I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0101

ESTIMATION OF INTEGRATED SQUARED DENSITY DERIVATIVES FROM A CONTAMINATED SAMPLE

A. DELAIGLE AND I. GIJBELS

http://www.stat.ucl.ac.be

Estimation of integrated squared density derivatives from a contaminated sample*

A. Delaigle and I. Gijbels

Institut de Statistique, Université catholique de Louvain, 20 Voie du Roman Pays, Louvain-la-Neuve, Belgium.

E-mail: gijbels@stat.ucl.ac.be

January 2001

Abstract

In this paper we propose a kernel estimator of integrated squared density derivatives, from a sample that has been contaminated by random noise. We derive asymptotic expressions for the bias and the variance of the estimator and show that the squared bias term dominates the variance term. This coincides with results available in the case of non-contaminated observations. We then discuss the selection of the bandwidth parameter when estimating integrated squared density derivatives based on contaminated data. We propose a data-driven bandwidth selection procedure of the plug-in type and investigate its finite sample performance via a simulation study.

Key words and phrases: characteristic function, deconvolution, errors-in-variables, integrated squared derivatives, kernel density estimation, normal reference rule, plug-in bandwidth selection.

1 Introduction

Estimation of integrated squared density derivatives has important practical use in, for example, the context of kernel density estimation. There, the optimal bandwidth, defined as the bandwidth which minimizes the asymptotic mean integrated squared error of a kernel density estimator, depends on the integrated squared second derivative of the unknown density function. See for example Silverman (1986). The problem of estimating integrated squared density derivatives received already considerable attention in the case

* This research was supported by 'Projet d'Actions de Recherche Concertées' (No. 98/03–217) from the Belgian government.

of non-contaminated samples. See Hall and Marron (1987), Aldershof (1991) and Jones and Sheather (1991), among others, for kernel-based methods. In practice however it may happen that the only data available have been observed with errors. This problem arises in many different fields such as microfluorimetry, electrophoresis and biostatistics. See for example Fan (1991a). See also Mendelsohn and Rice (1982) or Carroll, Ruppert and Stefanski (1995) who deal with analyzing such data. In case the data are corrupted with random error we need an estimator which adapts to this. In this paper we propose a kernel estimator of integrated squared density derivatives, which takes into account the measurement error. The idea of the estimator goes back to Hall and Marron (1987) who studied this estimation problem in case of non-contaminated samples.

We derive asymptotic expressions for the bias and variance of the proposed estimator. The asymptotic behavior of the estimator depends on the type of error distribution, and a distinction should be made between supersmooth densities (such as a normal density) and ordinary smooth densities (such as a Gamma density). This distinction is by now a standard one in kernel density estimation in the deconvolution problem context. Another important finding is that the asymptotic squared bias dominates the asymptotic variance. This corresponds to a similar finding in the case of non-contaminated samples. See Jones and Sheather (1991).

This paper also deals with the practical problem of how to choose the bandwidth when estimating integrated squared density derivatives. A possible approach is to focus on the theoretical bandwidth which minimizes the asymptotic mean squared error of the estimator. Since the asymptotic variance is of lower order than the asymptotic squared bias, this results into minimizing the asymptotic squared bias. We propose a plug-in bandwidth selection procedure and compare its performance with a bandwidth based on refering to a normal density (a so-called normal reference bandwidth selector). The performance of the bandwidth selection procedure and of the estimator of the integrated squared density derivative is evaluated via a simulation study, which reveals that the procedure performs quite well, even in the more difficult cases.

The paper is organized as follows. In Section 2 we introduce our estimator of the integrated squared density derivative based on deconvolving kernel density estimation techniques; we also recall the definition of supersmooth and ordinary smooth error densities. Section 3 contains the main results. There we establish asymptotic expressions for the bias and the variance of the estimator, and discuss choices of the bandwidth parameter. A simulation study investigating the performance of the proposed bandwidth selection procedures as well as the performance of the resulting estimator of the integrated squared density derivative is provided in Section 4. The proofs of the results are defered to the Appendix.

2 The estimator

Suppose we have an i.i.d. sample of size n from a random variable Y of unknown density f_Y, verifying $Y = X + Z$, where X is a random variable distributed according to f_X, and Z is a random variable representing the measurement error, independent of X, and distributed according to a known density f_Z. Then for a density f_X with a continuous rth order derivative function $f_X^{(r)}$ which is square integrable (i.e. $f_X^{(r)} \in L_2$), we are interested in estimating

$$\theta_r = \int \left\{ f_X^{(r)}(x) \right\}^2 dx, \tag{2.1}$$

for a positive integer r.

The estimator we propose is obtained by replacing f_X in (2.1) by its deconvolving kernel estimator, defined by

$$\widehat{f}_X(x) = (nh)^{-1} \sum_{i=1}^{n} K_Z\left(\frac{x - Y_i}{h}; h\right), \tag{2.2}$$

with

$$K_Z(x; h) = (2\pi)^{-1} \int e^{-itx} \varphi_K(t)/\varphi_Z(t/h)\, dt, \tag{2.3}$$

where K is a so-called kernel function, supported on $\mathbb{R}$ and real-valued, with Fourier transform (or characteristic function in case K is a density) $\varphi_K(t) = \int e^{itx} K(x) dx$, $h > 0$ is a smoothing parameter called the bandwidth, depending on n, such that $h \to 0$ as $n \to \infty$, and φ_Z is the characteristic function of Z. Throughout this paper we suppose that for all $t \in \mathbb{R}$, $\varphi_Z(t) \neq 0$. Detailed studies of the estimator (2.2) can be found in Carroll and Hall (1988) and Stefanski and Carroll (1990), among others. The estimator of θ_r is then defined by

$$\widehat{\theta}_r = \int \left\{ \widehat{f}_X^{(r)}(x) \right\}^2 dx \tag{2.4}$$

where

$$\widehat{f}_X^{(r)}(x) = (nh^{r+1})^{-1} \sum_{i=1}^{n} K_Z^{(r)}\left(\frac{x - Y_i}{h}; h\right)$$

and

$$K_Z^{(r)}(x; h) = (2\pi)^{-1} \int (-it)^r e^{-itx} \varphi_K(t)/\varphi_Z(t/h)\, dt, \tag{2.5}$$

Throughout this paper we make the assumption that $K_Z(\cdot; h)$ is supported on $\mathbb{R}$, is real-valued and $|K_Z^{(j)}(\cdot; h)|$ is integrable, for $j = 0, \ldots, r$.

From previous studies on the deconvolving kernel density estimator it is known that the kernel should rather be chosen among densities whose characteristic function has a compact support. See for example Stefanski and Carroll (1990), Wand and Jones (1995), Wand (1998), Delaigle (1999) and Hesse (1999), among others. Although this assumption can easily be relaxed in the case of ordinary smooth densities (see Definition 1 below), we will in this paper restrict to compactly supported φ_K for simplicity. For ease of presentation we also assume the support to be symmetric.

Our estimator is a generalization of the estimator introduced by Hall and Marron (1987) in the context of error-free observations, and studied by Jones and Sheather (1991), among others. Note that the estimator in (2.4) can be written as

$$\begin{aligned}
\widehat{\theta}_r =&(n^2h^{2r+2})^{-1}\sum_{i,j=1}^{n}\int K_Z^{(r)}\left(\frac{x-Y_i}{h};h\right)K_Z^{(r)}\left(\frac{x-Y_j}{h};h\right)\,dx\\
=&(n^2h^{2r+2})^{-1}\sum_{i=1}^{n}\int\left[K_Z^{(r)}\left(\frac{x-Y_i}{h};h\right)\right]^2\,dx\\
&\qquad+(n^2h^{2r+2})^{-1}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int K_Z^{(r)}\left(\frac{x-Y_i}{h};h\right)K_Z^{(r)}\left(\frac{x-Y_j}{h};h\right)\,dx\\
=&(nh^{2r+1})^{-1}R(K_Z^{(r)}(\cdot;h))\\
&\qquad+(n^2h^{2r+2})^{-1}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int K_Z^{(r)}\left(\frac{x-Y_i}{h};h\right)K_Z^{(r)}\left(\frac{x-Y_j}{h};h\right)\,dx, \qquad (2.6)
\end{aligned}$$

since $K_Z^{(r)}(\cdot;h)$ is supported on $\mathbb{R}$ and where for a function g in L_2, $R(g)=\int\{g(u)\}^2\,du$. In case of non-contaminated data, Hall and Marron (1987) proposed to drop the first (non-stochastic) term in an attempt to decrease the bias of the estimator. However the resulting diagonals-out estimator can be negative, whereas the diagonals-in estimator is always positive. Jones and Sheather (1991) also argue that it is preferable to keep this diagonal term. See also Aldershof (1991) and Section 3.

In the next section we propose a method for selecting the bandwidth h, based on the mean squared error of the estimator $\widehat{\theta}_r$. Similar to previous findings in the context of deconvolution problems we will see that the asymptotic behavior of the estimator strongly depends on the error distribution. See for example Carroll and Hall (1988), Stefanski (1990), Stefanski and Carroll (1990) and Fan (1991a,1991b,1992). One distinguishes between two categories of error distributions (see Fan (1991a)):

Definition 1. *The distribution of a random variable Z is said to be*

(i) supersmooth of order β if its characteristic function $\varphi_Z(t)$ satisfies:

$$d_0|t|^{\beta_0}\exp(-|t|^{\beta}/\gamma)\leq|\varphi_Z(t)|\leq d_1|t|^{\beta_1}\exp(-|t|^{\beta}/\gamma)\quad as\ t\to\infty, \qquad (2.7)$$

for some positive constants d_0, d_1, β, γ and constants β_0 and β_1;

(ii) ordinary smooth of order β if its characteristic function $\varphi_Z(t)$ satisfies:

$$d_0|t|^{-\beta} \leq |\varphi_Z(t)| \leq d_1|t|^{-\beta} \quad \text{as } t \to \infty, \tag{2.8}$$

for some positive constants d_0, d_1 and β.

Examples of supersmooth distributions are normal and Cauchy distributions while Gamma and Laplace distributions are ordinary smooth.

3 Mean Squared Error and bandwidth selection

In this section we establish the asymptotic expressions for the bias and the variance of the estimator $\widehat{\theta}_r$. Recall that a kernel function K is said to be of order k (with $k > 0$) if it satisfies

$$\mu_{K,j} \equiv \int x^j K(x)\, dx = \begin{cases} 1 & \text{for } j = 0 \\ 0 & \text{for } j = 1, \cdots, k-1 \\ c & \text{for } j = k, \end{cases}$$

with c some finite constant. In this paper we restrict to the case that k is even. This restriction is not necessary but is introduced for simplicity.

Let $f * g$ denote the convolution between two functions f and g, i.e. $f * g(x) = \int f(x-y)g(y)\, dy$. Further we denote by A_h the quantity $\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\, dt$ and by B_h the quantity $\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\, dt$. Then the following conditions will be needed.

Condition A:

(A1). K is a kth order symmetric kernel, r times differentiable;

(A2). $\int |y|^{k+1}|K(y)|\, dy < \infty$;

(A3). φ_K is supported on $[-B, B]$, with $0 < B < \infty$.

Condition B:

(B1). $\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\, dt$ is of order $o(nh^{2r+1})$;

(B2). $\int |K * K_Z(s;h)|\, ds$ and $\left\{\int |K * K_Z(s;h)|\, ds\right\}^2$ are of order $o(n^{-1}h^{-4r-1}B_h)$;

(B3). $\int |z| \left\{K_Z^{(r)} * K_Z^{(r)}(z;h)\right\}^2 dz$ exists and is of order $O(B_h)$.

Condition C:

(C1). $\{f_X^{(2r)}\}^2 f_Y$ and $|f_X^{(j)}|$ are integrable, for $j = 0, \dots, \max(r+k+1, 2r+1)$;

(C2). $f_X^{(j)}$ is bounded, for $j = 0, \dots, r+k+1$.

Condition D:

(D1). $\lim_{|x|\to\infty} f_X^{(i)}(x) f_X^{(j)}(x) = 0$ for all (i,j) such that $i + j = 2r + k - 1$;

(D2). For all $x \in I\!R$, $f_X^{(i)}(x - hy)K^{(j)}(y)$ is integrable as a function of y, for all (i,j) such that $i + j = r$,

For all $x \in I\!R$, $\lim_{|y|\to\infty} f_X^{(i)}(x - hy)K^{(j)}(y) = 0$, for all (i,j) such that $i + j = r - 1$;

(D3). For all $x \in I\!R$, $f_X^{(r+i)}(x + hy)K_Z^{(j)}(y)$ is integrable as a function of y, for all (i,j) such that $i + j = r$,

For all $x \in I\!R$, $\lim_{|y|\to\infty} f_X^{(r+i)}(x + hy)K_Z^{(j)}(y) = 0$, for all (i,j) such that $i + j = r - 1$.

These conditions look rather involved and are in fact mainly technical. It is possible to weaken them but to the extent of increasing the technical difficulty in deriving the results. Therefore we prefer to keep this set of conditions. Some discussion about the verification of the assumptions is provided in the next section. See Remark 3.

3.1 Asymptotic Mean Squared Error of the estimator

The estimator of θ_r defined in (2.4) requires the choice of the bandwidth parameter h. A possibility is to choose h such that the Mean Squared Error (MSE) of the estimator is minimized. However this quantity is rather involved and hence we focus on minimizing an approximation of the MSE, valid for large samples. Since the MSE can be decomposed into the sum of the squared bias and the variance, we first establish asymptotic expressions of those two quantities. This is done in Propositions 1 and 2. We pay special attention to the particular cases of ordinary smooth and supersmooth error densities f_Z, in Corollaries 1 and 2. Finally, in Theorem 1 we provide an asymptotic expression for the mean squared error of $\widehat{\theta}_r$. The proofs of the results are given in the Appendix.

Proposition 1. *Under Conditions A, C, (B1), (D1) and (D2) we have*

$$\mathrm{Bias}[\widehat{\theta}_r] = \mathrm{ABias}[\widehat{\theta}_r] + O(h^{k+1} + n^{-1}),$$

where

$$\mathrm{ABias}[\widehat{\theta}_r] = (-1)^{k/2} \frac{2h^k}{k!} \mu_{K,k} \theta_{r+k/2} + (2\pi n h^{2r+1})^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\, dt. \quad (3.1)$$

From this the asymptotic behavior of the bias in the case of supersmooth or ordinary smooth error distributions is easily obtained. The result is stated in the following corollary.

Corollary 1. *Under Conditions A, C, (B1), (D1) and (D2),*

(i) if the error is ordinary smooth and $h \sim n^{-\frac{1}{2r+2\beta+k+1}}$*, we have*

$$\mathrm{Bias}[\widehat{\theta}_r] \sim n^{-\frac{k}{2r+2\beta+k+1}}.$$

(ii) if the error is supersmooth, with $r \geq \beta_0$*, and* $h = dB(2/\gamma)^{1/\beta}(\ln n)^{-1/\beta}$ *with* $d > 1$*, we have*

$$\mathrm{Bias}[\widehat{\theta}_r] \sim (\ln n)^{-k/\beta}.$$

An approximation for the variance of $\widehat{\theta}_r$ is given below.

Proposition 2. *Under Conditions A, C, D, (B2) and (B3), we have*

$$\mathrm{var}[\widehat{\theta}_r] = \mathrm{Avar}[\widehat{\theta}_r] + o(n^{-2}h^{-4r-1}B_h),$$

where

$$\mathrm{Avar}[\widehat{\theta}_r] = R\,(f_Y)\,(\pi n^2 h^{4r+1})^{-1} \int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt. \tag{3.2}$$

The asymptotic behavior of the variance in the case of supersmooth or ordinary smooth error distributions is summarized in the next corollary.

Corollary 2. *Under Conditions A, C, D, (B2) and (B3),*

(i) if the error is ordinary smooth, and for h *as in Corollary 1(i), we have*

$$\mathrm{var}[\widehat{\theta}_r] \sim n^{-\frac{2k+1}{2r+2\beta+k+1}}. \tag{3.3}$$

(ii) if the error is supersmooth, with $r \geq \beta_0$*, and for* h *as in Corollary 1(ii), we have*

$$\mathrm{var}[\widehat{\theta}_r] = o(\ln n)^{-k/\beta}. \tag{3.4}$$

From Corollary 2 we conclude that with an appropriate choice of the bandwidth and in case of supersmooth or ordinary smooth error distributions, the variance of our estimator $\widehat{\theta}_r$ will be of inferior order than the squared bias. This is also expressed in a more general context in the next theorem which is an immediate consequence of Propositions 1 and 2.

Theorem 1. *Under Conditions A to D, and if*

$$\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt \ \ \textit{is of order} \ \ o(n^2h^{4r+2k+1}+h^{-1}A_h^2+nh^{2r+k}A_h),$$

then

$$\mathrm{MSE}[\widehat{\theta}_r]=\mathrm{AMSE}[\widehat{\theta}_r]+O(h^{2k+1})+O(h^kn^{-1})+O\left(A_h(nh^{2r-k})^{-1}\right)+O\left(A_h(n^2h^{2r+1})^{-1}\right),$$

where

$$\begin{aligned}\mathrm{AMSE}[\widehat{\theta}_r]=&\left\{\mathrm{ABias}[\widehat{\theta}_r]\right\}^2\\ =&4h^{2k}(k!)^{-2}\mu_{K,k}^2\theta_{r+k/2}^2+(4\pi^2n^2h^{4r+2})^{-1}\left\{\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt\right\}^2\\ &+(-1)^{k/2}2\mu_{K,k}\theta_{r+k/2}(\pi nh^{2r+1-k}k!)^{-1}\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt.\end{aligned} \tag{3.5}$$

An optimal choice of the bandwidth can then be based on minimization of the asymptotic squared bias. See Section 3.2.

Remark 1. Note that in practice it is not necessary that variance= $o(\mathrm{Bias}^2)$ in order to proceed with a bandwidth selector which minimizes only the squared bias. One just needs the variance to be considerably smaller than the squared bias, so that it makes sense to neglect it in the error criterion considered.

Remark 2. In the particular case of observations without measurement errors, i.e. when $Z=0$, it is easy to see that the expression for the asymptotic bias established in Proposition 1, coincides with Result 1 of Jones and Sheather (1991). For the variance of the estimator, Expression (A.16) in the proof of Proposition 2 becomes

$$\mathrm{var}[\widehat{\theta}_r]=\frac{R(f_X)}{\pi n^2h^{4r+1}}\int t^{4r}|\varphi_K(t)|^4\,dt+4n^{-1}\left\{\int\left[f_X^{(2r)}(u)\right]^2f_X(u)du-\theta_r^2\right\}+o(n^{-2}h^{-4r-1}),$$

which corresponds to Lemma 3.1 (a) of Hall and Marron (1987).

Remark 3. It is quite easy to check Conditions A, C and D, but Condition B is more involved and we will now look into this condition for ordinary smooth errors.
From the proof of Corollary 1, we know that $\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt \sim h^{-2\beta}$ and hence Condition (B1) is verified as soon as $nh^{2\beta+2r+1}\to\infty$ as $n\to\infty$. The bandwidth of Corollary 1(i), for example, satisfies this condition.
Similarly to the proof of Corollary 1(i), one can show that $|K_Z(\cdot;h)|$ and its derivatives are of order $O(h^{-\beta})$. Moreover, under sufficient conditions on K, one can prove that

$\int |K_Z(x;h)|\, dx$ is of order $O(h^{-\beta})$. Since by Condition A, $|K|$ is integrable, we deduce that $\int |K * K_Z(s;h)|\, ds$ exists and is of order $O(h^{-\beta})$. Now, from the proof of Corollary 2(i), $\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\, dt \sim h^{-4\beta}$, and thus Condition (B2) is satisfied as soon as $nh^{4r+2\beta+1} \to 0$ as $n \to \infty$. This again is the case if we consider the bandwidth of Corollary 1(i) with $k < 2r$.

Condition (B3) is satisfied if $\int |z| \left\{K_Z^{(r)} * K_Z^{(r)}(z;h)\right\}^2 dz = O(h^{-4\beta})$, which is easy to verify for a given kernel and error density.

3.2 Choice of the bandwidth

In order to approximate the theoretical optimal bandwidth, i.e. the bandwidth which minimizes the Mean Squared Error of the estimator, we search for the bandwidth h which minimizes the Asymptotic Mean Squared Error (AMSE) provided in expression (3.5). Recall that under the conditions of the theorem, the asymptotic mean squared error reduces to the asymptotic squared bias. Therefore the task is to minimize the latter quantity. A more careful look at the asymptotic bias term in (3.1) shows us that if we choose our kernel K such that $(-1)^{k/2}\mu_{K,k} < 0$, then the first term of (3.1) will be negative. Since the second term in (3.1) is always positive, minimization of the asymptotic (squared) bias results then into finding the bandwidth h for which the first term equals minus the second term in (3.1), i.e. the bandwidth which cancels the asymptotic bias. A similar remark has been made by Jones and Sheather (1991) in the case of error-free observations. The fact that one ends up at searching for h that cancels the asymptotic bias, also explains why it may be preferable to keep the first term of expression (2.6).

An example of a second-order kernel satisfying the condition $(-1)^{k/2}\mu_{K,k} < 0$ is given by

$$K(x) = \frac{48\cos x}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin x}{\pi x^5}\left(2 - \frac{5}{x^2}\right), \tag{3.6}$$

since for this kernel $k = 2$ and $\mu_{K,2} = 6$. See for example Delaigle (1999) for some properties of this kernel. See also Fan (1992), Wand (1998) or Hesse (1999), among others, for use of this kernel.

Note that for an ordinary smooth error density, the bandwidth that cancels the asymptotic bias, will be of the same order than the bandwidth of Corollaries 1(i) and 2(i). This is easily seen from Remark 3 and expression (3.1). Moreover, since, then, the leading term of the bias cancels, the bias will of order $O\left(n^{-\frac{k+1}{2r+2\beta+k+1}}\right)$ as can be seen from Proposition 1. Looking at Proposition 2, we can conclude that the leading term in the asymptotic mean squared error will be the leading term of the variance given in (3.2) and (3.3).

It is more difficult to say something specific for the case of a supersmooth error density. First of all, for a supersmooth error density it is not so easy to check the conditions of Theorem 1, and hence to check theoretically that the variance is negligible compared to the bias. In Section 4 we check in a simulation study that, for the examples considered, it makes sense to drop the asymptotic variance term in the expression for the asymptotic mean squared error. In the case of a supersmooth error though we cannot provide an analytic expression for the bandwidth that cancels the asymptotic bias.

Note that the asymptotic bias expression provided in (3.1) is of no direct practical use since it involves the unknown quantity $\theta_{r+k/2}$. Hence, in order to come up with a practical bandwidth selection procedure we need to estimate $\theta_{r+k/2}$. This quantity can be estimated in a similar way as we estimated θ_r, by using the estimator $\widehat{\theta}_{r+k/2}$ defined in (2.4). This estimator itself requires the choice of a bandwidth, which will be chosen as described above via cancellation of the asymptotic bias. Unfortunately, this will necessitate the estimation of $\theta_{r+2k/2}$, which itself requires the estimation of $\theta_{r+3k/2}$, and so on. Thus in theory this procedure should count an infinite number of stages. In practice one needs to stop at a certain point. More precisely, in practice we stop the iterative procedure at stage ℓ say (to be chosen in advance) and estimate $\theta_{r+(\ell-1)k/2}$ using a simple normal reference bandwidth selector. By this we mean that in order to estimate $\theta_{r+\ell k/2}$ we assume that the density f_X is a normal density, for which the quantity $\theta_{r+\ell k/2}$ reduces to

$$\theta_{r+\ell k/2} = \frac{(2(r+\ell k/2))!}{(2\sigma_X)^{2(r+\ell k/2)+1}(r+\ell k/2)!\pi^{1/2}}. \tag{3.7}$$

See for example Wand and Jones (1995). Hence, a simple estimator for the quantity $\theta_{r+\ell k/2}$ is then given by

$$\widehat{\theta}_{r+\ell k/2} = \frac{(2(r+\ell k/2))!}{(2\widehat{\sigma}_X)^{2(r+\ell k/2)+1}(r+\ell k/2)!\pi^{1/2}}, \tag{3.8}$$

where $\widehat{\sigma}_X^2 = \widehat{\sigma}_Y^2 - \sigma_Z^2$, with $\widehat{\sigma}_Y^2$ the sample variance of Y, calculated from the available observations $Y_1, \cdots, Y_n$. Our ℓ-stage plug-in iterative bandwidth selection procedure reads as follows:

Step 0. Estimate $\theta_{r+\ell k/2}$ via (3.8);

Step 1. Use $\widehat{\theta}_{r+\ell k/2}$ to select the bandwidth for estimating $\theta_{r+(\ell-1)k/2}$;

Step 2. Use $\widehat{\theta}_{r+(\ell-1)k/2}$ to select the bandwidth for estimating $\theta_{r+(\ell-2)k/2}$;

$\vdots$

Step $\ell - 1$. Use $\widehat{\theta}_{r+2k/2}$ to select the bandwidth for estimating $\theta_{r+k/2}$;

Step ℓ. Use $\widehat{\theta}_{r+k/2}$ to select the bandwidth for estimating θ_r, the quantity of interest.

Note that this procedure involves $\ell + 1$ steps. See Section 4 for details on the three-stage bandwidth selection procedure in the case of contaminated data. See also Aldershof

(1991) in the case of non-contaminated data. Note that the 1-stage procedure results into estimating $\theta_{r+k/2}$, the only unknown quantity in the asymptotic bias, via (3.8) assuming a normal density. We will refer to this as the normal reference estimator of the bandwidth, in compliance with what is called the normal reference rule-of-thumb bandwidth selector in the context of kernel density estimation. See, for example, Silverman (1986). Note that the 0-stage procedure is nothing but a purely parametric procedure where θ_r is estimated directly via (3.8), i.e. assuming a normal density. This type of parametric estimation of $\theta_2 = R(f_X'')$ is used when applying the normal reference rule in selecting the smoothing parameter in kernel density estimation. The nonparametric procedures discussed here for estimating θ_2, and more generally θ_r, improve upon this parametric estimation as soon as a normal model is no longer appropriate. Consequently, the use of nonparametric estimation of θ_2 in, for example, bandwidth selection in kernel density estimation will lead to a considerable to big improvement in the actual estimation of the density. See Delaigle and Gijbels (2001b) for a thorough study on this issue.

How to decide which ℓ to choose in the bandwidth selection procedure? A good strategy would be to take ℓ large enough for the 'last' normal reference estimation step to have as little influence as possible on the resulting bandwidth, but an increase of ℓ implies an increase in the variability of the estimator of θ_r, and thus a trade-off has to be made. See Aldershof (1991) for a theoretical study on the choice of ℓ in the case of error-free data. In that context he proves that $\ell = 3$ is a good choice.

In case the conditions of Theorem 1 are not satisfied, and hence one has no guarantee that the asymptotic squared bias is negligible in comparison with the asymptotic variance, one would stick to the bandwidth which minimizes the asymptotic mean squared error. This error equals the sum of the asymptotic squared bias, provided via Proposition 1, and the asymptotic variance given in Proposition 2. One can even opt for this approach in case the conditions of Theorem 1 are satisfied. Note that the asymptotic variance involves the unknown quantity $R(f_Y) = \int f_Y^2$, which can easily be estimated via the Jones and Sheather (1991) kernel estimate based on the observations $Y_1, \cdots, Y_n$, where the bandwidth is selected via an ℓ-stage procedure as above. In conclusion, when focusing on the asymptotic mean squared error, we consider an ℓ-stage bandwidth selection procedure as described above, but now with the bandwidth parameter selected via minimization of the asymptotic mean squared error. The only difference is the additional estimation of $R(f_Y)$ and its associated bandwidth.

4 Simulations

In this section we perform a simulation study, the aim of which is twofold. A first aim is to check empirically the property stated in Theorem 1, i.e. that choosing h to minimize the asymptotic mean squared error or to minimize the asymptotic squared bias is (asymptotically) equivalent. A second aim is to investigate the finite sample performance of the ℓ-stage bandwidth selection procedure described in Section 3. We conducted a far larger simulation study than the one we report on here. When appropriate we comment on results from these non-reported simulations.

Our study was based on three f_X densities, chosen because each reveals some typical features, and hence they represent some typical cases that can be encountered in practice. These densities, in increasing order of estimation difficulty, are:

1. Density #1: $X \sim$ N(0;1), the standard normal density.

2. Density #2: $X \sim \chi^2(8)$, chosen because it is skewed.

3. Density #3: $X \sim 0.5$ N(-3;1)+0.5 N(2;1), chosen for its two clearly separated modes.

From each of those densities, 500 samples of size $n = 50, 100, 200$ and 500 were generated, each of which was then contaminated by a sample from either a N(0;σ_Z^2) or a Laplace(σ_Z) error density. For each configuration, the parameter σ_Z was chosen such that the ratio σ_Z^2/σ_X^2 was equal to 0.25. Note however that the meaning of this ratio can be very different according to the density f_X considered, and hence does not allow to compare the difficulty of the estimation across densities. See also Wand (1998). In our simulation study we took as kernel function K, the second-order kernel defined in (3.6). The conclusions of our simulation study remain valid when using other kernels satisfying the conditions.

Our simulation study focuses on estimation of $\theta_2 = R(f_X'') = \int (f_X'')^2$ since this quantity plays a major role when dealing with bandwidth selection in kernel estimation of a density. See Wand and Jones (1995) or Delaigle and Gijbels (2001a). Although our simulation study restricted to estimation of θ_2, we believe that the conclusions of this study carry over to the estimation of general θ_r.

For each configuration we investigate the finite sample performance of four bandwidth selection procedures: the three-stage and one-stage bandwidth selection procedures, based on either the asymptotic squared bias only or based on the asymptotic mean squared error. The three-stage procedure reads as follows:

Step 0. Estimate θ_5 via (3.8);

Step 1. Use $\widehat{\theta}_5$ to select a bandwidth h_4 for estimating θ_4;

Step 2. Use $\widehat{\theta}_4$ to select a bandwidth h_3 for estimating θ_3;

Step 3. Use $\widehat{\theta}_3$ to select a bandwidth h_2 for estimating θ_2, the quantity of interest,

where h_i is the bandwidth found by minimization of the appropriate error criterion: the asymptotic mean squared error or the asymptotic squared bias. The simpler one-stage procedure consists of only two steps:

Step 0. Estimate θ_3 via (3.8);

Step 1. Use $\widehat{\theta}_3$ to select a bandwidth for estimating θ_2, the quantity of interest,

with a similar remark regarding the selection of the bandwidth. By $\widehat{\theta}_2^{\mathrm{AB},\ell}$ and $\widehat{\theta}_2^{\mathrm{AMSE},\ell}$ we denote the resulting estimator for θ_2 obtained via an ℓ-stage bandwidth selection procedure based on the asymptotic squared bias, respectively the asymptotic mean squared error. So, in our simulation study we compared the performances of $\widehat{\theta}_2^{\mathrm{AB},3}$, $\widehat{\theta}_2^{\mathrm{AMSE},3}$, $\widehat{\theta}_2^{\mathrm{AB},1}$, and $\widehat{\theta}_2^{\mathrm{AMSE},1}$.

We chose to report results on the three-stage selection procedure. As already mentioned in Section 3, increasing ℓ results into a decrease of the bias and an increase of the variance of the estimator $\widehat{\theta}_2$. Our larger simulation study confirmed this. Clearly, a best choice of ℓ depends both on the sample and the density considered. The comparison of the case $\ell = 3$ and the case $\ell = 1$ will already reveal the kind of improvement that can be expected by changing ℓ.

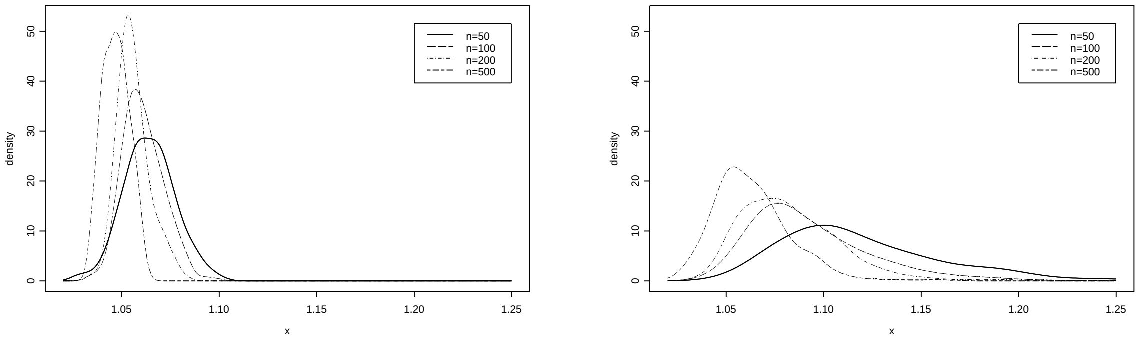

Figure 4.1: Kernel density estimates of the ratio's $\widehat{\theta}_2^{\mathrm{AB},1}/\widehat{\theta}_2^{\mathrm{AMSE},1}$ (left panel) and the ratios's $\widehat{\theta}_2^{\mathrm{AB},3}/\widehat{\theta}_2^{\mathrm{AMSE},3}$ (right panel) for a mixed normal density with Laplace error.

Figure 4.1 represents, for four different sample sizes, kernel density estimators of the ratio's $\widehat{\theta}_2^{\mathrm{AB},3}/\widehat{\theta}_2^{\mathrm{AMSE},3}$ (right panel) and $\widehat{\theta}_2^{\mathrm{AB},1}/\widehat{\theta}_2^{\mathrm{AMSE},1}$ (left panel) based on 500 replications of the estimators for the mixed normal density (density #3) contaminated by a Laplace error. All kernel density estimators in this study were calculated using a standard normal kernel and the normal reference bandwidth $h = 1.06\widehat{\sigma}n^{-1/5}$, where $\widehat{\sigma}$ denotes the standard deviation of the 500 values of the estimate. From Figure 4.1 we can see that as the sample size increases, the distribution of the ratio's $\widehat{\theta}_2^{\mathrm{AB},\ell}/\widehat{\theta}_2^{\mathrm{AMSE},\ell}$, for $\ell = 1, 3$, becomes less widespread and is centered in a point which gets closer to 1. This is not surprising since it just illustrates the theoretical result of Theorem 1, which asserts that for a Laplace error estimating θ_2 with a bandwidth obtained via minimization of the AMSE or of the squared ABias is asymptotically equivalent. See also Remark 3.

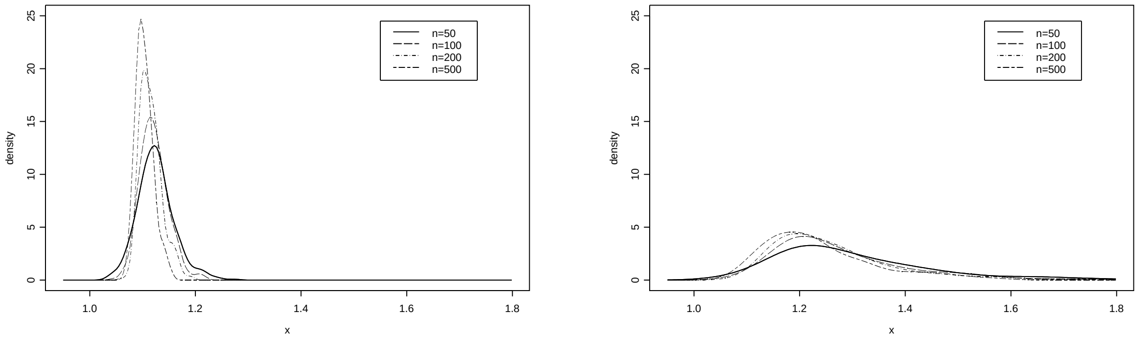

Figure 4.2: Kernel density estimates of the ratio's $\widehat{\theta}_2^{\text{AB},1}/\widehat{\theta}_2^{\text{AMSE},1}$ (left panel) and the ratios's $\widehat{\theta}_2^{\text{AB},3}/\widehat{\theta}_2^{\text{AMSE},3}$ (right panel) for a mixed normal density with gaussian error.

Recall that the conditions of Theorem 1 are difficult to check for supersmooth error densities, such as for example a normal error density. Our simulation results using a normal error will give us some insights here. Figure 4.2 is similar to Figure 4.1, but for the configuration with a normal error. From this figure we observe that even for a normal error distribution, estimation of θ_2 with a bandwidth selected by minimization of the AMSE seems to be asymptotically equivalent to estimation of θ_2 with a bandwidth obtained via minimization of the squared ABias. Note that in case of a normal error density, the asymptotics take place less rapidly. Of course, simulations can in no way constitute the proof of a theoretical assertion, but we believe that in most cases the two methods are indeed asymptotically equivalent, even for a gaussian error density. Similar results were obtained for the two other f_X densities.

Since the AMSE-based method requires more computations and the estimation of an additional quantity ($R(f_Y)$), and since both methods seem to give approximately the same results, we recommend the use of the ABias-based method.

Figure 4.3 shows, for four different sample sizes, kernel density estimates of the ratios's $\widehat{\theta}_2^{\text{AB},3}/\theta_2, \widehat{\theta}_2^{\text{AMSE},3}/\theta_2$, $\widehat{\theta}_2^{\text{AB},1}/\theta_2$ and $\widehat{\theta}_2^{\text{AMSE},1}/\theta_2$ in the case where f_X and f_Z are both gaussian. A good estimator should give rise to a density centered close to 1 and little widespread. Note that even if in this case (f_X gaussian) where the one-stage selection method is clearly the most adapted, the results on the three-stage procedures are very satisfactory. We also did simulations with Laplace errors and got similar results (not reported here): a very minor difference between the performance of the ABias-based and the AMSE-based bandwidth selection, and somewhat less widespread densities for the one-stage bandwidth selection procedure.

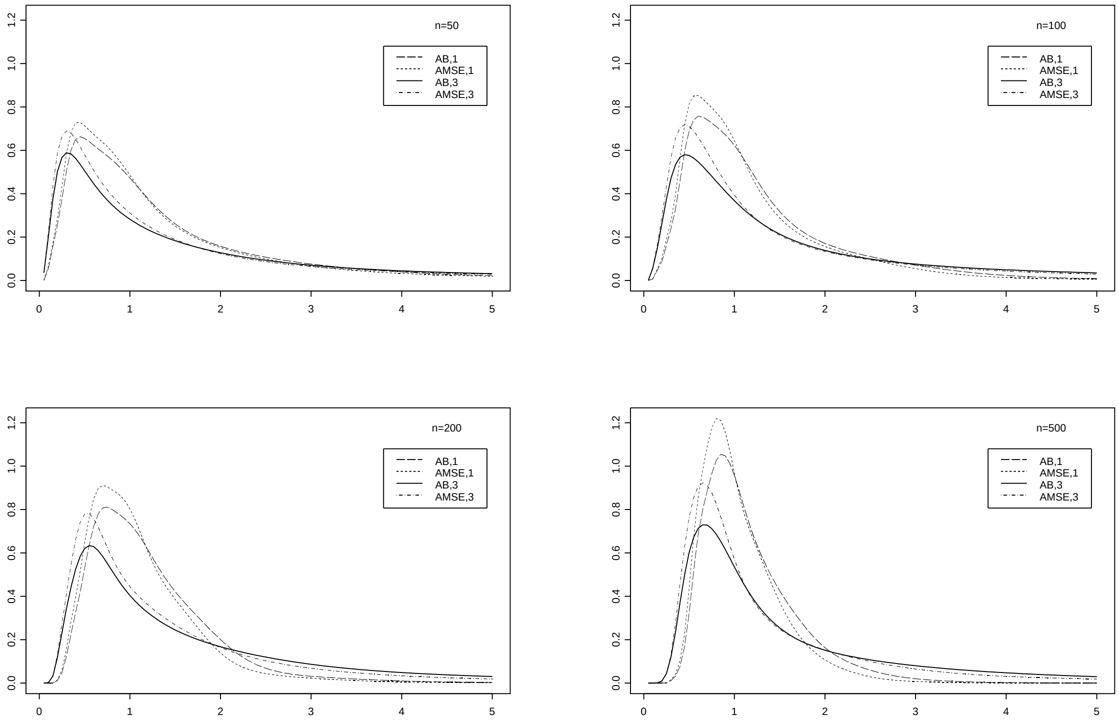

Figure 4.3: Comparison of four estimators $\widehat{\theta}_2^{\mathrm{AB},3}$, $\widehat{\theta}_2^{\mathrm{AMSE},3}$, $\widehat{\theta}_2^{\mathrm{AB},1}$, and $\widehat{\theta}_2^{\mathrm{AMSE},1}$ in case of a normal density f_X and a normal error density.

For the other two densities f_X, estimation based on a three-stage bandwidth selection procedure clearly outperforms estimation based on a one-stage bandwidth selection procedure. This can be seen from Figures 4.4 and 4.5 where we present kernel density estimators of the same four ratios as above. Figure 4.4 depicts the results for the mixed normal f_X density corrupted by a Laplace error density. It confirms that the methods based on the AMSE or the ABias are almost equivalent, but also that the one-stage bandwidth selection procedure performs quite badly. In this case, the three-stage procedure clearly outperforms the one-stage procedure. Note that for this configuration estimation of θ_2 seems to be a quite difficult case. Indeed, even for larger sample sizes (such as for example $n = 500$) the various bandwidth selection methods do not lead to a very good estimator of θ_2 (θ_2 was always underestimated). Similar conclusions, but with an even worse performance, can be drawn in case of a normal error density. Note that because of the particular structure of a mixed density the estimation of θ_2 could certainly be improved by choosing a local bandwidth rather than a global bandwidth. However this would require more computations and would be less simple to implement. Moreover, in general θ_2 is just part of a more complex estimation problem, such as in bandwidth selection for kernel density estimation. Hence we believe that in many cases a reasonably good estimator, such as ours, will do. See Delaigle and Gijbels (2001b) for an illustration of this in the case of density estimation.

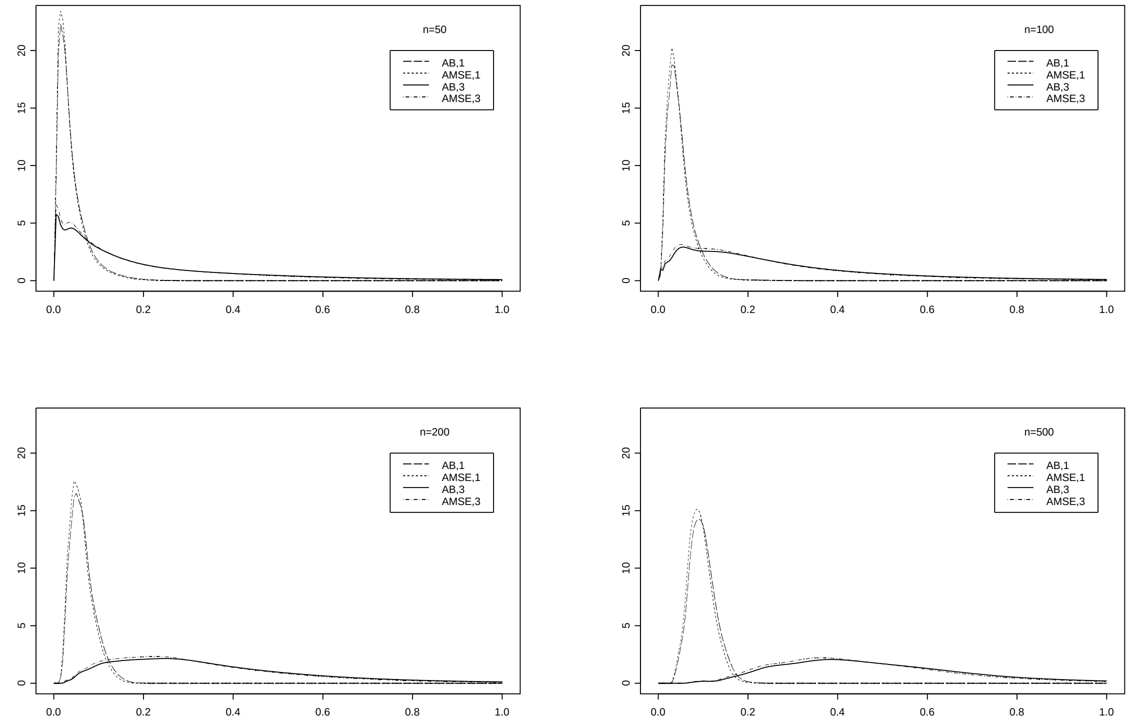

Figure 4.4: Comparison of four estimators $\widehat{\theta}_2^{\mathrm{AB},3}$, $\widehat{\theta}_2^{\mathrm{AMSE},3}$, $\widehat{\theta}_2^{\mathrm{AB},1}$, and $\widehat{\theta}_2^{\mathrm{AMSE},1}$ in case of a mixed normal density f_X and a Laplace error density.

In Figure 4.5 the results are shown for f_X a $\chi^2(8)$ density (density #2) and a gaussian error distribution. Similar remarks as before can be made, but for this configuration all four estimation procedures perform a bit better. This is not really surprising since a chi-squared density presents less particular features than a mixture of two normal densities and thus any f_X related quantity should be easier to estimate by using a global bandwidth. However θ_2 remains quite underestimated, which again reveals the fact that estimation of such a quantity is rather difficult.

In conclusion our simulation results on estimation of θ_2 based on a three-stage bandwidth selection method proved to be quite satisfactory, in comparison with the normal reference-based (one-stage) method. Our simulation study also reveals that dropping the variance term in the asymptotic mean squared error is a sensible thing to do, which confirms our theoretical findings of Section 3. The study also showed that the estimation task is more difficult in case of a supersmooth error density.

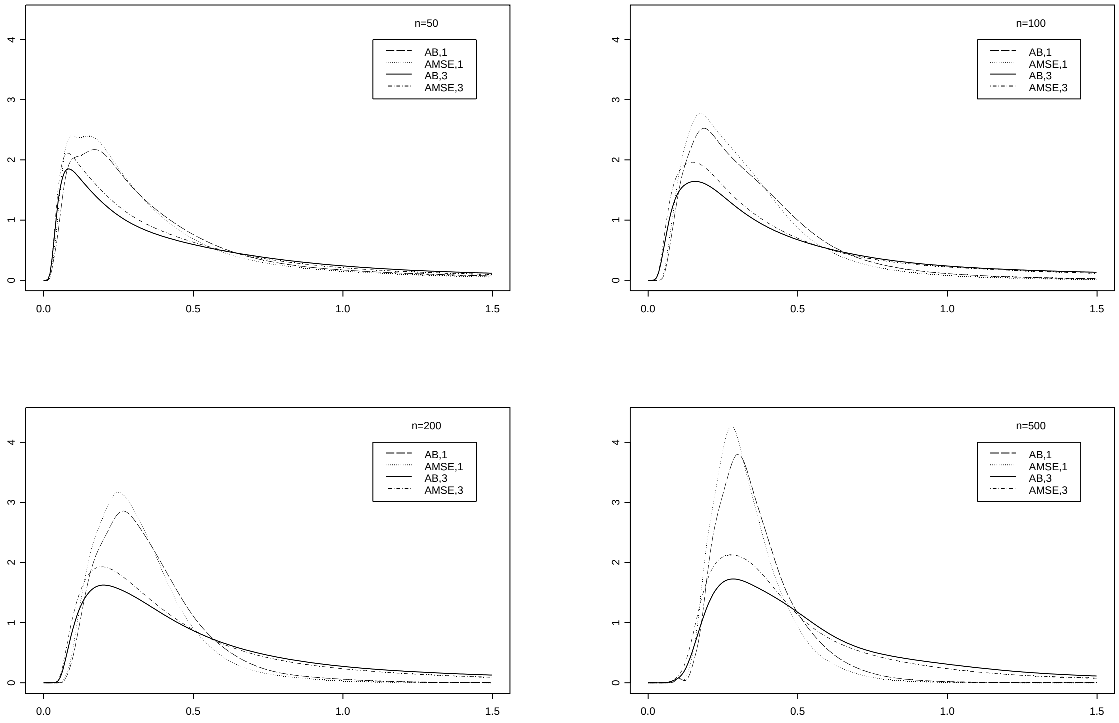

Figure 4.5: Comparison of four estimators $\widehat{\theta}_2^{\mathrm{AB},3}$, $\widehat{\theta}_2^{\mathrm{AMSE},3}$, $\widehat{\theta}_2^{\mathrm{AB},1}$, and $\widehat{\theta}_2^{\mathrm{AMSE},1}$ in case of a chi-squared density f_X and a normal error density.

A Appendix

In this section we provide the proofs for Propositions 1 and 2 and Corollaries 1 and 2. Before giving the proofs we establish some useful facts about various functionals involving the 'kernel' K_Z, defined in (2.3) and appearing in the estimator (2.2). These facts will be used in the proofs of the propositions and corollaries, presented in Section A.2.

A.1 Preliminary results

Lemma A.1. Under Condition (A3),

$$\varphi_{K_Z^{(r)}}(t;h) = (-it)^r \varphi_K(t)/\varphi_Z(t/h).$$

Proof. Recall from (2.5) that under (A3), we can write

$$\begin{aligned} K_Z^{(r)}(u;h) &= (2\pi)^{-1} \int (-it)^r e^{-itu} \varphi_K(t)/\varphi_Z(t/h)\, dt \\ &= (2\pi)^{-1} \int e^{itu} (it)^r \varphi_K(-t)/\varphi_Z(-t/h)\, dt \end{aligned}$$

$$= \quad (2\pi)^{-1}\varphi_g(u),$$

where $g(t) = (it)^r \varphi_K(-t)/\varphi_Z(-t/h)$. Thus $\varphi_g(u) = 2\pi K_Z^{(r)}(u;h)$.

Now by the Fourier inversion Theorem, we can write

$$g(t) = (2\pi)^{-1} \int e^{-itu}\varphi_g(u)\, du = \int e^{-itu} K_Z^{(r)}(u;h)\, du,$$

which is equivalent to $\varphi_{K_Z^{(r)}}(-t;h) = (it)^r \varphi_K(-t)/\varphi_Z(-t/h)$. □

Lemma A.2. Under Condition (A3),

$$R(K_Z^{(r)}(\cdot;h)) = (2\pi)^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\, dt.$$

Proof. Note that under (A3)

$$\begin{aligned} R(K_Z^{(r)}(\cdot;h)) = & \int \left\{ K_Z^{(r)}(u;h) \right\}^2 du = \int K_Z^{(r)}(u;h) K_Z^{(r)}(u;h)\, du \\ =& (2\pi)^{-1} \int \varphi_{K_Z^{(r)}}(t;h) \overline{\varphi}_{K_Z^{(r)}}(t;h)\, dt \\ =& (2\pi)^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\, dt, \end{aligned}$$

where we used Parseval's Theorem and Lemma A.1. □

The following lemma is a generalization of a result of Stefanski and Carroll (1990).

Lemma A.3. Under Conditions (A1) and (A3),

$$\mathrm{E}\left[K_Z^{(r)} \Big(\frac{x-Y}{h}; h \Big) \right] = \mathrm{E}\left[K^{(r)} \Big(\frac{x-X}{h} \Big) \right] \tag{A.1}$$

Proof. We have

$$\begin{aligned} \mathrm{E}\left[K_Z^{(r)} \Big(\frac{x-Y}{h}; h \Big) | X \right] =& (2\pi)^{-1} \int (-it)^r\, \mathrm{E}\left[e^{-it\frac{x-Y}{h}} | X \right] \varphi_K(t)/\varphi_Z(t/h)\, dt \\ =& (2\pi)^{-1} \int (-it)^r E\Big[e^{-it\frac{x-X}{h}} e^{it\frac{Z}{h}} | X \Big] \varphi_K(t)/\varphi_Z(t/h)\, dt \\ =& (2\pi)^{-1} \int (-it)^r e^{-it\frac{x-X}{h}} E\Big[e^{it\frac{Z}{h}} \Big] \varphi_K(t)/\varphi_Z(t/h)\, dt \\ =& (2\pi)^{-1} \int (-it)^r e^{-it\frac{x-X}{h}} \varphi_Z(t/h) \varphi_K(t)/\varphi_Z(t/h)\, dt \\ =& (2\pi)^{-1} \int (-it)^r e^{-it\frac{x-X}{h}} \varphi_K(t)\, dt \\ =& K^{(r)} \Big(\frac{x-X}{h} \Big), \end{aligned}$$

where we used Condition (A1) and the Fourier inversion Theorem. Statement (A.1) follows immediately. □

Lemma A.4. Under Condition (A3),

$$\int_{\mathcal{R}^3} K_Z^{(r)}(u-z;h)K_Z^{(r)}(u;h)K_Z^{(r)}(v-z;h)K_Z^{(r)}(v;h)\,du\,dv\,dz = (2\pi)^{-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt$$

Proof. We first focus on the integration with respect to the variable z. Under (A3),

$$\begin{aligned}
&\int K_Z^{(r)}(u-z;h)K_Z^{(r)}(v-z;h)\,dz \\
=&(2\pi)^{-2}\int\left(\int e^{-it(u-z)}(-it)^r\varphi_K(t)/\varphi_Z(t/h)\,dt\right)\left(\int e^{-is(v-z)}(-is)^r\varphi_K(s)/\varphi_Z(s/h)\,ds\right)\,dz \\
=&(2\pi)^{-2}\int\left(\int e^{-it(u-z)}(-it)^r\varphi_K(t)/\varphi_Z(t/h)\,dt\right)\left(\int e^{is(v-z)}(is)^r\varphi_K(-s)/\varphi_Z(-s/h)\,ds\right)\,dz \\
=&(2\pi)^{-2}\int \varphi_{g_1}(z)\bar{\varphi}_{g_2}(z)\,dz,
\end{aligned}$$

where

$$g_1(t) = e^{-itu}(-it)^r\varphi_K(t)/\varphi_Z(t/h) \qquad \text{and} \qquad \bar{g}_2(s) = e^{isv}(is)^r\varphi_K(-s)/\varphi_Z(-s/h).$$

Using Parseval's Theorem we find

$$\begin{aligned}
\int K_Z^{(r)}(u-z;h)K_Z^{(r)}(v-z;h)\,dz &= (2\pi)^{-1}\int g_1(t)\bar{g}_2(t)\,dt \\
&= (2\pi)^{-1}\int e^{it(v-u)}t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt.
\end{aligned}$$

Thus

$$\begin{aligned}
&\int_{\mathcal{R}^3} K_Z^{(r)}(u-z;h)K_Z^{(r)}(u;h)K_Z^{(r)}(v-z;h)K_Z^{(r)}(v;h)\,du\,dv\,dz \\
&= (2\pi)^{-1}\int_{\mathcal{R}^3} t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}e^{-itu}K_Z^{(r)}(u;h)e^{itv}K_Z^{(r)}(v;h)\,dt\,du\,dv \\
&= (2\pi)^{-1}\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\varphi_{K_Z^{(r)}}(-t;h)\varphi_{K_Z^{(r)}}(t;h)\,dt \\
&= (2\pi)^{-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt,
\end{aligned}$$

where we used Lemma A.1 and the fact that $\overline{\varphi}_g(t) = \varphi_g(-t)$, for any real-valued function g. □

Lemma A.5. Under Conditions (A1) and (A3),

$$\int K * K_Z(t;h)\,dt = 1$$

Proof. Immediate from

$$\int K * K_Z(t;h)\,dt = \int_{\mathcal{R}^2} K(t-x)K_Z(x;h)\,dt\,dx = \int_{\mathcal{R}^2} K(u)K_Z(x;h)\,du\,dx = 1,$$

by setting $u = t - x$, and since K_Z integrates to 1 as soon as K does (see for example Delaigle (1999)). $\square$

A.2 Proofs of the main results

We will now prove the main results stated in Section 3.

Proof of Proposition 1. By definition of $\widehat{\theta}_r$ (see (2.6)), we have

$$\mathrm{E}[\widehat{\theta}_r] = (nh^{2r+1})^{-1}R(K_Z^{(r)}(\cdot;h)) + (n-1)(nh^{2r+2})^{-1}\int \Big\{ \mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)\Big]\Big\}^2\,dx, \tag{A.2}$$

where by Lemma A.2 we know that

$$R(K_Z^{(r)}(\cdot;h)) = (2\pi)^{-1}\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt. \tag{A.3}$$

Now by Lemma A.3, a change of variable ($y = (x-z)/h$), r integrations by parts and by applying Condition (D2) we get

$$\begin{aligned}
&\mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y}{h};h\Big)\Big]\\
&\quad= \int K^{(r)}\Big(\frac{x-z}{h}\Big)f_X(z)\,dz\\
&\quad= h\int K^{(r)}(y)f_X(x-hy)\,dy\\
&\quad= h^{r+1}\int K(y)f_X^{(r)}(x-hy)\,dy\\
&\quad= h^{r+1}\big[f_X^{(r)}(x) + \frac{h^k}{k!}\mu_{K,k}f_X^{(r+k)}(x) - \frac{h^{k+1}}{k!}\iint_0^1 y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\big],
\end{aligned}$$

where we used Conditions (A1), (A2), (C2) and a kth order Taylor expansion of $f_X^{(r)}$ around x. Consequently, under Conditions C, we can write

$$\begin{aligned}
&h^{-2r-2}\int \Big\{ \mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y}{h};h\Big)\Big]\Big\}^2\,dx\\
&\qquad= R(f_X^{(r)}(x)) + h^{2k}(k!)^{-2}\mu_{K,k}^2 R(f_X^{(r+k)})\\
&\qquad\qquad + h^{2k+2}(k!)^{-2}\int\Big\{\iint_0^1 y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\Big\}^2\,dx
\end{aligned}$$

$$
\begin{aligned}
&+ 2h^k(k!)^{-1}\mu_{K,k}\int f_X^{(r)}(x)f_X^{(r+k)}(x)\,dx\\
&- 2h^{k+1}(k!)^{-1}\int_{\mathcal{R}^2}\int_0^1 f_X^{(r)}(x)y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\,dx\\
&- 2h^{2k+1}(k!)^{-2}\mu_{K,k}\int_{\mathcal{R}^2}\int_0^1 f_X^{(r+k)}(x)y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\,dx\\
=&\theta_r + h^{2k}(\{k!\})^{-2}\mu_{K,k}^2\theta_{r+k} + (-1)^{k/2}2h^k(k!)^{-1}\mu_{K,k}\theta_{r+k/2}\\
&+ h^{2k+2}(k!)^{-2}\int_{\mathcal{R}^3}\int_0^1\int_0^1 y^{k+1}z^{k+1}K(y)K(z)(1-\theta)^k(1-\xi)^k\\
&\qquad\qquad\times f_X^{(r+k+1)}(x-h\theta y)f_X^{(r+k+1)}(x-h\xi z)\,d\theta\,d\xi\,dy\,dz\,dx\\
&- 2h^{k+1}(k!)^{-1}\int_{\mathcal{R}^2}\int_0^1 f_X^{(r)}(x)y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\,dx\\
&- 2h^{2k+1}(k!)^{-2}\mu_{K,k}\int_{\mathcal{R}^2}\int_0^1 f_X^{(r+k)}(x)y^{k+1}K(y)(1-\theta)^k f_X^{(r+k+1)}(x-h\theta y)\,d\theta\,dy\,dx\\
\equiv&\theta_r + \frac{h^{2k}}{(k!)^2}\mu_{K,k}^2\theta_{r+k} + (-1)^{k/2}\frac{2h^k}{k!}\mu_{K,k}\theta_{r+k/2} + \frac{h^{2k+2}}{(k!)^2}(I) - 2\frac{h^{k+1}}{k!}(II) - 2\frac{h^{2k+1}}{(k!)^2}\mu_{K,k}(III),
\end{aligned}
$$

after $k/2$ integrations by parts, and by applying Conditions C and (D1).

In what follows we are going to provide upper bounds for the integrals (I) to (III) appearing in the above expression. This will allow us to give the order of the remaining terms of (A.2).

For the first integral we can write, applying Condition (C2) and a change of variable,

$$
\begin{aligned}
&|(I)|\\
&\le \int_{\mathcal{R}^3}\int_0^1\int_0^1 |yz|^{k+1}|K(y)||K(z)|(1-\theta)^k(1-\xi)^k|f_X^{(r+k+1)}(x-h\theta y)||f_X^{(r+k+1)}(x-h\xi z)|\,d\theta\,d\xi\,dy\,dz\,dx\\
&\le cst\Big\{\int |y|^{k+1}|K(y)|\,dy\Big\}^2\Big\{\int_0^1(1-\theta)^k\,d\theta\Big\}^2\int |f_X^{(r+k+1)}(u)|\,du\\
&= O(1),
\end{aligned}
$$

by using Conditions (A2) and (C1). We also have

$$
|(II)| \le cst\int_{\mathcal{R}^2}\int_0^1 |f_X^{(r)}(x)||y|^{k+1}|K(y)|(1-\theta)^k\,d\theta\,dy\,dx = O(1),
$$

using Conditions (A2) and C. Similarly we find that $|(III)| = O(1)$.

Consequently, under Conditions A, C, (D1) and (D2), we can write

$$
h^{-2r-2}\int\Big\{\mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)\Big]\Big\}^2\,dx = \theta_r + (-1)^{k/2}2h^k\mu_{K,k}\theta_{r+k/2}/k! + O(h^{k+1}). \quad \text{(A.4)}
$$

Finally, if we combine (A.2), (A.3) and (A.4), we find

$$
\mathrm{E}[\widehat{\theta}_r] \;=\; \theta_r + (-1)^{k/2}2h^k\mu_{K,k}\theta_{r+k/2}/k! + (2\pi n h^{2r+1})^{-1}\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt
$$

$$+O(h^{k+1})+O(n^{-1})$$

which leads to the asymptotic bias expression as stated in Proposition 1. □

Proof of Corollary 1. This proof uses techniques similar to those found in Fan (1991a).

From Expression (3.1) of Proposition 1, we see that the only term in the asymptotic bias that depends on the error distribution is the second one, given by

$$(2\pi nh^{2r+1})^{-1}\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt \equiv (2\pi nh^{2r+1})^{-1}.(I). \tag{A.5}$$

In order to establish the behavior of this term, we distinguish between the case of ordinary smooth errors and supersmooth errors.

(i). Ordinary smooth error case.

If the error is ordinary smooth, there exists an $M>0$ such that $|\varphi_Z(t)| \geq \frac{d_0}{2}|t|^{-\beta}$ for all $|t|>M$. Let $c=\min_{|t|\leq M}|\varphi_Z(t)|$, where c is strictly positive, and let $B_K=\max_{|t|\in I\!R}|\varphi_K(t)|$, where, by Condition A, $B_K<\infty$. Then, for n large enough, we have

$$\begin{aligned}
(I)=&\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}.1_{\{Mh\leq|t|\leq B\}}\,dt+\int_{-Mh}^{Mh}t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt\\
\leq&(2/d_0)^2h^{-2\beta}\int t^{2r+2\beta}|\varphi_K(t)|^2.1_{\{Mh\leq|t|\leq B\}}\,dt+\Big\{\min_{|t|\leq Mh}|\varphi_Z(t/h)|^2\Big\}^{-1}\int_{-Mh}^{Mh}t^{2r}|\varphi_K(t)|^2\,dt\\
\leq&(2/d_0)^2h^{-2\beta}\int t^{2r+2\beta}|\varphi_K(t)|^2.1_{\{Mh\leq|t|\leq B\}}\,dt+c^{-2}B_K^2\int_{-Mh}^{Mh}t^{2r}\,dt\\
=&c_1h^{-2\beta}+c_2(Mh)^{2r+1}\\
=&O(h^{-2\beta}),
\end{aligned}$$

with c_1 and c_2 positive constants. There exists also an $M>0$ such that $|\varphi_Z(t)|\leq 2d_1|t|^{-\beta}$, for all $|t|>M$. Hence we can write

$$\begin{aligned}
(I)=&\int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}.1_{\{Mh\leq|t|\leq B\}}\,dt+\int_{-Mh}^{Mh}t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt\\
\geq&(2d_1)^{-2}h^{-2\beta}\int t^{2r+2\beta}|\varphi_K(t)|^2\cdot 1_{\{Mh\leq|t|\leq B\}}\,dt\\
\sim&h^{-2\beta}.
\end{aligned}$$

Putting all this together, we have proven that $(I)\sim h^{-2\beta}$ and thus the term in (A.5) is of order $n^{-1}h^{-2r-2\beta-1}$. Hence the asymptotic bias if of order $c_3h^k+c_4n^{-1}h^{-2r-2\beta-1}$, with c_3 and c_4 constants. To minimize the bias we need to select a bandwidth so as to make the two terms asymptotically of the same order, i.e. choose h such that $h\sim n^{-\frac{1}{2r+2\beta+k+1}}$. The asymptotic order of the bias follows immediately.

(ii). Supersmooth error case.

If the error is supersmooth, there exists an $M > 0$ such that $|\varphi_Z(t)| \geq \frac{d_0}{2}|t|^{\beta_0}\exp(-|t|^\beta/\gamma)$ for all $|t| > M$. If we set $c = \min_{|t|\leq M}|\varphi_Z(t)|$, c a positive constant, we can write

$$\begin{aligned}(I) =& \int t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}.1_{\{Mh\leq|t|\leq B\}}\,dt + \int_{-Mh}^{Mh} t^{2r}|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt\\ \leq&(2/d_0)^2h^{2\beta_0}\exp(2|B/h|^\beta/\gamma)\int t^{2r-2\beta_0}|\varphi_K(t)|^2.1_{\{Mh\leq|t|\leq B\}}\,dt + c^{-2}B_K^2\int_{-Mh}^{Mh} t^{2r}\,dt\\ =&c_5h^{2\beta_0}\exp(2|B/h|^\beta/\gamma) + c_6(Mh)^{2r+1}\\ =&O\left\{h^{2\beta_0}\exp(2|B/h|^\beta/\gamma)\right\},\end{aligned}$$

with c_5 and c_6 positive constants. Finally the term in (A.5) is of order $O\left\{n^{-1}h^{-2r+2\beta_0-1}\exp(2|B/h|^\beta/\gamma)\right\}$. With the bandwidth given in Corollary 1(ii), we have $\exp(2|B/h|^\beta/\gamma) = \exp(d^{-\beta}\ln n) = n^{d^{-\beta}}$, and thus the term in (A.5) is of order $O\Big(n^{d^{-\beta}-1}(\ln n)^{\frac{2r-2\beta_0+1}{\beta}}\Big)$. With the same bandwidth, the first term of the asymptotic bias behaves like h^k, i.e. like $(\ln n)^{-k/\beta}$ and since $n^{d^{-\beta}-1}(\ln n)^{\frac{2r-2\beta_0+1}{\beta}}$ is $o\{(\ln n)^{-s}\}$ for any positive s, the asymptotic order of the bias follows.

□

Proof of Proposition 2. From Expression (2.6) we have

$$\begin{aligned}\mathrm{var}[\widehat{\theta}_r] =& n^{-4}h^{-4r-4}\,\mathrm{var}\Big[\sum_{i=1}^n\sum_{j\neq i}\int K_Z^{(r)}\Big(\frac{x-Y_i}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_j}{h};h\Big)\,dx\Big]\\ =&n^{-4}h^{-4r-4}\sum_{\substack{1\leq i\leq n, j\neq i\\ 1\leq i'\leq n, j'\neq i'}}\mathrm{cov}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_i}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_j}{h};h\Big)\,dx,\\ &\qquad\int K_Z^{(r)}\Big(\frac{x-Y_i'}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_j'}{h};h\Big)\,dx\Big]\\ =&2(n-1)n^{-3}h^{-4r-4}\,\mathrm{var}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big]\\ &+4(n-1)(n-2)n^{-3}h^{-4r-4}\,\mathrm{cov}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx,\\ &\qquad\int K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_3}{h};h\Big)\,dx\Big],\end{aligned}\tag{A.6}$$

since $\mathrm{cov}\left[\int K_Z^{(r)}\left(\frac{x-Y_1}{h};h\right)K_Z^{(r)}\left(\frac{x-Y_2}{h};h\right)\,dx, \int K_Z^{(r)}\left(\frac{x-Y_3}{h};h\right)K_Z^{(r)}\left(\frac{x-Y_4}{h};h\right)\,dx\right] = 0$, since Y_1, Y_2, Y_3 and Y_4 are independent. We will compute the variance and the covariance appearing in (A.6) separately. Note that the variance can be written as

$$\mathrm{var}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big]$$

$$= \mathrm{E}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big]^2 - \Big[\mathrm{E}\Big\{\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big\}\Big]^2$$

$$\equiv (I_V) - (II_V). \tag{A.7}$$

From Expression (A.4) we already know that under Conditions A, C, (D1) and (D2),

$$\int \Big\{\mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)\Big]\Big\}^2 dx = h^{2r+2}\theta_r + o(h^{2r+2}),$$

whence

$$(II_V) = \Big[\int \Big\{\mathrm{E}\Big[K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)\Big]\Big\}^2 dx\Big]^2 = h^{4r+4}\theta_r^2 + o(h^{4r+4}). \tag{A.8}$$

For the term (I_V) we obtain, after a first change of variable $u = (x - Y_2)/h$, a second change of variable $z = (x - y)/h$ and a Taylor expansion of order zero of f_Y around x,

$$\begin{aligned}
(I_V) =& h^2.\,\mathrm{E}\Big\{\int K_Z^{(r)}\Big(u - \frac{Y_1 - Y_2}{h};h\Big)K_Z^{(r)}(u;h)\,du\Big\}^2\\
=& h^2\int_{\mathcal{R}^2}\Big\{\int K_Z^{(r)}\Big(u - \frac{x-y}{h};h\Big)K_Z^{(r)}(u;h)\,du\Big\}^2 f_Y(x)f_Y(y)\,dx\,dy\\
=& h^3\int_{\mathcal{R}^2}\Big\{\int K_Z^{(r)}(u-z;h)K_Z^{(r)}(u;h)\,du\Big\}^2 f_Y(x)f_Y(x-hz)\,dx\,dz\\
=& h^3\int_{\mathcal{R}^2}\Big\{K_Z^{(r)} * K_Z^{(r)}(z;h)\Big\}^2 f_Y(x)[f_Y(x) - hz\int_0^1 f_Y'(x-h\theta z)\,d\theta]\,dx\,dz\\
=& h^3 R(f_Y)\int\Big\{K_Z^{(r)} * K_Z^{(r)}(z;h)\Big\}^2 dz\\
&- h^4\int_{\mathcal{R}^2}\int_0^1 z\Big\{K_Z^{(r)} * K_Z^{(r)}(z;h)\Big\}^2 f_Y(x)f_Y'(x-h\theta z)\,d\theta\,dx\,dz\\
=& h^3.R(f_Y)(2\pi)^{-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt + O(h^4B_h),
\end{aligned} \tag{A.9}$$

by Lemma A.4 and Conditions (C2) and (B3).

Finally from expressions (A.7), (A.8) and (A.9) we find

$$\mathrm{var}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big] = \frac{h^3.R(f_Y)}{2\pi}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt + O(h^4B_h). \tag{A.10}$$

The covariance in expression (A.6) can be written as

$$\begin{aligned}
&\mathrm{E}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx.\int K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_3}{h};h\Big)\,dx\Big]\\
&\quad - \mathrm{E}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx\Big].\,\mathrm{E}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_3}{h};h\Big)\,dx\Big]\\
&\equiv (I_C) - (II_C).
\end{aligned} \tag{A.11}$$

Similarly as for statement (A.8), we know that

$$(II_C) = h^{4r+4}\theta_r^2 + o(h^{4r+4}). \tag{A.12}$$

For the term (I_C), we have, by Lemma A.3, and the changes of variables $x = z/h$ and $y = v/h$, and further changes of variables $z = x - t/h$, $v = x - u/h$, $w = y - u/h$ and $g = y - s/h$

$$\begin{aligned}
(I_C) =& \mathrm{E}\Big[\int_{\mathcal{R}^2} K_Z^{(r)}\Big(\frac{z-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{z-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{v-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{v-Y_3}{h};h\Big)\,dz\,dv\Big] \\
=& \mathrm{E}\Big[\int_{\mathcal{R}^2} K^{(r)}\Big(\frac{z-X_1}{h}\Big)K_Z^{(r)}\Big(\frac{z-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{v-Y_2}{h};h\Big)K^{(r)}\Big(\frac{v-X_3}{h}\Big)\,dz\,dv\Big] \\
=& h^2\int_{\mathcal{R}^5} K^{(r)}\Big(x-\frac{t}{h}\Big)K_Z^{(r)}\Big(x-\frac{u}{h};h\Big)K_Z^{(r)}\Big(y-\frac{u}{h};h\Big)K^{(r)}\Big(y-\frac{s}{h}\Big)f_X(s)f_X(t)f_Y(u)\,ds\,dt\,du\,dx\,dy \\
=& h^4\int_{\mathcal{R}^5} K^{(r)}(z)K_Z^{(r)}(v;h)K_Z^{(r)}(w;h)K^{(r)}(g)f_X(u+hv-hz)f_X(u+hw-hg)f_Y(u)\,dg\,du\,dv\,dw\,dz \\
=& h^{4r+4}\int_{\mathcal{R}^5} K(z)K_Z(v;h)K_Z(w;h)K(g)f_X^{(2r)}(u+hv-hz)f_X^{(2r)}(u+hw-hg)f_Y(u)\,dg\,du\,dv\,dw\,dz,
\end{aligned}$$

after $4r$ integrations by parts, and by applying conditions (D2), (D3) and C. A further change of variables, $s = v - z$ and $t = w - g$ and using the symmetry of K leads to

$$\begin{aligned}
(I_C) =& h^{4r+4}\int_{\mathcal{R}^5} K(s-v)K_Z(v;h)K_Z(w;h)K(t-w)f_X^{(2r)}(u+hs)f_X^{(2r)}(u+ht)f_Y(u)\,ds\,dt\,du\,dv\,dw \\
=& h^{4r+4}\int_{\mathcal{R}^3} K*K_Z(s;h)\; K*K_Z(t;h)f_X^{(2r)}(u+hs)f_X^{(2r)}(u+ht)f_Y(u)\,ds\,dt\,du \\
=& h^{4r+4}\int\Big[f_X^{(2r)}(u)\Big]^2 f_Y(u)du \\
&+ h^{4r+5}\int_{\mathcal{R}^2}\int_0^1 sK*K_Z(s;h)f_X^{(2r+1)}(u+\theta hs)f_X^{(2r)}(u)f_Y(u)\,d\theta\,ds\,du \\
&+ h^{4r+5}\int_{\mathcal{R}^2}\int_0^1 tK*K_Z(t;h)f_X^{(2r)}(u)f_X^{(2r+1)}(u+\theta ht)f_Y(u)\,d\theta\,dt\,du \\
&+ h^{4r+6}\int_{\mathcal{R}^3}\int_0^1\int_0^1 stK*K_Z(s;h)\; K*K_Z(t;h)f_X^{(2r+1)}(u+\xi ht)f_X^{(2r+1)}(u+\theta hs)f_Y(u)\,d\theta\,d\xi\,ds\,dt\,du
\end{aligned} \tag{A.13}$$

after two Taylor expansions of order zero, of $f_X^{(2r)}$ around u, and by using (C1) and Lemma A.5. A change of variables $v = u + \theta hs, u + \theta ht$ or $u + \xi ht$ and $w = u + \theta hs$ gives

$$\begin{aligned}
(I_C) =& h^{4r+4}\int\Big[f_X^{(2r)}(u)\Big]^2 f_Y(u)du \\
&+ h^{4r+4}\int_{\mathcal{R}^2}\int_u^{u+hs} K*K_Z(s;h)f_X^{(2r+1)}(v)f_X^{(2r)}(u)f_Y(u)\,dv\,ds\,du \\
&+ h^{4r+4}\int_{\mathcal{R}^2}\int_u^{u+ht} K*K_Z(t;h)f_X^{(2r)}(u)f_X^{(2r+1)}(v)f_Y(u)\,dv\,dt\,du
\end{aligned}$$

$$+ h^{4r+4}\int_{\mathcal{R}^3}\int_u^{u+ht}\int_u^{u+hs} K*K_Z(s;h)K*K_Z(t;h)f_X^{(2r+1)}(v)f_X^{(2r+1)}(w)f_Y(u)\,dv\,dw\,ds\,dt\,du.$$

The second, third and fourth term in this expression can be bounded by noting that under Conditions (B2), and C

$$\begin{aligned}
&\Big|\int_{\mathcal{R}^2}\int_u^{u+hs} K*K_Z(s;h)f_X^{(2r+1)}(v)f_X^{(2r)}(u)f_Y(u)\,dv\,ds\,du\Big| \\
&\quad\le \int |K*K_Z(s;h)|\,ds\int |f_X^{(2r+1)}(v)|\,dv\int |f_X^{(2r)}(u)|f_Y(u)\,du \\
&\quad= O\Big(\int |K*K_Z(s;h)|\,ds\Big) = o(n^{-1}h^{-4r-1}B_h),
\end{aligned}$$

and

$$\begin{aligned}
&\Big|\int_{\mathcal{R}^3}\int_u^{u+ht}\int_u^{u+hs} K*K_Z(s;h)K*K_Z(t;h)f_X^{(2r+1)}(v)f_X^{(2r+1)}(w)f_Y(u)\,dv\,dw\,ds\,dt\,du\Big| \\
&\quad\le \Big\{\int |K*K_Z(s;h)|\,ds\Big\}^2\Big\{\int |f_X^{(2r+1)}(v)|\,dv\Big\}^2 \\
&\quad= O\Big(\Big\{\int |K*K_Z(s;h)|\,ds\Big\}^2\Big) = o(n^{-1}h^{-4r-1}B_h).
\end{aligned}$$

Hence

$$(I_C) = h^{4r+4}\int \Big[f_X^{(2r)}(u)\Big]^2 f_Y(u)du + o(n^{-1}h^3B_h). \tag{A.14}$$

From statements (A.11), (A.12) and (A.14) we deduce that

$$\begin{aligned}
&\mathrm{cov}\Big[\int K_Z^{(r)}\Big(\frac{x-Y_1}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)\,dx, \int K_Z^{(r)}\Big(\frac{x-Y_2}{h};h\Big)K_Z^{(r)}\Big(\frac{x-Y_3}{h};h\Big)\,dx\Big] \\
&\quad= h^{4r+4}\int\Big[f_X^{(2r)}(u)\Big]^2 f_Y(u)du - h^{4r+4}\theta_r^2 + o(n^{-1}h^3B_h)
\end{aligned}\tag{A.15}$$

Finally using (A.10) and (A.15), the variance expression in (A.6) may be written as

$$\begin{aligned}
\mathrm{var}[\widehat{\theta_r}] =& R(f_Y)(\pi n^2h^{4r+1})^{-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt + 4n^{-1}\Big\{\int\Big[f_X^{(2r)}(u)\Big]^2 f_Y(u)du - \theta_r^2\Big\} \\
&+ o(n^{-2}h^{-4r-1}B_h) \qquad\qquad (A.16)\\
=& R(f_Y)(\pi n^2h^{4r+1})^{-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt + o(n^{-2}h^{-4r-1}B_h).
\end{aligned}$$

□

Proof of Corollary 2. From Expression (3.2), we see that the asymptotic variance is of the same order as

$$n^{-2}h^{-4r-1}\int t^{4r}|\varphi_K(t)|^4|\varphi_Z(t/h)|^{-4}\,dt \equiv n^{-2}h^{-4r-1}.(I). \tag{A.17}$$

In order to study the behavior of the term (I) we distinguish between the cases of ordinary and supersmooth error densities.

(i). Ordinary smooth error case. The proof is similar to the proof of Corollary 1(i).

(ii). Supersmooth error case.

In this case, using techniques similar to the proof of Corollary 1(ii), we can show that $(I) = O\left\{h^{4\beta_0}\exp(4|B/h|^{\beta}/\gamma)\right\}$. If we choose h as stated in Corollary 2(ii), Expression (A.17) is of order $O\Big(n^{2d^{-\beta}-2}(\ln n)^{\frac{4r-4\beta_0+1}{\beta}}\Big)$, which is also $o(\ln n)^{-k/\beta}$.

$\square$

References

Aldershof, B. (1991). Estimation of integrated squared density derivatives. *PhD.thesis*, Department of Statistics, University of North Carolina at Chapel Hill, USA.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, **83**, 1184–1186.

Carroll, R.J., Ruppert, D. and Stefanski, L. (1995). *Measurement error in nonlinear models.* Chapman and Hall, London. *Journal of the American Statistical Association*, **83**, 1184–1186.

Delaigle, A. (1999). Bandwidth selection in kernel estimation of a density when the data are contaminated by errors. *Mémoire de DEA (Master thesis)*, Institut de Statistique, Université catholique de Louvain, Belgium.

Delaigle, A. and Gijbels, I. (2001a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Discussion Paper*, Institut de Statistique, Université catholique de Louvain, Belgium.

Delaigle, A. and Gijbels, I. (2001b). Comparison of several data-driven bandwidth selection procedures in kernel estimation of a density when the data are contaminated by random noise. *Discussion Paper*, Institut de Statistique, Université catholique de Louvain, Belgium.

Fan, J. (1991a). Global behaviour of deconvolution kernel estimates. *Statistica Sinica*, **1**, 541–551.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, **19**, 1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, **20**, 155–169.

Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, **6**, 109–115.

Hesse, C.H.(1999). Data-driven deconvolution. *Journal of Nonparametric Statistics*, **10**, 343–373.

Jones, M.C. and Sheather, S.J. (1991). Using non stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics and Probability Letters*, **6**, 511–514.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with B splines. *Journal of the American Statistical Association*, **77**, 748–753.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Stefanski, L.A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters*, **9**, 229–235.

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **2**, 169–184.

Wand, M.P. (1998). Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, **37**, 131–139.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall, London.