Universität Stuttgart

Fachbereich Mathematik

Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

Paola Gloria Ferrario, Harro Walk

Preprint 2011/022

Universität Stuttgart

Fachbereich Mathematik

Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

Paola Gloria Ferrario, Harro Walk

Preprint 2011/022

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 we consider first an estimator of the residual variance treated by Evans (and Jones) (2005, 2008) and by Liitiäinen et al. (2008, 2010), based on first and second nearest neighbors given an independent and identically distributed sample. Its strong consistency and strong Cesàro consistency are shown under mere boundedness and square integrability, respectively, of the dependent variable Y. Moreover, in view of the local variance, a correspondingly modified estimator of local averaging (partitioning) type is proposed, and strong L_1-consistency (for bounded Y) and rate of convergence (for bounded X and Y under Lipschitz continuity of the regression and the local variance function) are established.

Key words: regression function, residual variance, local variance, partitioning estimation, nearest neighbors, strong consistency, rate of convergence.

AMS Subject classification: 62G05, 62G20.

1 Introduction

Let Y be a square integrable real valued random variable and let X be a d-dimensional random vector, taking values in the space $\mathbb{R}^d$. The task of regression analysis is to estimate Y given X, i.e., to find a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, such that $f(X)$ is a "good approximation" of Y, that is, $|f(X) - Y|$ has to be "small". The "closeness" of $f(X)$ to Y is typically measured by the so-called **mean squared error** of f,

$$\boldsymbol{E}\{(Y - f(X))^2\}.$$

It is well known that the regression function m minimizes this error (where $m := \boldsymbol{E}\{Y|X = x\}$),

$$V := \min_f \boldsymbol{E}\{(Y - f(X))^2\} = \boldsymbol{E}\{(Y - m(X))^2\}. \tag{1}$$

V, the so-called residual variance, is a measure of how close we can get to Y using any measurable function f. It indicates how difficult a regression problem is. Since the distribution of m, and therefore m, are unknown, one is interested in estimating V by use of data observations

$$D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}, \tag{2}$$

which are independent copies of (X, Y).
A related interesting problem is the estimation of the local variance (or conditional variance), defined as

$$\sigma^2(x) := \boldsymbol{E}\{(Y - m(X))^2|X = x\} = \boldsymbol{E}\{Y^2|X = x\} - m^2(x). \tag{3}$$

It holds

$$V = \boldsymbol{E}\{\sigma^2(X)\}. \tag{4}$$

Liitiäinen et al. [12], with generalization in [13], investigated an estimator of the residual variance V, introduced by Evans (and Jones) [5, 6], which is based on first and second nearest neighbors. They obtained mean square convergence under bounded conditional fourth moment of Y and convergence order $O(n^{-2/d})$ for $d \geq 2$ under finite suitable moments of X and under Lipschitz continuity of m. It simplifies an estimator given in Devroye et al. [3], based on first nearest neighbors. References for the estimation of the local variance function, incl. the case of fixed design, are Müller and Stadtmüller [15, 16], Stadtmüller and Tsybakov [23], Ruppert et al. [21], Härdle and Tsybakov [9], Spokoiny [22], Pan and Wang [20], Hall et al. [8], Müller et al. [17], Neumann [19], Munk et al. [18], Kohler [10], Brown and Levine [1], and Cai et al. [2].
In this paper, first we show strong consistency of the (global) residual variance estimation sequence of Evans (and Jones) [5, 6] and Liitiäinen et al. [12, 13], under boundedness of Y and show strong consistency of the sequence of arithmetic means in the general case $\boldsymbol{E}\{Y^2\} < \infty$ (Section 2).
In Section 3 for the estimation of the local variance function σ^2 on the basis of data (2), we propose an estimation sequence (σ_n^2) of local averaging, namely partitioning, type. It is a modification of the (global) residual vairance estimator and uses again first and second nearest neighbors. We show strong L_1-consistency, that is, $\int |\sigma_n^2(x) - \sigma^2|\mu(dx) \to 0$ $a.s.$, under mere boundedness of Y (μ denoting the distribution of X).
Finally, in Section 4 we establish its rate, imposing Lipschitz conditions on σ^2 and on m together with boundedness of X and Y.

2 Residual Variance Estimation

In the literature different paradigms how to construct nonparametric estimates are treated. Beside the least squares approach, local averaging paradigms are used, especially kernel estimates, partitioning estimates and k-th nearest neighbor estimates. A reference is Györfi et al. [7].
For given $i \in \{1, \ldots, n\}$, the first nearest neighbor of X_i among $X_1, \ldots, X_{i-1}$, $X_{i+1}, \ldots, X_n$ is defined as $X_{[N,1]}$ with

$$N[i, 1] := N_n[i, 1] := \mathop{\arg\min}_{1 \leq j \leq n,\ j \neq i} \rho(X_i, X_j), \tag{5}$$

here ρ is a metric (typically the Euclidean one) in $\mathbb{R}^d$. The k-th nearest neighbor of X_i among $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ is defined as $X_{N[i,k]}$ via generalization of definition (5):

$$N[i,k] := N_n[i,k] := \underset{1 \le j \le n,\ j \ne i,\ j \notin \{N[i,1], \dots, N[i,k-1]\}}{\arg\min} \rho(X_i, X_j), \tag{6}$$

by removing the preceeding neighbors. If ties occur, a possibility to break them is given by taking the minimal index or by adding independent components Z_i, uniformly distributed on $[0,1]$, to the observation vectors X_i (see [7], pp. 86, 87). The latter possibility to break ties allow us to assume throughout the paper that ties occur with probability zero.
Hence, we get a reorder of the data according to increasing values of the distance of the variable X_j ($j \in \{1, \dots, n\} \setminus \{i\}$) from the variable X_i ($i = 1, \dots, n$). Correspondingly to that, we get also a new order for the variables Y_j :

$$\left(X_{N[i,1]},\ Y_{N[i,1]}\right),\ \dots, \left(X_{N[i,k]},\ Y_{N[i,k]}\right),\ \dots, \left(X_{N[i,n-1]},\ Y_{N[i,n-1]}\right).$$

In the following $N[i,1]$ and $N[i,2]$ will be used.
For the residual variance V, Evans (and Jones) [5, 6] introduced and Liitiäinen et al. [12, 13] analyzed (and generalized) the estimator

$$V_n = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - Y_{N[i,1]}\right) \left(Y_i - Y_{N[i,2]}\right), \tag{7}$$

in view of square mean consistency and rate of convergence.
We shall establish strong consistency.

Theorem 2.1 *If $|Y| \le L$ for some $L \in R_+$, then*

$$V_n \to V \quad a.s. \quad (n \to \infty).$$

The proof is based on the McDiarmid inequality (see, e.g., [7], Theorem A.2) and properties of nearest neighbors (see [7], Lemma 6.1 and Corollary 6.1 together with Lemma 6.3). To make the paper more self-contained, we state them in the following lemmas.

Lemma 2.2 *(McDiarmid inequality) Let $Z_1, \dots, Z_n$ be independent random variables taking values in a set A and assume that $f : A^n \to \mathbb{R}$ satisfies*

$$\sup_{z_1, \dots, z_n,\ z_i' \in A} |f(z_1, \dots, z_n) - f(z_1, \dots, z_{i-1}, z_i', z_{i+1}, \dots, z_n)| \le c_i, \quad 1 \le i \le n.$$

Then, for all $\epsilon > 0$,

$$\boldsymbol{P}\{f(Z_1, \dots, Z_n) - \boldsymbol{E} f(Z_1, \dots, Z_n) \ge \epsilon\} \le e^{-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}},$$

and

$$\boldsymbol{P}\{\boldsymbol{E} f(Z_1, \dots, Z_n) - f(Z_1, \dots, Z_n) \ge \epsilon\} \le e^{-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}}.$$

Lemma 2.3 *If $k_n/n \to 0$, then*

$$\|X_{N[1,k_n]} - X_1\| \to 0 \quad a.s.$$

Lemma 2.4 *Under the assumption that ties occur with probability zero,*

a)

$$\sum_{i=1}^{n} 1_{\{X \text{ is among the } k \text{ nearest neighbors of } X_i \text{ in } \{X_1, \dots, X_{i-1}, X, X_{i+1}, \dots, X_n\}\}}$$
$$\le k\gamma_d \quad a.s. \quad (k \le n),$$

b) for any integrable function f and any $k \leq n-1$,

$$\sum_{j=1}^{k} \boldsymbol{E}\{|f(X_{N[1,j]})|\} \leq k\gamma_d \boldsymbol{E}\{|f(X_1)|\},$$

Here $\gamma_d < \infty$ depends only on d.

Proof of Theorem 2.1 In the first step we show

$$\boldsymbol{E} V_n \to V \tag{8}$$

(asymptotic unbiasedness), using only square integrability of Y, compare [13], proof of Theorem 2.2.
With the notations

$$\begin{aligned} b_{i,j} &= m(X_i) - m(X_j) \\ r_i &= Y_i - m(X_i), \end{aligned} \tag{9}$$

we can write, according to [12] and [13]:

$$\begin{aligned} &\boldsymbol{E}\left\{Y_i - Y_{N[i,1]}\right)\left(Y_i - Y_{N[i,2]}\right\} = \\ &\boldsymbol{E}\left\{b_{i,N[i,1]}\left(r_i - r_{N[i,2]}\right)\right\} + \boldsymbol{E}\left\{b_{i,N[i,2]}\left(r_i - r_{N[i,1]}\right)\right\} \\ &+\boldsymbol{E}\left\{\left(r_i - r_{N[i,1]}\right)\left(r_i - r_{N[i,2]}\right)\right\} + \boldsymbol{E}\left\{b_{i,N[i,1]} b_{i,N[i,2]}\right\}. \end{aligned}$$

As shown in [12] and [13] via conditioning with respect to $X_1, \ldots, X_n$,

$$\boldsymbol{E}\left\{b_{i,N[i,1]}\left(r_i - r_{N[i,2]}\right)\right\} = \boldsymbol{E}\left\{b_{i,N[i,2]}\left(r_i - r_{N[i,1]}\right)\right\} = 0,$$

and

$$\boldsymbol{E}\left\{\left(r_i - r_{N[i,1]}\right)\left(r_i - r_{N[i,2]}\right)\right\} = \boldsymbol{E}\{r_i^2\} = \boldsymbol{E}\{(Y_i - m(X_i))^2\} = V.$$

Further

$$\left|\boldsymbol{E}\left\{b_{i,N[i,1]} b_{i,N[i,2]}\right\}\right| \leq \boldsymbol{E}\left\{|(m(X_i) - m(X_{N[i,1]})||m(X_i) - m(X_{N[i,2]})|\right\}.$$

Thus, because the X_i's are identically distributed,

$$\begin{aligned} &|\boldsymbol{E} V_n - V| \leq \boldsymbol{E}\left\{|(m(X_1) - m(X_{N[1,1]})||m(X_1) - m(X_{N[1,2]})|\right\} \\ \leq\ &\frac{1}{2}\boldsymbol{E}\left\{|m(X_1) - m(X_{N[1,1]})|^2\right\} + \frac{1}{2}\boldsymbol{E}\left\{|m(X_1) - m(X_{N[1,2]})|^2\right\}. \end{aligned}$$

Because the set of continuous functions on $\mathbb{R}^d$ with compact support is dense in $L_2(\mu)$ (see, e.g., [4], Chapter 4, Section 8.19, or [7], Theorem A.1), for an arbitrary $\epsilon > 0$ one can choose a continuous function $\widetilde{m}$ with compact support such that $\boldsymbol{E}\{|m(X_1) - \widetilde{m}(X_1)|^2\} \leq \epsilon$. Then

$$\begin{aligned} &\boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})|\} \\ \leq\ &3\boldsymbol{E}\{|(m-\widetilde{m})(X_1)|^2\} + 3\boldsymbol{E}\{|(m-\widetilde{m})(X_{N[1,1]})|^2\} \\ &+3\boldsymbol{E}\{|(\widetilde{m}(X_1) - \widetilde{m}(X_{N[1,1]})|^2\}. \end{aligned}$$

By Lemma 2.3 (with $k_n = 1$) and continuity of $\widetilde{m}$, one has

$$\widetilde{m}(X_{N[1,1]}) \to \widetilde{m}(X_1) \quad a.s.,$$

thus, by boundedness of $\widetilde{m}$,

$$\boldsymbol{E}\{|\widetilde{m}(X_1) - \widetilde{m}(X_{N[1,1]})|^2\} \to 0.$$

Further, by Lemma 2.4b,

$$\begin{aligned} &\boldsymbol{E}\{|(m-\widetilde{m})(X_{N[1,1]})|^2\} \\ \leq\ &\gamma_d \boldsymbol{E}\{|(m-\widetilde{m})(X_1)|\} \leq \gamma_d \epsilon. \end{aligned}$$

Therefore

$$\limsup_{n\to\infty} \boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})|^2\} \leq 3(1+\gamma_d)\epsilon,$$

thus

$$\boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})|^2\} \to 0.$$

Analogously one obtains $\boldsymbol{E}\{|m(X_1) - m(X_{N[1,2]})|^2\} \to 0$. Thus

$$\boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})||m(X_1) - m(X_{N[1,2]})|\} \to 0, \tag{10}$$

and (8) is obtained.
In the second step we show

$$V_n - \boldsymbol{E}V_n \to 0 \quad a.s. \tag{11}$$

Set

$$T_n := \sum_{i=1}^{n} (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}).$$

Now in view of an application of Lemma 2.2, let $(X_1, Y_1), \dots, (X_n, Y_n)$, $(X_1', Y_1'), \dots, (X_n', Y_n')$ be independent and identically distributed $(d+1)$-dimensional random vectors. For fixed $j \in \{1, \dots, n\}$ replace (X_j, Y_j) by (X_j', Y_j'), which leads to $T_{n,j}$. Noticing $|Y_i| \le L$, we have

$$|T_n - T_{n,j}| \le 8L^2 + 8L^2 \cdot 2 \cdot 2\gamma_d = 8(1 + 4\gamma_d)L^2, \tag{12}$$

where the first term of the right-hand side results from summand $i = j$ and the second term results from summands $i \in \{1, \dots, n\} \setminus \{j\}$, because replacement of X_j by X_j' has an influence on the first and second nearest neighbors of some, but at most $2\gamma_d$ (by Lemma 2.4 a), of the random vectors $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n$. By Lemma 2.2, for each $\epsilon > 0$ we obtain

$$\begin{aligned} & \boldsymbol{P}\{|V_n - \boldsymbol{E}V_n| \ge \epsilon\} \\ = \; & \boldsymbol{P}\{|T_n - \boldsymbol{E}T_n| \ge \epsilon n\} \\ \le \; & 2e^{-2\epsilon^2 n^2 / n(8(1+4\gamma_d)L^2)^2}, \end{aligned}$$

thus (11) by the Borel-Cantelli lemma.
(8) and (11) yield the assertion.

■

The following theorem states that the boundedness assumption in Theorem 2.1 on Y can be omitted if for estimation of V the sequence $((V_1 + \dots, V_n)/n)$ of arithmetic means insted of (V_n) is used.

Theorem 2.5 *In the general case* $\boldsymbol{E}\{Y^2\} < \infty$,

$$\frac{V_1 + \dots, V_n}{n} \to V \quad a.s.$$

(strong Cesàro consistency of (V_n)).

It remains an open problem whether $V_n \to V$ *a.s.* if $\boldsymbol{E}\{Y^2\} < \infty$.
For the proof of Theorem 2.5 we shall use an Efron-Stein inequality (Lemma 2.6, compare [7], Theorem A.3).

Lemma 2.6 *Let $Z_1, \dots, Z_n, \widetilde{Z}_1, \dots, \widetilde{Z}_n$ be independent m-dimensional random vectors where the two random vectors Z_k and $\widetilde{Z}_k$ have the same distribution ($k = 1, \dots, n$). For measurable $f : \mathbb{R}^{m \cdot n} \to \mathbb{R}$ assume that $f(Z_1, \dots, Z_n)$ is square integrable. Then*

$$\begin{aligned} & \boldsymbol{Var}\{f(Z_1, \dots, Z_n)\} \\ \le \; & \frac{1}{2} \sum_{k=1}^{n} \boldsymbol{E}\left\{|f(Z_1, \dots, Z_k, \dots, Z_n) - f(Z_1, \dots, \widetilde{Z}_k, \dots, Z_n)|^2\right\}. \end{aligned}$$

Proof of Theorem 2.5 For a real random variable U we set

$$U^{[c]} := U1_{\{|U|\le c\}} + c1_{\{U>c\}} - c1_{\{U<-c\}}, \quad c > 0.$$

First we show

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_{N[i,1]}\right)\left(Y_i - Y_{N[i,2]}\right) - \frac{1}{n}\sum_{i=1}^{n} V_{n,i} \to 0 \quad a.s.,$$

where

$$V_{n,i} := \left(Y_i^{[\sqrt{n}]} - Y_{N[i,1]}^{[\sqrt{n}]}\right)\left(Y_i^{[\sqrt{n}]} - Y_{N[i,2]}^{[\sqrt{n}]}\right).$$

Because $\boldsymbol{E}\{Y\}^2 < \infty$, $a.s.$ $Y_i = Y_i^{[\sqrt{i}]}$ for i sufficiently large, say, $i \ge M$ (random). For $i \in \{M, M+1, \ldots, n\}$, $a.s.$ $Y_i = Y_i^{[\sqrt{n}]}$. By Lemma 2.4a, for $p \in \{1, \ldots, M\}$ one has $N[i,1] = p$ for at most γ_d indices $\in \{1, \ldots, n\}$ and $N[i,2] = p$ for at most $2\gamma_d$ indices $i \in \{1, \ldots, n\}$. Thus $a.s.$

$$\left(Y_i - Y_{N[i,1]}\right)\left(Y_i - Y_{N[i,2]}\right) \ne \left(Y_i^{[\sqrt{n}]} - Y_{N[i,1]}^{[\sqrt{n}]}\right)\left(Y_i^{[\sqrt{n}]} - Y_{N[i,2]}^{[\sqrt{n}]}\right)$$

for at most $(1+3\gamma_d)M$ indices $\in \{1, \ldots, n\}$, which yields the assertion.
Therefore it suffices to show

$$\frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{l}\sum_{i=1}^{l} V_{l,i}\right) \to V \quad a.s. \tag{13}$$

In the second step we show

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{E} V_{n,i} \to V. \tag{14}$$

With $m^{(n)}(x) := \boldsymbol{E}\{Y^{[\sqrt{n}]} | X = x\}$ we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{E} V_{n,i} = \boldsymbol{E} V_{n,1} \\
= \;& \boldsymbol{E}\left\{(Y^{[\sqrt{n}]} - m^{(n)}(X))^2\right\} \\
+ \;& \boldsymbol{E}\left\{\left(m^{(n)}(X_1) - m^{(n)}(X_{N[1,1]})\right)\left(m^{(n)}(X_1) - m^{(n)}(X_{N[1,2]})\right)\right\},
\end{aligned}$$

the latter according to Liitiäinen et al. [12, 13]. By $\boldsymbol{E}\{Y^2\} < \infty$ and the dominated convergence theorem, $\int |m^{(n)}(x) - m(x)|^2 \mu(dx) \to 0$ and thus $\boldsymbol{E}\{(Y^{[\sqrt{n}]} - m^{(n)}(X))^2\} \to V$. Further m and also $m^{(n)}$ can be approximated by a continuous function $\widetilde{m}$ with compact support such that for each $\epsilon > 0$ an index $n_0(\epsilon)$ exists with $\boldsymbol{E}\{|m(X) - \widetilde{m}(X)|^2\} \le \epsilon$ and also

$$\boldsymbol{E}\{|m^{(n)}(X) - \widetilde{m}(X)|^2\} \le \epsilon \text{ for } n \ge n_0(\epsilon).$$

Then we obtain

$$\begin{aligned}
&\boldsymbol{E}\left\{\left|m^{(n)}(X_1) - m^{(n)}(X_{N[1,1]})\right|^2\right\} \\
\le \;& 3\boldsymbol{E}\{|(m^{(n)} - \widetilde{m})(X_1)|^2\} + \\
& 3\boldsymbol{E}\{|(m^{(n)} - \widetilde{m})(X_{N[1,1]})|^2\} + 3\boldsymbol{E}\{|\widetilde{m}(X_1) - \widetilde{m}(X_{N[1,1]})|^2\} \\
\le \;& 3\epsilon + 3\gamma_d\epsilon + o(1),
\end{aligned}$$

the latter as in the proof of Theorem 2.1. Therefore

$$\boldsymbol{E}\left\{\left|m^{(n)}(X_1) - m^{(n)}(X_{N[1,1]})\right|^2\right\} \to 0$$

and correspondingly

$$\boldsymbol{E}\left\{\left|m^{(n)}(X_1) - m^{(n)}(X_{N[1,2]})\right|^2\right\} \to 0,$$

thus

$$\boldsymbol{E}\left\{\left(m^{(n)}(X_1) - m^{(n)}(X_{N[1,1]})\right)\left(m^{(n)}(X_1) - m^{(n)}(X_{N[1,2]})\right)\right\} \to 0,$$

and (14) is obtained as well as

$$\frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{l}\sum_{i=1}^{l}\boldsymbol{E}V_{l,i}\right)\to V. \tag{15}$$

In the second step we show

$$\frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{l}\sum_{i=1}^{l}(V_{l,i}-\boldsymbol{E}V_{l,i})\right)\to 0 \quad a.s. \tag{16}$$

It suffices to show

$$\sum\frac{\boldsymbol{Var}\left\{\sum_{i=1}^{n}V_{n,i}\right\}}{n^3}<\infty, \tag{17}$$

for this implies

$$\sum\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}(V_{n,i}-\boldsymbol{E}V_{n,i})\right)^2<\infty \quad a.s.$$

and, by the Cauchy-Schwarz inequality and the Kronecker lemma,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{l}\sum_{i=1}^{l}\left(V_{l,i}-\boldsymbol{E}V_{l,i}\right)\right)\right|^2\\ \leq\ &\frac{1}{n}\sum_{l=1}^{n}\left|\frac{1}{l}\sum_{i=1}^{l}\left(V_{l,i}-\boldsymbol{E}V_{l,i}\right)\right|^2\to 0 \quad a.s.\end{aligned}$$

We shall show

$$\boldsymbol{Var}\left\{\sum_{i=1}^{n}V_{n,i}\right\}\leq cn\boldsymbol{E}\left\{\left(Y^{[\sqrt{n}]}\right)^4\right\},\ n\in\mathbb{N} \tag{18}$$

for a suitable finite constant c. This, together with $\boldsymbol{E}\{Y^2\}<\infty$, implies (17), because, as is well known (see, e.g., [14], Section 17.3), $\boldsymbol{E}|U|<\infty$ for a real variable U implies $\sum\boldsymbol{E}\left\{\left(U^{[n]}\right)^2\right\}/n^2<\infty$.

We prove (18) by using the Efron-Stein inequality (Lemma 2.6).

Let $n\geq 2$ be fixed. Replacement of (X_j,Y_j) by (X'_j,Y'_j) for fixed $j\in\{1,\dots,n\}$ (where $(X_1,Y_1),\dots(X_n,Y_n),(X'_1,Y'_1),\dots,(X'_n,Y'_n)$ are independent and identically distributed) leads from $T_n:=\sum_{i=1}^{n}V_{n,i}$, $N[j,1]$ and $N[j,2]$ to $T_{n,j}$, $N'[j,1]$ and $N'[j,2]$, respectively.

We obtain

$$|T_n-T_{n,j}|\leq A_{n,j}+B_{n,j}+C_{n,j}+D_{n,j}+E_{n,j}+F_{n,j}$$

where with $Z_i=Y_i^{[\sqrt{n}]}$, $Z'_j=Y_j'^{[\sqrt{n}]}$, $Z=Y^{[\sqrt{n}]}$

$$A_{n,j}=\sum_{\substack{l,\ q\in\{1,\dots,n\}\setminus\{j\}\\ l\neq q}}|Z_j-Z_l||Z_j-Z_q|1_{\{N[j,1]=l\}}1_{\{N[j,2]=q\}},$$

$$B_{n,j}=\sum_{\substack{l,\ q\in\{1,\dots,n\}\setminus\{j\}\\ l\neq q}}|Z'_j-Z_l||Z'_j-Z_q|1_{\{N'[j,1]=l\}}1_{\{N'[j,2]=q\}},$$

$$C_{n,j}=\sum_{\substack{i,\ q\in\{1,\dots,n\}\setminus\{j\}\\ i\neq q}}|Z_i-Z_j||Z_i-Z_q|1_{\{N[i,1]=j\}}1_{\{N[i,2]=q\}},$$

$$D_{n,j}=\sum_{\substack{i,\ q\in\{1,\dots,n\}\setminus\{j\}\\ i\neq q}}|Z_i-Z'_j||Z_i-Z_q|1_{\{N'[i,1]=j\}}1_{\{N'[i,2]=q\}},$$

$$E_{n,j}=\sum_{\substack{i,\ l\in\{1,\dots,n\}\setminus\{j\}\\ i\neq l}}|Z_i-Z_l||Z_i-Z_j|1_{\{N[i,1]=l\}}1_{\{N[i,2]=j\}},$$

$$F_{n,j} = \sum_{\substack{i,\ l \in \{1,\dots,n\}\setminus\{j\} \\ i \neq l}} |Z_i - Z_l||Z_i - Z'_j| 1_{\{N'[i,1]=l\}} 1_{\{N'[i,2]=j\}}.$$

Thus

$$|T_n - T_{n,j}|^2 \leq 6(A_{n,j}^2 + B_{n,j}^2 + C_{n,j}^2 + D_{n,j}^2 + E_{n,j}^2 + F_{n,j}^2).$$

By the Cauchy-Schwarz inequality applied to the sums defining $A_{n,j}, \dots,$
$F_{n,j}$ and by the inequality $|a-b|^2|a-c|^2 \leq 8(a^4 + b^4 + c^4)$ we obtain

$$\boldsymbol{E} \sum_{j=1}^{n} |T_n - T_{n,j}|^2 \leq 6 \cdot 6 \cdot 8 \boldsymbol{E} \sum_{\substack{j,\ l,\ q \in \{1,\dots,n\} \\ j \neq l,\ l \neq q,\ q \neq j}} (Z_j^4 + Z_l^4 + Z_q^4) 1_{\{N[j,1]=l\}} 1_{\{N[j,2]=q\}}.$$

As to the term concerning Z_j^4 we sum with respect to l and q and for the corresponding expected final sum we obtain the bound $n\boldsymbol{E}\left\{Z^4\right\}$. As to the term Z_l^4 we sum with respect to q, then with respect to j using Lemma 2.4a, and for the corresponding expected final sum we obtain the bound $\gamma_d n \boldsymbol{E}\left\{Z^4\right\}$. As to the term Z_q^4 we sum with respect l, then with respect to j using Lemma 2.4b and for the corresponding expected final sum we obtain the bound $2\gamma_d n \boldsymbol{E}\left\{Z^4\right\}$.
Therefore, by Lemma 2.6,

$$\boldsymbol{Var}(T_n) \leq \frac{1}{2} \cdot 6 \cdot 6 \cdot 8 \cdot (1 + 3\gamma_d) n \boldsymbol{E}\left\{\left(Y^{[\sqrt{n}]}\right)^4\right\},$$

i.e., (18). Thus (16) is obtained, which together with (15) implies (13). ∎

3 Local Variance Estimation: Strong Consistency

V_n in (7) as an estimator of $V = \boldsymbol{E}\{(Y - m(X))^2\}$ was treated in Section 2. In this section our aim is to give an estimator of the local variance function σ^2 in (3). Recall the relation between the residual and the local variance function in (4).
Our proposal for an appropriate estimator of σ^2 is

$$\sigma_n^2(x) := \frac{\sum_{i=1}^{n} (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_n(x)}(X_i)}{\sum_{i=1}^{n} 1_{A_n(x)}(X_i)},\ x \in \mathbb{R}^d \tag{19}$$

where $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \dots\}$ is a partition of $\mathbb{R}^d$ consisting of Borel sets $A_{n,j} \subset \mathbb{R}^d$, and where the notation $A_n(x)$ is used for the $A_{n,j}$ containing x. In this sense we localize the global expression in V_n by local averaging, in particular by partitioning. Analogously a kernel type estimator could be treated. The next theorem deals with strong consistency of the local variance estimator.

Theorem 3.1 *Let $(\mathcal{P}_n)_{n\in\mathbb{N}}$ with $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \dots\}$ be a sequence of partitions of $\mathbb{R}^d$ such that for each sphere S centered at the origin*

$$\lim_{n\to\infty} \max_{j: A_{n,j} \cap S \neq \emptyset} \mathit{diam}\ A_{n,j} \to 0 \tag{20}$$

and, for some $\rho = \rho(S) \in (0, \frac{1}{2})$

$$\#\{j : A_{n,j} \cap S \neq \emptyset\} \sim n^{\rho}. \tag{21}$$

Finally, let $|Y| \leq L$ for some $L \in \mathbb{R}_+$. Then

$$\int |\sigma_n^2(x) - \sigma^2(x)| \mu(dx) \to 0 \quad a.s.$$

Set now

$$\sigma_n^{2*}(x) := \frac{\sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_n(x)}(X_i)}{n\mu(A_n(x))}. \tag{22}$$

For the proof of Theorem 3.1 we need Lemma 3.3, which is based on Lemma 3.2 and the McDiarmid inequality (Lemma 2.2). Lemma 3.2 itself is based on the Efron-Stein inequality in Lemma 2.6.

Lemma 3.2 *Under (20) and (21), for each sphere S centered at* 0

$$\boldsymbol{E}\left\{\int_S |\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx)\right\} \to 0.$$

Proof One has

$$\begin{aligned}
& \boldsymbol{E}\left\{\int_S |\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx)\right\} \\
\leq\ & \int_S |\sigma^2(x) - \boldsymbol{E}\sigma_n^{2*}(x)|\mu(dx) + \boldsymbol{E}\left\{\int_S |\boldsymbol{E}\sigma_n^{2*}(x) - \sigma_n^{2*}(x)|\mu(dx)\right\} \\
\leq\ & K_n + M_n
\end{aligned}$$

First $K_n \to 0$ will be shown. According to Liitiäinen et al. [12, 13] one has

$$\begin{aligned}
& \boldsymbol{E}\{(Y_1 - Y_{N[1,1]})(Y_1 - Y_{N[1,2]})|X_1 = z\} \\
=\ & \sigma^2(z) + \boldsymbol{E}\{(m(X_1) - m(X_{N[1,1]}))(m(X_1) - m(X_{N[1,2]}))|X_1 = z\},
\end{aligned}$$

thus

$$\begin{aligned}
& \boldsymbol{E}\sigma_n^{2*}(x) \\
=\ & \int \frac{\sigma^2(z) 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz) \\
& + \int \frac{\boldsymbol{E}\{(m(X_1) - m(X_{N[1,1]}))(m(X_1) - m(X_{N[1,2]}))|X_1 = z\} 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz).
\end{aligned}$$

Notice

$$\begin{aligned}
& \int \left[\int \frac{\boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})||m(X_1) - m(X_{N[1,2]})||X_1 = z\} 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz)\right]\mu(dx) \\
=\ & \int \left[\int \frac{\boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]})||m(X_1) - m(X_{N[1,2]})||X_1 = z\} 1_{A_n(z)}(x)}{\mu(A_n(x))}\mu(dx)\right]\mu(dz) \\
\leq\ & \boldsymbol{E}\{|m(X_1) - m(X_{N[1,1]}||m(X_1) - m(X_{N[1,2]})|\} \\
& \to 0
\end{aligned}$$

by (10). Moreover,

$$\int \left|\sigma^2(x) - \int \frac{\sigma^2(z) 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz)\right|\mu(dx) \to 0.$$

For, because of $\int \sigma^2(x)\mu(dx) < \infty$, as in the proof of Theorem 2.1 for each $\epsilon > 0$ one can choose a continuous function $\widetilde{\sigma}^2$ with compact support such that

$$\int |\sigma^2(x) - \widetilde{\sigma}^2(x)|\mu(dx) < \epsilon,$$

further

$$\begin{aligned}
& \int \left|\int \frac{\sigma^2(z) 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz) - \int \frac{\widetilde{\sigma}^2(z) 1_{A_n(x)}(z)}{\mu(A_n(x))}\mu(dz)\right|\mu(dx) \\
\leq\ & \int \left|\sigma^2(z) - \widetilde{\sigma}^2(z)\right|\mu(dz) < \epsilon,
\end{aligned}$$

and one then notices

$$\int_S \left| \widetilde{\sigma}^2(x) - \int \frac{\widetilde{\sigma}^2(z) 1_{A_n(x)}(z)}{\mu(A_n(x))} \mu(dz) \right| \mu(dx) \to 0$$

because of uniform continuity of $\widetilde{\sigma}$ and (20). Therefore $K_n \to 0$.
Now M_n will be treated. Set $J_n := \{j : \ A_{n,j} \cap S \neq \emptyset\}$ and $l_n := \# J_n$.

$$\begin{aligned}
M_n &= \sum_{j \in J_n} \boldsymbol{E} \left\{ \int_{A_{n,j}} \left| \frac{\sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_{n,j}}(X_i)}{n\mu(A_{n,j})} \right.\right. \\
&\quad \left.\left. - \boldsymbol{E} \frac{\sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_{n,j}}(X_i)}{n\mu(A_{n,j})} \right| \mu(dx) \right\} \\
&\leq \frac{1}{n} \sum_{j \in J_n} \boldsymbol{E} \left\{ \left| \sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_{n,j}}(X_i) \right.\right. \\
&\quad \left.\left. - \boldsymbol{E} \sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_{n,j}}(X_i) \right| \right\} \\
&\leq \frac{1}{n} \sum_{j \in J_n} \sqrt{\boldsymbol{Var} \left\{ \sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_{n,j}}(X_i) \right\}} \\
&\leq \frac{l_n}{n} \sqrt{\frac{n}{2} (8L^2 + 8L^2 \cdot 2 \cdot 2\gamma_d)^2} \\
&\quad \text{(by Lemma 2.6 and the derivation of (12))} \\
&\leq 4\sqrt{2}(1 + 4\gamma_d) L^2 \frac{l_n}{\sqrt{n}} \to 0 \quad \text{(by (11))}.
\end{aligned}$$

Thus the assertion is obtained. ■

Lemma 3.3 *Assume (20) and (21). Let S be an arbitrary sphere centered at 0. Then a constant $c > 0$ exists such that for each $\epsilon > 0$*

$$\boldsymbol{P} \left\{ \int_S |\sigma^2(x) - \sigma_n^{2*}(x)| \mu(dx) > 2\epsilon \right\} \leq e^{-\epsilon^2 c n^{1-2\rho}}$$

for n sufficiently large.

Proof We follow the argument in the proof of Lemma 23.2 in [7]. One has

$$\begin{aligned}
& |\sigma^2(x) - \sigma_n^{2*}(x)| \\
= \ & \boldsymbol{E}|\sigma^2(x) - \sigma_n^{2*}(x)| + (|\sigma^2(x) - \sigma_n^{2*}(x)| - \boldsymbol{E}|\sigma^2(x) - \sigma_n^{2*}(x)|).
\end{aligned}$$

But $\int_S \boldsymbol{E}|\sigma^2(x) - \sigma_n^{2*}(x)| \mu(dx) \to 0$ due to Lemma 3.2.
Now, in view of an application of McDiarmid's inequality (Lemma 2.2) replacing (X_i, Y_i) by (X_i', Y_i') as in the proof of Theorem 2.1 leads from $\sigma_n^{2*}(x)$ to $\sigma_{n,j}^{2*}(x)$, $(j \in \{1, \ldots, n\})$, where, correspondingly to (12),

$$|\sigma_n^{2*}(x) - \sigma_{n,j}^{2*}(x)| \leq \frac{8(1 + 4\gamma_d)L^2}{n\mu(A_n(x))}.$$

Thus

$$\begin{aligned}
& \left| \int_S |\sigma^2(x) - \sigma_n^{2*}(x)| \mu(dx) - \int |\sigma^2(x) - \sigma_{n,j}^{2*}(x)| \mu(dx) \right| \\
= \ & \left| \int_S \left(|\sigma^2(x) - \sigma_n^{2*}(x)| - |\sigma^2(x) - \sigma_{n,j}^{2*}(x)| \right) \mu(dx) \right|
\end{aligned}$$

$$
\begin{aligned}
&\le \int_S |\sigma_n^2(x) - \sigma_{n,j}^{2*}(x)|\mu(dx) \quad (j = 1, \ldots, n) \\
&\quad \text{(due to the triangle inequality } |a-b| \ge ||a| - |b||) \\
&\le \frac{8(1+4\gamma_d)L^2}{n} \int_S \frac{1}{\mu(A_n(x))}\mu(dx) \\
&\le \frac{8(1+4\gamma_d)L^2}{n} l_n,
\end{aligned}
$$

where $l_n := \#\{j : A_{n,j} \cap S \neq \emptyset\}$.
Now, using Lemma 2.2, for arbitrary $\epsilon > 0$

$$
\begin{aligned}
&\boldsymbol{P}\left\{\int_S \left(\int |\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx) - \boldsymbol{E}\int_S |\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx)\right) > \epsilon\right\} \\
&\le e^{-2\epsilon^2/n\frac{[8(1+4\gamma_d)L^2]^2}{n^2}l_n^2} \\
&\le e^{-\epsilon^2 c n^{1-2\rho}}.
\end{aligned}
$$

with some $c > 0$. Therefore, because of $\int_S \boldsymbol{E}|\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx) < \epsilon$ for n large enough,

$$
\boldsymbol{P}\left\{\int_S |\sigma^2(x) - \sigma_n^{2*}(x)|\mu(dx) > 2\epsilon\right\} \le e^{-\epsilon^2 c n^{1-2\rho}}
$$

for n sufficiently large. ■

Proof of Theorem 3.1 Because Y is bounded, for an arbitrary $\epsilon > 0$ one can choose a sphere S centered at 0, such that

$$
\int_{S^c} |\sigma_n^2(x) - \sigma^2(x)|\mu(dx) \le \epsilon.
$$

Therefore it sufficies to show $\int_S |\sigma_n^{2*}(x) - \sigma^2(x)|\mu(dx) \to 0$ *a.s.* for each sphere S centered at 0. One obtains

$$
\begin{aligned}
&\int_S |\sigma_n^2(x) - \sigma^2(x)|\mu(dx) \\
&\le \int_S |\sigma_n^2(x) - \sigma_n^{2*}(x)|\mu(dx) + \int_S |\sigma_n^{2*}(x) - \sigma^2(x)|\mu(dx) \\
&\le G_n + D_n.
\end{aligned}
$$

But $D_n \to 0$ due to Lemma 3.3 and the Borel-Cantelli lemma.
Now, concerning G_n, similarly to the argument in [7], p. 465,

$$
\begin{aligned}
&\int |\sigma_n^{2*}(x) - \sigma_n^2(x)|\mu(dx) \\
&\le \int \left| \frac{\sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_n(x)}(X_i)}{n\mu(A_n(x))} \right. \\
&\quad \left. - \frac{\sum_{i=1}^n (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}) 1_{A_n(x)}(X_i)}{\sum_{i=1}^n 1_{A_n(x)}(X_i)} \right| \mu(dx) \\
&\le 4L^2 \int \sum_{i=1}^n 1_{A_n(x)}(X_i) \left| \frac{1}{n\mu(A_n(x))} - \frac{1}{\sum_{i=1}^n 1_{A_n(x)}(X_i)} \right| \mu(dx) \\
&\le 4L^2 \left| \sum_{i=1}^n \frac{1_{A_n(x)}(X_i)}{n\mu(A_n(x))} - 1 \right| \mu(dx) \to 0 \quad a.s. \\
&\quad \text{(due to (20) and (21)).}
\end{aligned}
$$

■

4 Rate of Convergence

In this section we establish a rate of convergence for the estimate of the local variance defined in (19). The rate corresponds to the rate obtained in classical regression estimation ([7], Theorems 4.3 and 3.2).

Theorem 4.1 *Let $\mathcal{P}_n$ be a cubic partition of $\mathbb{R}^d$ with side length h_n of the cubes ($n \in \mathbb{N}$). Assume that X and Y are bounded. Moreover, assume the Lipschitz conditions*

$$|\sigma^2(x) - \sigma^2(t)| \leq C\|x - t\|, \quad x, t \in \mathbb{R}^d, \tag{23}$$

and

$$|m(x) - m(t)| \leq D\|x - t\|, \quad x, t \in \mathbb{R}^d \tag{24}$$

($C,\ D \in \mathbb{R}_+$, $\|\ \|$ denoting the Euclidean norm).
Then, with

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

for the estimate (19) one gets

$$\boldsymbol{E}\int |\sigma_n^2(x) - \sigma(x)|\mu(dx) = O\left(n^{-\frac{1}{d+2}}\right).$$

For the proof of Theorem 4.1 the following lemma will be used.

Lemma 4.2 *Assume that X is bounded. Then for some finite constant c,*

$$\boldsymbol{E}\{\|X_{N[1,1]} - X_1\|^2\} \leq cn^{-2/\max\{d,2\}},$$

$$\boldsymbol{E}\{\|X_{N[1,2]} - X_1\|^2\} \leq cn^{-2/\max\{d,2\}} \quad (n \in \mathbb{N}).$$

This lemma in its first part is stated for $d \geq 3$ in Györfi et al. [7], Lemma 6.4, and implies the second part according to [7], p. 95. For $d = 2$ (and then obviously also for $d = 1$) it immediately follows from Liitiäinen et al. [12], 3.2 (with reference to [11]) and [13], Theorem 3.2.
For our purpose the weaker bound $cn^{-1/(d+2)}$ would suffice.
Proof of Theorem 4.1 Choose $L \in [0, \infty)$ such that $|Y_i| \leq L$ and denote by l_n the number of cubes of the partition $\mathcal{P}_n$ that cover the bounded support of μ. It holds $l_n = O(h_n^{-d})$. $c_1,\ c_2, \ldots$ will be suitable constants. Set

$$W_{n,i} := (Y_i - Y_{N[i,1]})(Y_i - Y_{N[i,2]}).$$

First, according to [7], p. 465, we note

$$\begin{aligned} &\left|\frac{\sum_{i=1}^n W_{n,i} 1_{A_n(x)}(X_i)}{\sum_{i=1}^n 1_{A_n(x)}(X_i)} - \frac{\sum_{i=1}^n W_{n,i} 1_{A_n(x)}(X_i)}{n\mu(A_n(x))}\right| \\ \leq\ & 4L^2 \left|\frac{\sum_{i=1}^n 1_{A_n(x)}(X_i)}{n\mu(A_n(x))} - 1\right|, \end{aligned} \tag{25}$$

further

$$\begin{aligned} &\boldsymbol{E}\int \left|\frac{\sum_{i=1}^n 1_{A_n(x)}(X_i) - n\mu(A_n(x))}{n\mu(A_n(x))}\right| \mu(dx) \\ \leq\ & \int \frac{\sqrt{\boldsymbol{Var}(\sum_{i=1}^n 1_{A_n(x)}(X_i))}}{n\mu(A_n(x))} \mu(dx) \\ \leq\ & \frac{1}{\sqrt{n}} \int \frac{1}{\sqrt{\mu(A_n(x))}} \mu(dx) \\ \leq\ & \frac{1}{\sqrt{n}} \sqrt{\int \frac{1}{\mu(A_n(x))} \mu(dx)} \\ \leq\ & \sqrt{l_n/n} \\ \leq\ & c_1 n^{-\frac{1}{2}} h_n^{-\frac{d}{2}}. \end{aligned} \tag{26}$$

In the second step we show

$$\int \left| \frac{\sum_{i=1}^n \boldsymbol{E}\{W_{n,i} 1_{A_n(x)}(X_i)\}}{n\mu(A_n(x))} - \sigma^2(x) \right| \mu(dx) \leq c_2 \left(h_n + n^{-2/\max\{d,2\}} \right), \tag{27}$$

i.e.

$$\int \left| \frac{\boldsymbol{E}\{W_{n,1} 1_{A_n(x)}(X_1)\}}{\mu(A_n(x))} - \sigma^2(x) \right| \mu(dx) \leq c_2 \left(h_n + n^{-2/\max\{d,2\}} \right). \tag{28}$$

According to Liitiäinen et al. [12], proof of Theorem 3, or [13], Appendix, via conditioning with respect to $X_1, \ldots, X_n$, we have

$$\begin{aligned}
& \boldsymbol{E}\{W_{n,1} 1_{A_n(x)}(X_1)\} \\
= \; & \boldsymbol{E}\{(Y_1 - m(X_1))^2 1_{A_n(x)}(X_1)\} \\
& + \boldsymbol{E}\left\{ \left(m(X_1) - m(X_{N[1,1]})\right) \left(m(X_1) - m(X_{N[1,2]})\right) 1_{A_n(x)}(X_1) \right\}.
\end{aligned}$$

Then

$$\begin{aligned}
& \int \left| \frac{\boldsymbol{E}\{(Y_1 - m(X_1))^2 1_{A_n(x)}(X_1)\}}{\mu(A_n(x))} \} - \sigma^2(x) \right| \mu(dx) \\
= \; & \int \left| \frac{\int \sigma^2(t) 1_{A_n(x)}(t) \mu(dt)}{\mu(A_n(x))} - \sigma^2(x) \right| \mu(dx) \\
\leq \; & \int \frac{\int [\sigma^2(t) - \sigma^2(x)] 1_{A_n(x)}(t) \mu(dt)}{\mu(A_n(x))} \mu(dx) \\
\leq \; & C \int \frac{\int \|t - x\| 1_{A_n(x)}(t) \mu(dt)}{\mu(A_n(x))} \mu(dx) \\
& \text{(by (23))} \\
\leq \; & C\sqrt{d} h_n \int \frac{\int 1_{A_n(x)}(t) \mu(dt)}{\mu(A_n(x))} \mu(dx) \leq C\sqrt{d} h_n.
\end{aligned}$$

Further

$$\begin{aligned}
& \int \left| \frac{\boldsymbol{E}\left\{ \left(m(X_1) - m(X_{N[1,1]})\right) \left(m(X_1) - m(X_{N[1,2]})\right) 1_{A_n(x)}(X_1) \right\}}{\mu(A_n(x))} \right| \mu(dx) \\
\leq \; & \frac{1}{2} \int \frac{\boldsymbol{E}\left\{ \left| m(X_1) - m(X_{N[1,1]}) \right|^2 1_{A_n(x)}(X_1) \right\}}{\mu(A_n(x))} \mu(dx) \\
& + \frac{1}{2} \int \frac{\boldsymbol{E}\left\{ \left| m(X_1) - m(X_{N[1,2]}) \right|^2 1_{A_n(x)}(X_1) \right\}}{\mu(A_n(x))} \mu(dx) \\
\leq \; & \frac{1}{2} D^2 \left[\boldsymbol{E}\left\{ \|X_{N[1,1]} - X_1\|^2 \right\} + \boldsymbol{E}\left\{ \|X_{N[1,2]} - X_1\|^2 \right\} \right] \\
& \text{(by (24) and } \int \left[1_{A_n(x)}(t) / \mu(A_n(x)) \right] \mu(dx) \leq 1 \text{ for each } t \in \mathbb{R}^d) \\
\leq \; & c_3 n^{-2/\max\{d,2\}}
\end{aligned}$$

by Lemma 4.2. Thus (28) and (27) are obtained.
In the third step we show

$$\int \boldsymbol{E} \left| \frac{\sum_{i=1}^n \left[W_{n,i} 1_{A_n(x)}(X_i) - \boldsymbol{E}\{W_{n,i} 1_{A_n(x)}(X_i)\} \right]}{n\mu(A_n(x))} \right| \mu(dx) \leq c_4 n^{-\frac{1}{2}} h_n^{-\frac{d}{2}}. \tag{29}$$

The left-hand side is bounded by

$$\int \frac{\sqrt{\boldsymbol{Var}\{\sum_{i=1}^n W_{n,i} 1_{A_n(x)}(X_i)\}}}{n\mu(A_n(x))} \mu(dx).$$

As in the proof of Theorem 2.5 we apply the Efron-Stein inequality (Lemma 2.6) and obtain, compare (18),

$$\boldsymbol{Var}\left\{\sum_{i=1}^{n} W_{n,i} 1_{A_n(x)}(X_i)\right\} \leq c_5 n L^4 \boldsymbol{E}\{1_{A_n(x)}(X)\} = c_6 n \mu(A_n(x)).$$

Further

$$\int \frac{\sqrt{\mu(A_n(x))}}{\mu(A_n(x))}\mu(dx) \leq \sqrt{\int \frac{1}{\mu(A_n(x))}\mu(dx)} \leq c_7\sqrt{l_n} \leq c_8 h_n^{-d/2}.$$

Thus (29) is obtained.
In the last step we gather (25), (26), (27), (29) and obtain

$$\begin{aligned} & \boldsymbol{E}\left\{\int |\sigma_n^2(x) - \sigma^2(x)|\mu(dx)\right\} \\ \leq \quad & c_9\left(n^{-\frac{1}{2}} h_n^{-\frac{d}{2}} + h_n + n^{-2/\max\{d,2\}}\right) \\ \leq \quad & c_{10} n^{-\frac{1}{d+2}} \end{aligned}$$

by the choice of (h_n). Thus the assertion is obtained. ■

References

[1] L.D. Brown and M. Levine, *Variance estimation in nonparametric regression via the difference sequence method*, Annals of Statistics **35** (2007), 2219–2232.

[2] T. Cai, M. Levine, and L. Wang, *Variance function estimation in multivariate nonparametric regression*, Journal of Multivariate Analysis **100** (2009), 126–136.

[3] L. Devroye, D. Schäfer, L. Györfi, and H. Walk, *The estimation problem of minimum mean squared error*, Statistics & Decisions **21** (2003), 15–28.

[4] N. Dunford and J.T. Schwartz, *Linear operators, general theory*, Wiley Classics Library, New York, 1958.

[5] D. Evans, *Estimating the variance of multiplicative noise*, 18th International Conference on Noise and Fluctuations, ICNF, in AIP Conference Proceedings **780** (2005), 99–102.

[6] D. Evans and A.J. Jones, *Non-parametric estimation of residual moments and covariance*, Proceedings of the Royal Society A **464** (2008), 2831–2846.

[7] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A distribution-free theory of nonparametric regression*, Springer, New York, 2002.

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

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

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

[11] E. Liitiäinen, F. Corona, and A. Lendasse, *Non-parametric residual variance estimation in supervised learning*, IWANN'07 Proceedings of the 9th International Work-Conference on Artificial Neural Networks. Lecture Notes in Computer Science: Computational and Ambient Intelligence **4507** (2007), 63–71.

[12] ———, *On nonparametric residual variance estimation*, Neural Processing Letters **28** (2008), 155–167.

[13] ———, *Residual variance estimation using a nearest neighbor statistic*, Journal of Multivariate Analysis **101** (2010), 811–823.

[14] M. Loève, *Probability theory*, 4th ed. Springer, Berlin, 1977.

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

[16] ———, *On variance function estimation with quadratic forms*, Journal of Statistical Planning and Inference **35** (1993), 213–231.

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

[18] A. Munk, N. Bissantz, T. Wagner, and G. Freitag, *On difference based variance estimation in nonparametric regression when the covariate is high dimensional*, Journal of the Royal Statistical Society: Series B **67** (2005), 19–41.

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

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

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

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

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

Paola Gloria Ferrario
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: paola.ferrario@mathematik.uni-stuttgart.de

Harro Walk
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: harro.walk@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2007/001

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

2011/022 *Ferrario, P.G.; Walk, H.:* Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

2011/021 *Eberts, M.; Steinwart, I.:* Optimal regression rates for SVMs using Gaussian kernels

2011/020 *Frank, R.L.; Geisinger, L.:* Refined Semiclassical Asymptotics for Fractional Powers of the Laplace Operator

2011/019 *Frank, R.L.; Geisinger, L.:* Two-term spectral asymptotics for the Dirichlet Laplacian on a bounded domain

2011/018 *Hänel, A.; Schulz, C.; Wirth, J.:* Embedded eigenvalues for the elastic strip with cracks

2011/017 *Wirth, J.:* Thermo-elasticity for anisotropic media in higher dimensions

2011/016 *Höllig, K.; Hörner, J.:* Programming Multigrid Methods with B-Splines

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.; Höfert, 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