cross–validation criterion fails, even when using robust estimators. Taking into account that the classical cross–validation criterion tries to measure both bias and variance, it would be sensible to introduce a new measure that establishes a trade–off between robust measures of bias and variance. Let μ_n and σ_n denote robust estimators of location and scale, respectively. A robust cross–validation criterion can be defined by minimizing on h

$$\Upsilon_2(h) \quad = \quad \mu_n^2\left(\widehat{\epsilon}_{i,w}(h)\right) + \sigma_n^2\left(\widehat{\epsilon}_{i,w}(h)\right) \ ,$$

where $\widehat{\epsilon}_{i,w}(h)$ indicates that when computing the robust location and scale estimators each residual $\widehat{\epsilon}_i(h)$ is weighted according to $w\left(t_i\right)$. As location estimator, μ_n, one can consider the median while σ_n can be taken as the bisquare a–scale estimator or the Huber $\tau-$scale estimator.

4 Monte Carlo Study

This section contains the results of a simulation study, in dimension $p = 1$, designed to evaluate the performance, under a partly linear model, of the robust bandwidth selectors defined in Section 3. For the plug–in bandwidth, we have used both the differentiation approach and the local polynomial approximation to estimate the derivatives of the regression functions. The aims of this study are

- to compare the behavior of the bandwidth selectors and of the regression estimators under contamination and under normal samples.
- to study the relationship between the bandwidth selection method and the initial smoothing parameter, when considering plug–in bandwidths.

4.1 General Description

The simulation study was carried out in Splus. The S–code is available at
`http://www.ic.fcen.uba.ar/`

In the smoothing procedure, we have used the Gaussian kernel with standard deviation $0.25/0.675 = 0.37$ such that the interquartile range is 0.5.

As mentioned in Section 3, in order to estimate the optimal bandwidth, using a plug–in approach, we need initial estimators of the regression parameter and the regression function, so that we can estimate the error's variance and the derivatives of the function g and ϕ.

Plug–in Bandwidth: Initial estimators of the parameter and the regression function.
The behavior of the least squares estimates was compared with that obtained by smoothing with a local $M-$estimate with bisquare score function, with constant 4.685, which gives a 95% efficiency. As initial estimate in the iterative procedure to compute the local $M-$estimate, we have considered the local median. Several choices for the initial bandwidth from 0.25 to 0.45 were considered to study the dependence on the choice of the initial bandwidth.

As initial estimate for the regression parameter, we have considered a $GM-$estimate defined by (13) with score function on the residuals $\psi_1(r) = \psi_{\text{H},c_1}(r) = \max(-c_1, \min(r, c_1))$, i.e, the Huber function, and weight function w_2

$$w_2(\eta) = W\left[((\eta - \mu_\eta)/\sigma_\eta)^2\right] \tag{21}$$

where $W(t) = \psi_{\text{H},c_2}(t)/t$. The tunning constants were chosen as $c_1 = 1.6$ and $c_2 = \chi_{1,0.975}$ while $\mu_\eta = \underset{1\le i\le n}{\text{median}}(\widehat{\eta}_i)$ and $\sigma_\eta = \underset{1\le i\le n}{\text{mad}}(\widehat{\eta}_i)/0.6754$ with $\widehat{\eta}_i = x_i - \widehat{\phi}_{\text{R}}(t_i)$.

Cross–validation selector

The performance of the plug–in bandwidth selector was also compared with that of the cross-validation criterion described in Section 3.4. We have consider μ_n as the median and σ_n as the Huber $\tau-$scale estimator. For this preliminary study, the search for the bandwidth parameter was performed searching, in a first step, over a grid of 16 points on the interval $[0.05, 0.8]$, and then, the search was refined around the minimum with a step of 0.01. So, too small or too large bandwidths are not allowed in this procedure as we do in the plug–in one.

Final estimators of the parameter and the regression function

Once the data–driven bandwidth was computed, the behavior of the least squares estimates using the classical plug–in or the L^2 cross–validation selector, was compared with that of the three step estimators described in Section 3.1. The local $M-$estimate was computed using the robust plug–in or the robust cross–validation bandwidth, respectively.

After smoothing the response variable y and the regression covariates x, the following robust regression estimates of β were computed:

- the $GM-$estimates with Huber function with $c_1 = 1.6$ on the residuals and with weight function (21) on the covariates where $c_2 = \chi_{1,0.975}$.
- the least trimmed with 33% trimmed observations, as introduced in Rousseeuw (1984).

We also computed two other estimators: the least median of squares estimator and a one–step estimator based on it. The results are not reported here, since they are quite similar to those obtained with the GM and the least trimmed estimators.

In all the tables and figures LS denotes the least squares estimate, GM and LTS denote the robust alternatives using the GM and the least trimmed estimates, respectively.

The performance of an estimate $\widehat{g}$ of g is measured using two measures:

$$\begin{aligned} \text{MSE}(\widehat{g}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{g}(t_i) - g(t_i)\right]^2 \\ \text{MedSE}(\widehat{g}) &= \text{median}\left(\left[\widehat{g}(t_i) - g(t_i)\right]^2\right) . \end{aligned}$$

Due to the expensive computing time of the cross–validation criterion, we performed 500 replications generating independent samples of size $n = 100$ according to the following model

$$y_i = x_i + 1 + 10t_i^2 + \epsilon_i \quad 1 \le i \le n$$
$$x_i = 1/\log(5)\exp\{\log(5)t_i\} + \eta_i \quad 1 \le i \le n\ ,$$

where $t_i = (i - 0.5)/n$. Thus, $g(t) = 1 - 10t^2$, $\beta = 1$ and $\phi(t) = 1/\log(5)\exp\{\log(5)t\}$. This model was considered by Linton (1995) and corresponds to a $\nu = 2$ degree of smoothness. To isolate the comparison between the competitors from any border effect, data were in fact generated at design points outside the interval $[0, 1]$ as well.

The non–contaminated case, indicated by C_0, correspond to (ϵ_i, η_i) i.i.d normal with mean 0 and standard deviation 1.

C_1 and C_2 will denote the following two contaminations.

- C_1: $\epsilon_i \sim 0.9\, N(0, 1) + 0.1\, \mathcal{C}(0, 1)$, where $\mathcal{C}(0, \sigma)$ indicates the distribution Cauchy centered in 0 with scale σ. This contamination corresponds to inflating the error and thus, will affect the variance of the regression estimates. It will also affect the performance of the plug–in bandwidth.
- C_2: $\epsilon_i \sim 0.9N(0, \sigma^2) + 0.1\, \mathcal{C}(0, 1)$ independent and artificially 10 observations of the carriers but not of the response variables, were modified to be equal to 20 at equally spaced values of t. This case corresponds to introduce high–leverage points besides inflating the errors. The aim of this contamination is to study changes in bias in the estimation of the regression parameter and on the bandwidth selector.

The following tables summarize the results of the simulations.

Tables 1 to 3 give means and standard deviations for the estimates $\widehat{h}$ of the optimal bandwidth using the differentiating, polynomial and cross–validation criteria, measured through the summary measures of $\log\left(\widehat{h}/h_{opt}\right)$. Note that for the regression functions considered h_{opt} equals 0.3581 for the classical least squares estimator while $h_{opt} = h_{\mathrm{R},opt} = 0.3071$ for the robust one. On the other hand, the asymptotically optimal bandwidth related to the cross–validation criterion considered, was computed numerically and equals 0.226, since it tries to fit not only the regression parameter but also the nonparametric component.

Table 4 give the mean and standard deviations for the regression estimates of β while Tables 5 and 6 show the mean of $\mathrm{MSE}(\widehat{g})$ and the median of $\mathrm{MedSE}(\widehat{g})$ over the 500 replications, when using the differentiating approach. Similar results are obtained by the other two methods and are not reported here. The bias of the regression estimators can be easily computed as the difference between the mean and 1.

Finally, Figures 1 and 3 gives the boxplots of $\log\left(\widehat{h}/h_{opt}\right)$ for the classical and robust data–driven bandwidths.

4.2 Simulation results

The simulation study confirms the inadequate behavior of the classical plug–in bandwidth selector under contamination and in particular, how it increases the mean square error of the estimates of β.

Table 1 to 3 shows that under contamination the robust estimator of the bandwidth is much more stable. Also, under C_2, an increasing bias appears for the classical selector as the pilot bandwidth increases, for both plug–in methods. Moreover, the best performances under C_0 are obtained for pilot bandwidths in the range of 0.38 to 0.45, both for the classical and robust estimator.

Figure 1 shows better how the pilot bandwidth influences the bandwidth selector. It explains also that the higher variability of the robust selector for normal errors, is not only due to some large estimates of the optimal bandwidth but also, when the pilot increases, to some very small bandwidth estimates, when considering the differentiation approach. It is worth noticing, that when using the robust polynomial method, larger biases are obtained as the pilot increases, but variability decreases in the same direction. On the other hand, Table 3 and Figure 3 show the advantage of the plug–in approach over robust cross–validation since they provide bandwidths with lower variability under C_0 and C_1. On the other hand, under C_2 plug–in methods show a better performance than cross–validation both in bias and variance while the plug–in bandwidth based on the differentiating approach shows a better performance than that based on polynomials, particularly, for small bandwidth (see also, Figure 3). This can be explained by the fact that the local $M-$regression approach considered can have a low local breakdown point (see the discussion given in Chapter 4, in Maronna, Martin and Yohai (2006)).

It is worth noticing that over the 500 replications, we get for the robust cross–validation criteria 66, 53 and 38 times bandwidths smaller than 0.1, under C_0, C_1 and C_2, respectively while only 2 and 13 times bandwidths larger than 0.7 are obtained under C_1 and C_2, respectively. On the other hand, for the least squares cross–validation criterion, 70 and 42 time bandwidths larger than 0.7 are obtained under C_1 and C_2, respectively. Besides, in 94, 46 and 27 of the 500 replications, we obtain bandwidths smaller than 0.1 under the studied contamination schemes. This shows that the main problem with cross–validation is its well–known problem of leading to small bandwidths. On the other hand, over the 500 replications the plug–in procedure with pilot 0.40 lead to no bandwidth estimates smaller than 0.1 under C_0, C_1 and C_2, respectively.

Table 4 confirms, as expected, the increased variance of the least squares estimate under contamination and the better performance in bias under C_2 of the LTS estimators. However, the LTS estimators have a higher standard deviation under C_1 and C_2, than the $GM-$estimator. Finally, it is worth noticing that the final regression estimate is quite stable with respect to the pilot selection.

With respect to the estimation of the regression function Tables 5 and 6 show the better performance of the $GM-$estimator which lead to almost the half MSE or $MedSE$ than the least trimmed estimator, even under contamination. Moreover, these measures seem to be also quite stable with respect to the initial bandwidth. Moreover, a comparison between Table 5 and and 6 allows to conclude that for some design points, t_i, the classical estimator does a bad job in estimating under contamination. Under normal errors, all estimators perform similarly, however the $GM-$estimator is more efficient than the least trimmed estimator.

5 Empirical Influence of the Bandwidth Selector

One of the aims of a robust procedure is to produce estimates less sensitive to outliers than the classical ones. The influence function is a measure of robustness with respect to single outliers. Statistical diagnostics and graphical displays for detecting outliers can be built based on empirical influence functions. In parametric models this topic is widely developed, however, less attention has been given in the nonparametric literature. A smoothed functional approach to nonparametric kernel estimators was introduced by Aït Sahalia (1995) and used by Tamine (2002) to define a smoothed influence function in nonparametric regression. However, this approach assumes that the bandwidth h is fixed and not data–driven. On the other hand, Manchester (1996) introduced a graphical method to display sensitivity of a scatter plot smoother. To measure the influence of outlying observations on the bandwidth selector, we will follow an approach similar to that given by Manchester (1996) and we will consider the finite–sample version of the influence function introduced by Tukey (1977), called the empirical influence function. Given a data set $\{(t_i, \mathbf{x}_i, y_i)\}_{1\leq i\leq n}$ which satisfies (11), let $\widehat{h}_n$ be a bandwidth selector based on this data set. Assume that $\mathbf{z} = (t_0, \mathbf{x}_0, y_0)$ represents a contaminating point with $t_0 \in [0, 1]$ and denote $\widehat{h}_{\mathbf{z}}$ the bandwidth selector based on the augmented data set $\{(t_1, \mathbf{x}_1, y_1), \dots (t_n, \mathbf{x}_n, y_n), \mathbf{z}\}$. In order to detect if a contaminating point produces undersmoothing, i.e., bandwidths approaching to 0, we can define the empirical influence surface as

$$\mathrm{EIF}(t_0, \mathbf{x}_0, y_0) = (n+1)\left|\log\left(\widehat{h}_{\mathbf{z}}\right) - \log\left(\widehat{h}_n\right)\right| . \tag{22}$$

Since the range of t is the interval $[0, 1]$ bandwidths approaching to 1 or larger than 1 lead to oversmoothing and so are useless. The measure defined in (22) does not allow us to visualize easily this type of *breakdown*, therefore we introduce another empirical influence function

$$\mathrm{EIF}_1(t_0, \mathbf{x}_0, y_0) = (n+1)\left|\log\left(\widehat{h}_{\mathbf{z}}/(1-\widehat{h}_{\mathbf{z}})\right) - \log\left(\widehat{h}_n/(1-\widehat{h}_n)\right)\right| . \tag{23}$$

A surface plot can be constructed for each value of t varying the values of $(\mathbf{x}, y)$ to see how outliers and leverage points $(\mathbf{x})$ affect the bandwidth at different places of the range of t.

As an example, we have generated, as in Section 4, a data set of size $n = 100$ following the model

$$\begin{aligned} y_i &= x_i + 1 + 10t_i^2 + \epsilon_i \quad 1 \leq i \leq n \\ x_i &= 1/\log(5)\exp\{\log(5)t_i\} + \eta_i \quad 1 \leq i \leq n\ , \end{aligned}$$

where $t_i = (i - 0.5)/n$. The data set is shown in Figure 4 together with the nonparametric component g and the regression function $\gamma(t) = g(t) + \beta\phi(t)$ in dashed and solid lines, respectively. We have considered three values for t_0, $t_0 = 0.10$, 0.50 and 0.90. For each of them we have computed $\mathrm{EIF}(t_0, x, y)$ and $\mathrm{EIF}_1(t_0, x, y)$ over a grid of 1600 equispaced points in $[-40,40]\times[-40,40]$. The resulting plots for $t_0 = 0.10$ are given in Figures 5 and 6. Similar plots are obtained for $t_0 = 0.50$ and $t_0 = 0.90$ are given in Figures 7 to 10. Figure 5 reveal the lack of robustness of the classical bandwidth. In particular, EIF_1 is not plotted for values near $x = 40$, since the bandwidth breaks–down giving values much larger than one. For other values of t, the same happens if we consider a larger range of values for x and y according to the point t. On the other hand, the empirical functions of the robust bandwidth are bounded and they show that the most influential points correspond to those having x between -3 and -1. Besides, large values of the empirical influence function are also obtained when y takes values between 5 and 30. However, these points do not yield to bandwidths in the boundary of the interval $[0,1]$. In all cases, the initial bandwidth was taken equal to 0.45 both for the differentiating approach and for the local polynomial one. Similar results are obtained for other initial bandwidths.

An influential study can be performed by the inclusion of several outliers in the neighborhood of the point t. For the robust plug–in bandwidth selectors, Figures 11 to 13 plot, as a function of t, the effect of adding k outliers, $\mathbf{z}_1,\dots,\mathbf{z}_k$, when $k = 1$, 3, 5, 7, 9 and 11. To be more precise, for a fixed point $0 < t < 1$ and a fixed amount of k outliers, we have added to the sample the points $\mathbf{z}_\ell = (t_\ell, 10, 10)$, $1 \le \ell \le k$, with $t_\ell = t + (2\ell-1)/(2n)$ if $t + (2\ell-1)/(2n) < 1$ and $t_\ell = t - (2\ell-1)/(2n)$, otherwise. This configuration was choosen so that the outliers were inserted between adjacent pair of design points to increase the impact on the estimator. The solid lines correspond to $\text{EIF}(t,10,10)$ while the dashed ones with empty circles to $\text{EIF}_1(t,10,10)$. For the cross-validation procedure the search was made first in the interval $[0.05,2]$ with a step of 0.05 and then in the same interval, around the local minimum with a step of 0.01, so that bandwidths smaller than 0.05 were never choosen not allowing implosion of the bandwidths. However, it should be noted that with the inclusion of more than 7 outliers, half of the times the obtained bandwidth was 0.05 showing the bad performance of the cross–validation criterion and explaining the large values of $\text{EIF}(t,10,10)$. As we can see, the robust selectors do not explode with this outlier configuration, however, the bandwidth selector is sensitive to the inclusion of $k = 11$ outliers at the boundary. Note that this amount of outliers represents locally more than 10% of contamination. An exception is the robust local polynomial selector when including 7 outliers that explodes at $t = 0.255$, giving a bandwidth larger than 1, due to the non–convergence of the algorithm in 20 iterations. Note that, at the boundary, the effect of adding outliers increases with the number of outliers. Moreover, the robust cross–validation procedure is much more sensitive than the robust plug–in selectors, leading as mentioned before, to small bandwidths when anomalous observations are present. On the other hand, the differentiating approach performs better than the polynomial one as the number of outliers increase. Moreover, as it can be seen from the plots, the dashed lines with empty circles are over the solid lines, showing that the main problem when introducing several outliers is that large bandwidths can be obtained, leading to oversmoothing. The worst situation arises with the robust plug–in selector based on the polynomial approach. Effectively, as shown in Figure 12, where when considering 7 outliers, the maximum of EIF_1 equals 419.25 corresponding to a bandwidth $h_{\mathbf{z}} = 0.9734$.

Our influential study shows that the robust procedures seem stable with the inclusion of one isolated outlier. However, even if they do not breakdown, they are quite sensitive to the inclusion of several outliers in the neighborhood of a fixed point. Moreover, the robust cross–validation criterion seems to perform worst than the robust plug–in procedures introduced.

6 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, Marriott and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997), Cantoni and Ronchetti (2001) and Leung (2005). The same conclusions hold under a partly linear regression model. Our proposals tends to overcome the sensitivity of the classical selectors by considering robust estimators of the derivatives of the regression function or a robust cross–validation criteria, under a partly linear regression model.

The problem of defining the influence function of the smoothing parameter is still an outstanding issue. We introduced an empirical influence measure that allows to evaluate on a given data set the sensitivity of the bandwidth selector to anomalous data. It turns out that, under a partly linear model, the classical plug–in bandwidth defined in Linton (1995) is not robust, since it leads to unbounded empirical influence functions. On the other hand, our proposals have bounded empirical influence even when introducing several outliers. The best performance, in all cases, for the considered model and the studied contaminations is attained by the plug–in rules, even they are all influenced by multiple outliers. In particular, the differentiating approach lead to smaller influence functions than that based on polynomials when dealing with more than one outlier.

Acknowledgements

The authors would like to thank the Referee for its valuable comments and suggestions that lead to improve the presentation of the paper. This research was partially supported by Grants X-094 from the Universidad de Buenos Aires, PID 5505 from CONICET and PAV 120 and PICT 21407 from ANPCYT, Argentina and by a grant of the Fundación Antorchas at Buenos Aires, Argentina.

P Appendix: Proofs.

PROOF (14). In order to derive (14), we need to assume that **A.1**, **A.5**, **B.1** to **B.5** and that the score function Ψ defining the local $M-$smoothers satifies **B.5 a)**.

Note that $\widehat{\phi}_0(t)$ is the solution of

$$(nh)^{-1}\sum_{i=1}^{n} K\left((t_i-t)/h\right)\Psi\left((y_i-\widehat{\phi}_0(t))/\sigma_0\right)=0$$

then, using a Taylor's expansion, we have that

$$\widehat{\phi}_0(t)=(nh)^{-1}\sum_{i=1}^{n} K\left((t_i-t)/h\right)\left(\phi_0(t_i)+u_i^*\right)+O_p((nh)^{-1})\ .$$

Hence

$$\widehat{\phi}_0(t)-\phi_0(t)=(nh)^{-1}\sum_{i=1}^{n} K\left((t_i-t)/h\right)\phi_0(t_i)-\phi_0(t)+(nh)^{-1}\sum_{i=1}^{n} K\left((t_i-t)/h\right)u_i^*+O_p((nh)^{-1})\ .$$

Denote $w_{ij}=(nh)^{-1}K\left((t_i-t_j)/h\right)$, $\mathbf{u}^*=(u_1^*,\ldots,u_n^*)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\phi}}_0=(\widehat{\phi}_0(t_1),\ldots,\widehat{\phi}_0(t_n))^{\mathrm{T}}$ then $\widehat{\boldsymbol{\phi}}_0-\boldsymbol{\phi}_0=(\mathbf{W}-\mathbf{I})\boldsymbol{\phi}_0+\mathbf{W}\mathbf{u}^*+O_p((nh)^{-1})$. In a similar way, we get that $\widehat{\boldsymbol{\phi}}_j-\boldsymbol{\phi}_j=(\mathbf{W}-\mathbf{I})\boldsymbol{\phi}_j+\mathbf{W}\boldsymbol{\eta}^{*^{(j)}}+O_p((nh)^{-1})$ with $\boldsymbol{\eta}^{*^{(j)}}=(\eta_{1j}^*,\ldots,\eta_{nj}^*)^{\mathrm{T}}$. Denote $\boldsymbol{\eta}_i^*=(\eta_{i1}^*,\ldots,\eta_{ip}^*)^{\mathrm{T}}$, $\boldsymbol{\phi}=(\boldsymbol{\phi}(t_1),\ldots,\boldsymbol{\phi}(t_n))^{\mathrm{T}}$ and $\boldsymbol{\phi}(t_i)=(\phi_1(t_i),\ldots,\phi_p(t_i))^{\mathrm{T}}$.

Using the expansion in Bianco and Boente (2004), we get that $n^{1/2}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)-\boldsymbol{\beta})=\sigma_\epsilon\mathbf{A}^{-1}\widehat{L}_n(\sigma_\epsilon,\boldsymbol{\beta})+o_p(n^{-2\mu})$ with $\mu=(4\nu-1)/(2(4\nu+1))$ and $\widehat{L}_n(\sigma_\epsilon,\boldsymbol{\beta})=n^{-1/2}\sum_{i=1}^n\psi_1\left((\widehat{r}_i-\widehat{\mathbf{z}}_i^{\mathrm{T}}\boldsymbol{\beta})/\sigma_\epsilon\right)w_2(\widehat{\mathbf{z}}_i)\widehat{\mathbf{z}}_i$.

Using a Taylor expansion, we have that $\widehat{L}_n(\sigma,\boldsymbol{\beta})=L_n(\sigma,\boldsymbol{\beta})+\sum_{i=1}^3 S_{in}+R_n$ where R_n has higher order than the other terms and

$$\begin{aligned}
L_n(\sigma_\epsilon,\boldsymbol{\beta}) &= n^{-1/2}\sum_{i=1}^{n}\psi_1\left((r_i-z_i^{\mathrm{T}}\boldsymbol{\beta})/\sigma_\epsilon\right)\psi_2\left(\boldsymbol{\eta}_i\right)\\
S_{1n} &= n^{-1/2}\sum_{i=1}^{n}\psi_1\left(\epsilon_i/\sigma_\epsilon\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\left(\boldsymbol{\phi}(t_i)-\widehat{\boldsymbol{\phi}}(t_i)\right)\\
S_{2n} &= (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^{n}\psi_1'\left(\epsilon_i/\sigma_\epsilon\right)\psi_2\left(\boldsymbol{\eta}_i\right)\left(g(t_i)-g^*(t_i)\right)\\
S_{3n} &= (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^{n}\psi_1'\left(\epsilon_i/\sigma_\epsilon\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\left(\boldsymbol{\phi}(t_i)-\widehat{\boldsymbol{\phi}}(t_i)\right)(g(t_i)-g^*(t_i))
\end{aligned}$$

with $g^*(t)=\widehat{\phi}_0(t)-\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\phi}}(t)$, $D_{\psi_2}\left(\mathbf{u}\right)$ stands for the matrix with (i,j) element $\frac{\partial}{\partial u_j}\psi_2(\mathbf{u})_i$.

Since the errors have symmetric distribution and ψ_1 is odd we get that $E(L_n(\sigma_\epsilon,\boldsymbol{\beta}))=0$. On the other hand, note that S_{1n} can be written as $S_{1n}=n^{-1/2}\Lambda_\eta\Lambda_\epsilon$ where $\Lambda_\epsilon=(\psi_1(\epsilon_1/\sigma_\epsilon),\ldots,\psi_1(\epsilon_n/\sigma_\epsilon))^{\mathrm{T}}$ and $\Lambda_\eta=\left(D_{\psi_2}\left(\boldsymbol{\eta}_1\right)\mathbf{v}_1,\ldots,D_{\psi_2}\left(\boldsymbol{\eta}_n\right)\mathbf{v}_n\right)$ with $\mathbf{v}_i$ the $i-$th row of $\mathbf{v}=(\mathbf{I}-\mathbf{W})\boldsymbol{\phi}-\mathbf{W}\boldsymbol{\eta}^*$ and $\boldsymbol{\eta}^*$ the matrix with $j-$th column $\boldsymbol{\eta}^{*^{(j)}}$. Thus, **B.1** and **B.3** entail that $E(S_{1n})=0$.

A similar expression is obtained for S_{2n} where $\psi_2(\boldsymbol{\eta}) = (\psi_2(\boldsymbol{\eta}_1), \dots, \psi_2(\boldsymbol{\eta}_n))^{\mathrm{T}}$

$$\begin{aligned} \sigma_\epsilon n^{\frac{1}{2}} S_{2n} &= \left\{\psi_2(\boldsymbol{\eta})^{\mathrm{T}}\Lambda'(\mathbf{I}-\mathbf{W})\boldsymbol{\phi}_0 - \psi_2(\boldsymbol{\eta})^{\mathrm{T}}\Lambda'\mathbf{W}\mathbf{u}^* + \psi_2(\boldsymbol{\eta})^{\mathrm{T}}\Lambda'\mathbf{W}\boldsymbol{\eta}^{*\mathrm{T}}\boldsymbol{\beta} - \psi_2(\boldsymbol{\eta})^{\mathrm{T}}\Lambda'(\mathbf{I}-\mathbf{W})\boldsymbol{\phi}\boldsymbol{\beta}\right\} \\ &= A_{1n} + A_{2n} + A_{3n} + A_{4n} \end{aligned}$$

with $\Lambda' = \mathrm{diag}(\psi_1'(\epsilon_1/\sigma_\epsilon), \dots, \psi_1'(\epsilon_n/\sigma_\epsilon))$. It is easy to see that $E(A_{1n}) = E(A_{3n}) = 0$ since the errors $\boldsymbol{\eta}$ have symmetric distribution and ψ_2 is odd. On the other hand, $E(A_{2n}) = 0$ since both ϵ and $\boldsymbol{\eta}$ have a symmetric distribution and ψ_1 and ψ_2 are odd functions. Finally, it is easy to show that $E(A_{4n}) = O((nh^2)^{-1/2})$.

Analogous arguments to those used in the classical setting allow to derive that

$$\begin{aligned} E(S_{3n}) &= n^{1/2}h^{2\nu}\, \alpha_\nu^2(K)\sigma_\epsilon^{-1}(\nu!)^{-2}\, E(\psi_1'(\epsilon/\sigma_\epsilon))E\left(D_{\psi_2}(\boldsymbol{\eta})\right) \int_0^1 g^{(\nu)}(t)\boldsymbol{\phi}^{(\nu)}(t)dt \\ &+ O(n^{-\mu}) + o\left(n^{(1-4\nu)/(2(4\nu+1))}\right) \end{aligned}$$

Then, if $\mathbf{c} \neq 0$ we get that

$$\begin{aligned} n^{1/2}E(\mathbf{c}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h) - \boldsymbol{\beta})) &= \sigma_\epsilon E(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\widehat{L}_n(\sigma_\epsilon, \boldsymbol{\beta})) + o_p(n^{2\nu}) \\ &= n^{1/2}h^{2\nu}\mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{1,\boldsymbol{\eta}}^{-1}E\left(D_{\psi_2}(\boldsymbol{\eta})\right)\alpha_\nu^2(K)(\nu!)^{-2}\int_0^1 g^{(\nu)}(t)\boldsymbol{\phi}^{(\nu)}(t)dt \ + o(n^{-\mu}) \end{aligned}$$

To conclude the proof, it is enough to obtain an expression for

$$\mathrm{VAR}\left(n^{1/2}\mathbf{c}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h) - \boldsymbol{\beta})/\sigma_{\mathrm{R}}\right) = \sigma_\epsilon^2/\sigma_{\mathrm{R}}^2\mathrm{VAR}\left(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\widehat{L}_n(\sigma_\epsilon, \boldsymbol{\beta})\right) + o_p(n^{-2\mu})\,.$$

Denote $\mathbf{W}^{(i)}$ the $i-$th row of $\mathbf{W}$. Let us consider the following expansion of $\widehat{L}_n(\sigma_\epsilon, \boldsymbol{\beta})$, $\widehat{L}_n(\sigma_\epsilon, \boldsymbol{\beta}) = L_n(\sigma_\epsilon, \boldsymbol{\beta}) + \mathbf{b}_n + \mathbf{c}_n + R_n$ where,

$$\begin{aligned} \mathbf{b}_n &= (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_i/\sigma_\epsilon\right)\psi_2\left(\boldsymbol{\eta}_i\right)\left(g(t_i) - \widetilde{g}(t_i)\right) + n^{-1/2}\sum_{i=1}^n \psi_1\left(\epsilon_i/\sigma_\epsilon\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\left(\boldsymbol{\phi}(t_i) - \widetilde{\boldsymbol{\phi}}(t_i)\right) \\ &- (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_1/\sigma_\epsilon\right)\left(g(t_i) - \widetilde{g}(t_i)\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\mathbf{W}^{(i)}\boldsymbol{\eta}^* \\ &+ (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_i/\sigma_\epsilon\right)\mathbf{W}^{(i)}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}^* - \mathbf{u}^*)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\left(\boldsymbol{\phi}(t_i) - \widetilde{\boldsymbol{\phi}}(t_i)\right) \\ &+ (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_1/\sigma_\epsilon\right)\left(g(t_i) - \widetilde{g}(t_i)\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\left(\boldsymbol{\phi}(t_i) - \widetilde{\boldsymbol{\phi}}(t_i)\right) \\ \mathbf{c}_n &= (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_i/\sigma_\epsilon\right)\mathbf{W}^{(i)}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}^* - \mathbf{u}^*)\psi_2\left(\boldsymbol{\eta}_i\right) - n^{-1/2}\sum_{i=1}^n \psi_1\left(\epsilon_i/\sigma_\epsilon\right)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\mathbf{W}^{(i)}\boldsymbol{\eta}^* \\ &- (\sigma_\epsilon\sqrt{n})^{-1}\sum_{i=1}^n \psi_1'\left(\epsilon_i/\sigma_\epsilon\right)\mathbf{W}^{(i)}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\eta}^* - \mathbf{u}^*)D_{\psi_2}\left(\boldsymbol{\eta}_i\right)\mathbf{W}^{(i)}\boldsymbol{\eta}^* \end{aligned}$$

$\widetilde{\boldsymbol{\phi}} = (\mathbf{I}-\mathbf{W})\boldsymbol{\phi}$ y $\widetilde{\mathbf{g}} = (\mathbf{I}-\mathbf{W})\mathbf{g}$. With regard to $\textsc{var}(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{c}_n)$, we have that

$$\begin{aligned} Var(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{c}_n) &= (h^2n^3)^{-1}\left\{\kappa_1\sum_i\sum_j K^2\left((t_i-t_j)/h\right)+\kappa_2\sum_i\sum_j\left[(K*K)\left((t_i-t_j)/h\right)\right]^2\right.\\ &\quad \left. - \; 2\kappa_3\sum_i\sum_j K\left((t_i-t_j)/h\right)(K*K)\left((t_i-t_j)/h\right)\right\} \end{aligned}$$

Now $(n^2h)^{-1}\sum_i\sum_j K^2((t_i-t_j)/h) \to \int K^2(u)du$ and arguing in an analogous way, with the other sums we get that,

$$Var(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{c}_n) = (nh)^{-1}\left\{\kappa_1\int K^2(u)du+\kappa_2\int(K*K)^2(u)du-2\kappa_3\int K(u)K*K(u)du\right\}$$

Similarly we get that $E(\mathbf{c}_n)=0$, $\textsc{var}(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{b}_n)=O(n^{-\mu})$ and $\textsc{cov}(\mathbf{b}_n,\mathbf{c}_n)=0$ which leads to

$$\begin{aligned} \textsc{var}\left(n^{1/2}\mathbf{c}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathrm{R}}(h)-\boldsymbol{\beta})/\sigma_{\mathrm{R}}\right) &= \sigma_\epsilon^2/\sigma_{\mathrm{R}}^2\textsc{var}\left(\mathbf{c}^{\mathrm{T}}\mathbf{A}^{-1}\widehat{L}_n(\sigma_\epsilon,\boldsymbol{\beta})\right)+o_p(n^{-2\mu})\\ &= \sigma_\epsilon^2/\sigma_{\mathrm{R}}^2(nh)^{-1}\left\{\kappa_1\int K^2(u)du+\kappa_2\int(K*K)^2(u)du\right.\\ &\quad \left.-2\kappa_3\int K(u)K*K(u)du\right\}+o_p(n^{-2\mu})\ , \end{aligned}$$

concluding the proof. □

Proof of Theorem 3.1. **A6** entails that $\sup_{t\in[h,1-2h]}|\widehat{\phi}_{j,\mathrm{R}}^{(\nu)}(t)-\phi_j^{(\nu)}(t)| \xrightarrow{a.s.} 0$, for $0\le j\le p$. Then, using (16), **A.2** and since $\int_0^h g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt+\int_h^{1-h} g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt$ converge to 0, we get that, for $1\le j\le p$

$$\int_h^{1-h}\widehat{g}_{\mathrm{R}}^{(\nu)}(t,h)\widehat{\phi}_{1,\mathrm{R}}^{(\nu)}(t,h)dt-\int_0^1 g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt \xrightarrow{a.s.} 0.$$

On the other hand, the strong consistency of $\widehat{\sigma}_\epsilon^2$, $\widehat{V}(\psi_1)$, $\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}$ $\widehat{\boldsymbol{\Sigma}}_{2,\boldsymbol{\eta}}$, $\widehat{\mathbf{D}}$ and $\widehat{\kappa}_\ell$ entail the desired result. □

Proof of Proposition 3.1. Using the continuity of the functional $\sigma^2(\cdot)$ and since the Strong Law of Large Numbers entails that $\Pi(P_n,P)\xrightarrow{a.s.}0$, it will be enough to show that

$$\Pi(\widehat{P}_n,P_n)\xrightarrow{a.s.}0\ , \qquad \text{(P.1)}$$

where Π stands for the Prohorov distance.

To prove (P.1), it will be enough to show that for any bounded and continuous function f, $|E_{\widehat{P}_n}(f)-E_{P_n}(f)|\xrightarrow{a.s.}0$. Let $\mathcal{C}=\mathcal{C}_1\times\mathcal{C}_2$ be such that $P(\mathcal{C})>1-\eta/(4\|f\|_\infty)$ with $\mathcal{C}_1=\{\|\mathbf{x}\|<C_1\}$, and $\mathcal{C}_2=\{|y|<C_2\}$, then

$$\begin{aligned} |E_{\widehat{P}_n}(f)-E_{P_n}(f)| &\le n^{-1}\sum_{i=1}^n|f(\widehat{\epsilon}_i)-f(\epsilon_i))|I_{\mathcal{C}}(\mathbf{x}_i,y_i)+2\|f\|_\infty n^{-1}\sum_{i=1}^n I_{\mathcal{C}^c}(\mathbf{x}_i,y_i)\\ &\le S_{1,n}+S_{2,n}\ . \end{aligned}$$

The Strong Law of Large Numbers implies that there exists a set $\mathcal{N}_1$ such that $P(\mathcal{N}_1) = 0$ and for any $w \notin \mathcal{N}_1$, $n^{-1} \sum_{i=1}^{n} I_{\mathcal{C}^c}(\mathbf{x}_i, y_i) \to P((\mathbf{x}, y) \in \mathcal{C}^c)$. Hence, for $w \notin \mathcal{N}_1$ and $n \geq n_1$, $|S_{2,n}| < \eta/2$.

On the other hand, let $\mathcal{U} = \{u : |u| \leq C_3\}$ where $C_3 = C_2 + C_1(\|\boldsymbol{\beta}\| + 1) + \|g\|_\infty + 1$. The uniformly continuity of f on $\mathcal{U}$ entail that there exists $\delta > 0$ such that for any $u_1, u_2 \in \mathcal{U}$, $|u_1 - u_2| < \delta \Rightarrow |f(u_1) - f(u_2)| < \eta/2$.

Using **A.2** we get that there exists a set $\mathcal{N}_2$ such that $P(\mathcal{N}_2) = 0$ and for any $w \notin \mathcal{N}_2$ and $n \geq n_2$ $\sup_{t \in [0,1]} |\widehat{g}(t) - g(t)| < \min(1, \delta/2)$ and $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}| < \min(1, \delta/(2\, C_1))$.

It is easy to see that $y_i - \widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_i - \widehat{g}(t_i) \in \mathcal{U}$ and $y_i - \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i - g(t_i) \in \mathcal{U}$, for $n \geq n_2$, when $(\mathbf{x}_i, y_i) \in \mathcal{C}$. Then, for $n \geq n_2$ and $i \in \mathcal{J} = \{i : (\mathbf{x}_i, y_i) \in \mathcal{C}\}$, we have that $\widehat{\epsilon}_i, \epsilon_i \in \mathcal{T}$ and $|\widehat{\epsilon}_i - \epsilon_i| < \delta$ implying that $|S_{1,n}| < \eta/2$. Thus, $|E_{\widehat{F}_n}(f) - E_{F_n}(f)| < \eta$ if $w \notin \mathcal{N}_1 \cup \mathcal{N}_2$ and $n \geq N = \max(n_1, n_2)$, concluding the proof of (P.1).

The proof of the consistency of $\widehat{V}(\psi_1)$, $\widehat{\boldsymbol{\Sigma}}_{1,\boldsymbol{\eta}}$, $\widehat{\boldsymbol{\Sigma}}_{2,\boldsymbol{\eta}}$ and $\widehat{\mathbf{D}}$ follows similar arguments as those considered in the proof of Lemma 2 in Bianco and Boente (2004). $\square$

References

[1] Aït Sahalia, Y. (1995). The delta method for nonlinear kernel functionals. PhD. dissertation, University of Chicago.

[2] Aneiros-Pérez, G. and Quintela del Río, G. (2002). Plug-in bandwidth choice in partial linear regression models with autoregressive errors.*J. Statist. Plann. Inference*, **57**, 23-48.

[3] Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plann. Inference*, **122**, 229-252.

[4] Bianco, A. and Boente, G. (2007). Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behavior and bandwidth selection. *J. Time Series Anal.*, **28**, 274-306.

[5] Boente, G. and Rodriguez, D. (2006). Robust estimators of high order derivatives of regression functions. *Statist. Probab. Lett.*, **76**, 1335-1344.

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

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

[8] Denby, L. (1986). Smooth regression functions. *Statistical Research Report* **26**, AT&T Bell Laboratories, Murray Hill.

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

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

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

[12] Gasser, Th., Müller, H. and Mammitzsch, V. (1985). Kernel for nonparametric curve estimation. *J. Royal Statist. Soc.*, Ser. B, **47**, 238-252.

[13] Härdle, W. and Gasser, Th. (1984). Robust non–parametric function fitting. *J. Royal Statist. Soc.*, **46**, 42-51.

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

[15] Härdle, W., Liang H. and Gao J. (2000). *Partially Linear Models.* Physica-Verlag

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

[17] Jiang, J. and Mack, Y. (2001). Robust local polynomial regression for dependent data. *Statistica Sinica*, **11**, 705-722.

[18] Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Ann. Statist.*, **33**, 2291-2310.

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

[20] Linton, O. (1995). Second Order Approximation in the partially linear regression model. *Econometrica*, **63**, 1079-112.

[21] Manchester, L. (1996). Empirical Influence for robust smoothing. *Austral. J. Statist.*, **38**, 275-296.

[22] Maronna, R., Martin, R. D. and Yohai, V. (2006). *Robust Statistics: Theory and Practice*. Wiley, New York.

[23] Rice, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.*, **4**, 203-208.

[24] Robinson, P. (1988). Root-n-consistent Semiparametric regression. *Econometrica*, **56**, 931-954.

[25] Ronchetti, E. and Staudte, R. (1994). A robust version of Mallow's C_p. *J. Amer. Statist. Assoc.*, **89**, 550-559.

[26] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.*, **79**, 871-881.

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

[28] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc.*, Ser. B, **50**, 413-436.

[29] Tamine, J. (2002). Smoothed influence function: another view at robust nonparametric regression. Discussion paper 62, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

[30] Tukey, J. (1977). *Exploratory Data Analysis.* Reading, MA: Addison–Wesley.

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

[32] Welsh, A. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statistics Sinica*, **6**, 347-366.

	initial bandwidth							
	0.25	0.30	0.35	0.38	0.40	0.43	0.45	
	classical estimator							
Mean	-0.2157	-0.1370	-0.1024	-0.0873	-0.0790	-0.0654	-0.0559	C_0
SD	0.2578	0.2180	0.1891	0.1451	0.1306	0.1182	0.1129	
Mean	-0.1192	-0.0169	0.0419	0.0712	0.0904	0.0969	0.1126	C_1
SD	0.2871	0.2547	0.2486	0.2576	0.2507	0.2305	0.2390	
Mean	0.0716	0.1588	0.22501	0.2695	0.2993	0.34734	0.3799	C_2
SD	0.2416	0.2619	0.2352	0.2437	0.2369	0.2466	0.2459	
	robust estimator							
Mean	-0.0957	-0.0024	0.0429	0.0719	0.0724	0.0802	0.0831	C_0
SD	0.2716	0.2535	0.2394	0.2499	0.2245	0.2271	0.2198	
Mean	-0.0910	-0.0033	0.0435	0.0749	0.0901	0.0855	0.0921	C_1
SD	0.2797	0.2561	0.2214	0.2419	0.2332	0.2077	0.2130	
Mean	-0.0269	0.0757	0.1252	0.1446	0.1515	0.1534	0.1509	C_2
SD	0.2489	0.2686	0.2389	0.2324	0.2342	0.2200	0.2029	

Table 1: Estimation of the optimal bandwidth. Summary measures of $\log\left(\widehat{h}/h_{opt}\right)$ using the differentiation approach.

	initial bandwidth							
	0.25	0.30	0.35	0.38	0.40	0.43	0.45	
	classical estimator							
Mean	-0.0303	0.0274	0.0156	0.0052	0.0007	-0.0030	-0.0049	C_0
SD	0.2453	0.2003	0.1316	0.1000	0.0844	0.0731	0.0647	
Mean	0.0015	0.1529	0.1564	0.1681	0.1602	0.1556	0.1534	C_1
SD	0.2479	0.2728	0.2479	0.2639	0.2236	0.2072	0.1983	
Mean	0.1183	0.2087	0.2898	0.3457	0.3787	0.4385	0.4489	C_2
SD	0.2781	0.25501	0.2415	0.2568	0.2493	0.2949	0.2354	
	robust estimator							
Mean	0.1169	0.1776	0.1658	0.1553	0.1499	0.1465	0.1441	C_0
SD	0.2467	0.2182	0.1414	0.1094	0.0899	0.0815	0.0716	
Mean	0.0615	0.1792	0.1813	0.1741	0.1605	0.1567	0.1522	C_1
SD	0.2793	0.2194	0.1661	0.1226	0.1051	0.1050	0.0756	
Mean	-0.6489	-0.5021	-0.3376	-0.2235	-0.14284	-0.0222	0.0499	C_2
SD	0.1517	0.16064	0.2115	0.2323	0.2372	0.2182	0.1946	

Table 2: Estimation of the optimal bandwidth. Summary measures of $\log\left(\widehat{h}/h_{opt}\right)$ using local polynomials.

	$\log\left(\frac{\widehat{h}}{h_{opt}}\right)$			$\widehat{\beta}$			
	LS	GM	LTS	LS	GM	LTS	
Mean	-0.0964	-0.0967	-0.0876	0.9929	0.9973	0.9888	C_0
SD	0.4587	0.5297	0.5639	0.1063	0.1927	0.1042	
Mean	0.2484	0.0044	0.0440	0.932	0.9839	0.9826	C_1
SD	0.5381	0.5236	0.5454	1.1457	0.1903	0.1134	
Mean	0.3881	0.2592	0.3387	0.0565	0.9341	0.8912	C_2
SD	0.5615	0.5906	0.5711	0.5007	0.2883	0.1216	

Table 3: Estimation of the optimal bandwidth and of the regression parameter β under C_0, C_1 and C_2, using cross–validation.

	initial bandwidth								
	0.25	0.30	0.35	0.38	0.40	0.43	0.45		
mean	0.9827	0.9823	0.9868	0.9819	0.9816	0.9811	0.9807	LS	
sd	0.1109	0.1032	0.1005	0.1017	0.1016	0.1015	0.1014		
mean	0.9905	0.9922	0.9957	0.9891	0.9901	0.9856	0.9842	LTS	C_0
sd	0.2171	0.2193	0.2212	0.2241	0.2121	0.2184	0.2138		
mean	0.9841	0.9825	0.9869	0.9787	0.9804	0.9792	0.9796	GM	
sd	0.1059	0.1043	0.1035	0.1053	0.1037	0.1054	0.1044		
mean	0.8492	0.8507	0.8215	0.8443	0.8617	0.8512	0.8499	LS	
sd	1.5911	1.5819	1.9095	1.9312	1.7182	1.7076	1.5221		
mean	0.9939	0.991	1.0125	1.003	0.9935	0.9877	0.9818	LTS	C_1
sd	0.2149	0.2176	0.2027	0.2054	0.2063	0.2107	0.2135		
mean	0.9792	0.9781	0.9921	0.9866	0.9842	0.9809	0.9756	GM	
sd	0.1156	0.1134	0.1069	0.1071	0.1108	0.1132	0.1124		
mean	0.0555	0.0561	0.0562	0.0562	0.0562	0.0562	0.056	LS	
sd	0.4103	0.4155	0.4211	0.4243	0.4264	0.4293	0.4311		
mean	0.9525	0.9395	0.9492	0.9537	0.9472	0.9553	0.9527	LTS	C_2
sd	0.2726	0.2874	0.2682	0.2618	0.2699	0.2677	0.2693		
mean	0.8957	0.8894	0.8901	0.8892	0.8883	0.8885	0.8891	GM	
sd	0.1181	0.122	0.1183	0.119	0.1194	0.1179	0.1178		

Table 4: Estimation of the regression parameter β under C_0, C_1 and C_2 when using the differentiating approach.

	initial bandwidth							
	0.25	0.30	0.35	0.38	0.40	0.43	0.45	
LS	0.0985	0.0979	0.1045	0.0954	0.0951	0.0955	0.0961	C_0
LTS	0.206	0.2168	0.2279	0.2389	0.2138	0.217	0.2112	
GM	0.0963	0.099	0.1044	0.1068	0.1022	0.1055	0.102	
LS	23.6493	18.6363	22.3226	21.3261	17.7332	18.1118	15.3747	C_1
LTS	0.222	0.2257	0.2262	0.2291	0.2332	0.2185	0.2203	
GM	0.1186	0.1175	0.1209	0.1291	0.129	0.1193	0.1176	
LS	28.5969	24.3241	21.7866	20.6437	19.9367	18.9953	18.3776	C_2
LTS	0.2926	0.3098	0.2841	0.2841	0.2915	0.2955	0.2938	
GM	0.1163	0.1212	0.1138	0.1128	0.1157	0.1151	0.1124	

Table 5: Estimation of the regression function g. Mean of $MSE(\widehat{g})$ when using the differentiating approach

	initial bandwidth							
	0.25	0.30	0.35	0.38	0.40	0.43	0.45	
LS	0.0385	0.0368	0.0371	0.0344	0.0359	0.037	0.0373	C_0
LTS	0.0591	0.0598	0.0714	0.0623	0.058	0.062	0.0623	
GM	0.036	0.0364	0.0413	0.0372	0.0372	0.038	0.0379	
LS	0.0875	0.0891	0.1271	0.1225	0.1054	0.0973	0.0941	C_1
LTS	0.0718	0.0699	0.0881	0.0735	0.073	0.0734	0.0672	
GM	0.0449	0.0439	0.0471	0.0483	0.0491	0.045	0.0415	
LS	1.2196	1.164	1.1079	1.0822	1.0454	1.0293	0.9688	C_2
LTS	0.0699	0.0697	0.0659	0.0701	0.0691	0.0677	0.0714	
GM	0.0466	0.044	0.0452	0.0463	0.0455	0.0458	0.0463	

Table 6: Estimation of the regression regression function g. Median of $MedSE(\widehat{g})$ when using the differentiating approach.

	$MSE(\widehat{g})$			$MedSE(\widehat{g})$		
	LS	GM	LTS	LS	GM	LTS
C_0	0.099	0.1848	0.1005	0.0392	0.0599	0.0405
C_1	9.698	0.1963	0.1209	0.0863	0.0692	0.0468
C_2	18.5513	0.3477	0.1203	1.3116	0.0857	0.0493

Table 7: Estimation of the regression regression function g. Mean of $MSE(\widehat{g})$ and median of $MedSE(\widehat{g})$ when using the cross-validation.

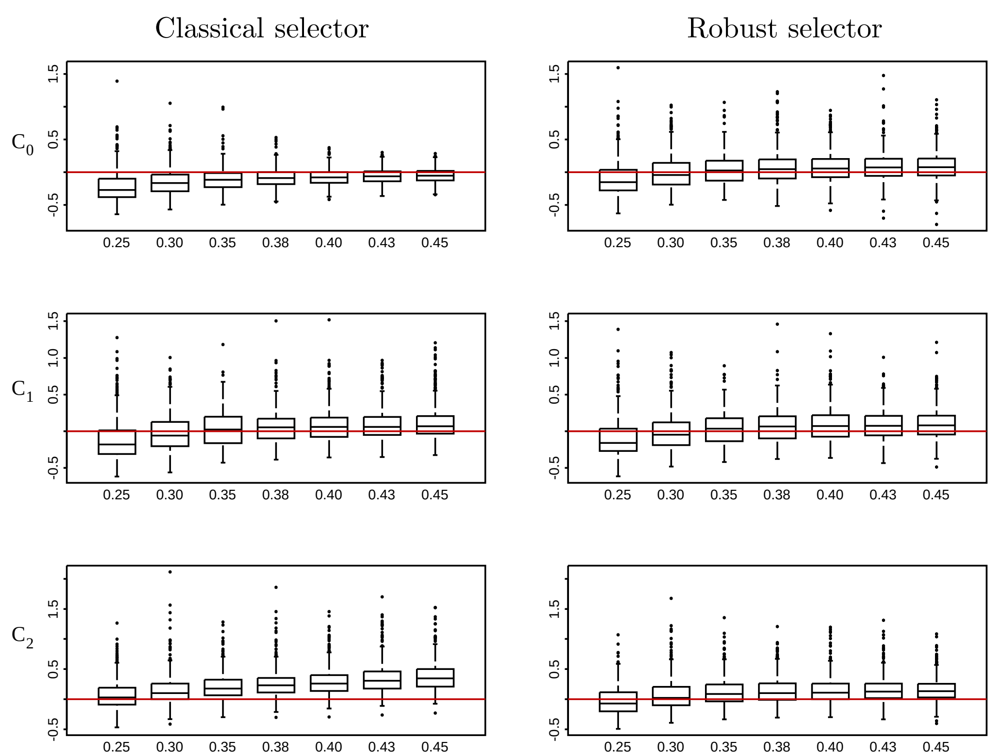


Figure 1: Boxplots of $\log\left(\widehat{h}/h_{opt}\right)$ using the differentiating approach.

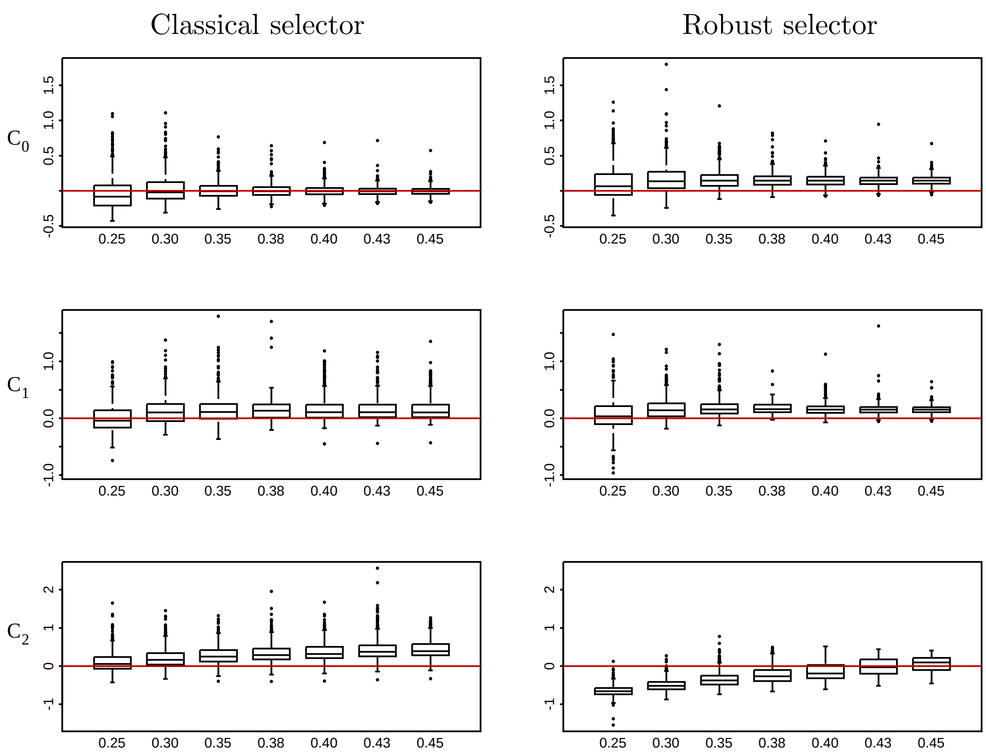


Figure 2: Boxplots of $\log\left(\widehat{h}/h_{opt}\right)$ using the robust local polynomials.

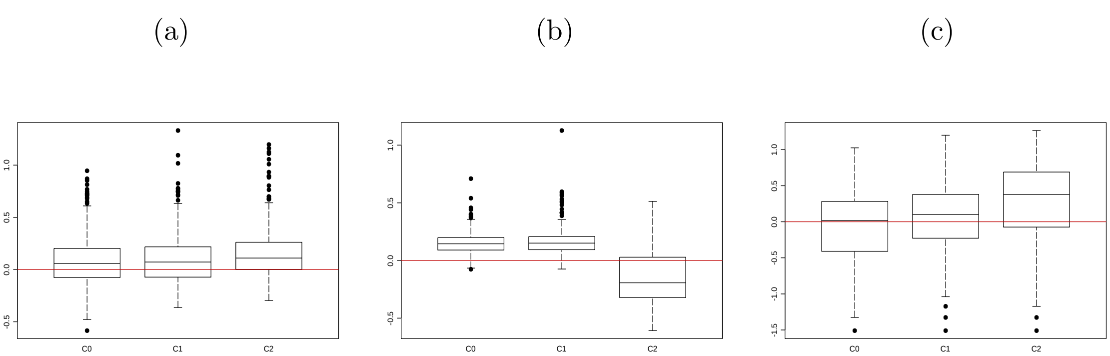


Figure 3: Boxplots of $\log\left(\widehat{h}/h_{opt}\right)$ for the robust data–driven bandwidths (a) and (b) plug–in bandwidths with initial bandwidth 0.4 (a) using the differentiating approach (b) using the local polynomial method and (c) robust cross–validation bandwidths.

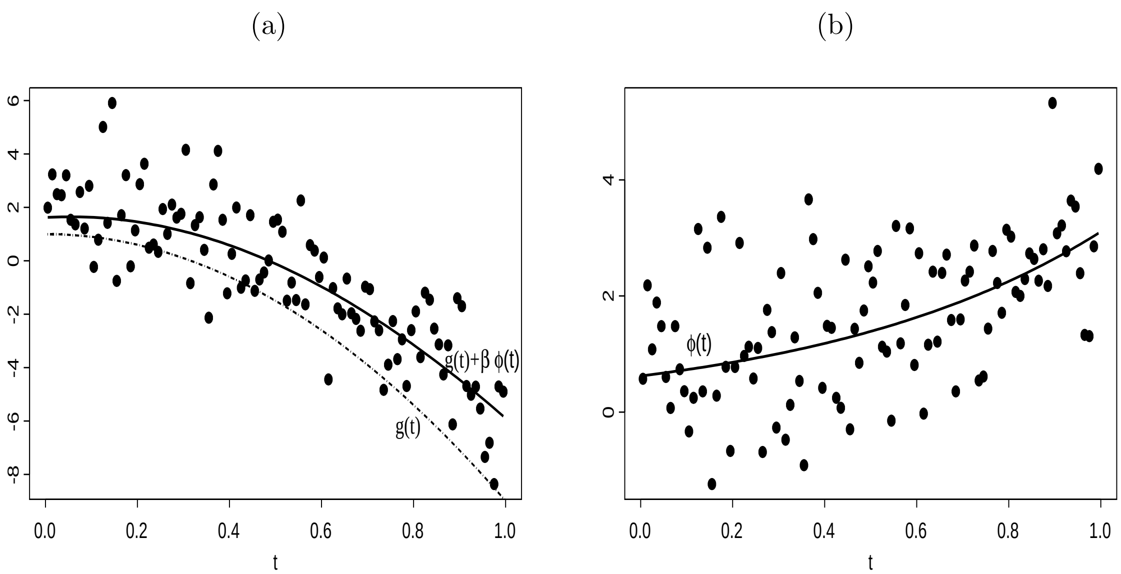


Figure 4: Generated Data Set. The dashed line corresponds to the nonparametric component g while the solid one to the regression function $\gamma(t) = g(t) + \beta\phi(t)$ in (a). In (b), the solid line corresponds to $\phi(t)$.

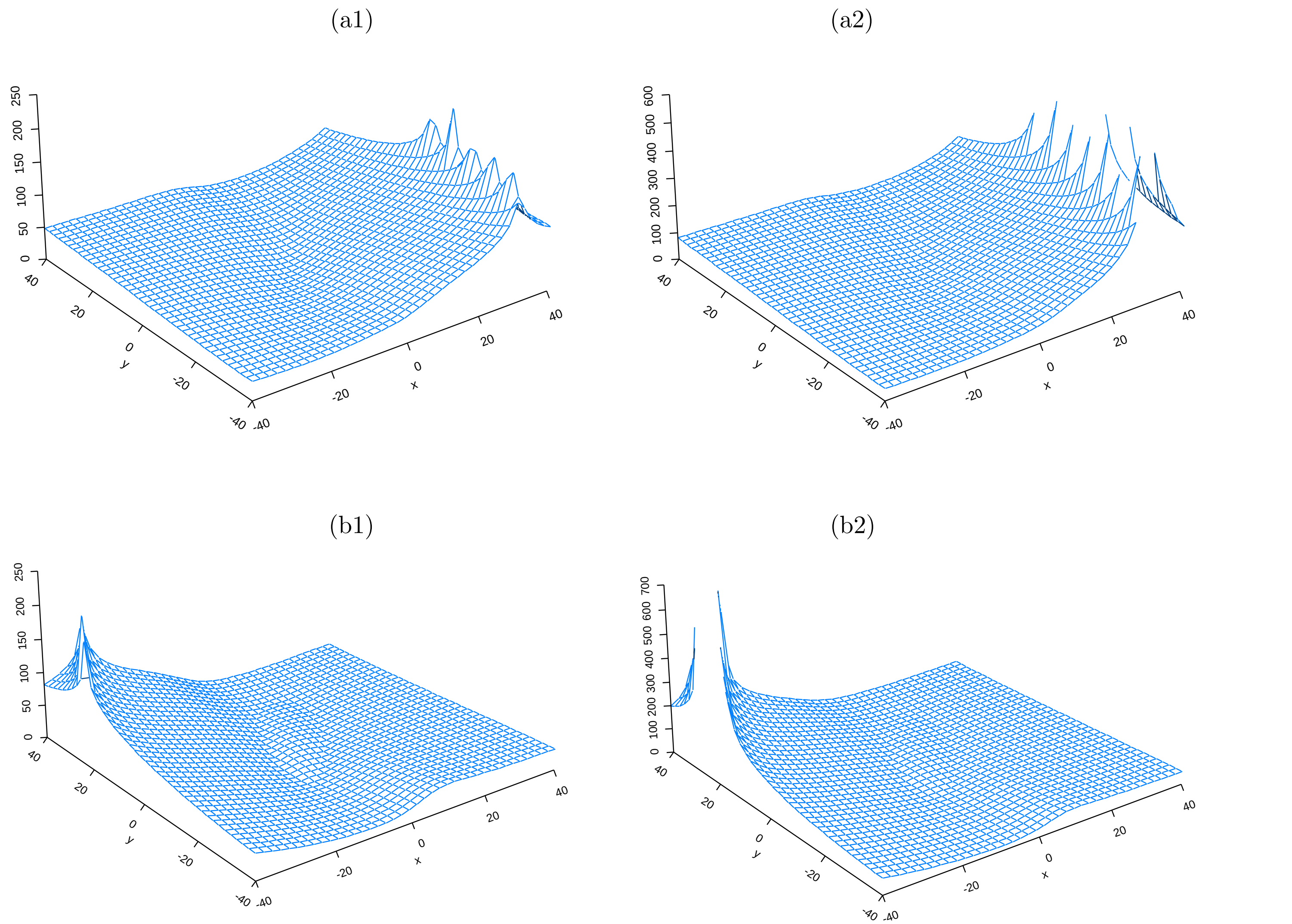


Figure 5: $EIF(0.10, x, y)$ and $EIF_1(0.10, x, y)$ for the classical bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

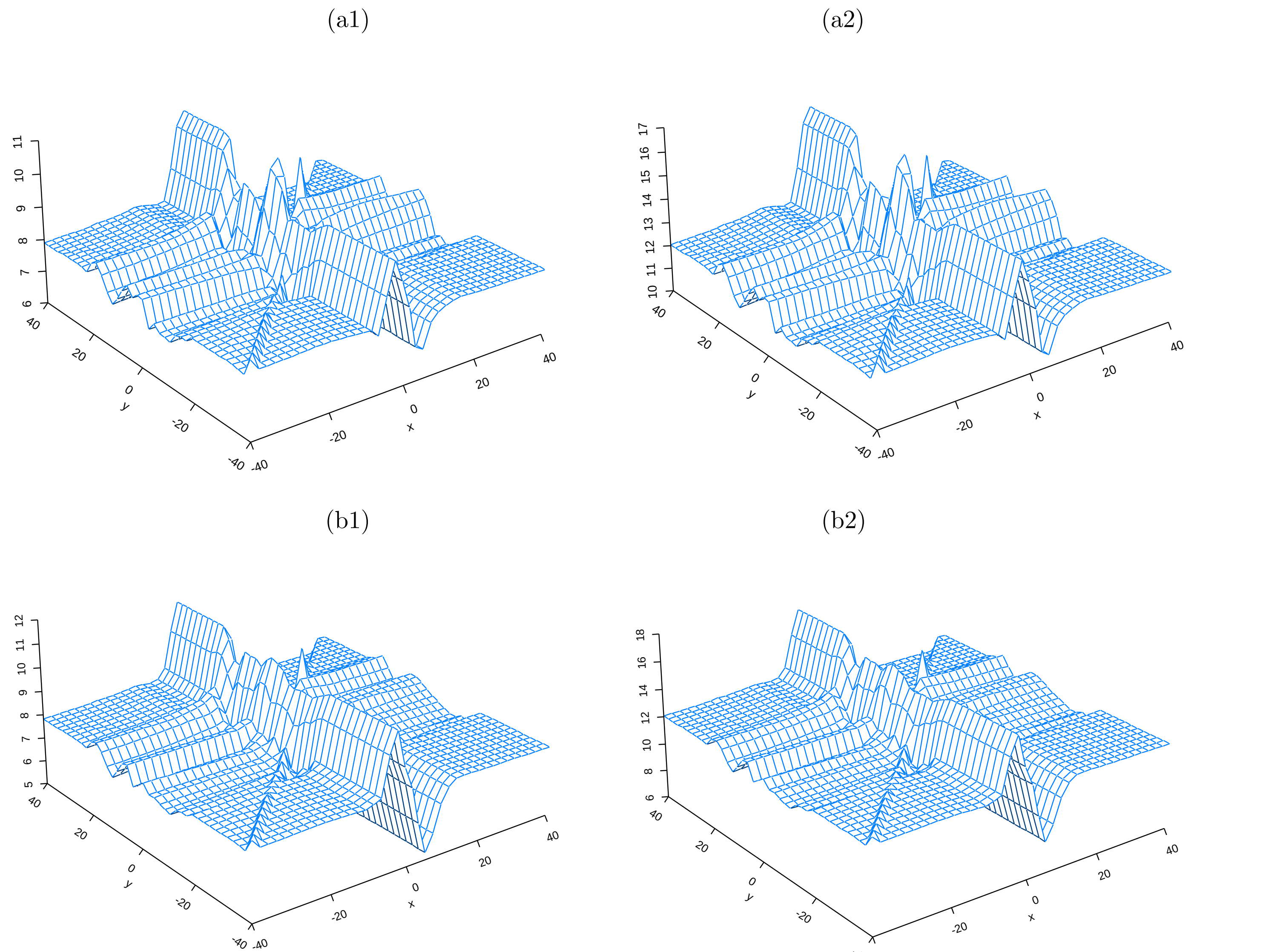


Figure 6: $EIF(0.10, x, y)$ and $EIF_1(0.10, x, y)$ for the robust bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

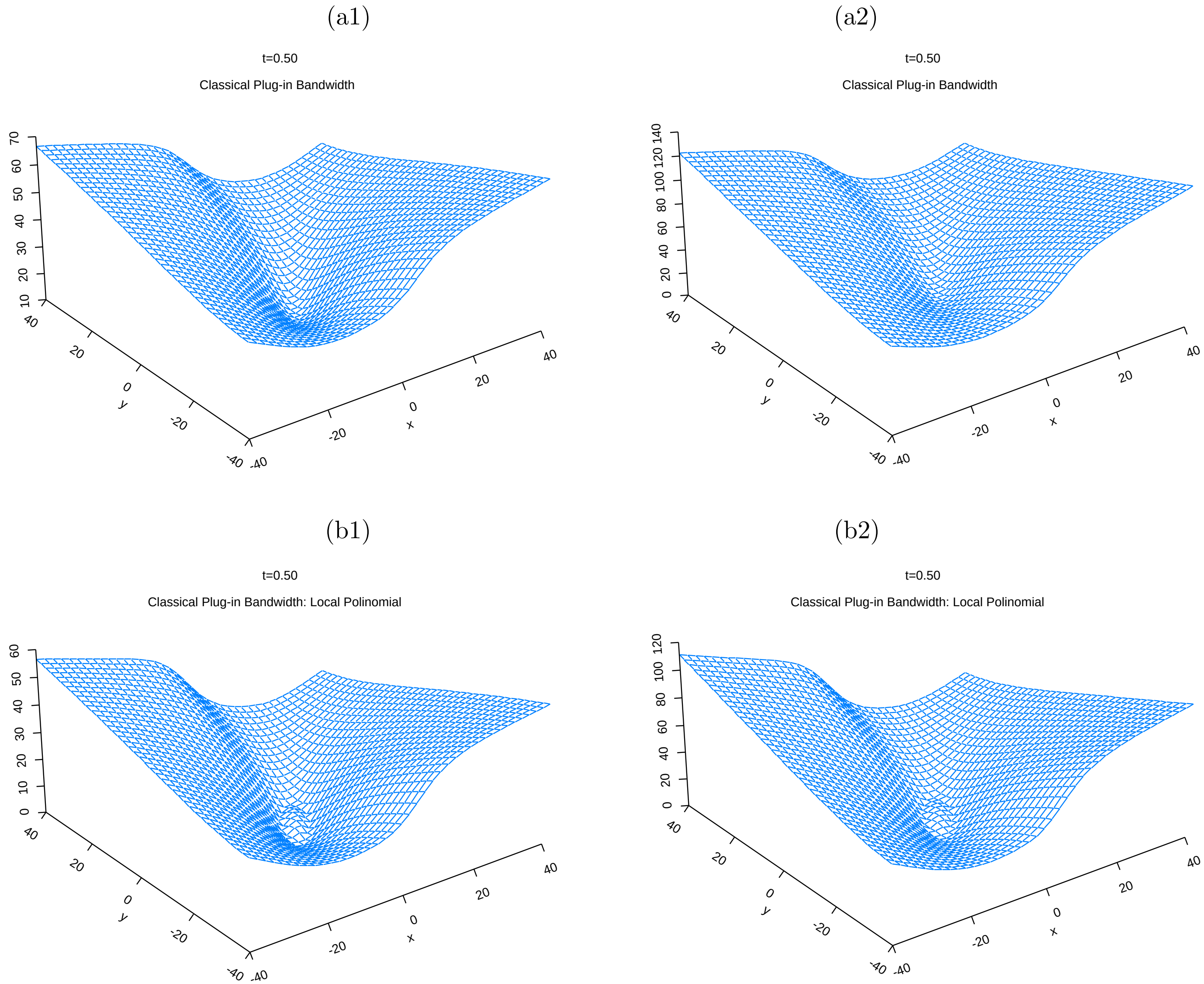


Figure 7: $EIF(0.50, x, y)$ and $EIF_1(0.50, x, y)$ for the classical bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

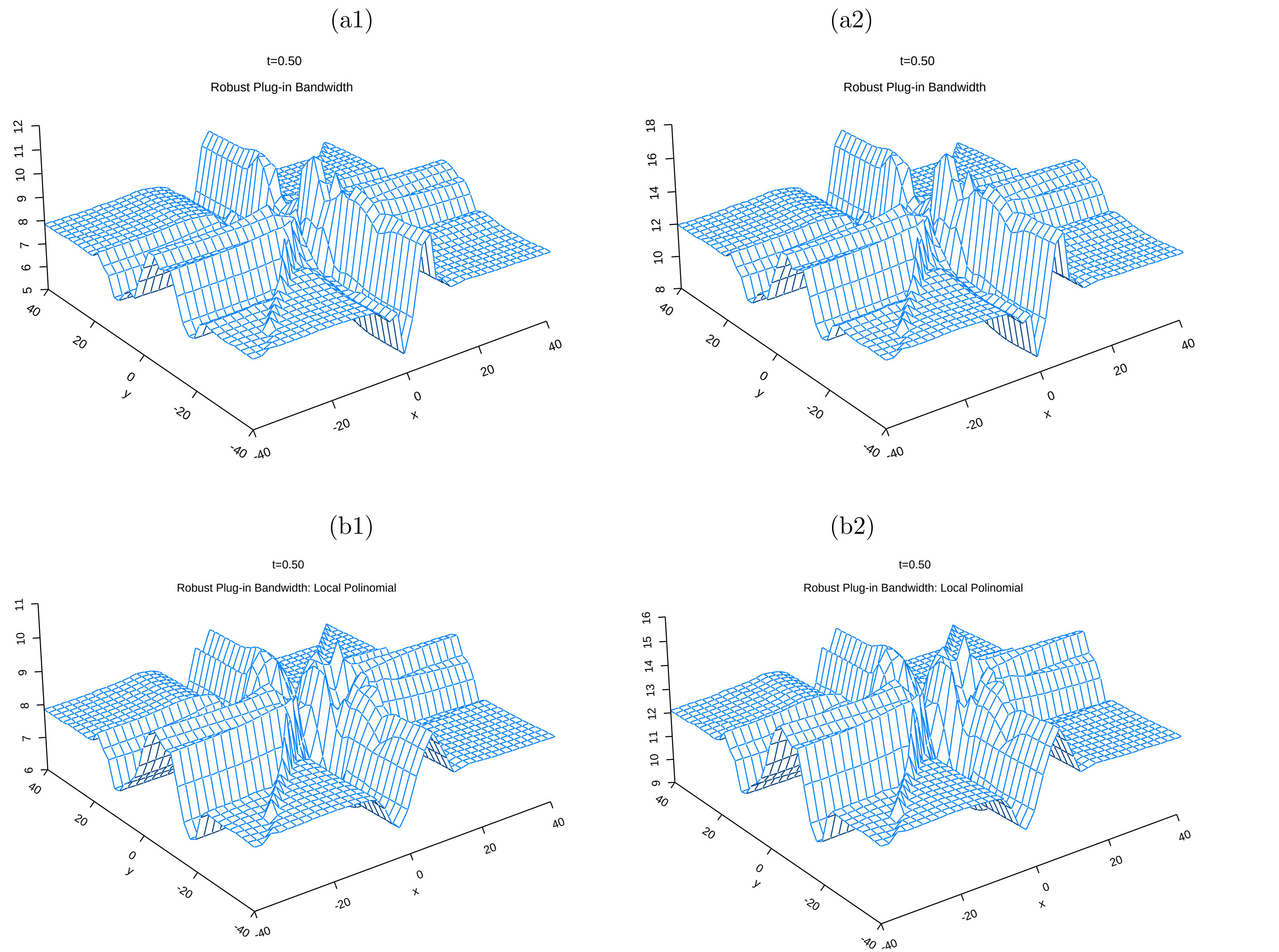
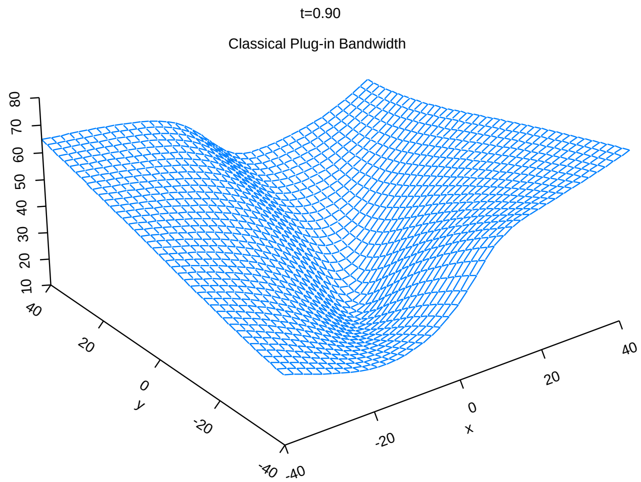
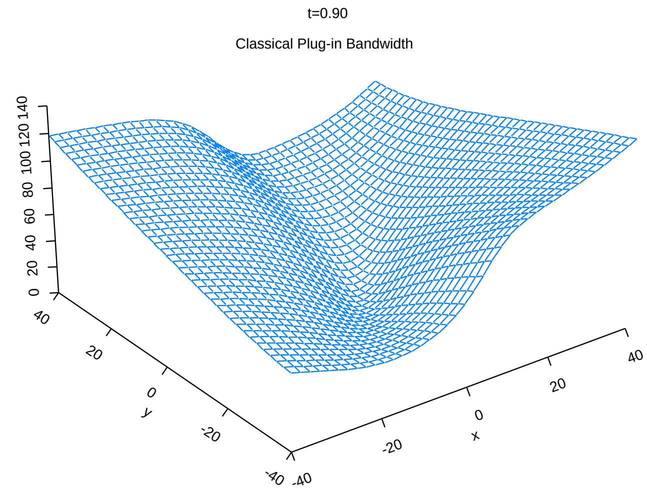


Figure 8: $EIF(0.50, x, y)$ and $EIF_1(0.50, x, y)$ for the robust bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

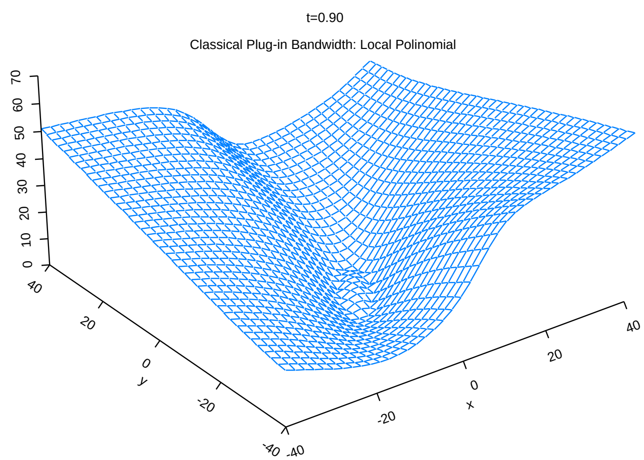
(a1)



(a2)



(b1)



(b2)

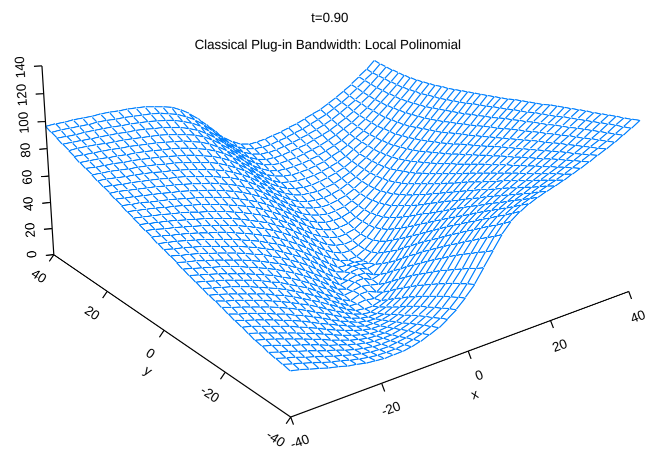


Figure 9: $EIF(0.90, x, y)$ and $EIF_1(0.90, x, y)$ for the classical bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

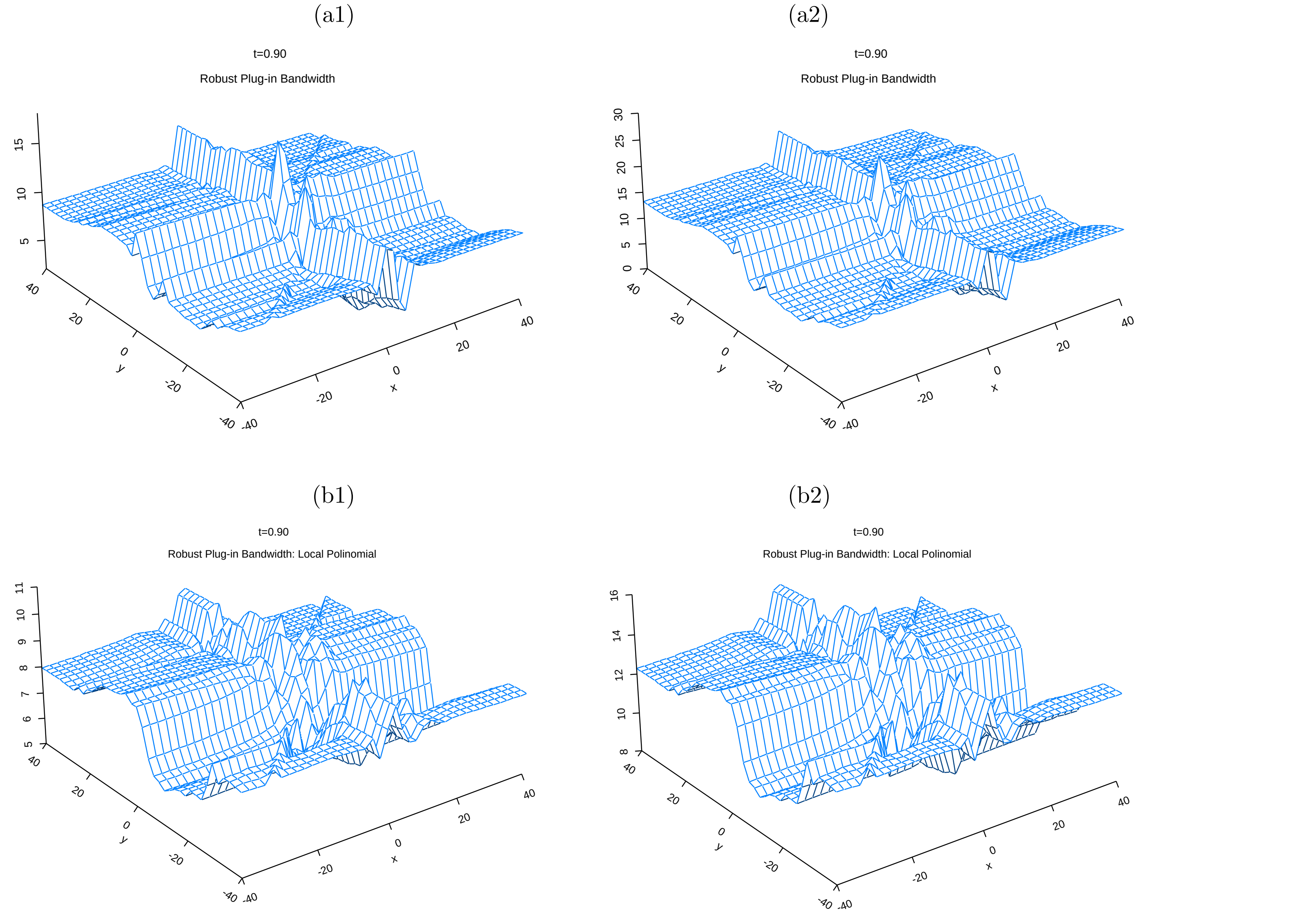


Figure 10: $EIF(0.90, x, y)$ and $EIF_1(0.90, x, y)$ for the robust bandwidth selector, using the differentiating approach, ((a1) and (a2), respectively) and using the local polynomial approach, ((b1) and (b2), respectively).

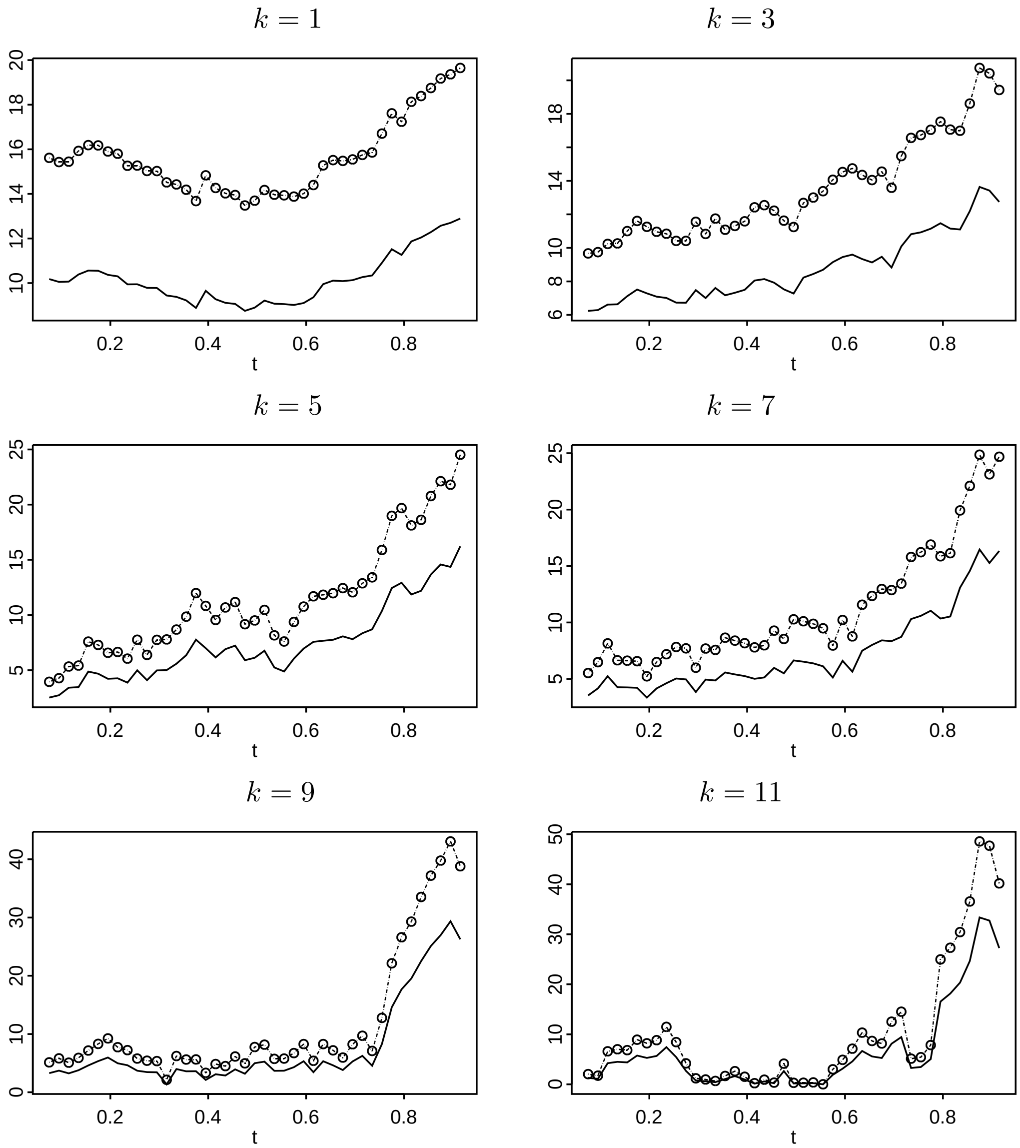


Figure 11: The solid lines correspond to $EIF(t, 10, 10)$ while the dashed lines $(- \cdot -)$ with empty circles to $EIF_1(t, 10, 10)$ for the robust plug–in selector based on the differentiating approach.

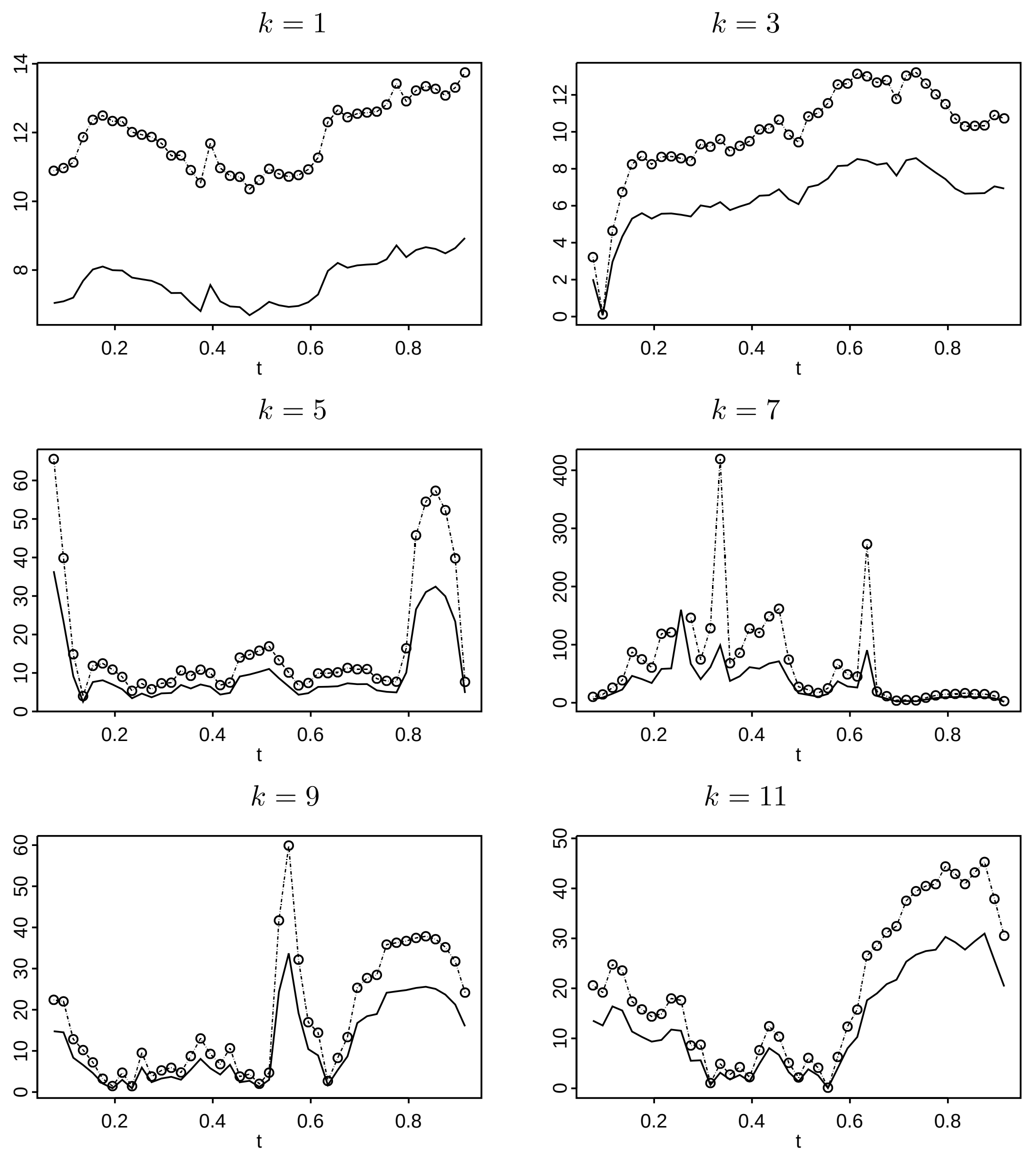


Figure 12: The solid lines correspond to $EIF(t, 10, 10)$ while the dashed lines $(- \cdot -)$ with empty circles to $EIF_1(t, 10, 10)$ for the robust plug–in selector based on the the robust local polynomial approach.

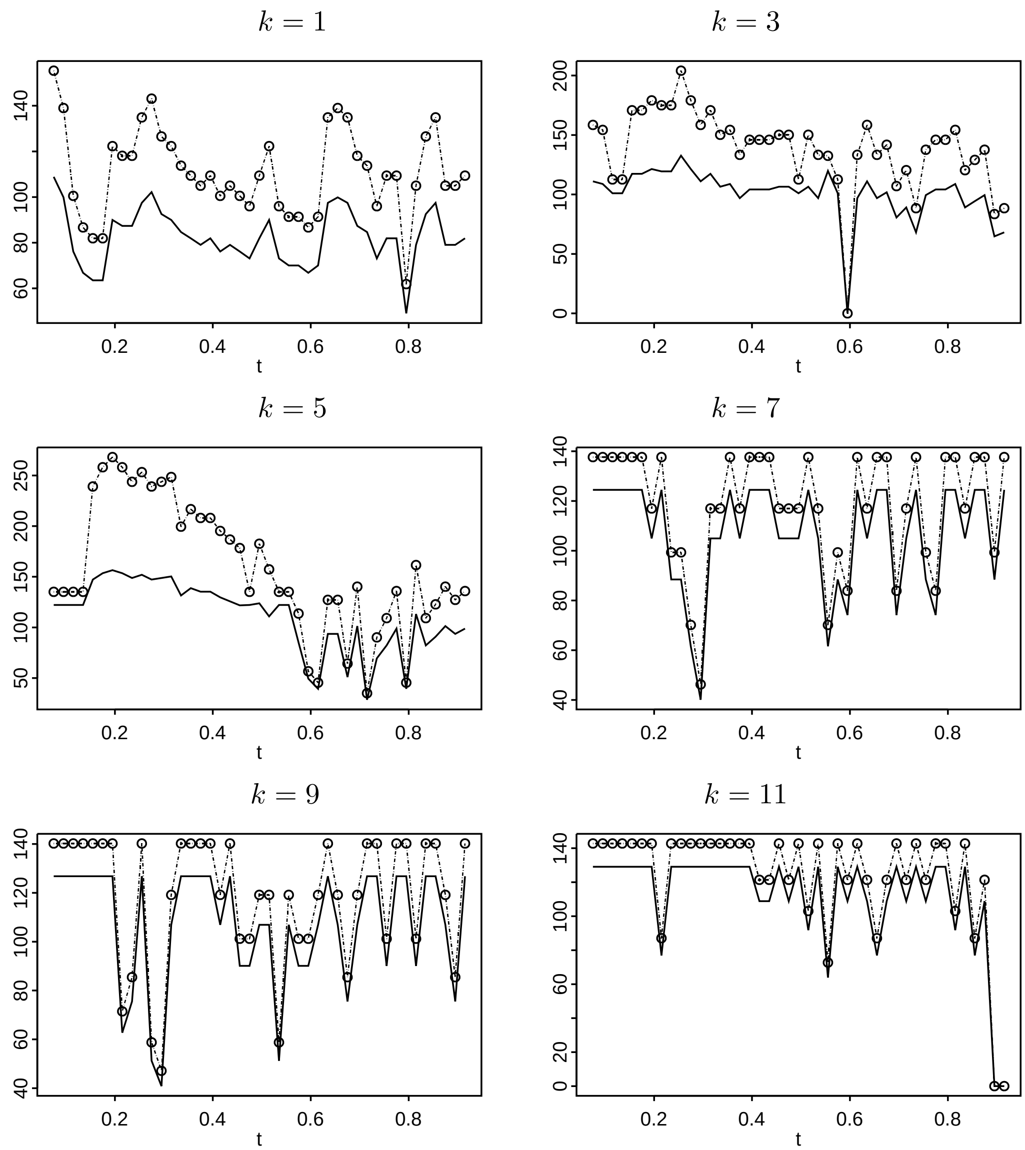


Figure 13: The solid lines correspond to $EIF(t, 10, 10)$ while the dashed lines $(- \cdot -)$ with empty circles to $EIF_1(t, 10, 10)$ for the robust cross–validation selector.