A convolution estimator for the density of nonlinear regression observations

Bård Støve Dag Tjøstheim

Department of Mathematics
University of Bergen
Johannes Brunsgate 12
5008 Bergen
Norway

Abstract

We present a convolution estimator for the density of the responses in a standard regression model. The rate of convergence for the variance of this estimator is examined and found to be of order n^{-1}. We also derive the bias of the new estimator and conduct simulation experiments to check the finite sample properties. The proposed estimator performs better than the kernel density estimator for well behaved noise densities.

Some key words: Convolution, Kernel function, Mean squared error, Nonparametric density estimation, Regression.

1 Introduction

There exists a vast literature on the problem of estimating an unknown density function $f(x)$ from a given sample $X_1, X_2, ..., X_n$ of independent and identically distributed random variables, see e.g.; the books by Härdle (1990), Wand & Jones (1995) and Simonoff (1996). The most used method is the kernel density estimator

$$f^*(x) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big), \tag{1}$$

where K is a kernel function and h is the bandwidth. It is well known that the asymptotic bias and variance of this estimator are of the order h^2 and $(nh)^{-1}$, respectively.

In this paper we consider the standard nonlinear regression model,

$$Y_i = g(X_i) + e_i, \tag{2}$$

where g is unknown and where $\{X_i\}$ and $\{e_i\}$ consist of independent and identically distributed random variables with $\{e_i\}$ independent of $\{X_i\}$. Denote the density of Y_i,

$f_Y(\cdot)$, this being the density of interest, and the densities of X_i and e_i, $f_X(\cdot)$ and $f_e(\cdot)$, respectively. For given observations of (X_i, Y_i) a standard method of estimating the density of Y_i is using the already mentioned kernel density estimator on $\{Y_i\}$. This estimator does not require the relationship (2) to hold, and if one is able to construct an estimator which takes this relationship into account, one should think that it would be possible to make an improvement. This idea was introduced in Støve & Tjøstheim (2005*b*) for nonparametric estimation of g. For that case, the asymptotic bias and variance were of the same order as the standard nonparametric regression estimators, but an asymptotic bias improvement was obtained. However, in the case of density estimation in equation (2), it will be seen that we manage to obtain a better convergence rate for the variance and often better bias properties, although asymptotically, the order of the bias is the same as for the kernel density estimator.

Other authors have also studied this convolution idea; Frees (1994) introduced density estimation for a symmetric function $Y = h(X_1, ..., X_m)$ of $m > 1$ independent and identically distributed variables. The density can be estimated at the rate $n^{-1/2}$, in contrast to the standard nonparametric rate of $(nh)^{-1/2}$, if this density and the conditional density of Y given X_1 are sufficiently smooth. This result generalizes to non-identically distributed random variables, and in particular to convoluted densities $f * l(y) = \int f(y-x)l(x)\mathrm{d}x$. Saavedra & Cao (2000) introduced this type of convolution-kernel estimator for the marginal density of a moving average process $Y_i = X_i - \theta X_{i-1}$ when θ is known, and proved that this estimator, for an appropriate choice of bandwidth, is $n^{1/2}$-consistent. The case when θ is unknown is examined in Saavedra & Cao (1999), and an analogous result is obtained, but in this case both θ and the innovations X_i have to be estimated. Further, Schick & Wefelmeyer (2004*a*) introduced a slightly simplified variant of this estimator and proved a stronger result of asymptotic normality. In Schick & Wefelmeyer (2004*b*) it is shown that the density of a sum of independent random variables can be estimated by the convolution of kernel estimators for the marginal densities, and that this estimator is $n^{1/2}$-consistent as well.

Note that we assume that the function $g(\cdot)$ and the error terms e_i are not known, and thus has to be estimated. This is in contradistinction to the models examined in Frees (1994) and Schick & Wefelmeyer (2004*b*), where the authors assume that the function g is known.

Our proposed estimator is presented in section 2, its asymptotic behaviour is examined in section 3, and some simulation results and a real data case are located in section 4. Some conclusions are given in section 5.

2 The estimator

From equation (2) using that $g(X_i)$ and e_i are independent, we get

$$f_Y(y) = \int f_e\big(y - g(u)\big) f_X(u)\mathrm{d}u, \tag{3}$$

where f_e is the density of the residuals. This motivates an estimator of the density f_Y given by the functional in (3), which can be written as

$$f_Y(y) = \mathrm{E}\big[f_e\big(y - g(X)\big)\big]. \tag{4}$$

Assume we have observations $(X_1, Y_1,), ..., (X_n, Y_n)$ of (X, Y). We introduce estimators of g, e, f_e and an estimate of the expectation in (4), thus

$$\hat{f}_Y(y) = \tilde{\mathrm{E}}\big[f^*_{\tilde{e}}\big(y - \tilde{g}(X)\big)\big]. \tag{5}$$

In this paper, $\tilde{g}$ is the Nadaraya-Watson estimator, see e.g. Härdle (1990), with bandwidth h_R, and kernel function $\tilde{K}_{x,h_R}(X_i) = (1/h_R)\tilde{K}\big((x - X_i)/h_R\big)$,

$$\tilde{g}(x) = \frac{\sum_{i=1}^n \tilde{K}_{x,h_R}(X_i)Y_i}{\sum_{i=1}^n \tilde{K}_{x,h_R}(X_i)}. \tag{6}$$

Other nonparametric regression estimators, such as the local linear estimator, see Fan & Gijbels (1996) page 20, for g can also be choosen. The estimate for e_i is

$$\tilde{e}_i = Y_i - \tilde{g}(X_i). \tag{7}$$

The estimate $f^*_{\tilde{e}}$ is by the kernel estimator with bandwidth h_D and kernel function $K(\cdot)$,

$$f^*_{\tilde{e}}(y) = \frac{1}{(n-1)h_D}\sum_{i=2}^{n} K\Big(\frac{y - \tilde{e}_i}{h_D}\Big). \tag{8}$$

Thus the final density estimator for f_Y is, using (5),

$$\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} f^*_{\tilde{e}}\big(y - \tilde{g}(X_i)\big) = \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{(n-1)h_D}\sum_{j=2}^{n} K\Big(\frac{y - \tilde{g}(X_i) - \tilde{e}_j}{h_D}\Big)\Big]. \tag{9}$$

3 Asymptotic properties

Asymptotic analysis of nonparametric estimators is usually based on the asymptotic bias and variance of the estimator. In this section we discuss these asymptotic properties of the estimator (9). The following assumptions are made, and a few others are introduced later when needed,

1. The kernel function K is a bounded non-negative, two times differentiable, symmetric function that integrates to 1, and hence satisfies

$$\int K'(z)\mathrm{d}z = 0 \quad \text{and} \quad \int z^2 K'(z)\mathrm{d}z = 0.$$

2. The function g is differentiable and its inverse exists.

3. The density f_X has compact support $S(X)$, is continous and two times differentiable on the support.

4. $\lim_{n\to\infty} h_D = 0$ and $\lim_{n\to\infty} n^2 h_D = \infty$.

Condition 1 is standard in nonparametric estimation, and observe that if the kernel function is the standard normal distribution, this condition is automatically fulfilled. Condition 2 is introduced to obtain simple expressions. We think it can be relaxed. Condition 3 is also standard, and the compact support is introduced for the sake of simplicity. It can probably be removed at the cost of lengthier arguments. Condition 4 is standard.

To study the mean squared error (MSE) of the estimator, it is useful to decompose the difference between the estimator and the true density in the following manner,

$$\hat{f}_Y(x) - f_Y(x) = \hat{f}_Y(x) - \tilde{f}_Y(x) + \tilde{f}_Y(x) - f_Y(x), \tag{10}$$

where

$$\tilde{f}_Y(x) = \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{(n-1)h_D}\sum_{j=2}^{n}K\Big(\frac{x-g(X_i)-e_j}{h_D}\Big)\Big], \tag{11}$$

that is, the proposed estimator with $g(\cdot)$ and e_j for $j = 1, ..., n$, known. We first consider this "estimator". Recall that the MSE can be decomposed as follows,

$$\mathrm{MSE}(\tilde{f}_Y(x)) = \mathrm{E}[(\tilde{f}_Y(x) - f_Y(x))^2] = \mathrm{var}(\tilde{f}_Y(x)) + [\mathrm{E}(\tilde{f}_Y(x)) - f_Y(x)]^2. \tag{12}$$

To ease notation, we now set $h_D = h$. Consider the bias term first. Since X_i and e_j are independent for all i and j,

$$\begin{aligned}\mathrm{E}\big(\tilde{f}_Y(x)\big) &= \frac{1}{n(n-1)h}\mathrm{E}\Big[\sum_{i=1}^{n}\sum_{j=2}^{n}K\big(\frac{x-g(X_i)-e_j}{h}\big)\Big] \\ &= \frac{1}{h}\mathrm{E}\Big[K\big(\frac{x-g(X)-e}{h}\big)\Big]. \end{aligned}\tag{13}$$

Further, by a change of variable, the convolution property and Taylor expansion, we get

$$\begin{aligned}\frac{1}{h}\mathrm{E}\Big[K\Big(\frac{x-g(X)-e}{h}\Big)\Big] &= \frac{1}{h}\iint K\Big(\frac{x-g(v)-u}{h}\Big)f_X(v)f_e(u)\mathrm{d}v\,\mathrm{d}u \\ &= \iint K(w)f_X(v)f_e(x-g(v)-hw)\mathrm{d}v\mathrm{d}w \\ = \int K(w)f_Y(x-hw)\mathrm{d}w &= f_Y(x) + \frac{h^2}{2}f_Y''(x)\int w^2K(w)\mathrm{d}w + O(h^4).\end{aligned}$$

Thus the bias can be written,

$$\mathrm{E}\big(\tilde{f}_Y(x)\big) - f_Y(x) = \frac{h^2}{2}f_Y''(x)\int w^2K(w)\mathrm{d}w + O(h^4). \tag{14}$$

This is in fact equal to the bias of the standard kernel density estimator, see e.g. Wand & Jones (1995) page 20.

The variance term can be decomposed into several covariance terms; see a similar argument in Saavedra & Cao (2000),

$$\mathrm{var}\big(\tilde{f}_Y(x)\big) = \frac{1}{n^2(n-1)^2h^2}\mathrm{var}\Big[\sum_{i=1}^{n}\sum_{j=2}^{n} K\Big(\frac{x-g(X_i)-e_j}{h}\Big)\Big]$$

$$= \frac{1}{n^2(n-1)^2h^2}\sum_{i=1}^{n}\sum_{j=2}^{n}\sum_{k=1}^{n}\sum_{l=2}^{n}\mathrm{cov}\Big(K\big(\frac{x-g(X_i)-e_j}{h}\big), K\big(\frac{x-g(X_k)-e_l}{h}\big)\Big)$$

$$= \frac{1}{n^2(n-1)^2h^2}\Big[(n-1)\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big) \quad (15)$$

$$+(n-1)(n-2)\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big) \quad (16)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big) \quad (17)$$

$$+2(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_3)-e_1}{h}\big)\Big) \quad (18)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_3)-e_2}{h}\big)\Big) \quad (19)$$

$$+(n-1)(n-2)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_2)-e_1}{h}\big)\Big) \quad (20)$$

$$+2(n-1)(n-2)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big), K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big) \quad (21)$$

$$+2(n-1)(n-2)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big), K\big(\frac{x-g(X_2)-e_1}{h}\big)\Big) \quad (22)$$

$$+(n-1)(n-2)(n-3)(n-4)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_3)-e_4}{h}\big)\Big) \quad (23)$$

$$+(n-1)(n-2)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big), K\big(\frac{x-g(X_2)-e_2}{h}\big)\Big) \quad (24)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big), K\big(\frac{x-g(X_2)-e_3}{h}\big)\Big) \quad (25)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_2)-e_1}{h}\big), K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)\Big]. \quad (26)$$

Each variance and covariance term can be studied using

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2, \quad (27)$$

$$\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y). \quad (28)$$

By independence the terms (18), (20), (23), (24), (25) and (26) are equal to zero, and we just have to examine the remaining terms.

Equation (15) is

$$\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big) = \mathrm{E}\Big(K^2\big(\frac{x-g(X_1)-e_1}{h}\big)\Big) - \Big[\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big)\Big]^2.$$

Consider the first term. By change of variables, convolution and Taylor expansion,

$$\mathrm{E}\Big(K^2\big(\frac{x-g(X_1)-e_1}{h}\big)\Big) = \iint K^2\big(\frac{x-g(v)-u}{h}\big) f_X(v) f_e(u) \mathrm{d}v\mathrm{d}u$$

$$= h\iint K^2(z) f_X(v) f_e(x-g(v)-hz)\mathrm{d}v\mathrm{d}z = h\int K^2(z) f_Y(x-zh)\mathrm{d}z$$

$$= h\int K^2(z)\big[f_Y(x) - hzf_Y'(x) + \frac{h^2z^2}{2}f_Y''(x)\big]\mathrm{d}z + O(h^4)$$

$$= hf_Y(x)\int K^2(z)\mathrm{d}z + \frac{h^3}{2}f_Y''(x)\int z^2K^2(z)\mathrm{d}z + O(h^4).$$

The second term in the variance expression is, using exactly the same techniques,

$$\Big[\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big)\Big]^2 = \Big[h\big[f_Y(x) + \frac{h^2}{2}f_Y''(x)\int z^2K(z)\mathrm{d}z + o(h^2)\big]\Big]^2.$$

In total,

$$(n-1)\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big)$$

$$= (n-1)\Big[hf_Y(x)\int K^2(z)\mathrm{d}z - h^2f_Y^2(x) + O(h^2)\Big]. \tag{29}$$

The second term, (16), is,

$$\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big) = \mathrm{E}\Big(K^2\big(\frac{x-g(X_1)-e_2}{h}\big)\Big) - \Big[\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\Big]^2.$$

This is equal to term (15), thus,

$$(n-1)(n-2)\mathrm{var}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)$$

$$= (n-1)(n-2)\Big[hf_Y(x)\int K^2(z)\mathrm{d}z - h^2f_Y^2(x) + O(h^2)\Big]. \tag{30}$$

The third term, equation (17), is

$$\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big), K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)$$

$$= \mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)$$

$$-\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big).$$

Further, by change of variables and Taylor expansion,

$$\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)$$

$$= \iiint K\big(\frac{x-g(v)-u_1}{h}\big)K\big(\frac{x-g(v)-u_2}{h}\big) f_X(v) f_e(u_1) f_e(u_2)\mathrm{d}v\mathrm{d}u_1\mathrm{d}u_2$$

$$= h^2\iiint K(z_1)K(z_2) f_X(v) f_e(x-g(v)-z_1h) f_e(x-g(v)-z_2h)\mathrm{d}v\mathrm{d}z_1\mathrm{d}z_2$$

$$= h^2\Big[\int f_X(v) f_e^2\big(x-g(v)\big)\mathrm{d}v + O(h^2)\Big].$$

As before,

$$\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)$$
$$= \Big[h\big[f_Y(x)+\frac{h^2}{2}f_Y''(x)\int z^2K(z)\mathrm{d}z+o(h^2)\big]\Big]^2 = h^2f_Y^2(x)+O(h^4).$$

In total this gives,

$$(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big),K\big(\frac{x-g(X_1)-e_3}{h}\big)\Big)$$
$$= (n-1)(n-2)(n-3)\Big[h^2\int f_X(v)f_e^2\big(x-g(v)\big)\mathrm{d}v-h^2f_Y^2(x)+O(h^4)\Big]. \qquad (31)$$

The fifth term, equation (19) is

$$\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big),K\big(\frac{x-g(X_3)-e_2}{h}\big)\Big). \qquad (32)$$

Using the assumption that the inverse of $g(\cdot)$ exists we obtain

$$\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)K\big(\frac{x-g(X_3)-e_2}{h}\big)\Big)$$
$$= \iiint K\big(\frac{x-g(v)-u}{h}\big)K\big(\frac{x-g(w)-u}{h}\big)f_e(u)f_X(v)f_X(w)\mathrm{d}u\mathrm{d}v\mathrm{d}w$$
$$= h^2\iiint K(z_1)K(z_2)f_X(v)l_X(g(v)+h(z_1-z_2))f_e(x-g(v)-hz_1)$$
$$\times r\big(g(v)+h(z_1-z_2)\big)\mathrm{d}v\mathrm{d}z_1\mathrm{d}z_2 = h^2\int r\big(g(v)\big)f_X(v)f_e\big(x-g(v)\big)l_X\big(g(v)\big)\mathrm{d}v+O(h^4),$$

where $(g^{-1})'=r$ and $f_X(g^{-1})=l_X$. Note that

$$r(v)=\frac{\mathrm{d}}{\mathrm{d}v}\big(g^{-1}(v)\big)=\frac{1}{g'\big(g^{-1}(v)\big)},$$

and $g^{-1}\big(g(v)\big)=v$, thus

$$r\big(g(v)\big)=\frac{1}{g'(v)}$$

and

$$l_X\big(g(v)\big)=f_X\big(g^{-1}(g(v))\big)=f_X(v).$$

As before,

$$\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\mathrm{E}\Big(K\big(\frac{x-g(X_3)-e_1}{h}\big)\Big)=h^2f_Y^2(x)+O(h^4).$$

In total,

$$(n-1)(n-2)(n-3)\mathrm{cov}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big),K\big(\frac{x-g(X_3)-e_2}{h}\big)\Big)$$
$$= (n-1)(n-2)(n-3)\Big[h^2\int\frac{f_X^2(v)f_e\big(x-g(v)\big)}{g'(v)}\mathrm{d}v-h^2f_Y^2(x)+O(h^4)\Big]. \qquad (33)$$

The seventh term, equation (21), is

$$\operatorname{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big),K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big). \tag{34}$$

We have

$$\begin{aligned}
&\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\\
&=\iiint K\big(\frac{x-g(v)-u}{h}\big)K\big(\frac{x-g(v)-w}{h}\big)f_X(v)f_e(u)f_e(w)\mathrm{d}v\mathrm{d}u\mathrm{d}w\\
&=h^2\iiint K(z_1)K(z_2)f_X(v)f_e(x-g(v)-hz_1)f_e(x-g(v)-hz_2)\mathrm{d}v\mathrm{d}z_1\mathrm{d}z_2\\
&=h^2\int f_X(v)f_e\big(x-g(v)\big)f_e\big(x-g(v)\big)\mathrm{d}v+O(h^4).
\end{aligned}$$

As before,

$$\begin{aligned}
&\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big)\Big)\mathrm{E}\Big(K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\\
&=h^2f_Y^2(x)+O(h^4).
\end{aligned}$$

In total,

$$\begin{aligned}
&2(n-1)(n-2)\operatorname{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big),K\big(\frac{x-g(X_1)-e_2}{h}\big)\Big)\\
&=2(n-1)(n-2)\Big[h^2\int f_X(v)f_e^2\big(x-g(v)\big)\mathrm{d}v\\
&\qquad -h^2f_Y^2(x)+O(h^4)\Big].
\end{aligned}\tag{35}$$

The eigth term, equation (22), is

$$\operatorname{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big),K\big(\frac{x-g(X_2)-e_1}{h}\big)\Big). \tag{36}$$

This is equal to the fifth term, (19), so in total,

$$\begin{aligned}
&2(n-1)(n-2)\operatorname{cov}\Big(K\big(\frac{x-g(X_1)-e_1}{h}\big),K\big(\frac{x-g(X_2)-e_1}{h}\big)\Big)\\
&=2(n-1)(n-2)\Big[h^2\int\frac{f_X^2(v)f_e\big(x-g(v)\big)}{g'(v)}\mathrm{d}v\\
&\qquad -h^2f_Y^2(x)+O(h^4)\Big].
\end{aligned}\tag{37}$$

Adding all the expressions stemming from (15)-(26), we get the total variance,

$$\begin{aligned}
\operatorname{var}\big(\tilde{f}_Y(x)\big)&=\frac{1}{n-1}\int f_X(v)f_e^2\big(x-g(v)\big)\mathrm{d}v\\
&+\frac{1}{n-1}\int\frac{f_X^2(v)f_e\big(x-g(v)\big)}{g'(v)}\mathrm{d}v-\frac{2}{n-1}f_Y^2(x)\\
&+\frac{1}{n(n-1)h}f_Y(x)\int K^2(z)\mathrm{d}z+O\big((n-1)^{-1}h^2\big).
\end{aligned}\tag{38}$$

Observe that from this expression the variance of $\tilde{f}_Y(x)$ has leading terms of order n^{-1}, this in contrast to the kernel density estimator which has leading terms $(nh)^{-1}$, see e.g. Wand & Jones (1995) page 21.

Thus the MSE for $\tilde{f}_Y(x)$ becomes,

$$\begin{aligned}\mathrm{MSE}\big(\tilde{f}_Y(x)\big) = \mathrm{E}\Big[\big(\tilde{f}_Y(x) - f_Y(x)\big)^2\Big] &= \frac{1}{4}h^4 f_Y''(x)^2\Big[\int w^2 K(w)\mathrm{d}w\Big]^2 \\ &+ \frac{1}{n-1}\int f_X(v) f_e^2\big(x - g(v)\big)\mathrm{d}v \\ &+ \frac{1}{n-1}\int \frac{f_X^2(v) f_e\big(x-g(v)\big)}{g'(v)}\mathrm{d}v - \frac{2}{n-1}f_Y^2(x) \\ &+ \frac{1}{n(n-1)h} f_Y(x)\int K^2(z)\mathrm{d}z + O\big((n-1)^{-1}h^2\big) + O(h^6). \qquad (39)\end{aligned}$$

If h is of order $n^{-1/4}$ it follows trivially that the MSE is of order $O(n^{-1})$.

Note that there are bias reducing techniques, using e.g. a higher order kernel, see e.g. Wand & Jones (1995) page 32, so that the bias can be reduced to $O(h^6)$ or $O(h^8)$, say, while still keeping the variance at $O(n^{-1})$. This means that there would be a wider choice of bandwidths for which the MSE is of order n^{-1}.

We next study the properties of the other term in equation (10), that is, $\hat{f}_Y(x) - \tilde{f}_Y(x)$. Here some additional assumptions are needed. Before we list them, let $f_{X,Y}(x,y)$ denote the joint distribution of (X,Y), define $m(x) = \int y f(x,y)\mathrm{d}y$:

5. $\mathrm{E}|Y|^s < \infty$ and $\sup_x \int |y|^s f(x,y)\mathrm{d}y < \infty$, $s \geq 2$.

6. m is continous on $S(X)$.

Consider the estimator $\hat{f}_Y(x)$ in (9). By substituting for $\tilde{e}_j$, Taylor expanding $K(\cdot)$ around $(x - g(X_i) - e_j)/h$ and using the mean value theorem, we obtain,

$$\begin{aligned}\hat{f}_Y(x) &= \frac{1}{n}\sum_{i=1}^n \Big[\frac{1}{(n-1)h}\sum_{j=2}^n K\big(\frac{x - \tilde{g}(X_i) - \tilde{e_j}}{h}\big)\Big] \\ = \frac{1}{n}\sum_{i=1}^n \Big[\frac{1}{(n-1)h}\sum_{j=2}^n K\big(&\frac{x - g(X_i) - e_j + \tilde{g}(X_j) - g(X_j) - (\tilde{g}(X_i) - g(X_i))}{h}\big)\Big] \\ = \frac{1}{n}\sum_{i=1}^n \Big[\frac{1}{(n-1)h}\sum_{j=2}^n \big[K\big(&\frac{x - g(X_i) - e_j}{h}\big) + K'\big(\frac{x - g(X_i) - e_j}{h}\big) \\ \times\big(&\frac{\tilde{g}(X_j) - g(X_j) - (\tilde{g}(X_i) - g(X_i))}{h}\big) + O_P\big(\big(\frac{\tilde{g}(X_j) - g(X_j) - (\tilde{g}(X_i) - g(X_i))}{h}\big)^2\big)\big]\Big]. \qquad (40)\end{aligned}$$

Thus,

$$\hat{f}_Y(x) - \tilde{f}_Y(x) =$$
$$\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{(n-1)h}\sum_{j=2}^{n}\big[K'\big(\frac{x-g(X_i)-e_j}{h}\big)\cdot\frac{\tilde{g}(X_j)-g(X_j)}{h}$$
$$+K'\big(\frac{x-g(X_i)-e_j}{h}\big)\cdot\frac{g(X_i)-\tilde{g}(X_i)}{h}\big]\Big]$$
$$+\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{(n-1)h}\sum_{j=2}^{n}O_P\big(\big(\frac{\tilde{g}(X_j)-g(X_j)-(\tilde{g}(X_i)-g(X_i))}{h}\big)^2\big)\Big]. \quad (41)$$

Let us first examine the second term in the Taylor expansion above. We observe that for $i = j$ it is zero. For $i \neq j$, if we denote by F_n the common empirical distibution function of X_i and X_j,

$$\frac{1}{h^3}\Big|\frac{1}{n}\sum_{i=1}^{n}\big[\frac{1}{(n-1)}\sum_{j=2}^{n}\big(\tilde{g}(X_j)-g(X_j)-(\tilde{g}(X_i)-g(X_i))\big)\big]^2\Big|$$
$$\leq \frac{1}{h^3}\iint \sup_{x_1\in S(X),x_2\in S(X)}\Big|\big(\tilde{g}(x_2)-g(x_2)-(\tilde{g}(x_1)-g(x_1))\big)^2\Big|\mathrm{d}F_n(x_1)\mathrm{d}F_n(x_2)$$
$$\leq \frac{1}{h^3}\sup_{x_1\in S(X),x_2\in S(X)}\Big|\big(\tilde{g}(x_2)-g(x_2)-(\tilde{g}(x_1)-g(x_1))\big)^2\Big|\iint \mathrm{d}F_n(x_1)\mathrm{d}F_n(x_2)$$
$$= \frac{1}{h^3}\sup_{x_1\in S(X),x_2\in S(X)}\Big|\big(g(x_1)-\tilde{g}(x_1)\big)^2+2\big(g(x_1)-\tilde{g}(x_1)\big)$$
$$\times\big(\tilde{g}(x_2)-g(x_2)\big)+\big(\tilde{g}(x_2)-g(x_2)\big)^2\Big|. \quad (42)$$

Using a uniform convergence result for the Nadarya-Watson estimator and making use of assumptions 5 and 6, see Mack & Silverman (1982),

$$\sup_{x\in S(X)}|\tilde{g}(x)-g(x)| = O_P\Big(\big[\frac{1}{nh_R}\log\big(\frac{1}{h_R}\big)\big]^{1/2}\Big), \quad (43)$$

and a well-known result from order in probability, we obtain that the order of the term in question is

$$O_P\Big(\frac{1}{nh_R\cdot h^3}\log\big(\frac{1}{h_R}\big)\Big), \quad (44)$$

where h_R is defined in (6). Using the same argument as when evaluating (42) it will also be seen that the mean of the absolute value and the standard deviation of this term is of the order given in (44). (See below for the existence of these quantities under the assumption $\inf_{x\in S(X)} f_X(x) > 0$.)

There is a potential to improve on (44), since the evalution is quite crude. An alternative would be to try to evaluate the order of the second order term directly, as will be done presently for the first order term using the convolution property, and then include a third order term which can be evaluated (crudely) as above, resulting in a term $O_p\Big(\big[\frac{1}{nh_Rh^3}\log\frac{1}{h_R}\big]^{3/2}\Big)$. For the crude estimate (44) to be of order $O(\frac{1}{\sqrt{n}})$, we must have

$$h_Rh^3 = O(n^{-1/2-\epsilon}) \text{ for some } \epsilon > 0. \quad (45)$$

Further, let ut examine the first term of the Taylor expansion of $\hat{f}_Y(x) - \tilde{f}_Y(x)$. If this is of larger order than (44), this will be the contributing term. We start by examining the expectation of this term. For the expectation of this term to exist we need the existence of $\mathrm{E}(\tilde{g}(X_i) - g(X_i))$, but using the definition of the Nadaraya-Watson estimator, this exists if the density function f_X is such that $\inf_{x \in S(X)} f_X(x) > 0$. This is assumed in the following. We again note that the expectation disappears for $i = j$, and for $i \neq j$ we have, using independence, for the first part of the first order term

$$\mathrm{E}\Big(\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=2}^{n} K'\big(\frac{x - g(X_i) - e_j}{h}\big) \cdot \big(\frac{\tilde{g}(X_j) - g(X_j)}{h}\big)\Big)$$

$$\sim \frac{1}{h^2} \iiint K'\big(\frac{x - g(x_1) - u}{h}\big)\big(\mathrm{E}(\tilde{g}(x_2)) - g(x_2)\big) f_e(u) f_X(x_1) f_X(x_2) \mathrm{d}u \mathrm{d}x_1 \mathrm{d}x_2. \quad (46)$$

Now Taylor expanding and using a convolution argument, we obtain

$$\begin{aligned}
&\frac{1}{h} \iiint K'(z_1)\big(\mathrm{E}(\tilde{g}(x_2)) - g(x_2)\big) f_e(x - g(x_1) - z_1 h) f_X(x_1) f_X(x_2) \mathrm{d}z_1 \mathrm{d}x_1 \mathrm{d}x_2 \\
&\qquad = \frac{1}{h} \iint K'(z_1)\big(\mathrm{E}(\tilde{g}(x_2)) - g(x_2)\big) f_Y(x - z_1 h) f_X(x_2) \mathrm{d}z_1 \mathrm{d}x_2 = \\
&\qquad\qquad -f_Y'(x) \int z_1 K'(z_1) \mathrm{d}z_1 \int \big(\mathrm{E}(\tilde{g}(x_2)) - g(x_2)\big) f_X(x_2) \mathrm{d}x_2 \\
&\qquad + \frac{h}{2} f_Y''(x) \int z_1^2 K'(z_1) \mathrm{d}z_1 \int \big(\mathrm{E}(\tilde{g}(x_2)) - g(x_2)\big) f_X(x_2) \mathrm{d}x_2 + O(h^4). \quad (47)
\end{aligned}$$

Observe that since the kernel is symmetric, $\int z_1^2 K'(z_1) \mathrm{d}z_1 = 0$, so there is no term of order $O(h^3)$. The whole term is of order $O(h^2)$ through the dependence on $\mathrm{E}\big(\tilde{g}(x_2) - g(x_2)\big)$.

Examining the second part of the first order term in (41), by similar arguments

$$-\mathrm{E}\Big(\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=2}^{n} K'\big(\frac{x - g(X_i) - e_j}{h}\big) \cdot \big(\frac{\tilde{g}(X_i) - g(X_i)}{h}\big)\Big)$$

$$\sim \int z_2 K'(z_2) \mathrm{d}z_2 \int \big(\mathrm{E}(\tilde{g}(x_1)) - g(x_1)\big) f_e'\big(x - g(x_1)\big) f_X(x_1) \mathrm{d}x_1 + O(h^4). \quad (48)$$

In total, the terms (47) and (48) are of order h^2, but can be reduced by higher order kernels.

Further, we examine the variance of the first order term in (41). The condition $\inf_{x \in S(X)} f_X(x) > 0$ again guarantees the existence of this variance. The calculations are similar to the calculations where we found the variance of $\tilde{f}_Y(x)$. The variance in question is,

$$\mathrm{var}\Big(\frac{1}{n(n-1)h^2} \sum_{i=1}^{n} \sum_{j=2}^{n} \Big[K'\big(\frac{x - g(X_i) - e_j}{h}\big)\big(g(X_i) - \tilde{g}(X_i) + \tilde{g}(X_j) - g(X_j)\big)\Big]\Big)$$

$$= \frac{1}{n^2(n-1)^2 h^4} \sum_{i=1}^{n} \sum_{j=2}^{n} \sum_{k=1}^{n} \sum_{l=2}^{n} \mathrm{cov}\Big[K'\big(\frac{x - g(X_i) - e_j}{h}\big)\big(g(X_i) - \tilde{g}(X_i) + \tilde{g}(X_j) - g(X_j)\big),$$

$$K'\big(\frac{x - g(X_k) - e_l}{h}\big)\big(g(X_k) - \tilde{g}(X_k) + \tilde{g}(X_l) - g(X_l)\big)\Big]. \quad (49)$$

The evaluation of these terms can be found in the appendix. It turns out that they are of order $O(h^4/n)$, thus they will only contribute higher order effects to the overall variance of $\hat{f}_Y(x)$.

In total the bias of $\hat{f}_Y(x)$ will consists of the terms (14), (47) and (48). This is of order h^2, as for the kernel density estimator, but as we will see in the next section, a bias improvement may actually occur in some cases. If the bandwidth condition (45) is fulfilled, the total variance has a leading term given by equation (38), i.e. $O(n^{-1})$; it is in fact the rate of the variance for a parametric estimation problem.

This result may seem striking. However, observe that the density of Y is expressed as a smooth function of the densities of X and e. This suggests that the density of Y can be estimated by plugging in estimators of the unknown densities and the unknown function g, in the functional. By the plug-in principle we can expect that this estimator converges at the parametric rate, even though the estimators being plugged in have a slower rate of convergence. Some references for smooth functionals of densities are e.g; Hall & Marron (1987), Birgè & Massart (1995) and Efromovich & Samarov (2000). In these cases the parametric convergence rate $n^{-1/2}$ for the estimatated functionals are obtained.

4 Evaluating the convolution estimator

To evaluate the finite sample properties of the proposed estimator, (9), we use simulation experiments to compare the convolution estimates with the estimates from the classical kernel estimator in (1).

To avoid looking at separate sets of points, the comparisons are based on the mean integrated squared error (MISE) of the two estimators. The MISE for the convolution estimator is

$$\mathrm{MISE}(\hat{f}) = \mathrm{E}\Big[\int_{-\infty}^{\infty} (\hat{f} - f)^2(x)\mathrm{d}x\Big], \tag{50}$$

and likewise for the kernel density estimator. We have used 100 simulated realizations with sample sizes from 100 to 5000 of the model (2), with different choices of the function $g(\cdot)$ and distributions of X and e. The value of MISE is approximated as an average of the ISE (integrated squared error) of the 100 realizations, and ISE is estimated by numerical integration. If the true density f_Y is not known analytically, we have based our comparisons on an estimated kernel density computed from 1 000 000 generated observations of (X_i, Y_i), using this as the "true" density.

The choice of bandwidth in non-parametric estimation has a considerable impact on the accuracy of the estimator. The bandwidth, h_D, used in the kernel density estimation in our simulation study, is the Solve-the-Equation Plug-in Apparoach proposed in Sheather & Jones (1991), and is the same for all the $(n-1)$ density estimations in equation (9). For ease of computation the bandwidth for the kernel smoothing of g is the rule-of-thumb, see e.g. Härdle (1990) page 91, $1.06\min(\hat{\sigma}, R/1.34)n^{-1/5}$, where R is the interquartile range, $\hat{\sigma}$ is the variance of all observations $X_1, ..., X_n$. Thus we may actually obtain better results using a more optimal bandwidth for the non-parametric regression.

The following models are considered:

1. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim \mathrm{N}(0, 0.1)$.
2. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim \mathrm{N}(0,1)$.
3. $g(x) = 3x,\ X \sim \mathrm{N}(1,1),\ e \sim \mathrm{N}(0,1)$.
4. $g(x) = x,\ X \sim \chi^2(3),\ e \sim (\chi^2(3) - 3)$.
5. $g(x) = x^2,\ X \sim \mathrm{U}[0,2],\ e \sim \mathrm{N}(0,1)$.
6. $g(x) = (0.5 + 4e^{-x^2})x,\ X \sim \mathrm{U}[-2,2],\ e \sim \mathrm{N}(0,1)$.
7. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim (\exp(1) - 1)$.
8. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim (\frac{1}{2}\mathrm{N}(-3/2, 1/2) + \frac{1}{2}\mathrm{N}(3/2, 1/2))$.
9. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim \text{Double exponential}(0,1)$.
10. $g(x) = x,\ X \sim \mathrm{N}(1,1),\ e \sim \sum_{l=0}^{2} \frac{2}{7}\mathrm{N}(\frac{12l-15}{7}, \frac{2}{7}) + \sum_{l=8}^{10} \frac{1}{21}\mathrm{N}(\frac{2l}{7}, \frac{1}{21})$.
11. $g(x) = x^2,\ X \sim \mathrm{U}[0,2],\ e \sim (\exp(1) - 1)$.
12. $g(x) = x^2,\ X \sim \mathrm{U}[0,2],\ e \sim (\frac{1}{2}\mathrm{N}(-3/2, 1/2) + \frac{1}{2}\mathrm{N}(3/2, 1/2))$.

Models 1-4 are linear models, with error terms that can be encountered in practice, and model 5 and 6 are non-linear with normally distributed error terms. Models 7-12 are rather unusual and difficult, and seldom met in practice, but we would like to see how the estimator performs in some extreme cases. In most cases of the examples the compactness assumption on f_X is not fulfilled. Actually, we do not believe that this assumption is necessary, and we wanted to check performances in cases where it is violated.

The simulation results are given in table 1. In some simulations only 30 realizations are performed, and those are marked. The table shows the percentage change by using the convolution estimator compared with the kernel density estimator. For the MISE, this change is calculated by

$$\frac{\mathrm{MISE}(f_Y^*) - \mathrm{MISE}(\hat{f}_Y)}{\mathrm{MISE}(f_Y^*)} \cdot 100. \tag{51}$$

for the squared bias

$$\frac{[\mathrm{Ave}(f_Y^* - f_Y)]^2 - [\mathrm{Ave}(\hat{f}_Y - f_Y)]^2}{[\mathrm{Ave}(f_Y^* - f_Y)]^2} \cdot 100, \tag{52}$$

where

$$[\mathrm{Ave}(f_Y^* - f_Y)]^2 = \frac{1}{k}\sum_{j=1}^{k}\big[\big(\frac{1}{100}\sum_{i=1}^{100} f_Y^{*i}(x_j)\big) - f_Y(x_j)\big]^2, \tag{53}$$

and similar for the convolution estimator. In (53) k denotes the number of gridpoints for which the estimators are calculated, usually $k = 500$. Thus $f_Y^{*i}(x_j)$ is the calculated

kernel estimate for the ith realization in gridpoint x_j. Further, $f_Y(x_j)$ denotes the true distribution in gridpoint x_j.

The variance change is calculated as

$$\frac{\hat{\text{var}}(f_Y^*) - \hat{\text{var}}(\hat{f}_Y)}{\hat{\text{var}}(f_Y^*)} \cdot 100, \tag{54}$$

where

$$\hat{\text{var}}(f_Y^*) = \frac{1}{k}\sum_{j=1}^{k}\Big[\frac{1}{99}\Big(\sum_{i=1}^{100}\big(f_Y^{*i}(x_j) - \text{Ave}\{f_Y^*(x_j)\}\big)^2\Big)\Big] \tag{55}$$

and similarly for the convolution estimator. Here $\text{Ave}\{f_Y^*(x_j)\}$ denotes the average of all 100 (or 30) kernel estimates in gridpoint x_j.

A minus sign in the table, thus indicates that the kernel density estimator performs better than the convolution estimator.

In many cases the MISE is smallest for the convolution estimator, however, in cases where the variance of the error terms is relatively small, especially for model 1, the kernel density estimator is best. This is not unexpected, since the convolution effect will not be large here.

In the non-linear models 5 and 6 with normally distributed error terms, the convolution estimator is clearly better. But introducing asymmetric and multimodal distributions on the error terms, as in model 7, 8 and 10, the convolution estimator seems to have some trouble. However, the results from model 11 and 12 also indicates that different g functions may have some influence on the estimator's behaviour, since in these cases the convolution estimator starts to perform better as sample size increases, even though we have asymmetric and bimodal error term distributions.

The variance for the convolution estimator is smaller for many simulations, and when the sample size increases, the improvements are also increasing, as expected. The squared bias is smallest almost for all cases for the convolution estimator. This is somewhat suprising since from the asymptotic analysis it is of the same order as the kernel estimator. Figure 1 shows the estimated variance and bias for the two estimators from the simulation in model 2 with sample size 100. The upper plot shows that the variance for the convolution estimator is smallest, as expected. The bias for the kernel density estimator is,

$$\text{E}\big(f_Y^*(x)\big) - f_Y(x) = \frac{h^2}{2}f_Y''(x)\int w^2K(w)\text{d}w + O(h^2), \tag{56}$$

and it behaves as we would expect, since it is proportional to the second derivative of the density in question, here a normal distribution with mean equal to one. The bias for the convolution estimator behaves quite differently, and overall it is considerably smaller. This difference can in fact be explained by the following reasoning.

Since f_X and f_e are normal distributions, it also means that the true f_Y will be normal with mean equal to one and variance equal to two. From this information it is possible to calculate the exact expressions for the bias of the convolution estimator and compare it to the observed bias in figure 1. The bias of the convolution estimator

consists of three terms, (14), (47) and (48). Equation (14) is now

$$\mathrm{E}\big(\tilde{f}_Y(x)\big) - f_Y(x) = \frac{h^2}{2} f_Y''(x) \int w^2 K(w)\mathrm{d}w$$

$$= \frac{h^2}{2} \int w^2 K(w)\mathrm{d}w \Big[- \frac{1}{4\sqrt{\pi}} \exp\{-1/4(x-1)^2\}(1 - \frac{(x-1)^2}{16\sqrt{\pi}}) \Big]. \qquad (57)$$

This expression is identical to the bias for the kernel density estimator.

In equation (47) and (48), the bias of the Nadaraya-Watson estimator of $g(x)$ is a part of the expression. This bias is well-known and is

$$\mathrm{E}\big(\tilde{g}(x)\big) - g(x) = h^2 \Big(\frac{1}{2} g''(x) + \frac{g'(x) f_X'(x)}{f_X(x)} \Big) \int u^2 K(u)\mathrm{d}u. \qquad (58)$$

Since $\int u^2 K(u)\mathrm{d}u = 1$, $g(x) = x$ and f_X is a normal distribution, this expression will be equal to $-2(x-1)h^2$. Inserting this in equation (47) and again using the fact that f_X is normal with mean and variance equal to one gives, for the leading term in (47),

$$-f_Y'(x) \int z_1 K'(z_1)\mathrm{d}z_1 \int \big(- 2(x_2 - 1)h^2 \frac{1}{\sqrt{2\pi}} \exp(-(x_2-1)^2/2) \big)\mathrm{d}x_2. \qquad (59)$$

Observe that the last integral above is equal to zero. Thus we have no contribution to the bias from this expression.

Further, equation (48) yields,

$$h^2 \int z_2 K'(z_2)\mathrm{d}z_2 \int \big((-2(x_1-1)(\frac{-(x-x_1)}{\sqrt{2\pi}} \exp(-(x-x_1)^2/2)) \frac{1}{\sqrt{2\pi}} \exp(-(x_1-1)^2/2) \big)\mathrm{d}x_1. \qquad (60)$$

If we choose to use a Gaussian kernel function with mean zero and variance one, then $\int z_2 K'(z_2)\mathrm{d}z_2$ is equal to minus one. Thus, the total bias of the convolution estimator will in this case be

$$\mathrm{E}\big(\hat{f}_Y(x)\big) - f_Y(x) = \frac{h^2}{2} \Big[- \frac{1}{4\sqrt{\pi}} \exp\{-1/4(x-1)^2\}(1 - \frac{(x-1)^2}{16\sqrt{\pi}}) \Big]$$

$$-h^2 \int \big((-2(x_1 - 1)(\frac{-(x-x_1)}{\sqrt{2\pi}} \exp(-(x-x_1)^2/2)) \frac{1}{\sqrt{2\pi}} \exp(-(x_1-1)^2/2) \big)\mathrm{d}x_1. \qquad (61)$$

This expression is plotted in figure 2, with a reasonable choice for the bandwidth, $h = 0.3$. And taking the different scaling into account, this graph is comparable to the empirical bias from figure 1. Similar explanations are possible for the other models, although some problems arise in the computation in the cases where the true density is not known.

As suggested in section 2, other nonparametric regression estimators may be used to estimate $g(x)$. In table 2, results from simulations from model 2, using the local linear estimator for estimating $g(x)$ are given. These results are in most cases better than the corresponding results using the Nadaraya-Watson estimator, given in table 1.

A real data set has been studied for completeness. The data is the motorcycle data set, from Härdle (1990) page 70. The X-values represent time after a simulated impact

Model	Sample size	Squared bias	Variance	MISE
1	100	61.5 %	-665.0 %	-504.7 %
1	500	79.9 %	-464.7 %	-346.4 %
1	5000 (30)	-638.8 %	-265.9 %	-314.4 %
2	100	79.2 %	23.5 %	34 %
2	500	97.7 %	48.2 %	59.2 %
2	5000 (30)	97.8 %	77.8 %	81.9 %
3	100	93.4 %	-116.0 %	-80.2 %
3	500	93.1 %	-40.6 %	-13.9 %
3	5000 (30)	43.7 %	15.9 %	19.6 %
4	100	62.1 %	19.1 %	27.5 %
4	500	68.2 %	1.6 %	17.7 %
4	5000	80.3 %	21.9 %	32.5 %
5	100	94.6 %	28.0 %	34.9 %
5	500	83.9 %	42.4 %	50.9 %
5	5000	79.4 %	60.2 %	64.4 %
6	100	87.6 %	-4.2 %	15.5 %
6	500	83.4 %	33.3 %	44.9 %
6	5000 (30)	79.1 %	62.2 %	64.3 %
7	100	68.6 %	-44.8 %	-30.0 %
7	500	60.5 %	-58.3 %	-42.1 %
7	5000	69.3 %	-117.4 %	-88.5 %
8	100	-64.3 %	-15.7 %	-25.7 %
8	500	25.7 %	-62.5 %	-43.1 %
8	5000	62.4 %	-69.9 %	-43.1 %
9	100	97.5 %	25.0 %	37.2 %
9	500	96.4 %	31.5 %	43.4 %
9	5000	86.8 %	47.5 %	55.9 %
10	100	-18.5 %	1.7 %	-3.1 %
10	500	13.6 %	-130.3 %	-96.9 %
10	5000	54.5 %	-319.4 %	-259.6 %
11	100	49.1 %	-37.9 %	-9.1 %
11	500	58.1 %	-33.2 %	-0.8 %
11	5000	57.5 %	-38.5 %	-20.1 %
12	100	-49.3 %	-15.3 %	-25.5 %
12	500	-23.1 %	-14.9 %	-16.9 %
12	5000	70.1 %	-5.8 %	2.0 %

Table 1: Percentage improvements in estimations using the convolution density estimator compared with the kernel density estimator. 100 realizations (in some cases 30, as indicated). The MISE, squared bias and variance are explained in formula (51), (52), (54), and below. A minus sign indicates that the kernel density estimator performs best.

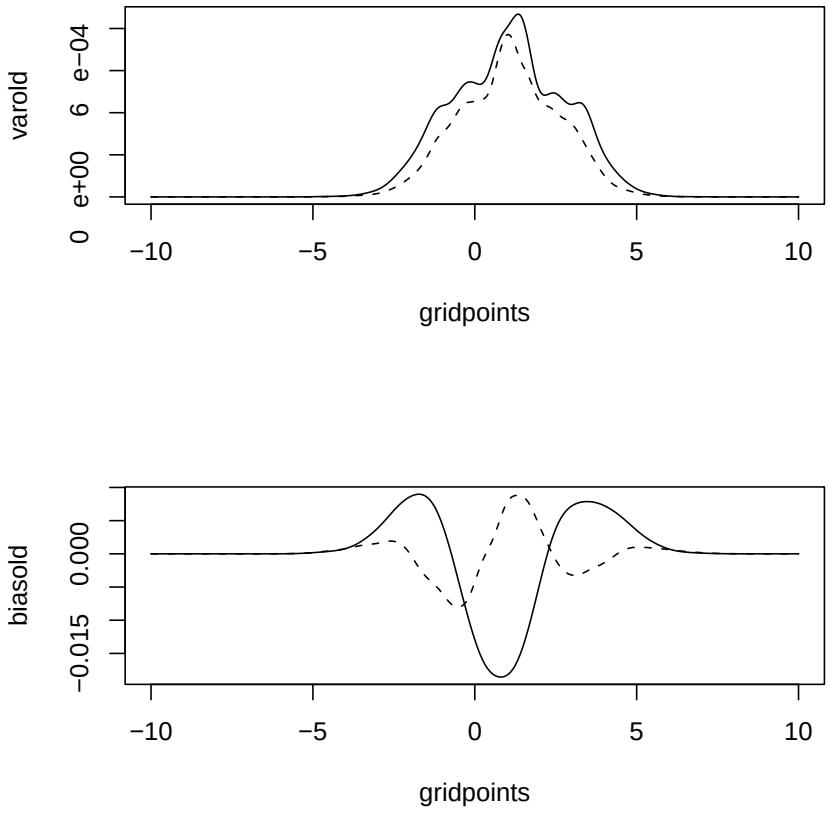


Figure 1: The estimated variance (top) and bias for model 2 (dashed line - convolution estimator, solid line - kernel density estimator).

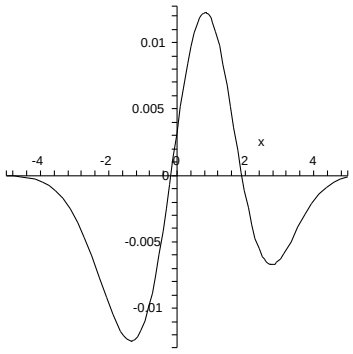


Figure 2: The theoretical bias of the convolution estimator in model 2.

Sample size	Squared bias	Variance	MISE
100	92.6 %	36.7 %	49.8 %
500	89.2 %	58.9 %	66.2 %
5000	85.8 %	70.9 %	74.3 %

Table 2: Percentage improvements in estimations using convolution density estimator with local linear estimator, compared with kernel density estimator. Model 2.

with motorcycles and the response variable Y is the head acceleration of a post human test object. The density of the response Y has been estimated by the kernel density estimator, where the bandwidth is the rule-of-thumb given in Härdle (1990), and the convolution estimator. The estimated densities are given in figure 3. It seems that the convolution estimator smooths more than the kernel density estimator, but both estimators seem to give reasonable results.

5 Conclusions

The proposed convolution density estimator seems to outperform the usual kernel estimator in many situations, especially if the error term density function is smooth and has a relatively large variance. The cases where it does not perform so well are not necessarily the most important ones in practice.

One should expect that if the g-function is more correctly estimated, then a better density estimate will be obtained. Thus using e.g. local polynomial regression may improve the density estimation. Also, by selecting the bandwidth parameters in the convolution estimator more correctly, by e.g. a cross-validation technique, we could possibly improve the estimates even more. We also believe that this estimator can

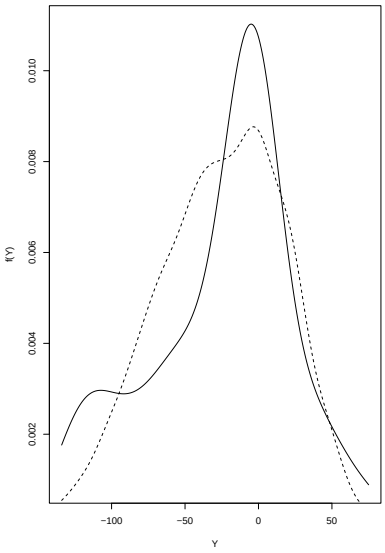


Figure 3: The estimated densities of a real data set (solid line - kernel and dashed line - convolution)

be used in a more general time-series setting, $X_t = g(X_{t-1}) + e_t$, where the marginal density of the process X_t is of interest. Some simulation experiments indicate that the convolution estimator will outperform the kernel density estimator, and that the order of the variance of the convolution estimator will also be n^{-1} in this situation, see Støve & Tjøstheim (2005*a*).

References

Birgè, L. & Massart, P. (1995), 'Estimation of integral functionals of a density', *Annals of Statistics* **23**, 11–29.

Efromovich, S. & Samarov, A. (2000), 'Adaptive estimation of the integral of squared regression derivatives', *Scandinavian Journal of Statistics* **27**, 335–351.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall.

Frees, E. W. (1994), 'Estimating densities of functions of observations', *Journal of the American Statistical Association* **89**, 517–525.

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

Härdle, W. (1990), *Smoothing Techniques: With Implementation in S*, Springer-Verlag.

Mack, Y. P. & Silverman, B. W. (1982), 'Weak and strong uniform consistency of kernel regression estimates', *Zeitschrift für Wahrscheinlichkeitstheorie verw. Gebiete* **61**, 405–415.

Saavedra, A. & Cao, R. (1999), 'Rate of convergence of a convolution-type estimator of the marginal density of a MA(1) process', *Stochastic Processes and their Applications* **80**, 129–155.

Saavedra, A. & Cao, R. (2000), 'On the estimation of the marginal density of a moving average process', *The Canadian Journal of Statistics* **28**, 799–815.

Schick, A. & Wefelmeyer, W. (2004*a*), 'Root n consistent and optimal density estimators for moving average processes', *Scandinavian Journal of Statistics* **31**, 63–78.

Schick, A. & Wefelmeyer, W. (2004*b*), 'Root n consistent density estimators for sums of independent random variables', *Nonparametric Statistics* **16**, 925–935.

Sheather, S. J. & Jones, M. C. (1991), 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society. Series B* **53**, 683–690.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, Springer-Verlag.

Støve, B. & Tjøstheim, D. (2005*a*), 'A convolution density estimator for nonlinear time series: Simulations and some preliminary analysis', Work in progress.

Støve, B. & Tjøstheim, D. (2005*b*), 'A new convolution estimator for nonparametric regression', Submitted.

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

Appendix

The variance in (49) can be classified into the following type of terms (see also Saavedra & Cao (2000)):

$$\frac{1}{n^2(n-1)^2h^4}\Big[(n-1)\mathrm{var}\Big(K'\big(\frac{x-g(X_1)-e_1}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_1\ )-g(X_1)\big)\Big) \quad (62)$$

$$+(n-1)(n-2)\mathrm{var}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big)\Big) \quad (63)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_1)-e_3}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_3)-g(X_3)\big)\Big) \quad (64)$$

$$+2(n-1)(n-2)(n-3)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_3)-e_1}{h}\big)\big(g(X_3)-\tilde{g}(X_3)+\tilde{g}(X_1)-g(X_1)\big)\Big) \quad (65)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_3)-e_2}{h}\big)\big(g(X_3)-\tilde{g}(X_3)+\tilde{g}(X_2)-g(X_2)\big)\Big) \quad (66)$$

$$+(n-1)(n-2)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_2)-e_1}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_1)-g(X_1)\big)\Big) \quad (67)$$

$$+2(n-1)(n-2)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_1}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_1)-g(X_1)\big),$$
$$K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big)\Big) \quad (68)$$

$$+2(n-1)(n-2)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_1}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_1)-g(X_1)\big),$$
$$K'\big(\frac{x-g(X_2)-e_1}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_1)-g(X_1)\big)\Big) \quad (69)$$

$$+(n-1)(n-2)(n-3)(n-4)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_3)-e_4}{h}\big)\big(g(X_3)-\tilde{g}(X_3)+\tilde{g}(X_4)-g(X_4)\big)\Big) \quad (70)$$

$$+(n-1)(n-2)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_1}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_1)-g(X_1)\big),$$
$$K'\big(\frac{x-g(X_2)-e_2}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_2)-g(X_2)\big)\Big) \quad (71)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K'\big(\frac{x-g(X_1)-e_1}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_1)-g(X_1)\big),$$
$$K'\big(\frac{x-g(X_2)-e_3}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_3)-g(X_3)\big)\Big) \quad (72)$$

$$+(n-1)(n-2)(n-3)\mathrm{cov}\Big(K'\big(\frac{x-g(X_2)-e_1}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_1)-g(X_1)\big),$$
$$K'\big(\frac{x-g(X_1)-e_3}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_3)-g(X_3)\big)\Big)\Big]. \quad (73)$$

The terms (70), (71) and (72) are neglible by independence, and the terms (62), (68) and (69) are automatically zero. Hence we study the remaining six terms, using $\mathrm{var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2$ and $\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.

Examining the term (63),

$$(n-1)(n-2)\mathrm{var}\Big(K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big)\Big) \doteq (n-1)(n-2)\mathrm{var}(U). \quad (74)$$

Starting by examining $\mathrm{E}(U^2)$, using the same techniques as when deriving (47) and (48), and since $\int z\big(K'(z)\big)^2\mathrm{d}z = 0$, we obtain,

$$\begin{aligned}\mathrm{E}(U^2) \sim \iiint K'\big(\frac{x-g(v)-u}{h}\big)^2\mathrm{E}\big[\big(g(v)-\tilde{g}(v)+\tilde{g}(w)-g(w)\big)^2\big] \\ \times f_e(u)f_X(v)f_X(w)\mathrm{d}u\mathrm{d}v\mathrm{d}w = h\iiint K'(z)^2\mathrm{E}\big[\big(g(v)-\tilde{g}(v)+\tilde{g}(w)-g(w)\big)^2\big] \\ \times f_X(v)f_X(w)f_e(x-g(v)-zh)\mathrm{d}v\mathrm{d}w\mathrm{d}z = h\iiint K'(z)^2 \\ \times\mathrm{E}\big[\big(g(v)-\tilde{g}(v)+\tilde{g}(w)-g(w)\big)^2\big]f_X(v)f_X(w)f_e(x-g(v))\mathrm{d}v\mathrm{d}w\mathrm{d}z + O(h^5). \quad (75)\end{aligned}$$

Next, using the same technique,

$$\begin{aligned}\big(\mathrm{E}(U)\big)^2 \sim h^2\Big[\iiint K'(z)\mathrm{E}\big[g(v)-\tilde{g}(v)+\tilde{g}(w)-g(w)\big] \\ \times f_X(v)f_X(w)f_e(x-g(v))\mathrm{d}v\mathrm{d}w\mathrm{d}z\Big]^2 + O(h^5). \quad (76)\end{aligned}$$

Thus the leading term of (63) will be $O((n-1)(n-2)h^4)$, since the expectation in the term (76) is the bias of the Nadarya-Watson estimator, which is $O(h^2)$, see e.g. Härdle (1990) page 135.

Examining the term (64)

$$\begin{aligned}(n-1)(n-2)(n-3)\mathrm{cov}\Big[K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big), \\ K'\big(\frac{x-g(X_1)-e_3}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_3)-g(X_3)\big)\Big] \\ \doteq (n-1)(n-2)(n-3)\mathrm{cov}(U,V). \quad (77)\end{aligned}$$

Calculating $\mathrm{E}(UV)$, by conditioning, substituting, Taylor expanding, and since $\int K'(z)\mathrm{d}z =$

0, we obtain,

$$\begin{aligned}
\mathrm{E}(UV) \sim \iiiint \int K'\big(\frac{x-g(v)-u}{h}\big)\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(w)\big)-g(w)\Big) \\
\times K'\big(\frac{x-g(v)-y}{h}\big)\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(a)\big)-g(a)\Big) \\
\times f_e(u)f_e(y)f_X(v)f_X(w)f_X(a)\mathrm{d}u\mathrm{d}y\mathrm{d}v\mathrm{d}w\mathrm{d}a \\
= h^2 \iiiint \int K'(z_1)K'(z_2)\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(w)\big)-g(w)\Big) \\
\times\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(a)\big)-g(a)\Big)f_X(v)f_X(w)f_X(a) \\
\times f_e(x-g(v)-z_1h)f_e(x-g(v)-z_2h)\mathrm{d}v\mathrm{d}w\mathrm{d}a\mathrm{d}z_1\mathrm{d}z_2 \\
= h^2 \iiiint \int K'(z_1)K'(z_2)\Big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}\big(\tilde{g}(w)\big)-g(w)\Big) \\
\times\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(a)\big)-g(a)\Big)f_X(v)f_X(w)f_X(a) \\
\times\big(f_e(x-g(v))-z_1hf_e'(x-g(v))+\frac{z_1^2h^2}{2}f_e''(x-g(v))+O(h^2)\big) \\
\times\big(f_e(x-g(v))-z_2hf_e'(x-g(v))+\frac{z_2^2h^2}{2}f_e''(x-g(v))+O(h^2)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}a\mathrm{d}z_1\mathrm{d}z_2 \\
= h^4 \iiiint \int z_1K'(z_1)z_2K'(z_2)\Big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\Big) \\
\times\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(a)\big)-g(a)\Big)f_X(v)f_X(w)f_X(a) \\
\times f_e'\big(x-g(v)\big)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}a\mathrm{d}z_1\mathrm{d}z_2+O(h^9). \qquad (78)
\end{aligned}$$

Further,

$$\begin{aligned}
\mathrm{E}(U) \sim -h^2 \iiint z_1K'(z_1)\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(w)\big)-g(w)\Big) \\
\times f_X(v)f_X(w)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1+O(h^5), \qquad (79)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(V) \sim -h^2 \iiint z_2K'(z_2)\Big(g(v)-\mathrm{E}\big(\tilde{g}(v)\big)+\mathrm{E}\big(\tilde{g}(a)\big)-g(a)\Big) \\
\times f_X(v)f_X(a)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}a\mathrm{d}z_2+O(h^5). \qquad (80)
\end{aligned}$$

In total, the leading term of (64) is $O((n-1)(n-2)(n-3)h^8)$, again using the order of the bias of the Nadaraya-Watson estimator.

Examining the term (65),

$$\begin{aligned}
2(n-1)(n-2)(n-3)\mathrm{cov}\Big[K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big), \\
K'\big(\frac{x-g(X_3)-e_1}{h}\big)\big(g(X_3)-\tilde{g}(X_3)+\tilde{g}(X_1)-g(X_1)\big)\Big] \\
\doteq 2(n-1)(n-2)(n-3)\mathrm{cov}(U,V). \qquad (81)
\end{aligned}$$

We have

$$\begin{aligned}\mathrm{E}(UV) \sim \iiiint \int & K'\big(\frac{x-g(v)-u}{h}\big)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)\\ & \times K'\big(\frac{x-g(a)-y}{h}\big)\big(g(a)-\mathrm{E}(\tilde{g}(a))+\mathrm{E}(\tilde{g}(v))-g(v)\big)\\ & \times f_e(u)f_e(y)f_X(v)f_X(w)f_X(a)\mathrm{d}u\mathrm{d}y\mathrm{d}v\mathrm{d}w\mathrm{d}a\\ = h^4 \iiiint \int & z_1K'(z_1)z_2K'(z_2)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)\\ & \times\big(g(a)-\mathrm{E}(\tilde{g}(a))+\mathrm{E}(\tilde{g}(v))-g(v)\big)f_X(v)f_X(w)f_X(a)\\ & \times f_e'\big(x-g(v)\big)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}a\mathrm{d}z_1\mathrm{d}z_2+O(h^9), \end{aligned} \tag{82}$$

$$\begin{aligned}\mathrm{E}(U) \sim -h^2 \iiint & z_1K'(z_1)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)\\ & \times f_X(v)f_X(w)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1+O(h^5), \end{aligned} \tag{83}$$

and

$$\begin{aligned}\mathrm{E}(V) \sim -h^2 \iiint & z_2K'(z_2)\big(g(a)-\mathrm{E}(\tilde{g}(a))+\mathrm{E}(\tilde{g}(b))-g(b)\big)\\ & \times f_X(a)f_X(b)f_e'\big(x-g(a)\big)\mathrm{d}a\mathrm{d}b\mathrm{d}z_2+O(h^5). \end{aligned} \tag{84}$$

The above calculations give that the order of the term (65) is $O((n-1)(n-2)(n-3)h^8)$.

Examining the term (73),

$$\begin{aligned}2(n-1)(n-2)(n-3)\mathrm{cov}\Big[& K'\big(\frac{x-g(X_2)-e_1}{h}\big)\big(g(X_2)-\tilde{g}(X_2)+\tilde{g}(X_1)-g(X_1)\big),\\ & K'\big(\frac{x-g(X_1)-e_3}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_3)-g(X_3)\big)\Big]\\ & \doteq 2(n-1)(n-2)(n-3)\mathrm{cov}(U,V). \end{aligned} \tag{85}$$

As before,

$$\begin{aligned}\mathrm{E}(UV) \sim h^4 \iiiint \int & z_1K'(z_1)z_2K'(z_2)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)\\ & \times\big(g(w)-\mathrm{E}(\tilde{g}(w))+\mathrm{E}(\tilde{g}(b))-g(b)\big)f_X(v)f_X(w)f_X(b)\\ & \times f_e'\big(x-g(w)\big)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}b+O(h^9). \end{aligned} \tag{86}$$

Further,

$$\begin{aligned}\mathrm{E}(U) \sim -h^2 \iiint & z_1K'(z_1)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)\\ & \times f_X(v)f_X(w)f_e'\big(x-g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1+O(h^5). \end{aligned} \tag{87}$$

and

$$\begin{aligned}\mathrm{E}(V) \sim -h^2 \iiint & z_2K'(z_2)\big(g(a)-\mathrm{E}(\tilde{g}(a))+\mathrm{E}(\tilde{g}(b))-g(b)\big)\\ & \times f_X(a)f_X(b)f_e'\big(x-g(a)\big)\mathrm{d}a\mathrm{d}b\mathrm{d}z_2+O(h^5). \end{aligned} \tag{88}$$

Thus the term (73) is of order $O((n-1)(n-2)(n-3)h^8)$.

Examining the term (66),

$$2(n-1)(n-2)(n-3)\mathrm{cov}\Big[K'\big(\frac{x-g(X_1)-e_2}{h}\big)\big(g(X_1)-\tilde{g}(X_1)+\tilde{g}(X_2)-g(X_2)\big),$$
$$K'\big(\frac{x-g(X_3)-e_2}{h}\big)\big(g(X_3)-\tilde{g}(X_3)+\tilde{g}(X_2)-g(X_2)\big)\Big]$$
$$\doteq 2(n-1)(n-2)(n-3)\mathrm{cov}(U,V). \quad (89)$$

We have

$$\mathrm{E}(UV)\sim\iiiint K'\big(\frac{x-g(v)-u}{h}\big)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)$$
$$\times K'\big(\frac{x-g(a)-u}{h}\big)\big(g(a)-\mathrm{E}(\tilde{g}(a))+\mathrm{E}(\tilde{g}(w))-g(w)\big)$$
$$\times f_e(u)f_X(v)f_X(w)f_X(a)\mathrm{d}u\mathrm{d}v\mathrm{d}w\mathrm{d}a. \quad (90)$$

Introducing $(g^{-1})'=r$ and $f_X(g^{-1})=l_X$ as in the derivation of (33), and noting that

$$r'\big(g(v)\big)=-\frac{g''(v)}{g'(v)^2} \quad (91)$$

and

$$l_X'\big(g(v)\big)=f_X'(v), \quad (92)$$

we obtain by using $\int K'(z)\mathrm{d}z=0$,

$$\mathrm{E}(UV)\sim h^2\iiiint K'(z_1)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)K'(z_2)$$
$$\times\Big[g\Big(g^{-1}\big(g(v)+h(z_1-z_2)\big)\Big)-\mathrm{E}\Big(\tilde{g}\big(g^{-1}(g(v)+h(z_1-z_2))\big)\Big)$$
$$+\mathrm{E}(\tilde{g}(w))-g(w)\Big]f_e(x-g(v)-z_1h)f_X(v)f_X(w)l_X\big(g(v)+h(z_1-z_2)\big)$$
$$\times r\big(g(v)+h(z_1-z_2)\big)\mathrm{d}z_1\mathrm{d}v\mathrm{d}w\mathrm{d}z_2=h^4\iiiint K'(z_1)$$
$$\times\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)K'(z_2)$$
$$\times\Big[g\big(g^{-1}(g(v)+h(z_1-z_2))\big)-\mathrm{E}\Big(\tilde{g}\big(g^{-1}(g(v)+h(z_1-z_2))\big)\Big)$$
$$+\mathrm{E}(\tilde{g}(w))-g(w)\Big]f_X(v)f_X(w)\Big\{z_1z_2r'\big(g(v)\big)l_X\big(g(v)\big)f_e'\big(x-g(v)\big)$$
$$-2z_1z_2r'\big(g(v)\big)f_e\big(x-g(v)\big)l_X'\big(g(v)\big)$$
$$+z_1z_2f_e'\big(x-g(v)\big)r\big(g(v)\big)l_X'\big(g(v)\big)\Big\}\mathrm{d}v\mathrm{d}w\mathrm{d}z_1\mathrm{d}z_2+O(h^9)=$$
$$h^4\iiiint K'(z_1)\big(g(v)-\mathrm{E}(\tilde{g}(v))+\mathrm{E}(\tilde{g}(w))-g(w)\big)K'(z_2)$$
$$\times\Big[g\big(g^{-1}(g(v)+h(z_1-z_2))\big)-\mathrm{E}\Big(\tilde{g}\big(g^{-1}(g(v)+h(z_1-z_2))\big)\Big)+\mathrm{E}(\tilde{g}(w))-g(w)\Big]$$
$$\times f_X(v)f_X(w)\Big\{-z_1z_2\frac{g''(v)}{g'(v)^2}f_X(v)f_e'\big(x-g(v)\big)+2z_1z_2\frac{g''(v)}{g'(v)^2}f_X'(v)f_e\big(x-g(v)\big)$$
$$+z_1z_2f_e'\big(x-g(v)\big)\frac{f_X'(v)}{g'(v)}\Big\}\mathrm{d}v\mathrm{d}w\mathrm{d}z_1\mathrm{d}z_2+O(h^9). \quad (93)$$

Moreover,

$$\mathrm{E}(U) \sim -h^2 \iiint z_1 K'(z_1)\big(g(v) - \mathrm{E}(\tilde{g}(v)) + \mathrm{E}(\tilde{g}(w)) - g(w)\big) \times f_X(v) f_X(w) f_e'\big(x - g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1 + O(h^5). \tag{94}$$

and

$$\mathrm{E}(V) \sim -h^2 \iiint z_2 K'(z_2)\big(g(a) - \mathrm{E}(\tilde{g}(a)) + \mathrm{E}(\tilde{g}(b)) - g(b)\big) \times f_X(a) f_X(b) f_e'\big(x - g(a)\big)\mathrm{d}a\mathrm{d}b\mathrm{d}z_2 + O(h^5). \tag{95}$$

Thus the term (66) is of order $O((n-1)(n-2)(n-3)h^8)$.

Finally, examining the term (67),

$$\begin{aligned}(n-1)(n-2)\mathrm{cov}\Big(K'\big(\frac{x - g(X_1) - e_2}{h}\big)\big(g(X_1) - \tilde{g}(X_1) + \tilde{g}(X_2) - g(X_2)\big),\\ K'\big(\frac{x - g(X_2) - e_1}{h}\big)\big(g(X_2) - \tilde{g}(X_2) + \tilde{g}(X_1) - g(X_1)\big)\Big)\\ \doteq (n-1)(n-2)\mathrm{cov}(U, V).\end{aligned} \tag{96}$$

Starting with $\mathrm{E}(UV)$,

$$\begin{aligned}\mathrm{E}(UV) \sim h^4 \iiiint z_1 K'(z_1) z_2 K'(z_2)\big(g(v) - \mathrm{E}(\tilde{g}(v)) + \mathrm{E}(\tilde{g}(w)) - g(w)\big)\\ \times\big(g(w) - \mathrm{E}(\tilde{g}(w)) + \mathrm{E}(\tilde{g}(v)) - g(v)\big) f_X(v) f_X(w)\\ \times f_e'\big(x - g(v)\big) f_e'\big(x - g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1\mathrm{d}z_2 + O(h^9).\end{aligned} \tag{97}$$

As before,

$$\mathrm{E}(U) \sim -h^2 \iiint z_1 K'(z_1)\big(g(v) - \mathrm{E}(\tilde{g}(v)) + \mathrm{E}(\tilde{g}(w)) - g(w)\big) \times f_X(v) f_X(w) f_e'\big(x - g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_1 + O(h^5). \tag{98}$$

and

$$\mathrm{E}(V) \sim -h^2 \iiint z_2 K'(z_2)\big(g(w) - \mathrm{E}(\tilde{g}(w)) + \mathrm{E}(\tilde{g}(v)) - g(v)\big) \times f_X(w) f_X(v) f_e'\big(x - g(v)\big)\mathrm{d}v\mathrm{d}w\mathrm{d}z_2 + O(h^5). \tag{99}$$

Thus, the leading term of (67) is $O((n-1)(n-2)h^8)$.

All of these calculations give us that the leading terms of the variance from equation (49) will be of order $O(h^4/n)$.