A Design-Adaptive Local Polynomial Estimator for the Errors-in-Variables Problem [*]

Aurore Delaigle, Jianqing Fan, and Raymond J. Carroll

Abstract: Local polynomial estimators are popular techniques for nonparametric regression estimation and have received great attention in the literature. Their simplest version, the local constant estimator, can be easily extended to the errors-in-variables context by exploiting its similarity with the deconvolution kernel density estimator. The generalization of the higher order versions of the estimator, however, is not straightforward and has remained an open problem for the last 15 years, since the publication of Fan and Truong (1993). We propose an innovative local polynomial estimator of any order in the errors-in-variables context, derive its design-adaptive asymptotic properties and study its finite sample performance on simulated examples. We provide not only a solution to a long-standing open problem, but also provide methodological contributions to error-in-variable regression, including local polynomial estimation of derivative functions.

Keywords: Bandwidth selector, Deconvolution, Inverse problems, Local polynomial, Measurement errors, Nonparametric Regression, Replicated measurements.

1 Introduction

Local polynomial estimators are popular techniques for nonparametric regression estimation. Their simplest version, the local constant estimator, can be easily extended to the errors-in-variables context by exploiting its similarity with the deconvolution kernel density estimator. The generalization of the higher order versions of the estimator, however, is not straightforward and has remained an open problem for the last 15 years, since the publication of Fan and Truong (1993). The purpose of this paper is to describe a solution to this long-standing open problem: we also make methodological contributions to errors-in-variable regression, including local polynomial estimation of derivative functions.

Suppose we have an i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ distributed like (X, Y) and we want to estimate the regression curve $m(x) = E(Y|X = x)$ or its νth derivative $m^{(\nu)}(x)$. Let K be a kernel function and $h > 0$ a smoothing parameter called the bandwidth. *When X is observable*, at each point x, the local polynomial estimator of order p approximates the function m by a pth order polynomial $m_p(z) \equiv \sum_{k=0}^{p} \beta_{x,k}(z - x)^k$, where the local parameters $\beta_x = (\beta_{x,0}, \ldots, \beta_{x,p})$ are fitted locally by a weighted least squares regression problem, via minimization of

$$\sum_{j=1}^{n} \left[Y_j - m_p(X_j) \right]^2 K_h(X_j - x), \tag{1.1}$$

where $K_h(x) = h^{-1}K(x/h)$. Then $m(x)$ is estimated by $\widehat{m}_p(x) = \widehat{\beta}_{x,0}$ and $m^{(\nu)}(x)$ is estimated by $\widehat{m}_p^{(\nu)}(x) = \nu!\widehat{\beta}_{x,\nu}$; see Fan and Gijbels (1996). Local polynomial estimators of order $p > 0$ have many advantages over other nonparametric estimators such as the Nadaraya-Watson estimator ($p = 0$). One of their attractive features is their capacity to adapt automatically to the boundary of the design points, thereby offering the potential of bias reduction with no or little variance increase.

In this paper, we consider the more difficult errors-in-variables problem, where the

1

goal is still to estimate the curve $m(x)$ or its derivative $m^{(\nu)}(x)$, but the only observations available are an i.i.d. sample $(W_1, Y_1), \ldots, (W_n, Y_n)$ distributed like (W, Y), where $W = X + U$ with U independent of X and Y. Here, X *is not observable* and instead we observe W, which is a version of X contaminated by a measurement error U with density f_U. In this context, when $p = 0$, $m_p(X_j) = \beta_{x,0}$ and a consistent estimator of m can simply be obtained after replacing the weights $K_h(X_j - x)$ in (1.1) by appropriate weights depending on W_j; see Fan and Truong (1993). For $p > 0$, however,

$$m_p(X_j) = \sum_{k=0}^{p} \beta_{x,k}(X_j - x)^k$$

depends on the unobserved X_j. As a result, despite the popularity of the measurement error problem, no one has yet been able to extend the minimization problem (1.1) and the corresponding local pth order polynomial estimators for $p > 0$ to the case of contaminated data. An exception is the recent paper by Zwanzig (2007), who constructed a local linear estimator of m in the context where the U_i's are normally distributed, the density of the X_i's is known to be uniform $U[0, 1]$, and the curve m is supported on $[0, 1]$.

We propose a solution to the general problem and thus generalize local polynomial estimators to the errors-in-variable case. The methodology consists of constructing simple unbiased estimators of the terms depending on X_j which are involved in the calculation of the usual local polynomial estimators. Our approach also provides an elegant estimator of the derivative functions in the errors-in-variables setting.

The errors-in-variables regression problem has been considered by many authors in both the parametric and the nonparametric context. See for example Fan and Masry (1992), Cook and Stefanski (1994), Stefanski and Cook (1995), Ioannides and Alevizos (1997), Koo and Lee (1998), Carroll, Maca and Ruppert (1999), Stefanski (2000), Taupin (2001), Berry, Carroll and Ruppert (2002), Carroll and Hall (2004), Stau-

denmayer and Ruppert (2004), Liang and Wang (2005), Comte and Taupin (2007), Delaigle and Meister (2007), Hall and Meister (2007) and Delaigle, Hall and Meister (2008); see also Carroll et al. (2006) for an exhaustive review of this problem.

2   Methodology

In this section, we will first review local polynomial estimators in the error-free case, in order to show exactly what has to be solved in the measurement error problem. After that, we give our solution.

2.1   Local polynomial estimator in the error-free case

In the usual error-free case (i.e. when the X_i's are observable), the local polynomial estimator of $m^{(\nu)}(x)$ of order p can be written in matrix notation as

$$\widehat{m}_p^{(\nu)}(x) = \nu! \boldsymbol{e}_{\nu+1}^\top (\boldsymbol{X}^\top \boldsymbol{K} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{K} \boldsymbol{y},$$

where $\boldsymbol{e}_{\nu+1}^\top = (0, , \ldots, 0, 1, 0, \ldots, 0)$ with 1 on the $(\nu+1)$th position, $\boldsymbol{y}^\top = (Y_1, \ldots, Y_n)$, $\boldsymbol{X} = \{(X_i - x)^j\}_{1 \le i \le n, 0 \le j \le p}$ and $\boldsymbol{K} = \mathrm{diag}\{K_h(X_j - x)\}$. See for example Fan and Gijbels (1996), page 59.

Using standard calculations, this estimator can be written in various equivalent ways. An expression that will be particularly useful in the context of contaminated errors, where we observe neither $\boldsymbol{X}$ nor $\boldsymbol{K}$, is the one used in Fan and Masry (1997), which follows from equivalent kernel calculations of Fan and Gijbels (1996, page 63). Let

$$\boldsymbol{S}_n = \begin{pmatrix} S_{n,0}(x) & \ldots & S_{n,p}(x) \\ \vdots & \ddots & \vdots \\ S_{n,p}(x) & \ldots & S_{n,2p}(x) \end{pmatrix}, \boldsymbol{T}_n = \begin{pmatrix} T_{n,0}(x) \\ \vdots \\ T_{n,p}(x) \end{pmatrix},$$

where

$$S_{n,k}(x) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{X_j - x}{h}\right)^k K_h(X_j - x), \quad T_{n,k}(x) = \frac{1}{n}\sum_{j=1}^{n} Y_j \left(\frac{X_j - x}{h}\right)^k K_h(X_j - x).$$

Then the local polynomial estimator of $m^{(\nu)}(x)$ of order p can be written as

$$\widehat{m}_p^{(\nu)}(x) = \nu! h^{-\nu} e_{\nu+1}^{\top} S_n^{-1} T_n.$$

2.2 Extension to the error-case

Our goal is to extend $\widehat{m}_p^{(\nu)}(x)$ to the errors-in-variables setting, where the data are a sample $(W_1, Y_1), \ldots, (W_n, Y_n)$ of contaminated i.i.d. observations coming from the model

$$Y_j = m(X_j) + \eta_j, \quad W_j = X_j + U_j, \quad \mathrm{E}(\eta_j | X_j) = 0,$$
$$\text{with } X_j \sim f_X \text{ and } U_j \sim f_U,$$

(2.1)

where U_j are the measurement errors, independent of (X_j, Y_j, η_j), and f_U is known.

For $p = 0$, a rate-optimal estimator has been developed by Fan and Truong (1993). Their technique is similar to the one employed in density deconvolution problems studied in Stefanski and Carroll (1990), see also Carroll and Hall (1988). It consists of replacing the unobserved $K_h(X_j - x)$ by an observable quantity $L_h(W_j - x)$ satisfying

$$E[L_h(W_j - x) | X_j] = K_h(X_j - x).$$

In the usual nomenclature of measurement error models, this means that $L_h(W_j - x)$ is an *unbiased score* for the kernel function $K_h(X_j - x)$.

Following this idea, we would like to replace $(X_j - x)^k K_h(X_j - x)$ in $S_{n,k}$ and $T_{n,k}$ by $(W_j - x)^k L_{k,h}(W_j - x)$, where $L_{k,h}(x) = h^{-1}L_k(x/h)$, and each L_k potentially depends on h and satisfies

$$E\left\{(W_j - x)^k L_{k,h}(W_j - x) \Big| X_j\right\} = (X_j - x)^k K_h(X_j - x).$$

(2.2)

4

That is, we propose to find unbiased scores for all components of the kernel functions. Thus, using the substitution principle, we propose to estimate $m^{(\nu)}(x)$ by

$$\widehat{m}_p^{(\nu)}(x) = \nu! h^{-\nu} \boldsymbol{e}_{\nu+1}^\top \widehat{\boldsymbol{S}}_n^{-1} \widehat{\boldsymbol{T}}_n, \tag{2.3}$$

where $\widehat{\boldsymbol{S}}_n = \{\widehat{S}_{n,j+\ell}(x)\}_{0 \leq j,\ell \leq p}$ and $\widehat{\boldsymbol{T}}_n = \{\widehat{T}_{n,0}(x), \ldots, \widehat{T}_{n,p}(x)\}^\top$ with

$$\widehat{S}_{n,k}(x) = n^{-1} \sum_{j=1}^n \left(\frac{W_j - x}{h}\right)^k L_{k,h}(W_j - x),$$

$$\widehat{T}_{n,k}(x) = n^{-1} \sum_{j=1}^n Y_j \left(\frac{W_j - x}{h}\right)^k L_{k,h}(W_j - x).$$

The method explained above seems relatively straightforward but its actual implementation is difficult, and this is the reason that the problem has remained unsolved. The main difficulty has been that it is very hard to find an explicit solution $L_{k,h}(\cdot)$ to the integral equation (2.2). In addition, a priori it is not clear that the solution will be independent of other quantities such as X_j, x, and other population parameters. Therefore, this problem has remained unsolved for more than 15 years.

The key to finding the solution is the Fourier transform. Instead of solving (2.2) directly, we solve its Fourier version

$$\mathrm{E}\left[\phi_{\{(W_j-x)^k L_{k,h}(W_j-x)\}}(t) \Big| X_j\right] = \phi_{\{(X_j-x)^k K_h(X_j-x)\}}(t), \tag{2.4}$$

where, for a function g, we let ϕ_g denote its Fourier transform, while for a random variable T, we let ϕ_T denote the characteristic function of its distribution.

We make the following basic assumptions:

Condition A:
$\int |\phi_X| < \infty$; $\phi_U(t) \neq 0$ for all t; $\phi_K^{(\ell)}$ is not identically zero and $\int |\phi_K^{(\ell)}(t)/\phi_U(t/h)| \, dt < \infty$ for all $h > 0$ and $0 \leq \ell \leq 2p$.

Condition A generalizes standard conditions of the deconvolution literature, where it is assumed to hold for $p = 0$. It is easy to find kernels which satisfy this condition. For example, kernels defined by $\phi_K(t) = (1 - t^2)^q \cdot 1_{[-1,1]}(t)$, with $q \geq 2p$, satisfy this condition.

Under these conditions, we show in the appendix that the solution to (2.4) is found by taking L_k (in the definition of $L_{k,h}$) equal to

$$L_k(u) = u^{-k} K_{U,k}(u),$$

with
$$K_{U,k}(x) = i^{-k} \frac{1}{2\pi} \int e^{-itx} \frac{\phi_K^{(k)}(t)}{\phi_U(-t/h)} \, dt.$$

In other words, our estimator is defined by (2.3), where

$$\widehat{S}_{n,k}(x) = n^{-1} \sum_{j=1}^{n} K_{U,k,h}(W_j - x) \quad \text{and} \quad \widehat{T}_{n,k}(x) = n^{-1} \sum_{j=1}^{n} Y_j K_{U,k,h}(W_j - x), \quad (2.5)$$

with $K_{U,k,h}(x) = h^{-1} K_{U,k}(x/h)$. Note that the functions $K_{U,k}$ depend on h, even though, to simplify the presentation, we did not indicate this dependence explicitly in the notations.

In what follows, for simplicity, we drop the p index from $\widehat{m}_p^{(\nu)}(x)$. It is also convenient to rewrite (2.3) as

$$\widehat{m}^{(\nu)}(x) = h^{-\nu} \nu! \sum_{k=0}^{p} \widehat{S}^{\nu,k}(x) \widehat{T}_{n,k}(x),$$

where $\widehat{S}^{\nu,k}(x)$ denotes the $(\nu + 1, k + 1)^{\text{th}}$ element of the inverse of the matrix $\widehat{S}_n$.

3 Asymptotic normality

3.1 Conditions

To establish asymptotic normality of our estimator, we need to impose some regularity conditions. Note that these conditions are stronger than those needed to define the

estimator, and to simplify the presentation, we allow overlap of some of the conditions, e.g., compare condition A and condition (B1) below.

As expected because of the unbiasedness of the score, and as we will show precisely, the asymptotic bias of the estimator, defined as the expectation of the limiting distribution of $\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x)$, is exactly the same as in the error-free case. Therefore, exactly the same as in the error-free case, the bias depends on the smoothness of m and f_X, and on the number of finite moments of Y and K. Define $\tau^2(u) = E\big[\{Y - m(x)\}^2 | X = u\big]$. Note that, to simplify the notation, we do not put an index x into the function τ, but it should be obvious that the function depends on the point x where we wish to estimate the curve m. We make the following assumptions:

Condition B:

(B1) K is a real and symmetric kernel such that $\int K(x)\, dx = 1$ and has finite moments of order $2p + 3$;

(B2) $h \to 0$ and $nh \to \infty$ as $n \to \infty$;

(B3) $f_X(x) > 0$ and f_X is twice differentiable such that $\|f_X^{(j)}\|_\infty < \infty$ for $j = 0, 1, 2$;

(B4) m is $p+3$ times differentiable, $\tau^2(\cdot)$ is bounded, $\|m^{(j)}\|_\infty < \infty$ for $j = 0, \ldots, p+3$, and for some $\eta > 0$, $E\big\{|Y_i - m(x)|^{2+\eta} | X = u\big\}$ is bounded for all u.

These conditions are rather mild and, apart from the assumptions on the conditional moments of Y, they are fairly standard in the error-free context. Boundedness of moments of Y are standard in the measurement error context; see Fan and Masry (1992) and Fan and Truong (1993).

The asymptotic variance of the estimator, defined as the variance of the limiting distribution of $\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x)$, differs from the error-free case since, as usual in deconvolution problems, it depends strongly on the type of measurement errors that

contaminate the X-data. Following Fan (1991a,b,c), we consider two categories of errors. An *ordinary smooth error* of order β is such that

$$\lim_{t\to+\infty} t^\beta \phi_U(t) = c \text{ and } \lim_{t\to+\infty} t^{\beta+1}\phi_U'(t) = -c\beta \qquad (3.1)$$

for some constants $c > 0$ and $\beta > 1$. A *supersmooth error* of order $\beta > 0$ is such that

$$d_0|t|^{\beta_0}\exp(-|t|^\beta/\gamma) \le |\phi_U(t)| \le d_1|t|^{\beta_1}\exp(-|t|^\beta/\gamma) \text{ as } |t| \to \infty, \qquad (3.2)$$

with d_0, d_1, γ, β_0 and β_1 some positive constants. For example, Laplace errors, Gamma errors and their convolutions are ordinary smooth, whereas Cauchy errors, Gaussian errors and their convolutions are supersmooth. Depending on the type of the measurement error, we need different conditions on K and U to establish the asymptotic behavior of the variance of the estimator. These conditions mostly concern the kernel function, which we can choose; they are fairly standard in deconvolution problems, and are easy to satisfy. See for example Fan (1991a,b,c) and Fan and Masry (1992). We assume:

Condition O (ordinary smooth case):
$||\phi_U'||_\infty < \infty$ and for $k = 0, \ldots, 2p+1$, $||\phi_K^{(k)}||_\infty < \infty$ and $\int \left[|t|^\beta + |t|^{\beta-1}\right] |\phi_K^{(k)}(t)|\,dt < \infty$ and, for $0 \le k, k' \le 2p$, $\int |t|^{2\beta}|\phi_K^{(k)}(t)| \cdot |\phi_K^{(k')}(t)|\,dt < \infty$.

Condition S (supersmooth case):
ϕ_K is supported on $[-1, 1]$ and, for $k = 0, \ldots, 2p$, $||\phi_K^{(k)}||_\infty < \infty$;

In the sequel, we let $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}_n$, $\mu_j = \int u^j K(u)\,du,$, $\boldsymbol{S} = \left(\mu_{k+\ell}\right)_{0\le k,\ell\le p}$, $\widetilde{\boldsymbol{S}} = \left(\mu_{k+\ell+1}\right)_{0\le k,\ell\le p}$, $\boldsymbol{\mu} = (\mu_{p+1}, \ldots, \mu_{2p+1})^\top$, $\widetilde{\boldsymbol{\mu}} = (\mu_{p+2}, \ldots, \mu_{2p+2})^\top$, and, for any square integrable function g, we define $R(g) = \int g^2$. Finally, we let $S^{i,j}$ denote the $(i+1, j+$

8

$1)^{\text{th}}$ element of the matrix $\boldsymbol{S}^{-1}f_X^{-1}(x)$ and, for c as in (3.1),

$$\boldsymbol{S}^* = \left((-1)^{k'}i^{k'-k}\frac{1}{2\pi c^2}\int |t|^{2\beta}\phi_K^{(k)}(t)\phi_K^{(k')}(t)\,dt\right)_{0\leq k,k'\leq p}.$$

Note that this matrix is always real since its $(k,k')^{\text{th}}$ element is zero when $k+k'$ is odd, and $i^{k'-k} = (-1)^{(k'-k)/2}$ otherwise.

3.2 Asymptotic results

Asymptotic properties of the estimator depend on the type of error that contaminates the data. The theorem below establishes asymptotic normality in the ordinary smooth error case.

Theorem 3.1. *Assume (3.1). Under Conditions A, B and O, if $nh^{2\beta+2\nu+1}\to\infty$ and $nh^{2\beta+4}\to\infty$, we have*

$$\frac{\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x) - \mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\}}{\sqrt{\mathrm{var}\{\widehat{m}^{(\nu)}(x)\}}} \xrightarrow{\mathcal{L}} N(0,1)$$

where $\mathrm{var}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^{\top}\boldsymbol{S}^{-1}\boldsymbol{S}^*\boldsymbol{S}^{-1}\boldsymbol{e}_{\nu+1}\dfrac{(\nu!)^2(\tau^2 f_X) * f_U(x)}{f_X^2(x)nh^{2\beta+2\nu+1}} + o\left(\dfrac{1}{nh^{2\beta+2\nu+1}}\right)$, *and*

(a) if $p-\nu$ is odd

$$\mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^{\top}\boldsymbol{S}^{-1}\boldsymbol{\mu}\,\frac{\nu!}{(p+1)!}m^{(p+1)}(x)h^{p+1-\nu} + o(h^{p+1-\nu});$$

(b) if $p-\nu$ is even

$$\mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^{\top}\boldsymbol{S}^{-1}\widetilde{\boldsymbol{\mu}}\,\frac{\nu!}{(p+2)!}\left[m^{(p+2)}(x) + (p+2)m^{(p+1)}(x)\frac{f_X'(x)}{f_X(x)}\right]h^{p+2-\nu}$$

$$- \boldsymbol{e}_{\nu+1}^{\top}\boldsymbol{S}^{-1}\widetilde{\boldsymbol{S}}\boldsymbol{S}^{-1}\boldsymbol{\mu}\,\frac{\nu!}{(p+1)!}m^{(p+1)}(x)\frac{f_X'(x)}{f_X(x)}h^{p+2-\nu} + o(h^{p+2-\nu}).$$

From this theorem we see that, as usual in nonparametric kernel deconvolution estimators, the bias of our estimator is exactly the same as the bias of the local

9

polynomial estimator in the error-free case and the errors-in-variables only affect the variance of the estimator. Compare the bias formulae above with Theorem 3.1 of Fan and Gijbels (1996), for example. In particular, our estimator has the design-adaptive property discussed in section 3.2.4 of that book.

The optimal bandwidth is found by the usual trade-off between the squared bias and the variance, which gives $h \sim n^{-1/(2\beta+2p+3)}$ if $p - \nu$ is odd and $h \sim n^{-1/(2\beta+2p+5)}$ if $p - \nu$ is even. The resulting convergence rates of the estimator are, respectively, $n^{-(p+1-\nu)/(2\beta+2p+3)}$ if $p-\nu$ is odd and $n^{-(p+2-\nu)/(2\beta+2p+5)}$ if $p-\nu$ is even. For $p = \nu = 0$, our estimator of m is exactly the estimator of Fan and Truong (1993), and has the same rate as there, that is $n^{-2/(2\beta+5)}$ (remember that we only assume that f_X is twice differentiable). For $p = 1$ and $\nu = 0$, our estimator is different from Fan and Truong (1993), but it converges at the same rate. For $p > 1$ and $\nu = 0$, our estimator converges at faster rates.

Remark 3.1. *Which p should one use?* This problem is essentially the same as in the error-free case (see Fan and Gijbels, 1996, section 3.3). In particular, although, in theory using higher values of p reduces the asymptotic bias of the estimator without increasing the order of its variance, the theoretical improvement for $p - \nu > 1$ is not generally noticeable in finite samples. In particular, the constant term of the dominating part of the variance can increase rapidly with p. However, the improvement from $p - \nu = 0$ to $p - \nu = 1$ can be quite significant, especially in cases where f_X or m are discontinuous at the boundary of their domain, in which case the bias for $p - \nu = 1$ is of smaller order than the bias for $p - \nu = 0$. In other cases, the biases of the estimators of orders p and $p + 1$, where $p - \nu$ is even, are of the same order. See sections 3.3 and 5.

Remark 3.2. *Higher order deconvolution kernel estimators.* As usual, it is possible to reduce the order of the bias by using higher order kernels, i.e., kernels that have

10

all their moments up to order k, say, vanishing, and imposing the existence of higher derivatives of f_X and m, as was done in Fan and Truong (1993). However, such kernels are not very popular, because it is well known that, in practice, they increase the variability of estimators and can make them quite unattractive. See for example Marron and Wand (1992). Similarly, one can also use the infinite order sinc kernel, which has the appealing theoretical property that it adapts automatically to the smoothness of the curves, see Diggle and Hall (1993) and Comte and Taupin (2007). However, the trick of the sinc kernel does not apply to the boundary setting. In addition, this kernel can only be used when $p = \nu = 0$, since $\phi_K^{(\ell)} \equiv 0 \; \forall \ell > 0$, where it can sometimes work poorly in practice, especially in cases where f_X or m have boundary points, see section 5 for illustration.

The next theorem establishes asymptotic normality in the supersmooth error case. In this case, the variance term is quite complicated and asymptotic normality can only be established under a technical condition, which generalizes Condition 3.1 (and Lemma 3.2) of Fan and Masry (1992). This condition is nothing but a refined version of the Lyapounov condition, and it essentially says that the bandwidth can not converge to zero too fast. It should be possible to derive more specific lower bounds on the bandwidth, but this would require considerable technical detail and therefore will be omitted here. In the next theorem we first give an expression for the asymptotic bias and variance of $\widehat{m}^{(\nu)}(x)$, defined as, respectively, the expectation and the variance of $h^{-\nu}\nu! Z_n$, the asymptotically dominating part of the estimator, and where Z_n is defined in (A.3). Then, under the additional assumption, we derive asymptotic normality.

Theorem 3.2. *Assume* (3.2). *Under conditions A, B and S, if* $h = d(2/\gamma)^{1/\beta}(\ln n)^{-1/\beta}$ *with* $d > 1$, *then*

11

(i) $\mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\}$ *is as in Theorem 3.1 and* $\mathrm{var}\{\widehat{m}^{(\nu)}(x)\} = o(\mathrm{Bias}^2\{\widehat{m}^{(\nu)}(x)\});$

(ii) *In addition, if, for* $U_{n,1}$ *defined in the appendix at equation* (A.5), *there exists* $r > 0$ *such that, for* $b_n = h^{\beta/(2r+10)}$, $E[U_{n,1}^2] \geq C_1^2 f_X^{-2}(x) h^{2\beta_0 \cdot 1\{\beta_0 < 1/2\} - 1} \exp\left(\frac{2 - 4\beta b_n}{h^\beta \gamma}\right)$, *with* $0 < C_1 < \infty$ *independent of* n, *then we also have*

$$\frac{\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x) - \mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\}}{\sqrt{\mathrm{var}\{\widehat{m}^{(\nu)}(x)\}}} \xrightarrow{\mathcal{L}} N(0,1).$$

When $h = d(2/\gamma)^{1/\beta}(\ln n)^{-1/\beta}$ with $d > 1$, as in the theorem, it is not hard to see that, as usual in the supersmooth error case, the variance is negligible compared to the squared bias and the estimator converges at the logarithmic rate $\{\log(n)\}^{-(p+1-\nu)/\beta}$ if $p - \nu$ is odd, and $\{\log(n)\}^{-(p+2-\nu)/\beta}$ if $p - \nu$ is even. Again, for $p = \nu = 0$, our estimator is equal to the estimator of Fan and Truong (1993) and thus has the same rate.

3.3 Behavior near the boundary

Since the bias of our estimator is the same as in the error-free case, it suffers from the same boundary effects when the design density f_X is compactly supported. Without loss of generality, suppose that f_X is supported on $[0, 1]$ and, for any integer $k \geq 0$ and any function g defined in $[0, 1]$ which is k times differentiable on $]0, 1[$, let $\widetilde{g}^{(k)}(x) = g^{(k)}(0^+) \cdot 1_{\{x=0\}} + g^{(k)}(1^-) \cdot 1_{\{x=1\}} + g^{(k)}(x) \cdot 1_{\{0<x<1\}}$. We derive asymptotic normality of the estimator under the following conditions, which are the same as those usually imposed in the error-free case:

Condition C:

(C1)–(C2) Same as (B1)–(B2);

(C3) $f_X(x) > 0$ for $x \in]0, 1[$ and f_X is twice differentiable such that $\|f_X^{(j)}\|_\infty < \infty$ for $j = 0, 1, 2$;

(C4) m is $p+3$ times differentiable on $]0,1[$, τ^2 is bounded on $[0,1]$ and continuous on $]0,1[$, $||m^{(j)}||_\infty < \infty$ on $[0,1]$ for $j = 0,\ldots,p+3$ and there exists $\eta > 0$ such that $E\{|Y_i - m(x)|^{2+\eta}|X = u\}$ is bounded for all $u \in [0,1]$.

We also define $\mu_k(x) = \int_{-x/h}^{(1-x)/h} x^k K(x)\,dx$, $\boldsymbol{S}_B(x) = \big(\mu_{k+k'}(x)\big)_{0 \le k,k' \le p}$, $\boldsymbol{\mu}(x) = \{\mu_{p+1}(x),\ldots,\mu_{2p+1}(x)\}^\top$ and

$$\boldsymbol{S}_B^*(x) = \left\{\int \widetilde{\tau}^2(x-u+hz)\widetilde{f}_X(x-u+hz)K_{U,k}(z)K_{U,k'}(z)f_U(u)\,du\,dz\right\}_{0 \le k,k' \le p}.$$

For brevity, we only show asymptotic normality in the ordinary smooth error case. Our results can be extended to the supersmooth error case: all our calculations for the bias are valid for supersmooth errors, and the only difference is the variance, which is negligible in that case.

The proof of the next theorem is similar to the proof of Theorem 3.1 and hence is omitted. It can be obtained from the sequence of Lemmas B.10 to B.13 of Delaigle, Fan and Carroll (2008). As for Theorem 3.2, a technical condition, which is nothing but a refined version of the Lyapounov condition, is required to deal with the variance of the estimator.

Theorem 3.3. *Assume (3.1). Under Conditions A, C and O, if $nh^{2\beta+2\nu+1} \to \infty$ and $nh^{2\beta+4} \to \infty$ as $n \to \infty$, and $\boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}_B^{-1}(x)\boldsymbol{S}_B^*(x)\boldsymbol{S}_B^{-1}(x)\boldsymbol{e}_{\nu+1} \ge C_1 h^{-2\beta}$ for some finite constant $C_1 > 0$, we have*

$$\frac{\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x) - \mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\}}{\sqrt{\mathrm{var}\{\widehat{m}^{(\nu)}(x)\}}} \xrightarrow{\mathcal{L}} N(0,1)$$

where $\quad \mathrm{Bias}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}_B^{-1}(x)\boldsymbol{\mu}(x)\dfrac{\nu!}{(p+1)!}\widetilde{m}^{(p+1)}(x)h^{p+1-\nu} + o(h^{p+1-\nu})$

and $\quad \mathrm{var}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}_B^{-1}(x)\boldsymbol{S}_B^*(x)\boldsymbol{S}_B^{-1}(x)\boldsymbol{e}_{\nu+1}\dfrac{(\nu!)^2}{\widetilde{f}_X^2(x)nh^{2\nu+1}} + o\Big(\dfrac{1}{nh^{2\beta+2\nu+1}}\Big).$

13

As before, the bias is the same as in the error-free case, and thus all well known results of the boundary problem extend to our context. In particular, the bias of the estimator for $p - \nu$ even is of order $h^{p+1-\nu}$, instead of $h^{p+2-\nu}$ in the case without a boundary, whereas the bias of the estimator for $p - \nu$ odd remains of order $h^{p+1-\nu}$, as in the no-boundary case. In particular, the bias of the estimator for $p - \nu = 0$ is of order h, whereas it is of order h^2 when $p - \nu = 1$. For this reason, local polynomial estimators with $p - \nu$ odd, and in particular with $p - \nu = 1$, are often considered to be more natural.

Note that since, in deconvolution problems, kernels are usually supported on the whole real line (see Delaigle and Hall, 2006), the presence of the boundary can affect every point of the type $x = ch$ or $x = 1 - ch$, with c a finite constant satisfying $0 \leq c \leq 1/h$. For $x = ch$, it can be shown that

$$\text{Bias}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_{B}^{-1}(x)\boldsymbol{\mu}(x) \frac{\nu!}{(p+1)!} m^{(p+1)}(0^{+})h^{p+1-\nu} + o(h^{p+1-\nu}),$$

whereas if $x = 1 - ch$, we have

$$\text{Bias}\{\widehat{m}^{(\nu)}(x)\} = \boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_{B}^{-1}(x)\boldsymbol{\mu}(x) \frac{\nu!}{(p+1)!} m^{(p+1)}(1^{-})h^{p+1-\nu} + o(h^{p+1-\nu}).$$

4 Generalizations

In this section, we show how our methodology can be extended to provide estimators in two important cases: (a) when the measurement error distribution is unknown; and (b) when the measurement errors are heteroscedastic. In the interest of space we focus on methodology and do not give detailed asymptotic theory.

4.1 Unknown measurement error distribution

In empirical applications, it can be unrealistic to assume that the error density is known. However, it is only possible to construct a consistent estimator of m if we are

able to consistently estimate the error density itself. Several approaches for estimating this density f_U have been considered in the nonparametric literature. Diggle and Hall (1993) and Neumann (1997) assume that a sample of observations from the error density is available and estimate f_U nonparametrically from those data. A second approach, applicable when the contaminated observations are replicated, consists in estimating f_U from the replicates. Finally, in some cases, if we have a parametric model for the error density f_U and additional constraints on the density f_X, it is possible to estimate an unknown parameter of f_U without any additional observation, see Butucea and Matias (2005) and Meister (2006).

We give details for the replicated data approach, which is by far the most commonly used. In the simplest version of this model, the observations are a sample of i.i.d. data (W_{j1}, W_{j2}, Y_j), $j = 1, \dots, n$, generated by the model

$$Y_j = m(X_j) + \eta_j, \quad W_{jk} = X_j + U_{jk}, k = 1, 2, \quad \mathrm{E}(\eta_j | X_j) = 0,$$
$$\text{with } X_j \sim f_X \text{ and } U_{jk} \sim f_U, \tag{4.1}$$

where the U_{jk}'s are independent, and independent of the (X_j, Y_j, η_j)'s.

In the measurement error literature, it is often assumed that the error density, $f_U(; \theta)$ is known up to a parameter θ which has to be estimated from the data. For example, if $\theta = \mathrm{var}(U)$, the unknown variance of U, a $\sqrt{n}$ consistent estimator is given by

$$\widehat{\theta} = \{2(n-1)\}^{-1} \sum (W_{j1} - W_{j2} - \overline{W}_1 + \overline{W}_2)^2, \tag{4.2}$$

where $\overline{W}_k = n^{-1} \sum_{j=1}^{n} W_{jk}$. See for example Carroll et al. (2006), equation (4.3). Taking $\phi_U(; \widehat{\theta})$ to be the characteristic function corresponding to $f_U(; \widehat{\theta})$, we can extend our estimator of $m^{(\nu)}$ to the unknown error case by replacing ϕ_U by $\phi_U(; \widehat{\theta})$ everywhere.

In the case where no parametric model for f_U is available, some authors suggest using a nonparametric estimator of f_U. For general settings, see Li and Vuong (1997),

15

Schennach (2004a,b) and Hu and Schennach (2008). In the common case where the error density f_U is symmetric, Delaigle, Hall and Meister (2008) propose to estimate $\phi_U(t)$ by $\widehat{\phi}_U(t) = \left| n^{-1} \sum_{j=1}^n \cos\{it(W_{j1} - W_{j2})\} \right|^{1/2}$. Following the approach they use for the case $p = \nu = 0$, we can extend our estimator of $m^{(\nu)}$ to the unknown error case by replacing ϕ_U by $\widehat{\phi}_U$ everywhere, adding a small positive number to $\widehat{\phi}_U$ when it gets too small. Detailed convergence rates of this approach have been studied by Delaigle, Hall and Meister (2008) in the local constant case ($p = 0$), where they show that the convergence rates of this version of the estimator is the same as that of the estimator with known f_U, as long as f_X is sufficiently smooth relative to f_U. Their conclusion can be extended to our setting.

4.2 Heteroscedastic measurement errors

Our local polynomial methodology can be generalized to the more complicated setting where the errors U_i are not identically distributed. In practice, this could happen when observations have been obtained in different conditions, for example if they were collected from different laboratories. Recent references on this problem include Delaigle and Meister (2007,2008) and Staudenmayer, Ruppert and Buonaccorsi (2008). In this context, the observations are a sample $(W_1, Y_1), \ldots, (W_n, Y_n)$ of i.i.d. observations coming from the model

$$
\begin{aligned}
Y_j = m(X_j) + \eta_j, \quad W_j = X_j + U_j, \quad \mathrm{E}(\eta_j | X_j) = 0, \\
\text{with } X_j \sim f_X \text{ and } U_j \sim f_{U_j},
\end{aligned}
\tag{4.3}
$$

where the U_j's are independent of the (X_j, Y_j)'s. Our estimator cannot be applied directly to such data because there is no common error density f_U, and therefore $K_{U,k}$ is not defined. Rather, we need to construct appropriate individual functions $K_{U_j,k}$

and then replace $\widehat{S}_{n,k}(x)$ and $\widehat{T}_{n,k}(x)$ in the definition of the estimator, by

$$\widehat{S}_{n,k}^{H}(x) = n^{-1}\sum_{j=1}^{n} K_{U_j,k,h}(W_j - x) \text{ and } \widehat{T}_{n,k}^{H}(x) = n^{-1}\sum_{j=1}^{n} Y_j K_{U_j,k,h}(W_j - x), \quad (4.4)$$

where we use the superscript H to indicate that we are treating the heteroscedastic case. As before, we require that $E\{\widehat{S}_{n,k}^{H}(x)|X_1,\dots,X_n\} = S_{n,k}(x)$ and $E\{\widehat{T}_{n,k}^{H}(x)|X_1, \dots, X_n\} = T_{n,k}(x)$.

A straightforward solution would be to define $K_{U_j,k}$ by

$$K_{U_j,k}(x) = i^{-k}\frac{1}{2\pi}\int e^{-itx}\phi_K^{(k)}(t)/\phi_{U_j}(-t/h)\,dt,$$

where ϕ_{U_j} is the characteristic function of the distribution of U_j. However, theoretical properties of the corresponding estimator are generally not good, because the order of its variance is dictated by the least favorable error densities, see Delaigle and Meister (2007, 2008). This problem can be avoided by extending the approach of those authors to our context, that is by taking

$$K_{U_j,k}(x) = i^{-k}\frac{1}{2\pi}\int e^{-itx}\frac{\phi_{U_j}(-t)\phi_K^{(k)}(t)}{n^{-1}\sum_{j=1}^{n}|\phi_{U_j}(t/h)|^2}\,dt.$$

Alternatively, in the case where the error densities are unknown but replicates are available as at (4.1) we can use instead

$$\widehat{S}_{n,k}^{H}(x) = n^{-1}\sum_{j=1}^{n} K_{U_j,k,h}(\overline{W}_j - x) \text{ and } \widehat{T}_{n,k}^{H}(x) = n^{-1}\sum_{j=1}^{n} Y_j K_{U_j,k,h}(\overline{W}_j - x), \quad (4.5)$$

where $\overline{W}_j = (W_{j1} + W_{j2})/2$ and

$$\tilde{K}_{U_j,k}(x) = i^{-k}\frac{1}{2\pi}\int e^{-itx}\frac{\phi_K^{(k)}(t)}{n^{-1}\sum_{j=1}^{n}e^{it(W_{j1}-W_{j2})/2}}\,dt,$$

adding a small positive number to the denominator when it gets too small. This is a generalization of the estimator of Delaigle and Meister (2008).

17

5 Finite sample properties

5.1 Simulation settings

Comparisons between kernel estimators and other methods have been carried out by many authors in various contexts, with or without measurement errors. One of their major advantages is that they are simple and can be easily applied to problems such as heteroscedasticity (see section 4.2), nonparametric variance or mode estimation and detection of boundaries. See also the discussion in Delaigle and Hall (2008). As for any method, in some cases, kernel methods outperform, and in other cases, are outperformed by, other methods. Our goal is not to re-derive these well known facts, but rather to illustrate the new results of our paper.

We applied our technique for estimating m and $m^{(1)}$ to several examples comprising curves with several local extrema and/or an inflection point, as well as monotonic, convex and/or unbounded functions. To summarize the work we are about to present, our simulations illustrate in finite samples:

- the gain that can be obtained by using a local linear estimator (LLE) in the presence of boundaries, in comparison with a local constant estimator (LCE);
- properties of our estimator of $m^{(1)}$ for $p = 1$ (LPE1) and $p = 2$ (LPE2);
- properties of our estimator when the error variance is estimated from replicates;
- the robustness of our estimator against misspecification of the error density;
- the gain obtained by using our estimators compared to their naive versions (denoted respectively by NLCE, NLLE, NLPE1 or NLPE2), which pretend there is no error in the data;
- the properties of the LCE using the sinc kernel (LCES) in the presence of boundary points.

We considered the following examples: (i) $m(x) = x^3 \exp(x^4/1000) \cos(x)$ and $\eta \sim N(0, 0.6^2)$; (ii) $m(x) = 2x \exp(-10x^4/81)$, $\eta \sim N(0, 0.2^2)$; (iii) $m(x) = x^3$, $\eta \sim N(0, 1.2^2)$; (iv) $m(x) = x^4$, $\eta \sim N(0, 4^2)$. In cases (i) and (ii) we took $X \sim 0.8X_1 + 0.2X_2$, where $X_1 \sim f_{X_1}(x) = 0.1875x^2 1_{[-2,2]}(x)$ and $X_2 \sim U[-1, 1]$. In cases (iii) and (iv) we took $X \sim N(0, 1)$.

In each case considered, we generated 500 samples of various sizes from the distribution of (W, Y), where $W = X + U$ with $U \sim$ Laplace or normal of zero mean, for several values of the noise-signal-ratio $\mathrm{var}(U)/\mathrm{var}(X)$. Except otherwise stated, we used the kernel whose Fourier transform is given by

$$\phi_K(t) = (1 - t^2)^8 \cdot 1_{[-1,1]}(t). \tag{5.6}$$

In order to illustrate the potential gain of using local polynomial estimators without confounding the effect of an estimator with that of the smoothing parameter selection, we used, for each method, the theoretical optimal value of h; that is, for each sample, we selected the value h minimizing the Integrated Squared Error ISE $= \int \{m^{(\nu)}(x) - \widehat{m}^{(\nu)}(x)\}^2 \, dx$, where $\widehat{m}^{(\nu)}$ is the estimator considered.

In the case where $\nu = 0$, a data-driven bandwidth procedure has been developed by Delaigle and Hall (2008). For example, for the LLE of m, N_W and D_W in section 3.1 of that paper are equal to $N_W = \widehat{T}_{n,2}\widehat{S}_{n,0} - \widehat{T}_{n,1}\widehat{S}_{n,1}$ and $D_W = \widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2$, respectively; see also Figure 2 in that paper. Note that the fully automatic procedure of Delaigle and Hall (2008) also includes the possibility of using a ridge parameter in cases where the denominator $D_W(x)$ gets too small. It would be possible to extend their method to cases where $\nu > 0$, by combining their SIMEX idea with data-driven bandwidths used in the error-free case, in much the same way as they combined their SIMEX idea with cross-validation for the case $\nu = 0$. Although not straightforward, this is an interesting topic for future research.
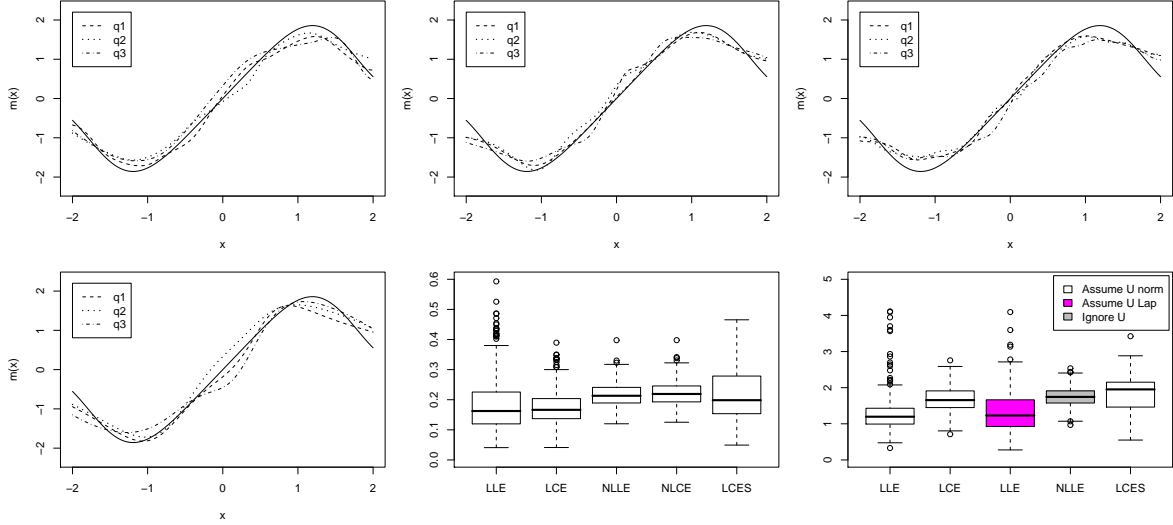
19

Figure 1: Estimated curves for case (ii) when U is Laplace, $\mathrm{var}(U) = 0.2\,\mathrm{var}(X)$ and $n = 500$. Estimators: local linear estimator (LLE, top left), local constant estimator (LCE, top center), the naive local linear estimator that ignores measurement error, (NLLE, top right) and the local constant estimators using the sinc kernel (LCES, bottom left); and box plots of the ISEs (bottom center). Bottom right: Box plots of the ISEs for case (i) when U is normal, $\mathrm{var}(U) = 0.2\,\mathrm{var}(X)$ and $n = 250$; data are averaged replicates and $\mathrm{var}(U)$ is estimated by (4.2).

5.2 Simulation results

In the figures below, we show box plots of the 500 calculated ISEs corresponding to the 500 generated samples. We also show graphs with the target curve (solid line) and three estimated curves (q_1, q_2 and q_3) corresponding to, respectively, the first, second and third quartiles of these 500 calculated ISEs for a given method.

Figure 1 shows the quantile estimated curves of m at (ii) and box plots of the ISEs, for samples of size $n = 500$, when U is Laplace and $\mathrm{var}(U) = 0.2\,\mathrm{var}(X)$. As expected by the theory, the LCE is more biased than the LLE near the boundary. Similarly, the LCES is more biased and variable than the LLE. As usual in measurement error
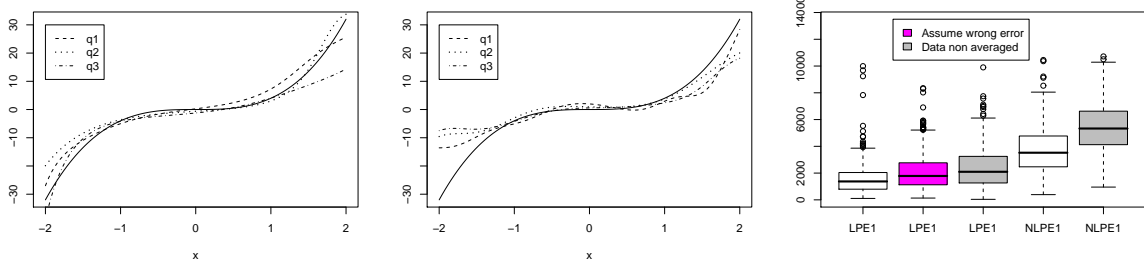
20

Figure 2: Estimates of the density function $m^{(1)}$ for curve (iv) when U is Laplace, $\text{var}(U) = 0.4\,\text{var}(X)$ and $n = 250$, using our local linear method (LPE1, left) and the naive local linear method that ignores measurement error (NLPE1, center), when data are averaged and $\text{var}(U)$ is estimated by (4.2). Right: box plots of ISEs, $\text{var}(U)$ estimated by (4.2). Data averaged except for boxes 3 and 5. Box 2 wrongly assumes Laplace error.

problems, the naive estimators which ignore the error are oversmoothed, especially near the modes and the boundary. The box plots show that the LLE tends to work better than the LCE, but also tends to be more variable. Except in a few cases, both outperform the LCES and the naive estimators.

Figure 1 also shows box plots of ISEs for curve (i) in the case where U is normal, $\text{var}(U) = 0.2\,\text{var}(X)$ and $n = 250$. Here we pretended $\text{var}(U)$ was unknown, but generated replicated data as at (4.1) and estimated $\text{var}(U)$ via (4.2). Since the error variance of the averaged data $\overline{W}_i = (W_{i1} + W_{i2})/2$ is half the original one, we applied each estimator with these averaged data, either assuming U was normal, or wrongly assuming it was Laplace, with unknown variance estimated. We found that the estimator was quite robust again error misspecification, as already noted by Delaigle (2008) in closely connected deconvolution problems. Like there, assuming Laplace distribution often worked reasonably well. See also Meister (2004). Except in a few cases, the LLE worked better than the LCE, and both outperformed the LCES and the NLLE (which itself outperformed the NLCE – not shown here).
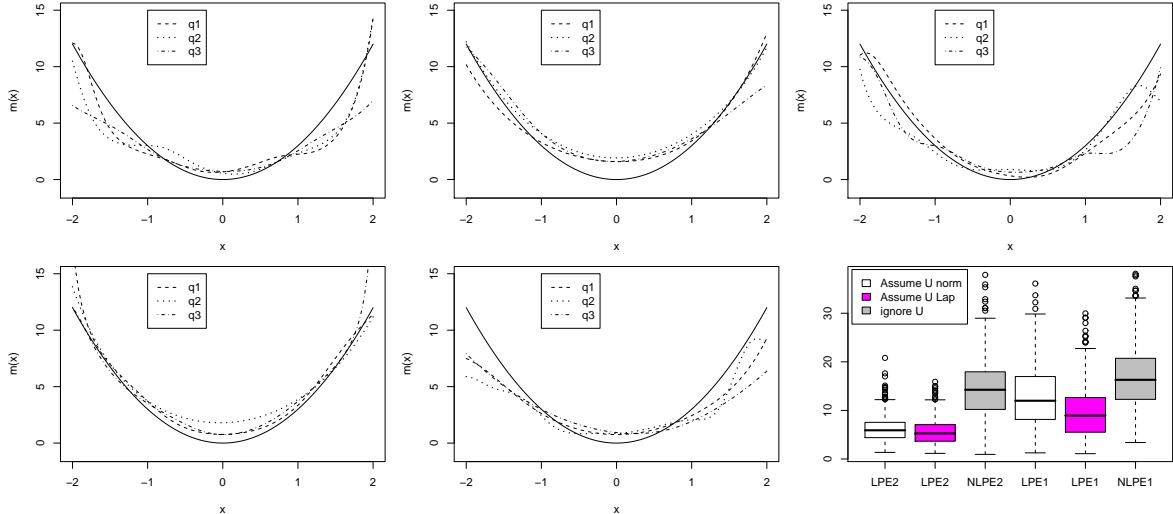
21

Figure 3: Estimates of the derivative function $m^{(1)}$ for case (iii) when U is normal with $\mathrm{var}(U) = 0.4\,\mathrm{var}(X)$ and $n = 500$, using our local linear method (LPE1, top left) and local quadratic method (LPE2, top center) assuming that U is normal, using the LPE1 (top right) and LPE2 (bottom left) wrongly assuming that the error distribution is Laplace, or using the NLPE2 (bottom center). Box plots of the ISEs (bottom right). Data are averaged replicates and $\mathrm{var}(U)$ is estimated by (4.2).

At Figure 2 we show results for estimating the derivative of curve (iv) in the case where U is Laplace, $\mathrm{var}(U) = 0.4\,\mathrm{var}(X)$ and $n = 250$. We assumed the error variance was unknown and we generated replicated data and estimated $\mathrm{var}(U)$ by (4.2). We applied the LPE1 on both the averaged data and the original sample of non averaged replicated data. For the averaged data, the errors distribution is that of a Laplace convolved with itself, and we took either that distribution, or wrongly assumed that the errors were Laplace. We compared our results with the naive NPLE1. Again, taking the error into account worked better than ignoring the error, even when a wrong error distribution was used, and whether we used the original data or the averaged data.

Finally Figure 3 concerns estimation of the derivative function $m^{(1)}$ in case (iii)

when U is normal with $\mathrm{var}(U) = 0.4\,\mathrm{var}(X)$ and $n = 500$. We generated replicates and calculated the LPE1 and LPE2 estimators assuming normal errors or wrongly assuming Laplace errors, and pretended the error variance was unknown and estimated it via (4.2). In this case as well, taking the measurement error into account gave better results than ignoring the errors, even when the wrong error distribution was assumed. The LPE2 worked better at the boundary than the LPE1, but at the interior it is the LPE1 that worked better.

6   Concluding Remarks

In the 20 years since the invention of the deconvoluting kernel density estimator and the 15 years of its use for local constant, Nadaraya-Watson kernel regression, the discovery of a kernel regression estimator for a function and its derivatives that has the same bias properties as in the no-measurement-error case has remained unsolved. By working with estimating equations and using the Fourier domain, we have shown how to solve this problem. The resulting kernel estimators are readily computed, and with the right degrees of the local polynomials, have the design adaptation properties that are so valued in the no-error case.

References

Berry, S., Carroll, R. J. and Ruppert,D. (2002). Bayesian Smoothing and Regression Splines for Measurement Error Problems. *J. Amer. Statist. Assoc.*, **97**, 160–169.

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, **11**, 309–340.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

Carroll, R.J. and Hall, P. (2004). Low-order approximations in deconvolution and regression with errors in variables. *J. Roy. Statist. Soc. Ser.B*, **66**, 31–46.

Carroll, R.J., Maca, J.D. and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika*, **86**, 541–554.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*, 2nd Edn. Chapman and Hall CRC Press, Boca Raton.

Comte, F. and Taupin, M.-L. (2007). Nonparametric estimation of the regression function in an errors-in-variables model. *Statist. Sinica*, **17**, 1065–1090.

Cook, J.R. and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.*, **89**, 1314–1328.

Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica*, **18**, 1025–1045.

Delaigle, A., Fan, J. and Carroll, R.J. (2008). Design-adaptive Local Polynomial Estimator for the Errors-in-Variables Problem. *Long version available from the authors.*

Delaigle, A. and Hall, P. (2006). On the optimal kernel choice for deconvolution. *Statist. and Proba. Letters*, **76**, 1594–1602.

Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Amer. Statist. Assoc.*, **103**, 280–287.

Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, **36**, 665–685.

Delaigle, A., and Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *J. Amer. Statist. Assoc.*, **102**, 1416–1426.

Delaigle, A., and Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli*, **14**, 562–579.

Diggle, P. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc.* Series B, **55**, 523–531.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya A*, **53**, 97–110.

Fan, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica*, **1**, 541–551.

Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272.

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, **21**, 600–610.

Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications*. Chapman & Hall, London.

Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multiv. Anal.*, **43**, 237–271.

Fan, J. and Masry, E. (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian J. Statist.*, **24**, 165–179.

Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.*, **21**, 1900–1925.

Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.*, **35**, 1535-1558.

Hu, Y. and Schennach, S.M. (2008). Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions. *Econometrica*, **76**, 195–216.

Ioannides, D. A. and Alevizos, P. D. (1997). Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.*, **32**, 35–43.

Koo, J-Y. and Lee, K-W. (1998). *B*-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.*, **40**, 57–66.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.*, **65**, 139–165.

Liang, H. and Wang, N. (2005). Large sample theory in a semiparametric partially linear errors-in-variables model. *Statist. Sinica*, **15**, 99–117.

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.*, **20**, 712–736.

Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Can. J. Statist.*, **32**, 439–449.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica*, **16**, 195–211.

Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparam. Statist.*, **7**, 307–330.

Schennach, S.M. (2004a). Estimation of nonlinear models with measurement error. *Econometrica*, **72**, 33–75.

Schennach, S.M. (2004b). Nonparametric regression in the presence of measurement error. *Econometric Theory*, **20**, 1046–1093.

26

Staudenmayer, J. and Ruppert, D. (2004). Local polynomial regression and simulation-extrapolation. *J. Roy. Statist. Soc., B.*, **66**, 17–30

Staudenmayer, J. and Ruppert, D. and Buonaccorsi, J. (2008). Density estimation in the presence of heteroscedastic measurement error. *J. Amer. Statist. Assoc.*, **103**, 726–736.

Stefanski, L. A. (2000). Measurement error models. *J. Amer. Statist. Assoc.*, **95**, 1353–1358.

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169–184.

Stefanski, L.A. and Cook, J.R. (1995). Simulation-extrapolation: The measurement error jackknife. *J. Amer. Statist. Assoc.*, **90**, 1247–1256.

Taupin, M.L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Statist.*, **29**, 66–93.

Zwanzig, S. (2007). On local linear estimation in nonparametric errors-in-variables models. *Theory of Stochastic Processes*, **13**, 316–327.

A   Appendix

A.1   Derivation of the estimator

We have that

$$
\phi_{\{(X_j-x)^k K_h(X_j-x)\}}(t) = \int e^{itx}(X_j - x)^k K_h(X_j - x)\, dx
$$
$$
= h^k e^{itX_j} \int e^{-ithu} u^k K(u)\, du
$$
$$
= i^{-k} h^k e^{itX_j} \phi_K^{(k)}(-ht),
$$

where we used the fact that $\phi_K^{(k)}(t) = i^k \int e^{itu} u^k \, K(u) \, du$. Similarly, we find that

$$\phi_{\{(W_j - x)^k L_{k,h}(W_j - x)\}}(t) = i^{-k} h^k e^{itW_j} \phi_{L_k}^{(k)}(-ht).$$

Therefore, from (2.4) and using $E[e^{itW_j}|X_j] = e^{itX_j} \phi_U(t)$, L_k satisfies

$$\phi_{L_k}^{(k)}(-ht) = \phi_K^{(k)}(-ht)/\phi_U(t).$$

From $\phi_{L_k}^{(k)}(t) = i^k \int e^{itu} u^k \, L_k(u) \, du$ and the Fourier inversion theorem, we deduce that

$$i^k u^k L_k(u) = \frac{1}{2\pi} \int e^{-itu} \phi_{L_k}^{(k)}(t) \, dt = \frac{1}{2\pi} \int e^{-itu} \phi_K^{(k)}(t)/\phi_U(-t/h) \, dt.$$

A.2 Proofs of the results of section 3

We show only the main results and refer to a longer version of this paper, Delaigle, Fan and Carroll (2008), for technical results which are straightforward extensions of results of Fan (1991a), Fan and Masry (1992) and Fan and Truong (1993).

In what follows, we will first give the proofs, referring to detailed Lemmas and Propositions that follow. A longer version with many more details is given by Delaigle, Fan and Carroll (2008).

Before we provide detailed proofs of the theorems, note that, since $\widehat{S}^{\nu,k}, k = 0, \cdots, p$ represents the $(\nu + 1)^{th}$ row of $\widehat{\boldsymbol{S}}_n^{-1}$, we have

$$\sum_{k=0}^p \widehat{S}^{\nu,k} \widehat{S}_{n,k+j} = 0, \text{ if } j \neq \nu \text{ and } 1 \text{ if } j = \nu.$$

Consequently, it is not hard to show that

$$h^\nu (\nu!)^{-1} [\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x)] = \sum_{k=0}^p \widehat{S}^{\nu,k}(x) \widehat{T}_{n,k}^*(x), \tag{A.1}$$

where

$$\widehat{T}_{n,k}^*(x) = \widehat{T}_{n,k}(x) - \sum_{j=0}^p h^j \frac{m^{(j)}(x)}{j!} \widehat{S}_{n,k+j}(x). \tag{A.2}$$

28

Proof of Theorem 3.1. We give the proof in the case where $p - \nu$ odd. The case $p - \nu$ even is treated similarly by replacing everywhere $S^{\nu,k}(x)$ by $S^{\nu,k}(x) - h\breve{S}^{\nu,k}$.

From (A.1) and Lemma A.1, we have

$$h^\nu(\nu!)^{-1}\{\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x)\} = Z_n(x) + O_P(h^{p+3}), \tag{A.3}$$

where

$$Z_n(x) = \sum_{k=0}^{p} S^{\nu,k}(x)\widehat{T}^*_{n,k}(x). \tag{A.4}$$

We deduce from Propositions A.1 and A.2 that the $O_P(h^{p+3})$ term in (A.3) is negligible. Hence, $h^\nu(\nu!)^{-1}[\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x)]$ is dominated by Z_n, and to prove asymptotic normality of $\widehat{m}^{(\nu)}(x)$, it suffices to show asymptotic normality of Z_n. In order to do this, write $Z_n = n^{-1}\sum_{i=1}^{n} U_{n,i}$, where

$$U_{n,i} \equiv P_{n,i} + Q_{n,i}, \tag{A.5}$$

$$P_{n,i} = \sum_{k=0}^{p} S^{\nu,k}(x)\{Y_i - m(x)\}K_{U,k,h}(W_i - x),$$

$$Q_{n,i} = -\sum_{k=0}^{p}\sum_{j=1}^{p} S^{\nu,k}(x)h^j \frac{m^{(j)}(x)}{j!}K_{U,k+j,h}(W_i - x).$$

As in Fan (1991a), to prove that

$$\frac{\sum_{j=1}^{n} U_{n,j} - nEU_{n,j}}{\sqrt{n\,\mathrm{var}\,U_{n,j}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{A.6}$$

it suffices to show that for some $\eta > 0$,

$$\lim_{n\to\infty} \frac{\mathrm{E}\,|U_{n,1}|^{2+\eta}}{n^{\eta/2}[\mathrm{E}\,U_{n,1}^2]^{(2+\eta)/2}} = 0. \tag{A.7}$$

Take η as in (B4) and let $\kappa(u) = E\{|Y - m(x)|^{2+\eta}|X = u\}$. We have

$$E|P_{n,j}|^{2+\eta} = \iint \kappa(v)\Big|\sum_{k=0}^{p} h^{-1}S^{\nu,k}(x)K_{U,k}\Big(\frac{x-u-v}{h}\Big)\Big|^{2+\eta} f_X(v)f_U(u)\,du\,dv$$

$$\leq Ch^{-1-\eta}\max_{0\leq k\leq p}||K_{U,k}||_\infty^\eta \cdot \int |K_{U,k}(u)|^2\,du \leq Ch^{-\beta(2+\eta)-1-\eta},$$

29

from Lemma B.5 of Delaigle, Fan and Carroll (2008), and where, here and below, C denotes a generic positive and finite constant. Similarly, we have

$$E|Q_{n,j}|^{2+\eta} = E\Big| \sum_{k=0}^{p} \sum_{j=1}^{p} S^{\nu,k}(x)h^j \frac{m^{(j)}(x)}{j!} K_{U,k+j,h}(W_i - x) \Big|^{2+\eta} = O(h^{1-\beta(2+\eta)}),$$

and thus $E|U_{n,j}|^{2+\eta} \le Ch^{-\beta(2+\eta)-1-\eta}$.

For the denominator, it follows from Proposition A.2 and Lemma B.7 of Delaigle, Fan and Carroll (2008) that

$$E[U_{n,j}^2] = h^{-2\beta-1} f_X^{-2}(x)(\tau^2 f_X) * f_U(x) \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}^{-1} \boldsymbol{S}^* \boldsymbol{S}^{-1} \boldsymbol{e}_{\nu+1} = Ch^{-2\beta-1}\{1 + o(1)\}.$$

We deduce that (A.7) holds and the proof follows from the expressions of $E(U_{n,j})$ and $\mathrm{var}(U_{n,j})$ given in Propositions A.1 and A.2. $\qquad \square$

Proof of Theorem 3.2. Using similar techniques as in the ordinary smooth error case and Lemmas B.8 and B.9 of the more detailed version of this paper (Delaigle, Fan and Carroll, 2008), it can be proved that (A.3) holds in the supersmooth case as well, and under the conditions of the theorem, the $O_P(h^{p+3})$ is negligible. Thus, as in the ordinary smooth error case, it suffices to show that, for some $\eta > 0$, (A.7) holds. With $P_{n,i}$ and $Q_{n,i}$ as in the ordinary smooth error case, we have

$$\begin{aligned} E|P_{n,i}|^{2+\eta} \le& Ch^{-1-\eta} \max_{0 \le k \le p} ||K_{U,k}||_\infty^\eta \cdot \int |K_{U,k}(u)|^2 \, du \\ \le& Ch^{\beta_2(2+\eta)-1-\eta} \exp\{(2+\eta)h^{-\beta}/\gamma\}, \\ E|Q_{n,j}|^{2+\eta} \le& Ch \max_{1 \le k \le 2p} \int |K_{U,k}(y)|^{2+\eta} \, dy \le Ch^{\beta_2(2+\eta)+1} \exp\{(2+\eta)h^{-\beta}/\gamma\}, \end{aligned}$$

with $\beta_2 = \beta_0 \cdot 1\{\beta_0 < 1/2\}$, where, here and below, C denotes a generic finite constant, and where we used Lemma B.9 of Delaigle, Fan and Carroll (2008). It follows that $E[U_{n,i}^{2+\eta}] \le Ch^{\beta_2(2+\eta)-1-\eta} \exp\{(2+\eta)h^{-\beta}/\gamma\}$. Under the conditions of the theorem, we conclude that (A.7) holds for any $\eta > 0$ and (A.6) follows. $\qquad \square$

Lemma A.1. *Under the conditions of Theorem 3.1, suppose that, for all $0 \le k \le 2p$, we have, when $p - \nu$ is odd, $(nh)^{-1/2}\{R(K_{U,k})\}^{1/2} = O(h^{p+1})$ and when $p - \nu$ is even, $(nh)^{-1/2}\{R(K_{U,k})\}^{1/2} = O(h^{p+2})$. Then,*

$$\sum_{k=0}^{p} \widehat{S}^{\nu,k}(x)\widehat{T}^*_{n,k}(x) = \sum_{k=0}^{p} R^{\nu,k}(x)\widehat{T}^*_{n,k}(x) + O_P(h^{p+3}), \tag{A.8}$$

where $R^{\nu,k}(x) = S^{\nu,k}(x)$ if $p - \nu$ is odd and $R^{\nu,k}(x) = S^{\nu,k}(x) - h\breve{S}^{\nu,k}(x)$ if $p - \nu$ is even, and where $\breve{S}^{i,j}$ denotes $f'_X(x)f_X^{-2}(x)(\boldsymbol{S}^{-1}\widetilde{\boldsymbol{S}}\boldsymbol{S}^{-1})_{i,j}$.

Proof. The arguments are an extension of the calculations of Fan and Gijbels (1996), pages 62 and 101–103. We have

$$\widehat{T}^*_{n,k}(x) = E\{\widehat{T}^*_{n,k}(x)\} + O_P\big[\sqrt{\operatorname{var}\{\widehat{T}^*_{n,k}(x)\}}\big].$$

By construction of the estimator, $E[\widehat{T}^*_{n,k}(x)]$ is equal to the expected value of its error-free counterpart $T^*_{n,k}(x) = T_{n,k}(x) - \sum_{j=0}^{p} h^j(j!)^{-1}m^{(j)}(x)\,S_{n,k+j}(x)$, which, by Taylor expansion, is easily found to be

$$\begin{aligned}
E[T^*_{n,k}(x)] =& \frac{m^{(p+1)}(x)}{(p+1)!}f_X(x)\mu_{k+p+1}\,h^{p+1} \\
&+ \Big[\frac{m^{(p+1)}(x)}{(p+1)!}f'_X(x) + \frac{m^{(p+2)}(x)}{(p+2)!}f_X(x)\Big]\mu_{k+p+2}\,h^{p+2} + o(h^{p+2}). \quad (A.9)
\end{aligned}$$

Since K is symmetric, it has zero odd moments, and thus $E[T^*_{n,k}(x)] \asymp h^{p+1}$ if $k + p$ is odd and $E[T^*_{n,k}(x)] \asymp h^{p+2}$ if $k + p$ is even. Moreover,

$$\begin{aligned}
\operatorname{var}\{\widehat{T}^*_{n,k}(x)\} =& \operatorname{var}\Big\{\widehat{T}_{n,k}(x) - m(x)\widehat{S}_{n,k}(x) - \sum_{j=1}^{p} h^j(j!)^{-1}m^{(j)}(x)\widehat{S}_{n,k+j}(x)\Big\} \\
=& O\Big[\operatorname{var}\{\widehat{T}_{n,k}(x) - m(x)\widehat{S}_{n,k}(x)\}\Big] + O\Big[\sum_{j=1}^{p} \operatorname{var}\{\widehat{S}_{n,k+j}(x)\}\Big] \\
=& O\{R(K_{U,k})/(nh)\} + O\Big\{\sum_{j=1}^{p} R(K_{U,j+k})/(nh)\Big\},
\end{aligned}$$

where we used

$$\operatorname{var}[\widehat{T}_{n,k}(x) - m(x)\widehat{S}_{n,k}(x)] \le (nh^2)^{-1}E\big[\{Y - m(x)\}^2 K^2_{U,k}\{(W - x)/h\}\big]$$

31

$$= (nh^2)^{-1} E\Big(E\big[\{Y - m(x)\}^2 K_{U,k}^2 \{(W-x)/h\}\big|X\big]\Big)$$

$$= (nh^2)^{-1} E\Big(\tau^2(X) K_{U,k}^2 \{(W-x)/h\}\big|X\big]\Big)$$

$$= (nh^2)^{-1} \iint \tau^2(y) K_{U,k}^2\{(y+u-x)/h\} f_X(y) f_U(u)\, dy\, du$$

$$\leq (nh)^{-1} \|\tau^2 f_X\|_\infty R(K_{U,k})$$

and results derived in the proof of Lemma B.1 of Delaigle, Fan and Carroll (2008).

Using our calculations above, we see that when $k + p$ is odd,

$$\widehat{T}_{n,k}^*(x) = c_1 h^{p+1} + o_P(h^{p+1})$$

whereas, for $k + p$ even,

$$\widehat{T}_{n,k}^*(x) = c_2 h^{p+2} + o_P(h^{p+2}),$$

where c_1 and c_2 denote some finite non zero constants (depending on x but not on n). Now, it follows from Lemmas B.1 and B.5 of Delaigle, Fan and Carroll (2008) that, under the conditions of the lemma, $\widehat{\boldsymbol{S}} = f_X(x)\boldsymbol{S} + h f_X'(x)\widetilde{\boldsymbol{S}} + O_P(h^2)$. Let $\boldsymbol{I}$ denote the identity matrix. By Taylor expansion, we deduce that

$$\widehat{\boldsymbol{S}}^{-1} = \big\{ f_X(x)\boldsymbol{S} + h f_X'(x)\widetilde{\boldsymbol{S}} \big\}^{-1} + O_P(h^2)$$

$$= (\boldsymbol{I} + h f_X'(x) f_X^{-1}(x)\boldsymbol{S}^{-1}\widetilde{\boldsymbol{S}})^{-1} \boldsymbol{S}^{-1} f_X^{-1}(x) + O_P(h^2)$$

$$= \boldsymbol{S}^{-1} f_X^{-1}(x) - h\boldsymbol{S}^{-1}\widetilde{\boldsymbol{S}}\boldsymbol{S}^{-1} f_X'(x) f_X^{-2}(x) + O_P(h^2).$$

Thus we have $\widehat{S}^{i,j} = S^{i,j} - h\breve{S}^{i,j} + O_P(h^2)$, where, due to the symmetry properties of the kernel, $S^{i,j} = 0$ when $i + j$ is odd, whereas $\breve{S}^{i,j} = 0$ when $i + j$ is even. This concludes the proof. $\qquad\square$

Proposition A.1. *Under Conditions A, B and O, we have, for $p - \nu$ odd*

$$E[\nu! h^{-\nu} Z_n] = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}^{-1} \boldsymbol{\mu}\, \frac{\nu!}{(p+1)!} m^{(p+1)}(x) h^{p+1-\nu} + o(h^{p+2-\nu})$$

and, for $p - \nu$ even,

$$E[\nu! h^{-\nu} Z_n] = \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}^{-1} \widetilde{\boldsymbol{\mu}} \frac{\nu!}{(p+2)!} \Big[(p+2) m^{(p+1)}(x) \frac{f_X'(x)}{f_X(x)} + m^{(p+2)}(x) \Big] h^{p+2-\nu}$$

$$- e_{\nu+1}^\top S^{-1} \widetilde{S} S^{-1} \mu \frac{\nu!}{(p+1)!} m^{(p+1)}(x) \frac{f_X'(x)}{f_X(x)} h^{p+2-\nu} + o(h^{p+2-\nu}).$$

Proof. From (A.4), we have $E[Z_n] = \sum_{k=0}^p R^{\nu,k}(x) E[T_{n,k}^*(x)]$, where $E[T_{n,k}^*(x)]$ is given at (A.9) . It follows that

$$
\begin{aligned}
E[Z_n] = & \sum_{k=0}^p S^{\nu,k}(x) \frac{m^{(p+1)}(x)}{(p+1)!} f_X(x) \mu_{k+p+1} h^{p+1} \\
& + \sum_{k=0}^p S^{\nu,k}(x) \left[\frac{m^{(p+1)}(x)}{(p+1)!} f_X'(x) + \frac{m^{(p+2)}(x)}{(p+2)!} f_X(x) \right] \mu_{k+p+2} h^{p+2} \\
& - \sum_{k=0}^p \breve{S}^{\nu,k}(x) \frac{m^{(p+1)}(x)}{(p+1)!} f_X(x) \mu_{k+p+1} h^{p+2} + o(h^{p+2}).
\end{aligned}
$$

Recall that $S^{\nu,k}(x) = 0$ unless $k + \nu$ is even and $\breve{S}^{\nu,k}(x) = 0$ unless $k + \nu$ is odd, and write $k + p = (k + \nu) + (p - \nu)$. If $k + \nu$ is even and $p - \nu$ is odd, or if $k + \nu$ is odd and $p - \nu$ is even, then $k + p$ is odd and thus $\mu_{k+p+2} = 0$. If $k + \nu$ is odd and $p - \nu$ is odd, or if $k + \nu$ is even and $p - \nu$ is even, then using similar arguments we find $\mu_{k+p+1} = 0$. $\qquad\square$

Proposition A.2. *Under Conditions A, B and O, we have*

$$\operatorname{var}(\nu! h^{-\nu} Z_n) = e_{\nu+1}^\top S^{-1} S^* S^{-1} e_{\nu+1} \frac{(\nu!)^2 (\tau^2 f_X) * f_U(x)}{f_X^2(x) n h^{2\beta+2\nu+1}} + o\left(\frac{1}{n h^{2\beta+2\nu+1}} \right).$$

Proof. Let U_n as in the proof of Theorem 3.1. We have

$$\operatorname{var}(U_{n,i}) = \operatorname{var}(P_{n,i}) + \operatorname{var}(Q_{n,i}) + 2 \operatorname{cov}(P_{n,i}, Q_{n,i}).$$

We split the proof in three parts.

(*i*) To calculate $\operatorname{var}(P_{n,i})$, note that

$$
\begin{aligned}
E[\{Y_i - m(x)\} K_{U,k,h}(W_i - x)] &= \int \{m(x + hv) - m(x)\} v^k K(v) f_X(x + hv) \, dv \\
&= O(h),
\end{aligned}
$$

and, noting that each $K_{U,k}$ is real, we have,

$$E\left[\{Y_i - m(x)\}^2 K_{U,k,h}(W_i - x) K_{U,k',h}(W_i - x) \right]$$

33

$$= \int E\big[\{Y_i - m(x)\}^2 \big| X = v\big] E\big[K_{U,k,h}(W_i - x)K_{U,k',h}(W_i - x)\big| X = v\big] f_X(v) \, dv$$

$$= \iint \tau^2(v) K_{U,k,h}(v + u - x) K_{U,k',h}(v + u - x) f_U(u) f_X(v) \, du \, dv$$

$$= h^{-1} \iint \tau^2(x + hz - u) K_{U,k}(z) K_{U,k'}(z) f_U(u) f_X(x + hz - u) \, du \, dz$$

$$= h^{-2\beta-1}(\tau^2 f_X) * f_U(x)(-1)^{k'} i^{-k-k'} \frac{1}{2\pi c^2} \int |t|^{2\beta} \phi_K^{(k)}(t) \phi_K^{(k')}(t) \, dt + o(h^{-2\beta-1}),$$

where we used (B.1) of Delaigle, Fan and Carroll (2008), which states that

$$\lim_{n \to \infty} h^{2\beta} \int K_{U,k}(v) K_{U,k'}(v) g(x - hv) \, dv$$

$$= i^{-k-k'}(-1)^{-k'} \frac{g(x)}{c^2} \frac{1}{2\pi} \int |t|^{2\beta} \phi_K^{(k)}(t) \phi_K^{(k')}(t) \, dt$$

with c as in (3.1). Finally

$$\mathrm{cov}\Big[\{Y_i - m(x)\} K_{U,k,h}(W_i - x), \{Y_i - m(x)\} K_{U,k',h}(W_i - x)\Big]$$

$$= h^{-2\beta-1}(\tau^2 f_X) * f_U(x)(-1)^{k'} i^{-k-k'} \frac{1}{2\pi c^2} \int |t|^{2\beta} \phi_K^{(k)}(t) \phi_K^{(k')}(t) \, dt + o(h^{-2\beta-1}),$$

and thus

$$\mathrm{var}(P_{n,i}) = h^{-2\beta-1}(\tau^2 f_X) * f_U(x) \sum_{k,k'=0}^{p} S^{\nu,k}(x) S^{\nu,k'}(x) S_{k,k'}^* + o(h^{-2\beta-1}).$$

Now $(S^{\nu,0}, \ldots, S^{\nu,p}) = \boldsymbol{e}_{\nu+1}^\top S^{-1} f_X^{-1}(x)$, which implies that

$$\mathrm{var}(P_{n,i}) = h^{-2\beta-1} f_X^{-2}(x)(\tau^2 f_X) * f_U(x) \boldsymbol{e}_{\nu+1}^\top \boldsymbol{S}^{-1} \boldsymbol{S}^* \boldsymbol{S}^{-1} \boldsymbol{e}_{\nu+1} + o(h^{-2\beta-1}).$$

(ii) To calculate $\mathrm{var}(Q_{n,i})$, note that $E\{K_{U,k+j,h}(W_i - x)\} = \int v^{k+j} K(v) f_X(x + hv) \, dv = O(1)$ and $|E\{K_{U,k+j,h}(W_i - x) K_{U,k'+j',h}(W_i - x)\}| = O(h^{-2\beta-1})$, by Lemma B.6 of Delaigle, Fan and Carroll (2008), which implies that

$$\mathrm{cov}\{K_{U,k+j,h}(W_i - x), K_{U,k'+j',h}(W_i - x)\} = O(h^{-2\beta-1})$$

and $\mathrm{var}(Q_{n,i}) = O(h^{-2\beta+1})$, which is negligible compared to $\mathrm{var}(P_{n,i})$.

(iii) We conclude from (i) and (ii) that $\mathrm{var}(U_{n,i}) = \mathrm{var}(P_{n,i})\{1 + o(1)\}$, which proves the result. $\qquad \square$

B Supplementary material of the long version of the paper − not for publication

B.1 Ordinary smooth errors

Lemma B.1. *Under Conditions A and B, if $R(K_{U,k})/(nh^3) \to 0$ as $n \to \infty$, we have*

$$\widehat{S}_{n,k}(x) = f_X(x)\mu_k + hf'_X(x)\mu_{k+1} + O_P(h^2) + O_p\{\sqrt{R(K_{U,k})/(nh)}\}.$$

Proof of Lemma B.1. By construction of the estimator, we have

$$E\big[\widehat{S}_{n,k}(x)\big] = \int (u-x)^k h^{-k} K_h(u-x) f_X(u)\, du$$

$$= \int f_X(x+hv) v^k K(v)\, dv$$

$$= f_X(x)\mu_k + hf'_X(x)\mu_{k+1} + O(h^2).$$

Let $V_k(x) = K_{U,k}\{(W-x)/h\}$. Then

$$\mathrm{var}[\widehat{S}_{n,k}(x)] = \frac{1}{nh^2} E[V_k^2(x)] - \frac{1}{nh^2}\{E[V_k(x)]\}^2,$$

where

$$E\big[V_k^2(x)\big] = h \int K_{U,k}^2(u) f_W(x+hu)\, du \le h\|f_W\|_\infty R(K_{U,k}).$$

We deduce that $\mathrm{var}\big[\widehat{S}_{n,k}(x)\big] = O\{R(K_{U,k})/(nh)\} + O(n^{-1})$ and

$$\widehat{S}_{n,k}(x) = E[\widehat{S}_{n,k}(x)] + O_P(\sqrt{\mathrm{var}[\widehat{S}_{n,k}(x)]})$$

$$= f_X(x)\mu_k + hf'_X(x)\mu_{k+1} + O(h^2) + O_P\{\sqrt{R(K_{U,k})/(nh)}\}.$$

$\square$

Note that, it follows from (3.1) that there exists $M > 0$ such that, for all $|t| > M$,

$$\frac{c}{2}|t|^{-\beta} \le |\phi_U(t)| \le 2c|t|^{-\beta}$$

$$\frac{c\beta}{2}|t|^{-\beta-1} \le |\phi'_U(t)| \le 2c\beta|t|^{-\beta-1}.$$

The next six lemmas are useful to prove the main results.

1

Lemma B.2. *Suppose that* $||\phi_K^{(k)}||_\infty < \infty$, $||\phi_K^{(k+1)}||_\infty < \infty$, $\phi_U(t) \neq 0$ *for all* t, $||\phi_U'||_\infty < \infty$ *and* $\int \left[|t|^\beta + |t|^{\beta-1}\right]\left[|\phi_K^{(k+1)}(t)| + |\phi_K^{(k)}(t)|\right] dt < \infty$. *Then, we have*

$$h^{2\beta}|K_{U,k}(x)|^2 \leq c\min(1, |x|^{-2}).$$

Proof. By integration by parts, we have, for all k

$$
\begin{aligned}
K_{U,k}(x) =& i^{-k}\frac{1}{2\pi}\int e^{-itx}\frac{\phi_K^{(k)}(t)}{\phi_U(-t/h)}\,dt \\
=& i^{-k}\frac{1}{2\pi ix}\int e^{-itx}\frac{\phi_K^{(k+1)}(t)}{\phi_U(-t/h)}\,dt + i^{-k}\frac{1}{2\pi ixh}\int e^{-itx}\frac{\phi_K^{(k)}(t)\phi_U'(-t/h)}{\phi_U(-t/h)^2}\,dt \\
\equiv& I + II,
\end{aligned}
$$

where

$$
\begin{aligned}
|I| =& \frac{1}{2\pi|x|}\left|\int e^{-itx}\frac{\phi_K^{(k+1)}(t)}{\phi_U(-t/h)}\,dt\right| \\
\leq& \frac{1}{2\pi|x|}\int_{|t|\leq Mh}\frac{|\phi_K^{(k+1)}(t)|}{|\phi_U(t/h)|}\,dt + \frac{1}{2\pi|x|}\int_{|t|>Mh}\frac{|\phi_K^{(k+1)}(t)|}{|\phi_U(t/h)|}\,dt \\
\leq& \inf_{|t|\leq M}|\phi_U(t)|^{-1}\frac{1}{2\pi|x|}\int_{|t|\leq Mh}|\phi_K^{(k+1)}(t)|\,dt + \frac{1}{2\pi|x|}\int_{|t|>Mh}\frac{|\phi_K^{(k+1)}(t)|}{\frac{c}{2}|t/h|^{-\beta}}\,dt \\
\leq& \frac{c_1}{|x|}\int_{|t|\leq Mh}|\phi_K^{(k+1)}(t)|\,dt + h^{-\beta}\frac{c_2}{|x|}\int_{|t|>Mh}|t|^\beta|\phi_K^{(k+1)}(t)|\,dt \\
\leq& h\frac{c_1}{|x|} + h^{-\beta}\frac{c_2}{|x|} \\
\leq& ch^{-\beta}\frac{c_2}{|x|}.
\end{aligned}
$$

For the second integral we have

$$
\begin{aligned}
|II| =& \frac{1}{2\pi|x|h}\left|\int e^{-itx}\frac{\phi_K^{(k)}(t)\phi_U'(-t/h)}{\phi_U(-t/h)^2}\,dt\right| \\
\leq& \frac{1}{2\pi|x|h}\int_{|t|\leq Mh}\frac{|\phi_K^{(k)}(t)|\cdot|\phi_U'(t/h)|}{|\phi_U(t/h)|^2}\,dt + \frac{1}{2\pi|x|h}\int_{|t|>Mh}\frac{|\phi_K^{(k)}(t)|\cdot|\phi_U'(t/h)|}{|\phi_U(t/h)|^2}\,dt \\
\leq& \inf_{|t|\leq M}|\phi_U(t)|^{-2}||\phi_U'(t)||_\infty\frac{1}{2\pi|x|h}\int_{|t|\leq Mh}|\phi_K^{(k)}(t)|\,dt
\end{aligned}
$$

2

$$+ \frac{1}{2\pi |x| h} \int_{|t|>Mh} \frac{|\phi_K^{(k)}(t)| \cdot 2c\beta |t/h|^{-\beta-1}}{\frac{c^2}{4}|t/h|^{-2\beta}} \, dt$$

$$\leq C|x|^{-1} + Ch^{-\beta}|x|^{-1}$$

$$\leq Ch^{-\beta}|x|^{-1},$$

where, here and below $C > 0$ denotes a generic finite constant. We conclude that

$$h^{2\beta}|K_{U,k}(x)|^2 \leq C|x|^{-2}.$$

On the other hand we also have that

$$|K_{U,k}(x)| \leq \frac{1}{2\pi} \int \frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|} \, dt \leq Ch^{-\beta}.$$

$\square$

Lemma B.3. *Suppose that* $||\phi_K^{(k)}||_\infty < \infty$, $\phi_U(t) \neq 0$ *for all* t, $\int |t|^\beta |\phi_K^{(k)}(t)| \, dt < \infty$. *Then, we have*

$$\lim_{n\to\infty} h^\beta K_{U,k}(x) = i^{-k} \frac{1}{2\pi c} \int e^{-itx} |t|^\beta \phi_K^{(k)}(t) \, dt,$$

Proof. We have

$$h^\beta \left| e^{-itx} \frac{\phi_K^{(k)}(t)}{\phi_U(t/h)} \right| \leq h^\beta \frac{\max_{|t|\leq Mh} |\phi_K^{(k)}(t)|}{\min_{|t|\leq M} |\phi_U(t)|} 1_{\{|t|\leq Mh\}} + \frac{2}{c} |t|^\beta |\phi_K^{(k)}(t)| 1_{\{|t|>Mh\}}$$

$$\leq C \frac{\max_{|t|\leq M} |\phi_K^{(k)}(t)|}{\min_{|t|\leq M} |\phi_U(t)|} 1_{\{|t|\leq M\}} + \frac{2}{c} |t|^\beta |\phi_K^{(k)}(t)|,$$

which is integrable, and where, here and below, $C > 0$ denotes a generic finite constant and with c given in (3.1). We deduce by the Lebesgue dominated convergence theorem that

$$\lim_{n\to\infty} h^\beta K_{U,k}(x) = \lim_{n\to\infty} i^{-k} h^\beta \frac{1}{2\pi} \int e^{-itx} \frac{\phi_K^{(k)}(t)}{\phi_U(-t/h)} \, dt$$

$$= i^{-k} \frac{1}{2\pi} \int e^{-itx} \lim_{n\to\infty} \frac{|t|^\beta \phi_K^{(k)}(t)}{(|t|/h)^\beta \phi_U(-t/h)} 1_{\{|t|>Mh\}} \, dt$$

$$= i^{-k} \frac{1}{2\pi c} \int e^{-itx} |t|^\beta \phi_K^{(k)}(t) \, dt.$$

$\square$

3

Lemma B.4. *Assume that the conditions of the above two lemmas are satisfied for k and k' and let $K_n(x) = h^{2\beta} K_{U,k}(x) K_{U,k'}(x)$ and g a bounded function. Then we have at any continuity point x of g*

$$\lim_{n\to\infty} h^{2\beta} \int K_{U,k}(v) K_{U,k'}(v) g(x - hv)\, dv$$

$$= i^{-k-k'} (-1)^{-k'} \frac{g(x)}{c^2} \frac{1}{2\pi} \int |t|^{2\beta}\, phi_K^{(k)}(t) \phi_K^{(k')}(t)\, dt \qquad (B.1)$$

with c as in (3.1).

Proof. It follows from Lemma B.3 that

$$\lim_{n\to\infty} K_n(x) = i^{-k-k'} \frac{1}{4\pi^2 c^2} \int e^{-itx} |t|^\beta \phi_K^{(k)}(t)\, dt \int e^{-itx} |t|^\beta \phi_K^{(k')}(t)\, dt.$$

We also have, from the last line of the proof of Lemma B.2 that

$$\sup_n |K_n(x)| \le K^*(x)$$

where $K^*(x) = C \min(1, |x|^{-2})$ is such that $\int K^*(x)\, dx < \infty$ and $\lim_{x\to\infty} |xK^*(x)| = 0$, and C denotes a generic finite constant. Therefore, from Lemma 2.1 of Fan (1991a), we have, at any point x of continuity of g

$$\lim_{n\to\infty} h^{2\beta} \int K_{U,k}(v) K_{U,k'}(v) g(x - hv)\, dv$$

$$= i^{-k-k'} \frac{g(x)}{c^2} \int \left(\frac{1}{2\pi} \int e^{-itx} |t|^\beta \phi_K^{(k)}(t)\, dt \right) \left(\frac{1}{2\pi} \int e^{-isx} |s|^\beta \phi_K^{(k')}(s)\, ds \right) dx$$

$$= i^{-k-k'} (-1)^{-k'} \frac{g(x)}{c^2} \frac{1}{2\pi} \int |t|^{2\beta} \phi_K^{(k)}(t) \phi_K^{(k')}(t)\, dt.$$

$\square$

Lemma B.5. *Suppose that $\|\phi_K^{(k)}\|_\infty < \infty$, $\phi_U(t) \ne 0$ for all t, $\int |t|^\beta |\phi_K^{(k)}(t)|\, dt < \infty$ and $\int |t|^{2\beta} |\phi_K^{(k)}(t)|^2\, dt < \infty$. Then, we have*

$$\|K_{U,k}\|_\infty \le C h^{-\beta} \quad and \quad \int K_{U,k}^2 \le C h^{-2\beta}$$

for some finite constant C.

4

Proof. We have

$$
\begin{aligned}
|K_{U,k}(x)| \leq & \frac{1}{2\pi} \int \frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|} \, dt \\
= & \frac{1}{2\pi} \int_{|t| \leq Mh} \frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|} \, dt + \frac{1}{2\pi} \int_{|t| > Mh} \frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|} \, dt \\
\leq & \inf_{|t| \leq M} |\phi_U(t)|^{-1} \frac{1}{2\pi} \int_{|t| \leq Mh} |\phi_K^{(k)}(t)| \, dt + \frac{1}{2\pi} \int_{|t| > Mh} \frac{|\phi_K^{(k)}(t)|}{\frac{c}{2}|t/h|^{-\beta}} \, dt \\
\leq & C \int_{|t| \leq Mh} |\phi_K^{(k)}(t)| \, dt + Ch^{-\beta} \int_{|t| > Mh} |t|^{\beta} |\phi_K^{(k)}(t)| \, dt \\
\leq & Ch + Ch^{-\beta} \\
\leq & Ch^{-\beta},
\end{aligned}
$$

with c as in (3.1) and where C denotes some generic finite constant. The proof for the other result is similar and requires application of Parseval's theorem. $\qquad\square$

Lemma B.6. *Suppose that, for $j = k$ and k', $\|\phi_K^{(j)}\|_\infty < \infty$, $\phi_U(t) \neq 0$ for all t, and $\int |t|^{2\beta} |\phi_K^{(j)}(t)|^2 \, dt < \infty$. Then, we have*

$$
\left| \int K_{U,k}(v) K_{U,k'}(v) \, dv \right| \leq Ch^{-2\beta}
$$

for some finite constant C.

Proof. Follows from Lemma B.5, if we note that

$$
\left| \int K_{U,k}(v) K_{U,k'}(v) \, dv \right|^2 \leq \int |K_{U,k}(v)|^2 \, dv \int |K_{U,k'}(v)|^2 \, dv = O(h^{-4\beta}).
$$

$\qquad\square$

Lemma B.7. *Suppose that K is real and symmetric and such that, for all $0 \leq k, k' \leq p$, $\int |t|^{2\beta} |\phi_K^{(k)}(t)| \cdot |\phi_K^{(k')}(t)| \, dt < \infty$ and $\left| \sum_{k=0}^{p} i^{-k} S^{\nu+1,k} \phi_K^{(k)}(t) \right|^2 > 0$ on an interval included in $[-1, 1]$. Then, if the error is ordinary smooth of order $\beta > 0$, we have*

$$
e_{\nu+1}^{\top} S^{-1} S^* S^{-1} e_{\nu+1} \geq C,
$$

for some finite positive constant C.

5

Proof. Under the conditions of the theorem, we have

$$\boldsymbol{e}_{\nu+1}^{\top}\boldsymbol{S}^{-1}\boldsymbol{S}^{*}\boldsymbol{S}^{-1}\boldsymbol{e}_{\nu+1} = \sum_{k,k'=0}^{p} S^{\nu+1,k}S^{\nu+1,k'}\frac{(-1)^{k'}}{2\pi c^2 i^{k+k'}}\int |t|^{2\beta}\phi_K^{(k)}(t)\phi_K^{(k')}(t)\,dt$$

$$= \frac{1}{2\pi c^2}\int |t|^{2\beta}\Big|\sum_{k=0}^{p} i^{-k}S^{\nu+1,k}\phi_K^{(k)}(t)\Big|^2 dt$$

$$> 0.$$

$\square$

B.2 Supersmooth errors

Lemma B.8. *Suppose that ϕ_K is supported on $[-1,1]$, $\|\phi_K^{(k)}\|_\infty < \infty$, and $\phi_U(t) \neq 0$ for all t. Then, we have*

$$\|K_{U,k}\|_\infty \leq Ch^{\beta_2}\exp(h^{-\beta}/\gamma) \text{ and } \int K_{U,k}^2 \leq Ch^{2\beta_2}\exp(2h^{-\beta}/\gamma)$$

for some finite constant $C > 0$ and where

$$\beta_2 = \beta_0 \cdot 1\{\beta_0 < 1/2\}. \tag{B.2}$$

Proof. We have, for a constant $M > 0$ large enough,

$$|K_{U,k}(x)| \leq \frac{1}{2\pi}\int_{|t|\leq Mh}\frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|}\,dt + \frac{1}{2\pi}\int_{Mh<|t|\leq 1}\frac{|\phi_K^{(k)}(t)|}{|\phi_U(t/h)|}\,dt$$

$$\leq \inf_{|t|\leq M}|\phi_U(t)|^{-1}\frac{1}{2\pi}\int_{|t|\leq Mh}|\phi_K^{(k)}(t)|\,dt + \frac{1}{2\pi}\int_{Mh<|t|\leq 1}\frac{|\phi_K^{(k)}(t)|}{\frac{d_0}{2}|t/h|^{\beta_0}\exp(-|t/h|^{\beta}/\gamma)}\,dt$$

$$\leq C\int_{|t|\leq Mh}|\phi_K^{(k)}(t)|\,dt + Ch^{\beta_0}\exp(h^{-\beta}/\gamma)\int_{Mh<|t|\leq 1}|t|^{-\beta_0}\,dt,$$

where $C \geq 0$ denotes some generic finite constant, and the conclusion follows from

$$\int_{Mh<|t|\leq 1}|t|^{-\beta_0}\,dt = \begin{cases} C\big[t^{-\beta_0+1}\big]_{Mh}^1 = O(1) & \text{if } \beta_0 < 1 \\ C\big[\ln(t)\big]_{Mh}^1 = O\{\ln(h)\} = O(h^{-\beta_0}) & \text{if } \beta_0 = 1 \\ C\big[t^{-\beta_0+1}\big]_{Mh}^1 = O(h^{-\beta_0+1}) & \text{if } \beta_0 > 1. \end{cases}$$

The proof for the other result is similar and requires Parseval's theorem. $\square$

6

Lemma B.9. *Suppose that ϕ_K is supported on $[-1,1]$, $\phi_U(t) \neq 0$ for all t, and, for $j = k$ and k', $||\phi_K^{(j)}||_\infty < \infty$. Then, we have, for β_2 as in (B.2),*

$$\left| \int K_{U,k}(v) K_{U,k'}(v) \, dv \right| \le ch^{2\beta_2} \exp(2h^{-\beta}/\gamma).$$

Proof. Follows from Lemma B.8, if we note that

$$\left| \int K_{U,k}(v) K_{U,k'}(v) \, dv \right|^2 \le \int |K_{U,k}(v)|^2 \, dv \int |K_{U,k'}(v)|^2 \, dv.$$

$\square$

B.3 Proofs of the results of section 3.3

Let $Z_n = \sum_{k=0}^{p} S^{\nu,k}(x) \widehat{T}_{n,k}^*(x)$, $\widehat{T}_{n,k}^*(x) = \widehat{T}_{n,k}(x) - \sum_{j=0}^{p} h^j (j!)^{-1} \widetilde{m}^{(j)}(x) \, \widehat{S}_{n,k+j}(x)$ and where $S^{i,j}(x)$ denotes the $(i+1, j+1)$th element of $\boldsymbol{S}_B^{-1}(x) f_X^{-1}(x)$.

Lemma B.10. *Under conditions A,C and O, if $\boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_B^{-1}(x) \boldsymbol{S}_B^*(x) \boldsymbol{S}_B^{-1}(x) \boldsymbol{e}_{\nu+1} \ge C_1 h^{-2\beta}$ for some finite constant $C_1 > 0$, we have*

$$\mathrm{var}(\nu! h^{-\nu} Z_n) = \boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_B^{-1}(x) \boldsymbol{S}_B^*(x) \boldsymbol{S}_B^{-1}(x) \boldsymbol{e}_{\nu+1} \frac{(\nu!)^2}{f_X^2(x) n h^{2\nu+1}} + o\left(\frac{1}{n h^{2\beta+2\nu+1}} \right).$$

Proof of Lemma B.10. We calculate the variance of $Z_n = n^{-1} \sum_{i=1}^{n} U_{n,i}$, where we used the notation $U_{n,i} = P_{n,i} + Q_{n,i}$, with

$$P_{n,i} = \sum_{k=0}^{p} S^{\nu,k}(x) \{ Y_i - \widetilde{m}(x) \} K_{U,k,h}(W_i - x)$$

$$Q_{n,i} = -\sum_{k=0}^{p} \sum_{j=1}^{p} S^{\nu,k}(x) h^j \frac{\widetilde{m}^{(j)}(x)}{j!} K_{U,k+j,h}(W_i - x).$$

We have $\mathrm{var}(U_{n,i}) = \mathrm{var}(P_{n,i}) + \mathrm{var}(Q_{n,i}) + 2 \mathrm{cov}(P_{n,i}, Q_{n,i})$. We split the proof into two parts.

(i) To calculate $\mathrm{var}(P_{n,i})$, note that

$$E[\{ Y_i - \widetilde{m}(x) \} K_{U,k,h}(W_i - x)] = h^{-k} \int \{ \widetilde{m}(u) - \widetilde{m}(x) \}(u - x)^k K_h(u - x) \widetilde{f}_X(u) \, du$$

7

$$= \int \{\widetilde{m}(x + hv) - \widetilde{m}(x)\} v^k K(v) \widetilde{f}_X(x + hv)\, dv$$

$$= O(h)$$

and

$$E\Big[\{Y_i - \widetilde{m}(x)\}]^2 K_{U,k,h}(W_i - x) K_{U,k',h}(W_i - x)\Big\}$$
$$= h^{-1} \iint \widetilde{\tau}^2(x + hz - u) K_{U,k}(z) K_{U,k'}(z) f_U(u) \widetilde{f}_X(x + hz - u)\, du\, dv.$$

Finally

$$\mathrm{cov}\Big[\{Y_i - m(x)\} K_{U,k,h}(W_i - x),\, \{Y_i - m(x)\} K_{U,k',h}(W_i - x)\Big]$$
$$= h^{-1} \iint \widetilde{\tau}^2(x + hz - u) K_{U,k}(z) K_{U,k'}(z) f_U(u) \widetilde{f}_X(x + hz - u)\, du\, dv\{1 + o(1)\}$$

and thus

$$\mathrm{var}(P_{n,i}) = h^{-1} \iint \widetilde{\tau}^2(x + hz - u) \Big[\sum_{k=0}^{p} S^{\nu,k}(x) K_{U,k}(z)\Big]^2 f_U(u) \widetilde{f}_X(x + hz - u)\, du\, dv\{1 + o(1)\}.$$

Now $(S^{\nu}, \ldots, S^{\nu+p}) = \boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_B^{-1}(x) \widetilde{f}_X^{-1}(x) + O(h)$, which implies that

$$\mathrm{var}(P_{n,i}) = \widetilde{f}_X^{-2}(x) \boldsymbol{e}_{\nu+1}^{\top} \boldsymbol{S}_B^{-1}(x) \boldsymbol{S}_B^*(x) \boldsymbol{S}_B^{-1}(x) \boldsymbol{e}_{\nu+1}\{1 + o(1)\}.$$

(ii) To calculate $\mathrm{var}(Q_{n,i})$, note that

$$E\{K_{U,k+j,h}(W_i - x)\} = \int v^{k+j} K(v) \widetilde{f}_X(x + hv)\, dv = O(1)$$

and

$$E\{K_{U,k+j,h}(W_i - x) K_{U,k'+j',h}(W_i - x)\} = h^{-1} \int K_{U,k+j}(v) K_{U,k'+j'}(v) f_W(x + hv)\, dv,$$

which implies that

$$\mathrm{cov}\{K_{U,k+j,h}(W_i - x)],\, K_{U,k'+j',h}(W_i - x)\}$$
$$= h^{-1} \int K_{U,k+j}(v) K_{U,k'+j'}(v) f_W(x + hv)\, dv\{1 + o(1)\}$$

8

and

$$\text{var}(Q_{n,i}) = \sum_{k,k'=0}^{p} \sum_{j,j'=1}^{p} S^{\nu,k}(x) S^{\nu,k'}(x) h^{j+j'} \frac{\widetilde{m}^{(j)}(x)}{j!} \frac{\widetilde{m}^{(j')}(x)}{j'!}$$

$$\times \text{cov}\{K_{U,k+j,h}(W_i - x)], K_{U,k'+j',h}(W_i - x)\}$$

$$= \sum_{k,k'=0}^{p} \sum_{j,j'=1}^{p} S^{\nu,k}(x) S^{\nu,k'}(x) h^{j+j'} \frac{\widetilde{m}^{(j)}(x)}{j!} \frac{\widetilde{m}^{(j')}(x)}{j'!}$$

$$\times h^{-1} \int K_{U,k+j}(v) K_{U,k'+j'}(v) f_W(x + hv) \, dv \{1 + o(1)\}$$

$$= O(h^{-2\beta+1}),$$

which is negligible compared to $\text{var}(P_{n,i})$.

$\square$

Lemma B.11. *Under conditions A and C, we have*

$$E[\widehat{T}^*_{n,k}(x)] = \frac{\widetilde{m}^{(p+1)}(x)}{(p+1)!} \widetilde{f}_X(x) \mu_{k+p+1}(x) \, h^{p+1} + o(h^{p+1}).$$

and

$$E(Z_n) = e_{\nu+1}^{\top} \mathbf{S}_B^{-1}(x) \mu(x) \frac{\widetilde{m}^{(p+1)}(x)}{(p+1)!} h^{p+1} + o(h^{p+1}).$$

Proof. Same as in the error-free case, see Fan and Gijbels (1996). $\square$

Lemma B.12. *Under the conditions of Theorem 3.3, suppose that, for all $0 \le k \le 2p$, we have $(nh)^{-1/2}\{R(K_{U,k})\}^{1/2} = O(h^{p+1})$. Then, we have*

$$\sum_{k=0}^{p} \widehat{S}^{\nu,k}(x) \widehat{T}^*_{n,k}(x) = \{\sum_{k=0}^{p} S^{\nu,k}(x) \widehat{T}^*_{n,k}(x)\}\{1 + o_P(1)\}.$$

Proof. Similar to the proof of Lemma A.1, except that, here, for all k, $\widehat{T}^*_{n,k}(x) = c_1 h^{p+1} + o_P(h^{p+1})$, with $c_1 \neq 0$ some finite constant. Therefore,

$$\sum_{k=0}^{p} \widehat{S}^{\nu,k}(x) \widehat{T}^*_{n,k}(x) = \sum_{k=0}^{p} S^{\nu,k}(x) \widehat{T}^*_{n,k}(x) + o_P\{\widehat{T}^*_{n,1}(x)\},$$

9

whereas

$$\sum_{k=0}^{p} S^{\nu,k}(x)\widehat{T}^*_{n,k}(x) = c_2\widehat{T}^*_{n,1}(x) + o_P\{\widehat{T}^*_{n,1}(x)\},$$

with $c_2 \neq 0$ some finite constant, which proves the result. $\qquad\square$

Lemma B.13. *Under Conditions A and C, if $R(K_{U,k})/nh \to 0$, we have*

$$\widehat{S}_{n,k}(x) = \widetilde{f}_X(x)\mu_k(x) + O_P(h) + O_p\{\sqrt{R(K_{U,k})/(nh)}\}.$$

Proof of Lemma B.13. We have

$$\begin{aligned}
E\{\widehat{S}_{n,k}(x)\} &= \int \widetilde{f}_X(x + hv)v^k K(v)\, dv \\
&= \int_{-x/h}^{(1-x)/h} \{\widetilde{f}_X(x) + O(h)\}v^k K(v)\, dv \\
&= \widetilde{f}_X(x)\mu_k(x) + O(h).
\end{aligned}$$

Take $V_k(x) = K_{U,k}\{(x - W)/h\}$. Then we have

$$E[\widehat{S}^2_{n,k}(x)] = \frac{1}{nh^2}E\{V_k^2(x)\} - \frac{1}{nh^2}[E\{V_k(x)\}]^2,$$

where

$$\begin{aligned}
E\{V_k^2(x)\} &= h \int K_{U,k}^2(u) f_W(x + hu)\, du \\
&\leq h \sup_x f_W(x) \int |K_{U,k}(u)|^2\, du \\
&= \frac{h}{2\pi} \sup_x f_W(x) \int \frac{|\phi_K^{(k)}(t)|^2}{|\phi_U(t/h)|^2}\, du \\
&= h\, O\{R(K_{U,k})\}
\end{aligned}$$

and we conclude as in the continuous case. $\qquad\square$

10