Journal of Numerical Mathematics and Stochastics*, 10** (1) : 20-31, 2018 © ***JNM@S
http://www.jnmas.org/jnmas10-2.pdf Euclidean Press, LLC
Online: ISSN 2151-2302

Kernel Adjusted Conditional Density Estimation

Z. CHAKHCHOUKH, and D. YAHIA

Laboratory of Applied Mathematics, Biskra University, POBox 145, Algeria, E-mail: yahia_dj@yahoo.fr

Abstract. *In this paper, we propose a new kernel adjusted method in nonparametric conditional density estimation. Asymptotic properties of the new estimator are determined and the optimal choice of the smoothing parameters is presented. In a simulation study, the proposed method shows a good performance and indicates that it is better than existing comparable methods for all sample sizes.*

Key words : Conditional Density Estimation, Kernel Estimation, Mean Squared Error.

AMS Subject Classifications : 62G07, 62G20

1. Introduction

Let $\left\{(X_i, Y_i) i \ \geq 1\right\}$ be a bivariate random variable (rv) with a common probability density function $f(.,.)$. Our interest in this paper is in the estimation of the conditional density function $f(y|x)$ of Y given X. This function describes a comprehensive relationship between responses and explanatory variables, and can be considered as general case of regression. We can give a simple definition to the conditional density function as follows :

$$f(y|x) = \frac{f(x,y)}{f(x)}, \quad \text{for } f_X(x) \neq 0. \tag{1}$$

where $f(.)$ is a marginal density of X, and conceive the conditional density function as a source for various statistical quantities such as : mean, prediction interval, moments, distribution, quantile and so on.

The conditional density estimation was introduced by Rosenblatt in [12], and a bias correction was proposed by Hyndman et al. in [8]. Fan et al. proposed in [4] a direct estimator based on local polynomial estimation. The natural kernel estimator of $f(y|x)$, see [15], is

$$f_n(y|x) = \frac{\hat{f}(x,y)}{\hat{f}(x)}, \tag{2}$$

where

$$f_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)H_h(y - Y_i)$$

is the kernel estimator of $f(x,y)$, and

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \quad (3)$$

is the kernel estimator of $f(x)$ and $K_h(.) := K(./h)$, $H_h(.) := H(./h)$, where $K(.)$ and $H(.)$ are known as Tow kernel functions (usually a bounded and symmetric pdf). Some examples of very common kernel functions are the Epanechnikov and the Gaussian kernel. The parameter $h := h_n$ is called the smoothing parameter, or the bandwidth, and it controls the smoothness of the resulting estimation. In practice, the value of h depends on the sample size and satisfies the condition $h \to 0$ and $nh \to \infty$ as $n \to \infty$. From the principal properties of estimator (2) that are

usually studied, we mention only some of them. We start with recalling that, under standard regularity conditions, the conditional density estimator has the asymptotic bias and variance:

$$bias(f_n(y|x)) = \frac{h^2}{2}\eta_K^2\left(2\,\frac{f'(x)}{f(x)}\frac{\partial f(y|x)}{\partial x} + \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{\partial^2 f(y|x)}{\partial y^2}\right) + O\left(h^4\right) + O(1/nh), \quad (4)$$

and

$$var(f_n(y|x)) = \frac{\mu(K)\,f(y|x)}{nh^2 f(x)}(\mu(K) - hf(y|x)) + O(1/h), \quad (5)$$

where $\mu(K) := \int K^2(t)dt$ and $\eta_K^2 := \int t^2 K(t)dt < \infty$.

Adding the squared bias (4) to the variance (5) gives the asymptotic mean square error $(AMSE)$:

$$AMSE(f_n(y|x)) := bias^2(f_n(y|x)) + var(f_n(y|x)).$$

The optimal choice of h minimizing the last expression satisfies $h_{opt} = cn^{-1/6}$ for $c > 0$.

The most important of the previous facts is how to get an h_{opt} , or in other words how to get an appropriate c. There are some methods as to how to obtain the value of c. The first possibility consists in getting the estimator of higher order derivatives of $f(x)$ and $f(y|x)$ by putting the value subjectively preliminary of h. The second possibility is to apply the method of cross-validation. It should be noted, in this respect, that Loader gave in [9] a brief discussion about cross-validation. The previous two methods are considered as solutions to the problem of searching for a suitable value for h, but they contain some shortcomings. In the first possibility, the estimator is unstable when the sample size is small or moderate (see, [5]). With all these difficulties, Srihera and Stute proposed and studied in [14] a new methodology on density estimation, to extend in approach of regression, in aim to avoid the higher order derivatives. Their methodology relies on the choice of h and also on the choice of K which plays a very important role in the efficiency of estimator when the kernel function satisfies the sufficient usual conditions. More precisely, this methodology of density estimation, [14], [3], was based on classical kernel density that will become a function of location scale family.

In this paper, we develop a new kernel type estimator of the conditional density function, in which the kernel is adapted to the data but not fixed. The new method loads to an adaptive choice of the smoothing parameters. The basic technique of construction of the proposed estimator is kind of a location-scale transformation, with reduced bias and variance, which produces good results in all studied situations.

In Section 2, the proposed estimator is reported and its bias, variance, mean squared error and distribution behavior are determined and presented. In Section 3, simulation studies are undertaken to test the performance of the proposed estimator, and compare it with the usual conditional density estimator.

2. Main Results

2.1. **The proposed estimator**

We follow the same methodology as in [14] and [3], but specialize in this study in the estimation of the conditional density function. Starting from the formula (2), we give a new definition of the conditional density estimator. The procedure followed is to make a change on the kernel function; let K_0 be a kernel from the location-scale family associated with the marginal density $f(.)$, that is

$$K_0(t) := K_0(t,\theta,\sigma) = \sigma f(\sigma t + \theta). \quad (6)$$

The scaling factor σ gives us more flexibility and the choice of the adjusted kernel K_0 is based on the minimization of the *AMSE*. Since the density f in (6) is not available, we have to replace it by the usual estimator f_n from above. Therefore, the classical density estimator (3) becomes (see, [4])

$$\hat{f}_n(x) = \frac{\sigma}{nh}\sum_{i=1}^{n} f_n\left(\sigma\frac{x-X_i}{h}+\theta\right)$$
$$= \frac{\sigma}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\sigma x-\sigma X_i+\theta h-hX_j}{h^2}\right).$$

The choice of $h,\ \theta$ and σ will be discussed later.

Next, we invoke formulas (1) and (2) to obtain our new estimator as follows

$$\hat{f}_n(y|x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f_n(\sigma x+\theta)\, H_h(y-Y_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n} f_n(\sigma x+\theta)}$$

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right) H_h\left(y - Y_i\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)} \quad . \tag{7}$$

2.2. **Asymptotic properties**

As usual, in order to study the asymptotic properties (i.e., bias, variance and *AMSE*) of the given estimator, we apply the following regularity assumptions:

(A_1) : $EY^2 < \infty$ and $EX^2 < \infty$.

(A_2) : X has a density f, which is continuously differentiable in a neighborhood of x.

(A_3) : $f(y|x)$ is twice continuously differentiable in a neighborhood of x.

(A_4) : K is a probability density satisfying $K(-u) = K(u)$ for all $u \in \mathbb{R}$. Furthermore, K has a finite third moments :$\int |u|^3 k(u)du < \infty$.

(A_5) : $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$.

Theorem 2.1. *Under* $(A_1) - (A_5)$, *we have*

$$\begin{aligned} bias\left(\hat{f}_n(y|x)\right) &= h^2 \frac{2 f'(x)\, f'(y|x) + f(x)\, f''(y|x)\; var(X)}{2\sigma^2 f(x)} + O\left(h^3\right) \\ &=: h^2 B(x) + O\left(h^3\right), \end{aligned} \tag{8}$$

and

$$\begin{aligned} var\left(\hat{f}_n(y|x)\right) &= \frac{1}{nh^2} \frac{\sigma\, f(y|x)}{f^2(x)}\; \mu(K) \int f^2(u)du + O(1/h) \\ &=: \frac{1}{nh^2} \rho^2(x) + O(1/h), \end{aligned} \tag{9}$$

where $\mu(K)$ *is the same as in (5). Accordingly, the asymptotic mean squared error* $(AMSE)$ *is*

$$AMSE\left(\hat{f}_n(y|x)\right) = h^4 B^2(x) + \frac{1}{nh^2} \rho^2(x).$$

Thus, the estimator is consistent provided $h \to 0$ *and* $nh^2 \to \infty$, *as* $n \to \infty$.

Proof. Firstly, we consider the bias part. Using formula (15), we may write

$$I_{2n} := \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x)\{f(y|X_i) - f(y|x)\}$$
$$= \frac{1}{\hat{f}_n(x)} \frac{\sigma}{n^2h^2} \sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)(f(y|X_i) - f(y|x)).$$

We know that $\hat{f}_n$ converge to f in probability (see, [14]). Moreover, the numerator is equivalent to

$$\frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)(f(y|z) - f(y|x)). \quad (10)$$

Then in the following, we may consider any fixed θ in (10) rather than θ_n. The expectation becomes

$$\frac{\sigma}{h^2} \iint K\left(\frac{\sigma x - \sigma z + \theta h - hy}{h^2}\right)(f(y|z) - f(y|x))f(y)f(z)dzdy$$
$$= \iint K(v)\left\{f\left(x + \frac{\theta h - vh^2 - hy}{\sigma}\right) - f(y|x)\right\}$$
$$\times f(y)\left\{f\left(x + \frac{\theta h - vh^2 - hy}{\sigma}\right)^2\right\}dydv.$$

By Taylor's expansion, the last integral, up to a $O(h^3)$ term, equals :

$$\iint K(v)\left\{f'(y/x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right) + \frac{1}{2}f''(y|x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)^2\right\}$$
$$\times f(y)\left\{f(x) + f'(x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right) + \frac{1}{2}f''(x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)^2\right\}dydv$$
$$= \iint K(v)f'(y|x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)dydv$$
$$+ \iint K(v)f'(y|x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)f(y)f'(x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)dydv$$
$$+ \iint K(v)f'(y|x)\left(\frac{\theta h - vh^2 - hy}{2\sigma}\right)f''(x)f(y)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)^2 dydv$$
$$+ \frac{1}{2}\iint K(v)f''(y|x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)^2 f(y)f(x)dvdy$$
$$+ \frac{1}{2}\iint K(v)f''(y|x)\acute{f}(x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)f(y)dvdy$$
$$+ \frac{1}{2}\iint K(v)f''(y|x)f(y)f''(x)\left(\frac{\theta h - vh^2 - hy}{\sigma}\right)^2 dydv.$$

For $\theta = E(X)$ and $\int u\, K(u)du = 0,$ the first integral is neglected. If we take $\theta_n = n^{-1}\sum_{i=1}^{n} X_i,$

the first, third, fifth and sixth integrals are $O_p(h^3)$, as it is known, $h = n^{-1/6}$. So we are left with the calculation of the second and the fourth integrals. The second integral, at $\theta = E(X)$, equals,

$$\frac{f'(x)f'(y|x)h^2}{\sigma^2}var(X) + o(h^3).$$

For θ_n, it equals :

$$\frac{f'(x)f'(y|x)h^2}{\sigma^2}\int(\theta_n - y)^2 f(dy) + O(h^3).$$

As for the fourth integral, we obtain the expansion

$$\frac{f(x)f''(y|x)h^2 var(X)}{\sigma^2} + O(h^3).$$

To conclude, when the variance of I_{2n} is $o(h^2)$, we get the bias term (8), i.e.

$$bias\left(\hat{f}_n(y|x)\right) = h^2\frac{2f'(x)f'(y|x) + f(x)f''(y|x)var(X)}{2\sigma^2 f(x)} + O(h^3).$$

Secondly, the variance part (9) is determined as follow. For each $h > 0$ and every $\sigma > 0$, from (15), we have

$$I_{1n} := \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x)\{H_h(y - Y_i) - f(y|X_i)\}.$$

The predictable quadratic variation of I_{1n} is given as

$$\sum_{i=1}^{n} W_{ni}^2(x)\sigma_1^2(X_i) = \frac{1}{\hat{f}_n^2(x)}\frac{\sigma^2}{(nh)^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_1^2(X_i)$$
$$\times\left\{\sum_{i=1}^{n} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)\right\}^2.$$

$\hat{f}_n \to f$ in probability, see here also [14]. Under the assumption (A_5), the numerator takes the value

$$\frac{\sigma^2}{n^2(n-1)(n-2)h^4}\sum_{i\neq j\neq k}\sigma_1^2(X_i)K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)$$
$$\times K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_k}{h^2}\right), \tag{11}$$

with $\theta = EX$. This is the U-statistic of degree three with a kernel depending on h, σ and θ.

Its prediction equals

$$\frac{\sigma^2}{h^4 n}\int\int\int\sigma_1^2(y)\, K\left(\frac{\sigma x - \sigma y + \theta h - hz}{h^2}\right)$$
$$\times K\left(\frac{\sigma x - \sigma y + \theta h - hu}{h^2}\right) f(y)f(z)f(u)dydzdu.$$

The last term is equivalent to

$$\frac{1}{nh^2}\frac{\sigma f(y|x)}{f^2(x)}\mu(K)\int f^2(u)du + O(1/h).$$

This completes the proof. ■

The following result concerns the distribution convergence of our estimator $\hat{f}_n(y|x)$.

Theorem 2.2. *Under the assumptions of theorem 2.1, if* $\theta = E(X)$ *and* $h = o(n^{-1/6})$, *then* $(nh^2)^{1/2}\left(\hat{f}_n(y|x) - f(y|x)\right) \to N(0, \rho^2(x))$ *in distribution, where as before*

$$\rho^2(x) = \frac{\sigma f(y|x)}{f^2(x)}\mu(K)\int f^2(u)du.$$

Proof. Note that, by [14], $\hat{f}_n$ converge to f in probability. Therefore, the study of the variance and the distribution behavior of $\hat{f}_n(y|x)$, is based on that of the joint density function $f(y,x)$. Hence we introduce its Hájek projection (see, e.g. [6]) $\hat{f}_n^0(y,x)$, defined by

$$\begin{aligned}
&\frac{\sigma}{nh^3}\sum_{i=1}^{n}\Big\{\int\int K\left(\frac{\sigma x - \sigma X_i + \theta h - hz}{h^2}\right)H(y-s)f(z,s)dzds \\
&+\int\int K\left(\frac{\sigma x - \sigma t + \theta h - hX_i}{h^2}\right)H(y-s)f(t,s)dtds \\
&-\int\int\int K\left(\frac{\sigma x - \sigma t + \theta h - hz}{h^2}\right)H(y-s)f(z,t)f(t)f(s)dzdtds\Big\}.
\end{aligned}$$

Note that $\hat{f}_n^0(y,x)$ is a sum of i. i. d. rv's with $E\left(\hat{f}_n^0(y,x)\right) = E\left(\hat{f}_n(y,x)\right)$. Its variance equals

$$\begin{aligned}
Var\left(\hat{f}_n^0(y,x)\right) = \frac{\sigma^2}{nh^6}\Big\{&\int\int K\left(\frac{\sigma x - \sigma X_i + \theta h - hz}{h^2}\right)H(y-s)f(z,s)dzds \\
&+\int\int K\left(\frac{\sigma x - \sigma t + \theta h - hX_i}{h^2}\right)H(y-s)f(t,s)dtds \\
&-2\int\int\int K\left(\frac{\sigma x - \sigma t + \theta h - hz}{h^2}\right)H(y-s) \\
&\times f(z,t)f(t)f(s)dzdtds\Big\}^2.
\end{aligned}$$

Then it follows from arguments similar to those used in the proof of Lemma 4.1. in [3], that

$$nh^2 Var\left(\hat{f}_n^0(y,x)\right) = \frac{\sigma f(y|x)}{f^2(x)}\mu(K)\int f^2(u)du + o(1). \tag{12}$$

Under the assumption of theorem 2.2 and by definition of $\hat{f}_n(y,x)$ and $\hat{f}_n^0(y,x)$, we have

$$\hat{f}_n(y,x) - \hat{f}_n^0(y,x) = \frac{\sigma}{n(n-1)h^2}\sum_{i\neq j}\Big\{K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)H(y-Y_i)$$

$$- \iint K\left(\frac{\sigma x - \sigma X_i + \theta h - hz}{h^2} \right) H(y - Y_i) f(x,z) dxdz$$
$$- \iint K\left(\frac{\sigma x - \sigma t + \theta h - hX_j}{h^2} \right) H(y - Y_i) f(x,t) dxdt$$
$$+ \iiint K\left(\frac{\sigma x - \sigma t + \theta h - hs}{h^2} \right) H(y - Y_i) f(y) f(s) f(t) dydsdt \Big\}$$
$$=: \frac{\sigma}{n(n-1)h^2} \sum_{i \neq j} L_h(X_i, X_j, Y_i).$$

It is readily seen that the last sum is a degenerate U-statistic of degree two, i.e.,

$$E[L_h(X_i, X_j, Y_i) L_h(X_k, X_j, Y_k) | X_j] = 0,$$

and

$$E[L_h(X_i, X_j, Y_i) L_h(X_i, X_k, Y_i) | X_i] = 0,$$

for $i \neq j \neq k.$ As a conclusion we get

$$E\left[\hat{f}_n(y/x) - \hat{f}_n^{\circ}(y/x)\right]^2 = \frac{\sigma^2}{n(n-1)h^3} E[L_h^2(X_1, X_2, Y_1)].$$

Note that, each terms in L_h admits a second moment of the order $O(h^2)$. Then

$$nh^2 E\left[\hat{f}n(y,x) - \hat{f}n^{\circ}(y,x)\right]^2 = O\left(n^{-1}h^{-2}\right) = o(1). \tag{13}$$

The convergence of $\hat{f}_n(y,x)$ to $f(y,x)$ follow immediately from (12) and (13). Also we can verify Lindeberg's condition (see, Corollary 1 in [1] by Brown and Eagleson). Considering the variance term in (15), the variance of (11) tends to zero as $n \to \infty$, and the predictable quadratic variation of I_{1n} equals

$$\frac{1}{nh^2} \frac{\sigma f(y|x)}{f^2(x)} \mu(K) \int f^2(u) du + o\left(\frac{1}{nh^2} \right) := \rho^2(x) + o(n^{-1}h^{-2}).$$

Then $(nh^2)^{1/2} I_{1n} \to N(0, \rho^2(x))$ in distribution *as* $n \to \infty$.
Here the proof completes. ■

2.3. **Optimal choice of smoothing parameters**

In the classical kernel estimation literature, the optimal bandwidth h_{opt} can be derived by minimizing the *AMSE* expression. This optimal bandwidth gives a trade-off between bias and variance (e.g.). In the case of adjusted kernel estimation, the *AMSE* depends on h and σ only. Therefore, the new method loads to an adaptive choice of both smoothing parameters h and σ. Firstly, the optimal bandwidth h_{opt} can be derived by differentiating the *AMSE* of $\hat{f}_n$ with respect to h and setting the derivatives to 0. Taking these derivatives and simplifying we obtain the following expression :

$$h_{opt} = \left(\frac{2}{3n} \frac{\rho^2(x)}{B^2(x)} \right)^{-1/6} =: cn^{-1/6}.$$

Next, we note that, if $h = o(n^{-1/6})$, the bias is negligible. In such a situation, the scale parameter σ should be chosen as small as possible to make $\rho^2(x)$ also small. Therefore, the optimal choice of σ minimizing the $AMSE\left(\hat{f}_n\right)$ may be achieved by setting $h = n^{-1/6}$, then

$$\sigma_{opt}^{1/5} = \frac{2\left\{2 f^{\prime}(x) f^{\prime}(y|x) + f(x) f^{\prime\prime}(y|x) var(X)\right\}^2}{f(y|x)\, \mu(K) \int f^2(u)\, du}. \tag{14}$$

The difference between this classical methodology and the new methodology becomes clear. The classical kernel focuses on h_{opt}^{1}, whereas the new methodology gives the principal role to the scale parameter σ.

Note also that our estimator can be rewritten as follow

$$\hat{f}_n(y|x) =: \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x) H_h(y - Y_i),$$

where $W_n(x) =: W_{ijn}(x)$, and

$$W_{ijn}(x) = K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\sigma x - \sigma X_i + \theta h - hX_j}{h^2}\right)\right\}^{-1}$$

are called the weights functions. They are non negative and satisfy :

$$\sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x) = 1, \quad \text{for all } x \in \mathbb{R}.$$

As a result we end up with the expression

$$\begin{aligned}\hat{f}_n(y|x) - f(y|x) &= \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x)\{H_h(y - Y_i) - f(y|X_i)\} \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x)\{f(y|X_i) - f(y|x)\}, \\ &=: I_{1n} + I_{2n}.\end{aligned} \tag{15}$$

To find an adaptive choice of σ, we don't focus on the analytic form of $B(x)$ and $\rho^2(x)$ in (8) and (9) respectively. As I_{1n} and $Y_i - f(y|X_i)$ are conditionally centred, and since $\sigma_1^2(X_i)$ is unknown, we give $\sum_{i=1}^{n}\sum_{j=1}^{n} W_n^2(x)\sigma_1^2(X_i)$ a predictable quadratic variation of I_{1n}, so as to replace it by $\{Y_i - f(y|X_i)\}^2$ with $h = n^{-1/6}$. As a conclusion, we can say that, the bias term of σ is

$$bias(\sigma) = \sum_{i=1}^{n}\sum_{j=1}^{n} W_n(x)\{f_n(y|X_i) - f(y|x)\}.$$

The variance of σ is found by

$$var(\sigma) = \sum_{i=1}^{n}\sum_{j=1}^{n} W_n^2(x)\{Y_i - f(y|X_i)\}^2.$$

As a result, we obtain the adaptive scale parameter σ by minimizing the *AMSE* expression : $AMSE(\sigma) = bias^2(\sigma) + var(\sigma)$.

3. Simulation Study

In this section we present a simulation study to illustrate the performance of our new estimator $\hat{f}_n$, given by Eq. (7), to establish the appropriate choice of the scale parameter σ. We will conduct the study by taking the data from the simple model $Y_j = \alpha X_j + \varepsilon_j, \quad j = 1,\ldots,n$, where (X_j) and (ε_j) are i. i.d. standard normal rv's. We will clarify the performance of our estimator $\hat{f}_n$ by assuring the present results by using the Monte Carlo simulation, for small and moderate sample sizes in certain points of x. The conditional density function of Y given X is given by

$$f(y|x) = \frac{1}{\left(\sqrt{-4\pi\alpha}\right)} \exp\left(\frac{1}{4\alpha}[(1+\alpha^2)x^2 - 2xy(1+\alpha) + y^2] + \frac{x^2}{2}\right).$$

Table 1: Bias, Var and AMSE values over the boundary region.

		$\hat{f}_n$			f_n		
n	x	$\lvert Bias\rvert$	Var	$AMSE$	$\lvert Bias\rvert$	Var	$AMSE$
	.001	.2030	.0412	.0825	.2464	.0607	.1215
25	.05	.2039	.0416	.0832	.2230	.0497	.0994
	.2	.2009	.0403	.0807	.2368	.0561	.1122
	.001	.2061	.0425	.0849	.2373	.0563	.1126
50	.05	.2048	.0419	.0839	.2527	.0638	.1277
	.2	.2036	.0414	.0829	.2306	.0532	.1064
	.001	.2076	.0431	.0862	.2625	.0689	.1378
100	.05	.2083	.0433	.0867	.2367	.0560	.1120
	.2	.2057	.0423	.0846	.2343	.0549	.1098
	.001	.2099	.0440	.0881	.2357	.0555	.1111
200	.05	.2096	.0439	.0879	.2458	.0604	.1208
	.2	.2062	.0425	.0851	.2399	.0575	.1151

We measure the performance of the estimators by the asymptotic mean squared error (*AMSE*). The simulation is based on 1000 replications, in each replication the sample sizes is: $n = 25, 50, 100$ and 200 was used. For the kernel, we choose a Gaussian kernel. The choice of bandwidth is very important for a good performance of any kernel estimator. In all cases, we consider the asymptotic optimal global bandwidth $h_{opt} = n^{-1/6}$, $\hat{\theta} = \bar{X}_n$ and the optimal scale parameter σ_{opt} is as defined in Eq.(14).

For each value of $x \in \{0.001,\ 0.05,\ 0.2\}$ we have calculated the absolute bias ($|Bias|$), variance (Var) and the *AMSE* values of the two considered estimators and have displayed the results in Table 1. The comparison shows that the proposed estimator exhibits the best performance. This is due to the fact that it is locally adaptive and produces good results in all studied situations. The main results of our simulation studies is that the proposed estimator can be recommended for bias and variance reduction and for improved boundary effects. Apparently the overall $\hat{f}_n$ is the best choice among the two estimators considered. The usual estimator f_n is clearly the worst, and this is undoubtedly due to the boundary effect.

Acknowledgments
The authors would like to thank an anonymous referee for his critical reading of the original typescript.

References

[1] B. M. Brown, and G. K. Eagleson, Martingale convergence to infinitely divisible laws with finite variances, *Transactions of the American Mathematical Society* **162**, (1971), 449–453.

[2] A. Das Gupta, *Asymptotic Theory of Statistics and Probability*, Springer Verlag, Berlin, 2008.

[3] G. Eichner,and W. Stute, Kernel adjusted nonparametric regression, *Journal of Statistical Planning and Inference* **142**(9), (2012), 2537-2544.

[4] J. Fan, Y. Qiwei , and T. Howell, Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, *Biometrika* **83**, (1996), 189-206.

[5] W. Feluch, and J. Koronacki, A note on modified cross-validation in density estimation, *Computational Statistics & Data Analysis* **13**, (1992), 143–151.

[6] J. Hajek, Asymptotic normality of simple linear rank statistics under alternatives, *The Annals of Mathematical Statistics* **39**, (1968), 325–343.

[7] W. Hardle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

[8] R. J. Hyndman, D. M. Bashtannyk, and G. K. Grunwald, Estimating and visualizing

conditional densities, *Journal of Computational and Graphical Statistics* **5**, (1996), 315–336.

[9] C. Loader, *Local Regression and Likelihood*, Springer, New York,1999.

[10] E. A. Nadaraya, Some new estimates for distribution functions, *Theory of Probability & Its Applications* **9**(3), (1964), 497-500.

[11] E. Parzen, On estimation of a probability density function and mode, *The Annals of Mathematical Statistics* **33**(3), (1962),1065-1076.

[12] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics* **27**(3), (1956), 832-837.

[13] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London,1986.

[14] R. Srihera, and W. Stute, Kernel adjusted density estimation, *Statistics & Probability Letters* **81**,(2011), 571–579.

[15] G. R.Terrell, and D. W. Scott, Variable kernel density estimation, *Annals of Statistics* **20**(3), (1992), 1236-1265.

[16] M. P. Wand, and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.

[17] G. S.Watson, Smooth regression analysis, *Sankhya*, *Series A* **26**, (1964), 359–372.

Article history: Submitted December, 08, 2017; Revised May, 15, 2018; Accepted September, 01, 2018.