Universität Stuttgart

Fachbereich Mathematik

Nonparametric Local Averaging Estimation of the Local Variance Function

Paola Gloria Ferrario

Preprint 2011/015

Universität Stuttgart

Fachbereich Mathematik

Nonparametric Local Averaging Estimation of the Local Variance Function

Paola Gloria Ferrario

Preprint 2011/015

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik.uni-stuttgart.de/preprints

ISSN **1613-8309**

Abstract

In this paper the problem of local variance estimation is considered. Given an independent and identically distributed sample, estimates of local averaging type, especially partitioning and kernel estimates, are investigated in view of consistency and rate of convergence. Furthermore the case of additional measurement errors in the dependent variables is treated.

Key words: regression function, local (or conditional) variance function, local averaging, partitioning, kernel, nearest neighbor estimates, least squares, measurement error, consistency, rate of convergence.

1 Introduction

Let $(X,Y),\ (X_1,Y_1),\ (X_2,Y_2),\ldots$ be independent and identically distributed $\mathbb{R}^d\times\mathbb{R}$-valued random vectors with $\boldsymbol{E}\{Y^4\}<\infty$. The regression function $m:\mathbb{R}^d\to\mathbb{R}$ is defined by

$$m(x):=\boldsymbol{E}\{Y|X=x\}.$$

m allows to predict a non-observable realization of Y on the basis of an observed realization x of X by $m(x)$. In competition with other measurable functions $f:\mathbb{R}^d\to\mathbb{R}$ the expression $\boldsymbol{E}\{(Y-f(X))^2\}$ is minimal for m, i.e.,

$$\boldsymbol{E}\left\{(Y-m(X))^2\right\}=\min_f \boldsymbol{E}\left\{(Y-f(X))^2\right\} \tag{1}$$

because of

$$\boldsymbol{E}\left\{|f(X)-Y|^2\right\}=\boldsymbol{E}\left\{(m(X)-Y)^2\right\}+\int|f(x)-m(x)|^2\mu(dx)$$

where μ denotes the distribution P_X of X.
However m is unknown if the distribution of (X,Y) is unknown. Nonparametric regression deals with the following problem: Given independent copies $(X_1,Y_1),\ldots,(X_n,Y_n)$ of (X,Y), an estimate m_n of the regression function shall be constructed, such that $\int|m_n(x)-m(x)|^2\mu(dx)$ is "small". Widespread principles of constructing m_n are local averaging and least squares estimations.
By local averaging the estimation of $m(x)$ is given by the weighted mean of those Y_i where X_i is in a certain sense close to x :

$$m_n(x)^{(LA)}=m_n^{(LA)}(x,X_1,Y_1,\ldots,X_n,Y_n)=\sum_{i=1}^n W_{n,i}(x)\cdot Y_i, \tag{2}$$

where the weights $W_{n,i}(x,X_1,\ldots,X_n)\in\mathbb{R}$, briefly written as $W_{n,i}(x)$, depend on $X_1,\ldots,X_n$ and are therefore non-deterministic. We have "small" (nonnegative) weights in the case that X_i is "far" from x. Depending on the definition of the weights, we distinguish between partitioning, kernel and nearest neighbor estimates.
By the least squares methods the idea is to minimize the empirical L_2-risk over an appropriate set of functions $\mathcal{F}_n$ and to choose the minimizing function(s) over $\mathcal{F}_n$ as regression estimate, that is

$$m_n(\cdot)^{(LS)}=m_n^{(LS)}(\cdot,X_1,Y_1,\ldots,X)=\arg\min_{f\in\mathcal{F}_n}\ \frac{1}{n}\sum_{i=1}^n|f(X_i)-Y_i|^2. \tag{3}$$

The quality of predicting by the regression function m is locally given by the local variance

$$\sigma^2(x):=\boldsymbol{E}\{(Y-m(X))^2|X=x\}=\boldsymbol{E}\{Y^2|X=x\}-m^2(x).$$

In the literature Kohler in [5], Section 3.1, deals with the estimation of local variance functions. We refer also the reader to Müller and Stadtmüller [6], Stadtmüller and Tsybakov [12], Ruppert et al. [10], Härdle and Tsybakov [4], Spokoiny [11], Pan and Wang [9], Hall et al. [3], Müller et al.

[7], Neumann [8] for the estimation of the local variance function also in the case of fixed design. In this paper we modify and extend the approach of Kohler in [5], Remark 5, in context of local averaging estimation of σ^2 by use of local averaging estimates m_n of m.
Note that we use the whole sample $(X_1, Y_1 \dots, X_n, Y_n)$ for the estimation of σ^2 as well as for the auxiliary estimate of m. The investigation of the asymptotic behavior regards the special structure of the standard local averaging methods used here. It is possible to obtain corresponding results for the general local averaging methods in Stone [13] (compare also Györfi et al. [2]) at the expense of splitting the sample ($n = n' + n''$) with weights $W^{(1)}_{n',j}$ and $W^{(2)}_{n'',j}$.
Section 2 deals with universal consistency of local averaging estimation of the local variance function, i.e., first mean convergence of the L_2-error $\int |\sigma_n^2(x) - \sigma^2(x)|^2 \mu(dx)$ to zero for general distribution (X, Y) under the optimal moment condition $\boldsymbol{E}\{Y^4\} < \infty$.
Section 3 is devoted to the rate of convergence in context of bounded Y and Lipschitz conditions on m and σ^2.
In Section 4 the concept of additional noise (induced because of the use of m_n instead of m in Sections 2 and 3) is extended to the case that Y_i's are noise-contaminated, again under uniform boundedness of the dependent variables.

2 Local Variance Estimation

One defines new random variables

$$Z := Y^2 - m^2(X)$$

and in context of observations in the case of known regression function

$$Z_i := Y_i^2 - m^2(X_i)$$

and of unknown regression function

$$Z_{n,i} := Y_i^2 - m_n^2(X_i). \tag{4}$$

(Notice that usually m is unknown and has to be estimated. In this way one has a plug-in method.)
Note that the local variance function is a regression on the pair (X, Z).
This motivates the construction of a family of estimates of the local variance that have the form

$$\sigma_n^2(x) := \sigma_n^2(x)^{(LA)} = \sum_{i=1}^{n} W_{n,i}(x) \cdot Z_{n,i}, \tag{5}$$

The weights $W_{n,i}(x)$ can take different forms. In the literature partitioning weights are used, defined by

$$W_{n,i}(x, X_1, \dots, X_n) = \frac{1_{A_n(x)}(X_i)}{\sum_{l=1}^{n} 1_{A_n(x)}(X_l)} \tag{6}$$

($A_n(x)$ denoting the $A_{n,j}$ of the partitioning sequence $\{A_{n,j}\}$ containing $x \in \mathbb{R}^d$), with $0/0 := 0$.
Further kernel weights are used, especially with symmetric kernel $K : \mathbb{R}^d \to [0, \infty)$, satisfying $1_{S_{0,R}} \geq K(x) \geq b 1_{S_{0,r}}(x)$ $(0 < r \leq R < \infty,\ b > 0)$, defined by

$$W_{n,i}(x, X_1, \dots, X_n) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{l=1}^{n} K\left(\frac{x - X_l}{h_n}\right)}, \tag{7}$$

with bandwidth $h_n > 0$ and $0/0 := 0$ again. $S_{0,r}$ denotes the sphere with radius $r > 0$ centered in 0.
Finally, nearest neighbor weights are also frequently used, defined by

$$W_{n,i}(x, X_1, \dots, X_n) = \frac{1}{k_n} 1_{\{X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \dots, X_n\}\}} \tag{8}$$

$(2 \leq k_n \leq n)$, here usually assuming that ties occur with probability 0. This can be obtained for example via tie-breaking by indices (compare [2], pp. 86, 87).
We distinguish local averaging methods in the auxiliary estimates m_n in (4) and in the estimates σ_n^2 in (5), indicating the weights by $W_{n,i}^{(2)}$ and $W_{n,i}^{(1)}$ (instead of $W_{n,i}$ in (5)), respectively. Thus

$$m_n(X_i) = \sum_{j=1}^{n} W_{n,j}^{(2)}(X_i, X_1, \ldots, X_n) Y_j \tag{9}$$

where

$$W_{n,j}^{(2)}(x, X_1, \ldots, X_n),$$

is of partitioning type, with partitioning sequence $\left\{A_{n,j}^{(2)}\right\}$, or of kernel type, with kernel $K^{(2)}$ and bandwidhts $h_n^{(2)}$, or of nearest neighbor type, with $k_n^{(2)}$ neighbors.
Now with $Z_{n,i} = Y_i^2 - m_n^2(X_i)$ we define a family of estimators of the local variance function by

$$\sigma_n^2(x) = \sum_{i=1}^{n} W_{n,i}^{(1)}(x) \cdot Z_{n,i}, \tag{10}$$

depending on weights

$$W_{n,i}^{(1)}(x) = W_{n,i}^{(1)}(x, X_1, \ldots, X_n),$$

that are of partitioning type, with partitioning sequence $\left\{A_{n,j}^{(1)}\right\}$, or of kernel type, with kernel $K^{(1)}$ and bandwidhts $h_n^{(1)}$. (Nearest neighbor weights will not be used for $W_{n,i}^{(1)}(x)$.)

Theorem 2.1 *Let (X, Y) have an arbitrary distribution with $\boldsymbol{E}\{Y^4\} < \infty$. For partitioning weights defined according to (6) assume that, for each sphere S centered at the origin*

$$\lim_{n\to\infty} \max_{A_{n,j}^{(l)} \cap S \neq \emptyset} diam(A_{n,j}^{(l)}) = 0, \quad l = 1, 2, \tag{11}$$

$$\lim_{n\to\infty} \frac{|\{j : A_{n,j}^{(l)} \cap S \neq \emptyset\}|}{n} = 0, \quad l = 1, 2. \tag{12}$$

For kernel weights defined according to (7) with kernels $K^{(l)}$ assume that the bandwidths satisfy

$$0 < h_n^{(l)} \to 0, \quad n h_n^{(l)d} \to \infty, \quad l = 1, 2, \tag{13}$$

($K^{(l)}$ symmetric, $1_{S_{0,R}}(x) \geq K^{(l)}(x) \geq b 1_{S_{0,r}}(x)$ $(0 < r \leq R < \infty,\ b > 0)$).
For nearest neighbor weights defined according to (8) assume that

$$2 \leq k_n^{(2)} \leq n, \quad k_n^{(2)} \to \infty, \quad \frac{k_n^{(2)}}{n} \to 0 \tag{14}$$

Then for the estimate (10) under the above assumptions

$$\lim_{n\to\infty} \boldsymbol{E} \int \left(\sigma_n^2(x) - \sigma^2(x)\right)^2 \mu(dx) = 0$$

holds. (Universal consistency of local averaging estimators of the local variance)

Theorem 2.1 will be proven by Lemmas 2.2 and 2.3.
The following Lemma 2.2 modifies Remark 5 in Kohler [5]. It is within the framework that the dependent variable Y can be observed only with supplementary, maybe correlated, measurement errors. Since it is not assumed that the means of these measurement errors are zero, these kinds of errors are not already included in standard models. Therefore, the dataset is of the form

$$\overline{D}_n = \{(X_1, \overline{Y}_{1,n}), \ldots, (X_n, \overline{Y}_{n,n})\},$$

Lemma 2.2 *Let $\overline{m}_n$ be local averaging estimators of m of the form*

$$\overline{m}_n(x) = \sum_{i=1}^{n} W_{n,i}(X)\overline{Y}_i$$

with $\overline{Y}_i = \overline{Y}_{i,n}$. Assume that the weights $W_{n,i}(x) = W_{n,i}(x, X_1, \ldots, X_n)$ are of partitioning type (6) or kernel type (7) with $1_{S_{0,R}}(x) \geq K(x) \geq b1_{S_{0,r}}$ for some $0 < r \leq R < \infty$, $b > 0$, satisfying (11) $\wedge$ (12) and (13), respectively. Further assume

$$\boldsymbol{E}\left\{Y^2\right\} < \infty, \quad \boldsymbol{E}\left\{\overline{Y}_i^2\right\} < \infty \quad (i = 1, \ldots, n)$$

and

$$\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|\overline{Y}_i - Y_i|^2\right\} \to 0. \tag{15}$$

Then

$$\boldsymbol{E}\left\{\int |\overline{m}_n(x) - m(x)|^2 \mu(dx)\right\} \to 0.$$

Proof As Kohler ([5], Remark 5) suggested,

$$\begin{aligned}
& \boldsymbol{E}\left\{\int \left(\sum_{i=1}^{n} W_{n,i}(x)\overline{Y}_i - m(x)\right)^2 \mu(dx)\right\} \\
= \ & \boldsymbol{E}\left\{\int \left(\sum_{i=1}^{n} W_{n,i}(x)\left[\overline{Y}_i - Y_i + Y_i\right] - m(x)\right)^2 \mu(dx)\right\} \\
\leq \ & 2\boldsymbol{E}\left\{\int \left(\sum_{i=1}^{n} W_{n,i}(x)Y_i - m(x)\right)^2 \mu(dx)\right\} \\
+ \ & 2\boldsymbol{E}\left\{\int \left(\sum_{i=1}^{n} W_{n,i}(x)\left[\overline{Y}_i - Y_i\right]\right)^2 \mu(dx)\right\} \\
= \ & 2K_{n,1} + 2K_{n,2}.
\end{aligned}$$

The term $K_{n,1}$ is simply the expected L_2-error of the local averaging estimate of m on the basis of observed $((X_1, Y_1), \ldots, (X_n, Y_n))$. By Theorem 4.2 and Theorem 5.1 in [2], respectively, $K_{n,1} \to 0$. It remains to show $K_{n,2} \to 0$. By the Cauchy-Schwarz inequality together with $W_{n,i}(X) \geq 0$, $\sum_{i=1}^{n} W_{n,i}(X) \leq 1$, one has

$$K_{n,2} \leq \boldsymbol{E}\left\{\int \left(\sum_{i=1}^{n} W_{n,i}(x)|\overline{Y}_i - Y_i|^2\right) \mu(dx)\right\}.$$

With

$$f_i(x) := \boldsymbol{E}\left\{|\overline{Y}_i - Y_i|^2 \middle| X_i = x\right\}, \quad x \in \mathbb{R} \quad (i \in \{1, \ldots, n\}),$$

there is, because of the special structure of $W_{n,i}$, a finite constant c such that, for all $i \in \{1, \ldots, n\}$

$$\begin{aligned}
& \int \boldsymbol{E}\left\{\sum_{i=1}^{n} W_{n,i}(x)|\overline{Y}_i - Y_i|^2\right\} \mu(dx) \\
= \ & \int \boldsymbol{E}\left\{\sum_{i=1}^{n} W_{n,i}(x) f_i(X_i)\right\} \mu(dx) \\
\leq \ & \frac{c}{n}\sum_{i=1}^{n} \int f_i(u)\mu(du)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{c}{n}\sum_{i=1}^{n}\boldsymbol{E} f_i(u)\mu(du) \\
&= \frac{c}{n}\boldsymbol{E}\left\{\sum_{i=1}^{n}|\overline{Y}_i - Y_i|^2\right\}
\end{aligned}
$$

where the inequality is obtained via the individual summands by the arguments in [2], p.62 (with $c = 1$) and pp. 74, 75, respectively. Thus, by the assumption (15),

$$
K_{n,2} \le c\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|\overline{Y}_i - Y_i|^2\right\} \to 0.
$$

■

Lemma 2.3 *Let m_n be local averaging estimators of m of the form*

$$
m_n(x) = \sum_{i=1}^{n} W_{n,i}(x)Y_i.
$$

Assume that the weights $W_{n,i}(x) = W_{n,i}(x, X_1, \ldots, X_n)$ are of partitioning type (6) or of kernel type (7) with $1_{S_{0,R}}(x) \ge K(x) \ge b1_{S_{0,r}}(x)$ for some $0 < r \le R < \infty$, $b > 0$, or of nearest neighbor type (8) (here under assumption that ties occur with probability 0), satisfying (11) $\wedge$ (12), (13) and (14), respectively. Further assume $\boldsymbol{E}\{Y^2\} < \infty$. Then

$$
\boldsymbol{E}\{|m_n(X_1) - m(X_1)|^2\} \to 0.
$$

If moreover $\boldsymbol{E}\{Y^4\} < \infty$, then

$$
\boldsymbol{E}\{|m_n(X_1) - m(X_1)|^4\} \to 0.
$$

Proof We first assume that $\boldsymbol{E}\{Y^2\} < \infty$ and that $W_{n,j}^{(2)}$ is of kernel type. Then

$$
\begin{aligned}
& \boldsymbol{E}\{|m_n(X_1) - m(X_1)|^2\} \\
= \;& \boldsymbol{E}\left\{\left|\frac{Y_1K(0) + \sum_{j=2}^{n} Y_j K\left(\frac{X_1 - X_j}{h_n}\right)}{K(0) + \sum_{j=2}^{n} K\left(\frac{X_1 - X_j}{h_n}\right)} - m(X_1)\right|^2\right\} \\
\le \;& 2K(0)^2\boldsymbol{E}\left\{\frac{\boldsymbol{E}\left\{Y^2|X\right\}}{\left(K(0) + \sum_{j=2}^{n} K\left(\frac{X - X_j}{h_n}\right)\right)^2}\right\} \\
+ \;& 2\boldsymbol{E}\left\{\left|\frac{\sum_{j=2}^{n} Y_j K\left(\frac{X - X_j}{h_n}\right)}{K(0) + \sum_{j=2}^{n} K\left(\frac{X - X_j}{h_n}\right)} - m(X)\right|^2\right\}.
\end{aligned}
$$

The second term of the right-hand side converges to 0 as in the proof of Theorem 5.1 in [2]. With a suitable finite constant c_1 the first term is bounded by

$$
\begin{aligned}
& c_1 \int \boldsymbol{E}\left\{Y^2\middle|X = x\right\}\boldsymbol{E}\left\{\frac{1}{\left(1 + \sum_{j=2}^{n} 1_{S_{x,rh_n}}(X_j)\right)^2}\right\}\mu(dx) \\
\le \;& c_1 \int \boldsymbol{E}\left\{Y^2|X = x\right\}\frac{1}{n\mu(S_{x,rh_n})}\mu(dx) \to 0
\end{aligned}
$$

by Lemma 4.1 (i) in [2] together with Lemma 24.6 in [2], $nh_n^d \to \infty$ in assumption (13), $\boldsymbol{E}\{Y^2\} < \infty$ and the dominated convergence theorem. The case that $W_{n,j}^{(2)}$ is defined via partitioning, is treated analogously by use of Theorem 4.2 in [2], and for $\epsilon > 0$

$$
\begin{aligned}
& \int \boldsymbol{E}\left\{Y^2|X=x\right\} \boldsymbol{E} \frac{1}{\left(1+\sum_{j=2}^n 1_{A_n(x)}(X_j)\right)^2} \mu(dx) \\
\leq & \int \boldsymbol{E}\{Y^2|X=x\} 1_{\{\boldsymbol{E}\{Y^2|X=x\}>L\}} \mu(dx) \\
& + \int \boldsymbol{E}\{Y^2|X=x\} 1_{\{\boldsymbol{E}\{Y^2|X=x\}\leq L\}} \boldsymbol{E}\left\{\frac{1}{1+\sum_{j=2}^n 1_{A_n(x)}}\right\} \mu(dx) \\
\leq & \ \epsilon + L\int_{S^c} \mu(dx) + L\int_S \frac{1}{n\mu(A_n(x))}\mu(dx) \\
& \text{(for suitable L} < \infty \text{ and by Lemma 4.1 (i) in [2])} \\
\leq & \ 2\epsilon + L\int_S \frac{1}{n\mu(A_n(x))}\mu(dx) \\
& \text{(for suitable sphere S centered at 0)} \\
= & \ 2\epsilon + o(1) \\
& \text{(by assumption (12))}.
\end{aligned}
$$

In the case that $W_{n,j}^{(2)}$ is defined via nearest neighbors, we write

$$
\begin{aligned}
& \boldsymbol{E}\{|m_n(X_1)-m(X_1)|^2\} \\
= & \ \boldsymbol{E}\left\{\left|\frac{Y_1+D_n}{1+(k_n-1)} - m(X_1)\right|^2\right\}
\end{aligned}
$$

with

$$
D_n = \sum_{j=2}^n Y_j 1_{\{X_j \text{ is among the } (k_n-1) \text{ NNs of } X_1 \text{ in } \{X_2,\ldots,X_n\}\}},
$$

and notice

$$
\boldsymbol{E}\left\{\left|\frac{D_n}{k_n-1} - m(X_1)\right|^2\right\} \to 0
$$

by Theorem 6.1 in [2]. Further

$$
\boldsymbol{E}\left\{\left|\frac{Y_1+D_n}{k_n} - \frac{D_n}{k_n-1}\right|^2\right\} \leq \frac{2}{k_n^2}\left(\boldsymbol{E}\left\{Y_1^2\right\} + \boldsymbol{E}\left\{\left(\frac{D_n}{k_n-1}\right)^2\right\}\right) \to 0
$$

because of

$$
\boldsymbol{E}\left\{Y_1^2\right\} < \infty, \quad \boldsymbol{E}\left\{\left(\frac{D_n}{k_n-1}\right)^2\right\} \to \boldsymbol{E}\{m(X_1)^2\} < \infty, k_n \to \infty.
$$

Now, we consider the case $\boldsymbol{E}\{Y^4\} < \infty$. The above proof shows that for $r = 2, 4$ one has the representation

$$
\begin{aligned}
J_n^{(r)} & := \boldsymbol{E}\{|m_n(X_1)-m(X_1)|^r\} \\
& = \boldsymbol{E}\left\{\left|\sum_{i=1}^n Y_i \overline{W}_{n,i}(X, X_2, \ldots, X_n) - m(X)\right|^r\right\}
\end{aligned}
$$

for some $\overline{W}_{n,i} \geq 0$ with $\sum_{i=1}^n \overline{W}_{n,i} = 1$, e.g., in the kernel case $\overline{W}_{n,i}$ with $K(0)$ instead of $K\left(\frac{\cdot - X_1}{h_n}\right)$ in $W_{n,i}$. Then by Györfi [1], Theorem 1 with proof (compare also [2], Lemma 23.3 with proof, and [14], last part of Lemma 8 with $\delta = 1$, $p = 2$ and convergence in probability instead of almost sure convergence), $J_n^{(2)} \to 0$ for $\boldsymbol{E}\{Y^2\} < \infty$ (already proven) implies $J_n^{(4)} \to 0$ for $\boldsymbol{E}\{Y^4\} < \infty$.

$\blacksquare$

Proof of Theorem 2.1 We apply Lemma 2.2 with Y_i, $\overline{Y}_i$, $W_{n,i}$, $\overline{m}_n$ and m replaced by $Y_i^2 - m^2(X_i)$, $Y_i^2 - m_n^2(X_i)$, $W_{n,i}^{(1)}$, σ_n^2 and σ^2 in Theorem 2.1, respectively. We notice

$$
\begin{aligned}
& \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|(Y_i^2 - m_n^2(X_i)) - (Y_i^2 - m^2(X_i))|^2\right\} \\
= \quad & \boldsymbol{E}\{|m_n^2(X_1) - m^2(X_1)|^2\} \\
& \text{(due to symmetry with respect to } (X_1, Y_1), \ldots, (X_n, Y_n)) \\
\leq \quad & \left(\boldsymbol{E}\left\{|m_n(X_1) + m(X_1)|^4\right\}\right)^{\frac{1}{2}} \left(\boldsymbol{E}\left\{|m_n(X_1) - m(X_1)|^4\right\}\right)^{\frac{1}{2}} \\
& \text{(because of the Cauchy-Schwarz inequality)} \\
\to \quad & 0.
\end{aligned}
$$

The latter is obtained by the triangle inequality

$$
\begin{aligned}
& \left(\boldsymbol{E}\left\{|m_n(X_1) + m(X_1)|^4\right\}\right)^{\frac{1}{4}} \\
\leq \quad & \left(\boldsymbol{E}\left\{|m_n(X_1) - m(X_1)|^4\right\}\right)^{\frac{1}{4}} + 2\boldsymbol{E}(\{|m(X_1)|^4\})^{\frac{1}{4}},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{E}\{m(X_1)^4\} &= \boldsymbol{E}\{m(X)^4\} \\
&= \boldsymbol{E}\{(\boldsymbol{E}(Y|X))^4\} \leq \boldsymbol{E}\{\boldsymbol{E}\{Y^4|X\}\} = \boldsymbol{E}\{Y^4\} < \infty
\end{aligned}
$$

because of Jensen's inequality for conditional expectations,

$$
\boldsymbol{E}\left\{|m_n(X_1) - m(X_1)|^4\right\} \to 0
$$

because of Lemma 2.3 with $W_{n,i}$ there replaced by $W_{n,i}^{(2)}$ in Theorem 2.1. Thus Lemma 2.2 yields the assertion. $\blacksquare$

3 Rate of Convergence

In this section we establish a rate of convergence for the estimate of the local variance defined in Section 2.

Theorem 3.1 *Let the estimate of the local variance σ^2 be given by* (10) *with weights $W_{n,i}^{(1)}(x)$ of cubic partition with side length $h_n^{(1)}$ or with naive kernel $1_{S_{0,1}^{(1)}}$ with bandwidths $h_n^{(1)}$, further for $m_n(X_i)$ given by (9) with weights $W_{n,i}^{(2)}(x)$ of cubic partition with side length $h_n^{(2)}$ or with naive kernel $1_{S_{0,1}}$ and bandwidths $h_n^{(2)}$ or with $k_n^{(2)}$-nearest neighbor (the latter for $d \geq 3$).*
Assume that

$$
|Y| \leq L \in [0, \infty),
$$

that

$$
|m(x) - m(z)| \leq C\|x - z\|, \quad x, z \in \mathbb{R}^d,
$$

and finally, that

$$
|\sigma^2(x) - \sigma^2(z)| \leq D\|x - z\|, \quad x, z \in \mathbb{R}^d,
$$

($\|\ \|$ denoting the Euclidean norm). Let X have a compact support S^*. Then, for

$$
h_n^{(1)} \sim n^{-\frac{1}{d+2}},
$$

and

$$
h_n^{(2)} \sim n^{-\frac{1}{d+2}}, \quad \textit{and} \quad k_n^{(2)} \sim n^{\frac{2}{d+2}}, \quad \textit{respectively,}
$$

$$
\boldsymbol{E}\int |\sigma_n^2(x) - \sigma^2(x)|\mu(dx) = O\left(n^{-\frac{2}{d+2}}\right).
$$

Proof As in the proof of Lemma 2.2 and by the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
& \boldsymbol{E}\left\{\int\left(\left(\sum_{i=1}^{n} W_{n,i}^{(1)}(x) Z_{n,i}\right)-\sigma^2(x)\right)^2 \mu(dx)\right\} \\
\leq\ & 2\boldsymbol{E}\left\{\int\left(\sum_{i=1}^{n} W_{n,i}^{(1)}(x)\left(Y_i^2-m^2(X_i)\right)-\sigma^2(x)\right)^2 \mu(dx)\right\} \\
& +2\boldsymbol{E}\left\{\int\left(\sum_{i=1}^{n} W_{n,i}^{(1)}(x)\left[m_n^2(X_i)-m^2(X_i)\right]\right)^2 \mu(dx)\right\} \\
\leq\ & 2\boldsymbol{E}\left\{\int\left(\sum_{i=1}^{n} W_{n,i}^{(1)}(x)\left(Y_i^2-m^2(X_i)\right)-\sigma^2(x)\right)^2 \mu(dx)\right\} \\
& +2\frac{c}{n}\sum_{i=1}^{n}\boldsymbol{E}\left\{\left|m_n^2(X_i)-m^2(X_i)\right|^2\right\} \\
\leq\ & 2\boldsymbol{E}\left\{\int\left(\sum_{i=1}^{n} W_{n,i}^{(1)}(x)\left(Y_i^2-m^2(X_i)\right)-\sigma^2(x)\right)^2 \mu(dx)\right\} \\
& +c^*\boldsymbol{E}\left\{\left|m_n(X_1)-m(X_1)\right|^2\right\} \\
=:\ & 2K_n+c^*L_n
\end{aligned}$$

with suitable $c^* \in [0,\infty)$ because of boundedness of Y and by symmetry. We have

$$K_n = O\left(n^{-\frac{2}{d+2}}\right)$$

by Theorems 4.3, 5.2 and 6.2 in [2], respectively.
According to the proof of Lemma 2.3 these theorems together with boundedness of Y (with sphere $S \supset S^*$, centered in 0),

$$\int_S \frac{1}{n\mu(A_n^{(2)}(x))}\mu(dx) = O\left(\frac{1}{nh_n^{(2)d}}\right) = O\left(n^{-\frac{2}{d+2}}\right)$$

(because the number of cubes in S is $O\left(\frac{1}{h_n^{(2)d}}\right)$) in the partitioning case,

$$\int_S \frac{1}{n\mu(S_{x,rh_n^{(2)}})}\mu(dx) = O\left(\frac{1}{nh_n^{(2)d}}\right) = O\left(n^{-\frac{2}{d+2}}\right)$$

(by (5.1) in [2]) in the kernel case,

$$\boldsymbol{E}\left\{\left|\frac{Y_1+D_n}{k_n^{(2)}}-\frac{D_n}{k_n^{(2)}-1}\right|^2\right\} = O\left(\frac{1}{k_n^{(2)2}}\right) = O\left(n^{-\frac{2}{d+2}}\right)$$

in the nearest neighbor case, yield

$$L_n = O\left(n^{-\frac{2}{d+2}}\right).$$

■

4 Local Variance Estimation with Additional Measurement Errors

We recall the important variable $Z := Y^2 - m^2(X)$ and their corresponding observations in the case of known regression function $Z_i := Y_i^2 - m^2(X_i)$. In the general case m is however to be estimated.

One can do that by use of least squares estimates $m_n = m^{(LS)}$ or by local averaging estimates $m_n = m_n^{(LA)}$ (as in Theorem 2.1) of partitioning type satisfying (11) $\wedge$ (12) with $l = 2$, additionally assuming that the partitioning sequence $(\{A_{n,j}^{(2)}\})$ is nested, i.e., $A_{n+1}^{(2)}(x) \subset A_n^{(2)}(x)$ for all $n \in \mathbb{N}$, $x \in \mathbb{R}^d$, or of kernel type ($K^{(2)}$ symmetric, $1_{S_{0,R}} \geq K^{(2)}(x) \geq b 1_{S_{0,r}}(x)$ $(0 < r \leq R < \infty,\ b > 0)$) with bandwidths h_n satisfying (13) with $l = 2$:

$$Z_{n,i} = Y_i^2 - m_n^2(X_i).$$

With additional noise these variables are taking the form

$$\overline{Z}_{n,i} = \overline{Y}_i^2 - \overline{m}_n^2(X_i).$$

with the noisy data $\overline{Y}_i$ used also in the corresponding definition of $\overline{m}_n = \overline{m}^{(LS)}$ and $\overline{m}_n = \overline{m}^{(LA)}$, respectively. We refer the reader to Kohler [5] for the topics concerning the regression estimates $m^{(LS)}$, by use of piecewise polynomials, see also [2], Chapter 19, especially Section 19.4 and Problems and Excercises.
Let a familiy of estimates of the local variance function in case of additional measurement errors for the dependent variable Y be given by

$$\overline{\sigma}_n^2(x) := \overline{\sigma}_n^2(x)^{(LA)} = \sum_{i=1}^{n} W_{n,i}^{(1)}(x) \cdot \overline{Z}_{n,i}, \tag{16}$$

with weights $W_{n,i}^{(1)}(x) = W_{n,i}^{(1)}(x, X_1, \ldots, X_n)$ of partitioning or of kernel type with kernel $K^{(1)}$ and bandwidths $h_n^{(1)}$, respectively, satisfying (11) $\wedge$ (12) and (13) with $l = 1$, respectively.

Theorem 4.1 *Let the assumptions of Theorem 2.1 hold and additionally let the difference between Y_i and the noisy data $\overline{Y}_i$ satisfy*

$$\frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \overline{Y}_i\right)^2 \xrightarrow{P} 0. \tag{17}$$

For Y_i and $\overline{Y}_i$ assume uniform boundedness: $|Y| \leq L$, $|\overline{Y}_i| \leq L$ for some $L \in [0, \infty)$.
Then, for the estimate (16) with $\overline{m}_n = \overline{m}_n^{(LA)}$

$$\lim_{n\to\infty} \boldsymbol{E} \int \left(\overline{\sigma}_n^2(x) - \sigma^2(x)\right)^2 \mu(dx) = 0$$

holds. (Consistency of the local averaging estimator of the local variance with additional measurements error in the response variable)

Proof We apply Lemma 2.2 with Y_i, $\overline{Y}_i$, $\overline{m}_n$ and m replaced by $Y_i^2 - m^2(X_i)$, $\overline{Y}_i^2 - \overline{m}_n^2(X_i)$, σ_n^2 and σ^2, respectively. It is enough to show

$$\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n} |(\overline{Y}_i^2 - \overline{m}_n(X_i)^2) - (Y_i^2 - m^2(X_i))|^2\right\} \to 0.$$

The left-hand side is bounded by

$$\begin{aligned} & 2\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n} |\overline{Y}_i^2 - Y_i^2|^2 + \frac{1}{n}\sum_{i=1}^{n} |\overline{m}_n(X_i)^2 - m^2(X_i)|^2\right\} \\ \leq\ & c'\left(\boldsymbol{E}\frac{1}{n}\sum_{i=1}^{n} |\overline{Y}_i - Y_i|^2 + \boldsymbol{E}\frac{1}{n}\sum_{i=1}^{n} |\overline{m}_n(X_i) - m(X_i)|^2\right), \end{aligned}$$

for some finite constant c' because of the uniform boundedness assumption.
Because of (17) and the uniform boundness assumption we have

$$\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n} |\overline{Y}_i - Y_i|^2\right\} \to 0 \tag{18}$$

by the dominated convergence theorem. For $\overline{m}_n^{(LA)}$ we notice that by [2], Lemma 24.11 and Lemma 24.7 (Hardy-Littlewood) and its extension (24.10) together with pp. 595, 503, 504, respectively, (for the empirical measure with respect to $(X_1, \ldots, X_n)$ and the function $x_i \to \overline{y}_i - y_i$ $(i = 1, \ldots, n)$ for the realizations $(x_i, y_i, \overline{y}_i)$ of $(X_i, Y_i, \overline{Y}_i)$, without sup)

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\frac{1}{n}\sum_{j=1}^{n}(\overline{Y}_j - Y_j)1_{A_n(X_j)}(X_i)}{\frac{1}{n}1_{A_n(X_j)}(X_i)}\right|^2 \leq c^* \frac{1}{n}\sum_{i=1}^{n}(\overline{Y}_i - Y_i)^2$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\frac{1}{n}\sum_{j=1}^{n}(\overline{Y}_j - Y_j)K\left(\frac{X_i - X_j}{h_n}\right)}{\frac{1}{n}K\left(\frac{X_i - X_j}{h_n}\right)}\right|^2 \leq c^* \frac{1}{n}\sum_{i=1}^{n}(\overline{Y}_i - Y_i)^2,$$

respectively, for some finite constant c^*, thus

$$\boldsymbol{E}\left\{\frac{1}{n}|\overline{m}_n^{(LA)}(X_i) - m_n^{(LA)}(X_i)|^2\right\} \leq c^* \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|\overline{Y}_i - Y_i|^2\right\} \to 0.$$

Further

$$\begin{aligned} & \boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|m_n^{(LA)}(X_i) - m(X_i)|^2\right\} \\ = \;& \boldsymbol{E}\left\{\frac{1}{n}|m_n^{(LA)}(X_1) - m(X_1)|^2\right\} \\ & \text{(by symmetry)} \\ \to \;& 0 \text{ (by Lemma 2.3).} \end{aligned}$$

Therefore

$$\boldsymbol{E}\left\{\frac{1}{n}\sum_{i=1}^{n}|\overline{m}_n^{(LA)}(X_i) - m(X_i)|^2\right\} \to 0 \tag{19}$$

■

Remark 4.1 *One can obtain a result analogous to Theorem 4.1 for $\overline{m}_n = \overline{m}_n^{(LS)}$ where besides (18) one gets (19) with $\overline{m}^{(LA)}$ replaced by $\overline{m}^{(LS)}$ via conditioning and Lemmas 2, 3, 4 in Kohler [5] together with [2], Section 19.4 and Problems and Excercises, for extension to $d > 1$.*

Acknowledgments

The author thanks Prof. em. Dr. Harro Walk for precious advice, guidance and encouragement.

References

[1] L. Györfi. *Universal consistencies of a regression estimate for unbounded regression functions.* In *Nonparametric Functional Estimation and Related Topics*, G. Roussas, ed., pages 329-338. Kluwer Academic Publishers, Dordrecht, 1991.

[2] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer, New York, 2002.

[3] P. Hall, J.W. Kay, and D.M. Titterington. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77(3):521–528, 1990.

[4] W. Härdle and A. Tsybakov. Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, 81:223–242, 1997.

[5] M. Kohler. Nonparametric regression with additional measurement errors in the dependent variable. *Journal of statistical planning and inference*, 136:3339–3361, 2006.

[6] H-G. Müller and U. Stadtmüller. Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics*, 15(2):610–625, 1987.

[7] U. Müller, A. Schick, and W. Wefelmeyer. Estimating the error variance in nonparametric regression by a covariate-matched u-statistic. *Statistics*, 37(3):179–188, 2003.

[8] M.H. Neumann. Fully data-driven nonparametric variance estimators. *Statistics*, 25:189–212, 1994.

[9] Z. Pan and X. Wang. A wavelet-based nonparametric estimator of the variance function. *Computational Economics*, 15:79–87, 2000.

[10] D. Ruppert, M.P. Wand, U. Holst, and O. Hossjer. Local polynomial variance-function estimation. *Technometrics*, 39(3):262–273, 1997.

[11] V. Spokoiny. Variance estimation for high-dimensional regression models. *Journal of Multivariate Analysis*, 82:111–133, 2002.

[12] U. Stadtmüller and A.B. Tsybakov. Nonparametric recursive variance estimation. *Statistics*, 27:55–63, 1995.

[13] C.J. Stone. Nonparametric regression. *The Annals of Statistics*, 5(4):595–645, 1977.

[14] H. Walk. *Strong laws of large numbers and nonparametric estimation.* In *Recent Developments in Applied Probability and Statistics.* L. Devroye et al., eds., pages 183-214. Physica-Verlag, Heidelberg, 2010.

Paola Gloria Ferrario
Department of Mathematics, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany
E-Mail: paola.ferrario@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2007/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2011/015 *Ferrario, P.:* Nonparametric Local Averaging Estimation of the Local Variance Function

2011/014 *Müller, S.; Dippon, J.:* k-NN Kernel Estimate for Nonparametric Functional Regression in Time Series Analysis

2011/013 *Knarr, N.; Stroppel, M.:* Unitals over composition algebras

2011/012 *Knarr, N.; Stroppel, M.:* Baer involutions and polarities in Moufang planes of characteristic two

2011/011 *Knarr, N.; Stroppel, M.:* Polarities and planar collineations of Moufang planes

2011/010 *Jentsch, T.; Moroianu, A.; Semmelmann, U.:* Extrinsic hyperspheres in manifolds with special holonomy

2011/009 *Wirth, J.:* Asymptotic Behaviour of Solutions to Hyperbolic Partial Differential Equations

2011/008 *Stroppel, M.:* Orthogonal polar spaces and unitals

2011/007 *Nagl, M.:* Charakterisierung der Symmetrischen Gruppen durch ihre komplexe Gruppenalgebra

2011/006 *Solanes, G.; Teufel, E.:* Horo-tightness and total (absolute) curvatures in hyperbolic spaces

2011/005 *Ginoux, N.; Semmelmann, U.:* Imaginary Kählerian Killing spinors I

2011/004 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D-Scales: Part II — Gain-Scheduled Control

2011/003 *Scherer, C.W.; Köse, I.E.:* Control Synthesis using Dynamic D-Scales: Part I — Robust Control

2011/002 *Alexandrov, B.; Semmelmann, U.:* Deformations of nearly parallel G_2-structures

2011/001 *Geisinger, L.; Weidl, T.:* Sharp spectral estimates in domains of infinite volume

2010/018 *Kimmerle, W.; Konovalov, A.:* On integral-like units of modular group rings

2010/017 *Gauduchon, P.; Moroianu, A.; Semmelmann, U.:* Almost complex structures on quaternion-Kähler manifolds and inner symmetric spaces

2010/016 *Moroianu, A.; Semmelmann,U.:* Clifford structures on Riemannian manifolds

2010/015 *Grafarend, E.W.; Kühnel, W.:* A minimal atlas for the rotation group $SO(3)$

2010/014 *Weidl, T.:* Semiclassical Spectral Bounds and Beyond

2010/013 *Stroppel, M.:* Early explicit examples of non-desarguesian plane geometries

2010/012 *Effenberger, F.:* Stacked polytopes and tight triangulations of manifolds

2010/011 *Györfi, L.; Walk, H.:* Empirical portfolio selection strategies with proportional transaction costs

2010/010 *Kohler, M.; Krzyżak, A.; Walk, H.:* Estimation of the essential supremum of a regression function

2010/009 *Geisinger, L.; Laptev, A.; Weidl, T.:* Geometrical Versions of improved Berezin-Li-Yau Inequalities

2010/008 *Poppitz, S.; Stroppel, M.:* Polarities of Schellhammer Planes

2010/007 *Grundhöfer, T.; Krinn, B.; Stroppel, M.:* Non-existence of isomorphisms between certain unitals

2010/006 *Höllig, K.; Hörner, J.; Hoffacker, A.:* Finite Element Analysis with B-Splines: Weighted and Isogeometric Methods

2010/005 *Kaltenbacher, B.; Walk, H.:* On convergence of local averaging regression function estimates for the regularization of inverse problems

2010/004 *Kühnel, W.; Solanes, G.:* Tight surfaces with boundary

2010/003 *Kohler, M; Walk, H.:* On optimal exercising of American options in discrete time for stationary and ergodic data

2010/002 *Gulde, M.; Stroppel, M.:* Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras

2010/001 *Leitner, F.:* Examples of almost Einstein structures on products and in cohomogeneity one

2009/008 *Griesemer, M.; Zenk, H.:* On the atomic photoeffect in non-relativistic QED

2009/007 *Griesemer, M.; Moeller, J.S.:* Bounds on the minimal energy of translation invariant n-polaron systems

2009/006 *Demirel, S.; Harrell II, E.M.:* On semiclassical and universal inequalities for eigenvalues of quantum graphs

2009/005 *Bächle, A, Kimmerle, W.:* Torsion subgroups in integral group rings of finite groups

2009/004 *Geisinger, L.; Weidl, T.:* Universal bounds for traces of the Dirichlet Laplace operator

2009/003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation

2009/002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces

2009/001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of E^3 and of the 3-torus

2008/006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps

2008/005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems

2008/004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities

2008/003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope

2008/002 *Hertweck, M.; Hofert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2,q)$

2008/001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term

2007/006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term

2007/005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field

2007/004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees

2007/003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides

2007/002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry

2007/001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions