WILD BOOTSTRAP VERSUS MOMENT-ORIENTED BOOTSTRAP.

Volker Sommerfeld*

September 10, 1997

Abstract

We investigate the relative merits of a "moment-oriented" bootstrap method of Bunke (1997) in comparison with the classical wild bootstrap of Wu (1986) in nonparametric heteroscedastic regression situations. The "moment-oriented" bootstrap is a wild bootstrap based on local estimators of higher order error moments that are smoothed by kernel smoothers. In this paper we perform an asymptotic comparison of these two different bootstrap procedures. We show that the moment-oriented bootstrap is in no case worse than the wild bootstrap. We consider the cases of bandwidths with MISE-optimal rates and of bandwidths with rates that perform an optimal bootstrap approximation. When the regression function has the same amount of smoothness as the second and the third order error moment, then it turns out that, in the former case, our method better approximates the distribution of the pivotal statistic than the usual wild bootstrap does. The reason for this behavior is the unavoidable bias in nonparametric regression estimation that permits only a suboptimal amount of smoothing in the classical wild bootstrap case. In the latter case we need more smoothness of the error moments to make the moment-oriented bootstrap better than wild bootstrap. These results are applied to the construction of pointwise confidence intervals where we prove that our bootstrap has a superior behavior for equal smoothness of the regression function and error moments.

1 Introduction

We consider the nonparametric regression model

$$Y_i = m(x_i) + \epsilon_i, \quad 1 \le i \le n, \tag{1.1}$$

with heteroskedastic errors ϵ_i. Throughout this paper we assume

(A1) a (fixed) equidistant design $x_1 < \cdots < x_n$ on the interval $[0,1]$ and finite error moments $\mu_2(x_i), \mu_3(x_i), \ldots$ of any order.

Note that all what follows holds also true for a non equidistant but regular design in the sense of Sacks and Ylvisaker (1970). They assumed that the design $\{x_i\}$ is generated by a design density f, that is

$$\int_0^{x_i} f(x)\,dx = \frac{i-1}{n-1}, \quad i = 1, \ldots, n,$$

*Institut für Mathematik, Humboldt-Universität zu Berlin, PSF 1297, D-10099 Berlin, Germany. The research was carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin. The paper was printed using funds made available by the Deutsche Forschungsgemeinschaft.

where $f \in Lip_1([0,1])$ is positive on $[0,1]$. Obviously, this implies that $max_{1\leq i\leq n}(x_i - x_{i-1}) = O(n^{-1})$. The assumption of equidistance is only made for a simpler notation. For sequences A_n and B_n we write $A_n \asymp B_n$ if $A_n = B_n(C + o(1))$ for some constant C and $A_n \sim B_n$ if additionally $C = 1$.

We assume that

(A2) m is (k+s)-times continously differentiable, $k, s \geq 2$, and

(A3) μ_2 and μ_3 are r-times continously differentiable, $r \geq 2$.

Furthermore, we denote by K_j, $j = 2, 3, \ldots$, a kernel of order j with compact support. Without loss of generality , we assume $supp(K_j) = [-1, 1]$. Then,

$$\hat{m}_h(x_0) = \sum_{i=1}^{n} w_{k,h}(x_0, x_i) Y_i \tag{1.2}$$

is a Gasser-Müller kernel estimator of the regression function at a fixed point x_0 with weights given by

$$w_{k,h}(x_0, x_i) := \frac{1}{h} \int_{s_{i-1}}^{s_i} K_k\left(\frac{x_0 - u}{h}\right) du$$

where $s_i := (x_i - x_{i-1})/2$.

The pivotal quantities considered here are derived from the quantity

$$S_n = S_{n,h} := \frac{\hat{m}_h(x_0) - m(x_0)}{V_n^{1/2}} \tag{1.3}$$

where $V_n = V_n(x_0) = \sum_{i=1}^n w_{k,h}^2(x_0, x_i)\mu_2(x_i)$ is the variance of $\hat{m}_h(x_0)$. To obtain an observable quantity we have to replace the unknown variance V_n by an appropriate estimate.

That is, we have to estimate the error variance $\mu_2(x_i)$. In this paper, this will be done in two ways. At first, we consider an estimator

$$\hat{\mu}_{2,1}(x_i) := \hat{\epsilon}_i^2 \tag{1.4}$$

which is based on unsmoothed residuals $\hat{\epsilon}_i = Y_i - \hat{m}(x_i)$. On the other hand, we could assume that the error variance is smooth. In this case we estimate it by smoothed local estimators. That is, we have the estimator

$$\hat{\mu}_{2,2}(x_i) := \sum_{j=1}^{n} w_{r,\lambda_2}(x_i, x_j)\tilde{\mu}_2(x_j) \tag{1.5}$$

with the local estimators

$$\tilde{\mu}_2(x_j) = \frac{(Y_j - Y_{j-1})^2}{2} \tag{1.6}$$

That gives the pivotal statistics

$$T_{\nu,n} = T_{\nu,n,h} := \frac{\hat{m}_h(x_0) - m(x_0)}{\hat{V}_{\nu,n}^{1/2}}, \quad \nu = 1, 2, \tag{1.7}$$

where

$$\begin{aligned}\hat{V}_{1,n} &:= \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\hat{\mu}_{2,1}(x_i) \\ &= \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\hat{\epsilon}_i^2 \qquad (1.8)\end{aligned}$$

and

$$\begin{aligned}\hat{V}_{2,n} &:= \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\hat{\mu}_{2,2}(x_i) \\ &= \sum_{i,j=1}^{n} w_{k,h}^2(x_0,x_i)w_{r,\lambda_2}(x_i,x_j)\tilde{\mu}_2(x_j) \qquad (1.9)\end{aligned}$$

are the corresponding estimates of V_n based on an unsmoothed and smoothed estimators of the error variance, respectively.

Then, we approximate the distributions of the pivotal statistics $T_{\nu,n}$, $\nu = 1,2$ by the corresponding distributions of the bootstrapped statistics

$$T_{\nu,n}^* = T_{\nu,n,h_*,g,\lambda_2}^* = \frac{\hat{m}_{h_*}^*(x_0) - \hat{m}_g(x_0)}{\left(\hat{V}_{\nu,n}^*\right)^{1/2}} \qquad (1.10)$$

where

$$\hat{V}_{\nu,n}^* = \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\hat{\mu}_{2,\nu}^*(x_i) \qquad (1.11)$$

are the bootstrap counterparts of $\hat{V}_{\nu,n}$. In what follows, we will choose the bootstrap bandwidth h_* as $h_* = h$. Note that we could also select h_* according to the bootstrap sample $Y_1^*,\ldots,Y_n^*$. This would be a more natural but also more computer intensive procedure. The assertions of this paper are easily generalized to the latter case. The bandwidth g will be specified later.

We remark that the bootstrap estimates $\hat{\mu}_{2,\nu}^*(\bullet)$ $(\nu = 1,2)$ are obtained by the following procedures. The unsmoothed estimator

$$\hat{\mu}_{2,1}(x_i)^* = (\hat{\epsilon}_i^*)^2 \qquad (1.12)$$

is based on the classical "wild bootstrap" of Wu (1986) (see also Härdle & Mammen, 1993, for its application to nonparametric regression) whereas the smoothed estimator

$$\hat{\mu}_{2,2}(x_i)^* = \sum_{j=1}^{n} w_{r,\lambda_2}(x_i,x_j)\tilde{\mu}_2^*(x_j) \qquad (1.13)$$

is based on a moment-oriented variant of wild bootstrap (see Bunke, 1997). More precisely, these bootstrap procedures are defined as follows.

- We denote by F_i the (unknown) distribution of the error ϵ_i, $i = 1,\ldots,n$. We approximate F_i by a bootstrap distribution $\hat{F}_{n,i}$ which has

– the first three central moments 0, $\hat{\epsilon}_i^2$ and $\hat{\epsilon}_i^3$ (wild bootstrap) or

– the first four central moments 0, $\hat{\mu}_2(x_i)$, $\hat{\mu}_3(x_i)$ and $\hat{\mu}_4(x_i)$ (moment-orinted bootstrap). Here, the estimators $\hat{\mu}_{j,2}(x_i)$, $j \leq 4$, are local estimators $\tilde{\mu}_j$ which are smoothed by a Gasser-Müller kernel smoother with kernel of order r and bandwidth λ_j, $j = 2, 3, 4$.

- Bootstrap observations are given by independent random variables (conditionally under the observations 1.1) $Y_i^* = \hat{m}_g(x_i) + \epsilon_i^*$ with $\epsilon_i^* \sim \hat{F}_{n,i}$.
- A bootstrap estimator $\hat{m}^*_{h_*}$ of $\hat{m}_h$ is obtained by a kernel smoothing of the bootstrap observations Y_i^* with bandwidth h_*.

Thereby, we define the local estimators of the error moments as follows.

- ERROR VARIANCE:

$$\tilde{\mu}_2(x_i) := \frac{1}{2}(Y_i - Y_{i-1})^2$$

- THIRD ERROR MOMENT:

$$\tilde{\mu}_3(x_i) := \frac{1}{6}(2Y_i - Y_{i-1} - Y_{i+1})^3$$

- FOURTH ERROR MOMENT:

$$\tilde{\mu}_4(x_i) := \frac{1}{12}(2Y_i - Y_{i-1} - Y_{i+1})^4 - \frac{1}{8}(Y_{i+2} + Y_{i-1} - Y_{i+1} - Y_i)^4.$$

In order to deal correctly with the bias of $\hat{m}_h(x_0)$ and its bootstrap counterpart, the bandwidth g has to be choosen as explained in the following. We denote by P_* the distribution of Y_i^* $(i = 1, \ldots, n)$ conditional under the observations $Y_1, \ldots, Y_n$. Furthermore, we denote the expectation with respect to P_* conditional on the observations $Y_1, \ldots, Y_n$ by E_*. Then we get from Gasser & Müller (1979), appendix 1, that

$$\left| E_* \hat{m}^*_{h_*}(x_0) - \hat{m}_g(x_0) - (-1)^k \frac{h_*^k}{k!} \hat{m}_g^{(k)}(x_0) \int u^k K_k(u)\, du \right| = O_P(n^{-1}) + o_P(h_*^k).$$

Hence, we should make sure that

$$\left| \hat{m}_g^{(k)}(x_0) - m^{(k)}(x_0) \right| = o_P(1)$$

in order to achieve the same asymptotic bias for the statistic T_n and their bootstrapped counterpart, respectively. In order to do that, for constants ν and η with $\eta \geq \nu + 2$ we denote by $K_{(\eta,\nu)}$ the ν-th derivative of ordinary $(\eta - \nu)$-th order kernel $K_{\eta-\nu}$. Then, according to Gasser & Müller (1984), the kernel $K_{(\eta,\nu)}$ satisfies

$$\int_{-\tau}^{\tau} K_{\eta,\nu}(u) u^j \, du = \begin{cases} 0, & j = 0, \ldots, \nu - 1, \nu + 1, \ldots, \eta - 1 \\ (-1)^\nu \nu!, & j = \nu \\ \beta, & j = \eta. \end{cases} \tag{1.14}$$

We estimate $m^{(k)}$ by

$$\hat{m}_g^{(k)}(x_0) := \sum_{i=1}^{n} w_{(k+s,k),g}(x_0, x_i) m(x_i) \tag{1.15}$$

with

$$w_{(k+s,k),g}(x_0, x_i) := \frac{1}{g} \int_{s_{i-1}}^{s_i} K_{(k+s,s)} \left(\frac{x_0 - u}{g} \right) du. \tag{1.16}$$

Then, according to Gasser & Müller (1984) the variance of $\hat{m}_g^{(k)}(x_0)$ is of order $O(n^{-1} g^{-(2k+1)})$ so that g has to tend slower to zero than $n^{-1/(2k+1)}$ to ensure the consistency of $\hat{m}_g^{(k)}(x_0)$. For example, we could use the optimal bandwidth g for the estimator $\hat{m}_g^{(k)}(x_0)$ of $m^{(k)}(x_0)$ which is of order $g \asymp n^{-1/(2(k+s)+1)} \gg n^{-1/(2k+1)}$. That is, we assume

(A4) $h/g \to 0$, $h_*/g \to 0$ and $h, g, \lambda_i \to 0$, $nh, ng, n\lambda_i \to \infty$ for $i = 2, 3, 4$.

In this paper we aim at compare the (asymptotic) behavior of these two bootstrap approximations. We make the following additional assumptions:

(A5) K_k, $K_{(k+s,k)}$ and K_r are k-, (k+s)- and r-times continously differentiable.

(A6) $\mu_s(t) \in Lip_1([0, 1])$ for $s = 1, \ldots, 6$.

In this paper it will be shown that for $h \gg \lambda_2 \asymp \lambda_3$ or $h \asymp \lambda_2 \asymp \lambda_3$ the wild bootstrap and tha moment-oriented bootstrap have the same rate of convergence. Therefore we assume $h \ll \lambda_2 \asymp \lambda_3$ in order to investigate the cases when the moment-oriented bootstrap performs better. For $h \asymp \lambda_2 \asymp \lambda_3$ we will partly analyze the corresponding constants.

Furthermore, as we indicated earlier, we assume for the bootstrap bandwidth h_* that $h_* = h$. Yet, it is easily seen that all calculations can be performed for any bootstrap bandwidth h_* with $(h_* - h)/h = O_P(n^{-\delta})$ for some $\delta > 0$.

This paper is organized as follows. In section 2 we derive Edgeworth expansions of the pivotal statistics $T_{\nu,n}$ ($\nu = 1, 2$) and their bootstrap counterparts. We show that these approximations of the Edgeworth series by the bootstrapped ones depend mainly on the variance differences $|\hat{V}_{\nu,n} - V_{\nu,n}|$. In section 3 we consider the convergence of the two bootstrap estimates $\hat{V}_{\nu,n}$ ($\nu = 1, 2$) of the variance of the regression function and calculate rates of convergence. Putting together the results of the sections 2 and 3, we give in section 4 rates of convergence of the Edgeworth expansions to their bootstrapped versions. In section 5, the results of section 4 are applied to bootstrap confidence intervals. The main results are stated in sections 4 and 5. Section 6 deals with some discussion of the obtained results. In section 7 we give the proofs and in section 8 we prove some technical lemmas.

2 Edgeworth expansions

Recall that we intend to construct confidence intervals for $m(x_0)$. In order to do that, we consider the pivotal statistics $T_{\nu,n}$ which are defined in 1.7. According to lemma 3.1 of Sommerfeld (1997) the bias corrected statistic $T_{\nu,n} - b_n/V_n^{1/2}$ converges in distribution to the standard normal distribution.

We have different possibilities to deal with the unknown bias term $b_n/V_n^{1/2}$ which is of order $O(h^k(nh)^{1/2})$. At first, we could undersmooth. That is, we choose the bandwidth $h \ll n^{-1/(2k+1)}$ smaller than the optimal one in order to ensure that the bias term asymptotically vanishes: $b_n/V_n^{1/2} = o(1)$. Another possibility is to correct $T_{\nu,n}$ by an estimator $\hat{b}_n/\hat{V}_{\nu,n}^{1/2}$ of $b_n/V_n^{1/2}$. This leads to a remaining bias which is of higher order. To be more precise, we denote the remaining bias after correction by $\tilde{b}_n := E\hat{b}_n - b_n$. Then it follows immediately from 8.50 and 8.54 that

$$\tilde{b}_n = O(h^k g^s) \ll b_n = O(h^k) \tag{2.1}$$

A third method that is investigated in this paper consists in performing the bias correction implicitely by the bootstrap as follows. Note that the bootstrap pivots $T^*_{\nu,n} - \hat{b}_n/\hat{V}_n^{1/2}$ $(\nu = 1, 2)$ converge in distribution to the standard normal distribution. Hence, we have the asymptotic equivalence

$$P(T_{\nu,n} \le t) - P_*(T^*_{\nu,n} \le t) \asymp \frac{b_n}{V_n^{1/2}} - \frac{\hat{b}_n}{\hat{V}_n^{1/2}} \asymp \frac{b_n - \hat{b}_n}{V_n^{1/2}}$$

where $b_n - \hat{b}_n \ll V_n^{1/2}$.

In this paper we consider the following two choices of the initial bandwidth h. At first, we can choose $h \asymp n^{-1/(2k+1)}$ with the MISE-optimal rate in order to perform later a data-driven selection of h. On the other hand, Neumann (1997) proved that for optimal rates of the coverage probability of bootstrap confidence intervals we have to choose some $h \ll n^{-1/(2k+1)}$, that is we undersmooth.

The derivation of Edgeworth series for these two choices of h is different because in the first case we have to consider a bias corrected version of $T_{\nu,n}$ whereas in the latter case we can derive the expansion for $T_{\nu,n}$ directly. Therefore we treat these two cases separately.

2.1 MISE-optimal bandwidth h

In this subsection we consider Edgeworth expansions for the bias corrected statistic $T_{\nu,nc} := T_{\nu,n} - b_n/V_{\nu,n}^{1/2}$ where

$$V_{1,n} := \sum_{i,j} w_{k,h}^2(x_0, x_i)\mu_2(x_i) \tag{2.2}$$

and

$$V_{2,n} := \sum_{i,j} w_{k,h}^2(x_0, x_i) w_{r,\lambda_2}(x_i, x_j)\mu_2(x_j). \tag{2.3}$$

We assume (A1) to (A6) and additionally

(A7) All error moments $\mu_j(\cdot)$ are continous on the interval $[0,1]$

and the Cramér type condition (see Neumann, 1997)

(A8) $\max_i \sup_{\|t\|>b} \left| E \exp\left\{ it' \binom{\epsilon_i}{\epsilon_i^2} \right\} \right| \le C_b < 1$ for all $b > 0$.

Note that (A1) and (A7) imply that all moments of the ϵ_i's are uniformly bounded. Then the following Edgeworth expansions hold true:

Lemma 2.1 *Assuming (A1) to (A8) we have for arbitrarily small $\gamma > 0$*

$$\begin{aligned} P(T_{1,nc} < t) &= \Phi(t) + \rho_{n3} \frac{2t^2+1}{6} \phi(t) \\ &\quad + \frac{1}{2} \frac{b_n}{V_n^{1/2}} \rho_{n3} t \phi(t) + O((nh)^{-1+\gamma}) \end{aligned} \tag{2.4}$$

with

$$\rho_{n3} = \rho_{n3}(x_0) = V_{1,n}^{-3/2} \sum_{i=1}^{n} w_{k,h}^3(x_0, x_i) \mu_3(x_i) = O((nh)^{-1/2}). \tag{2.5}$$

Furthermore, it holds true that

$$\begin{aligned} P(T_{2,nc} < t) &= \Phi(t) + \frac{1}{2} \frac{\sum_{i=1}^n w_{k,h}^3(x_0, x_i) \mu_3(x_i)}{V_{2,n}^{3/2}} \phi(t) \\ &\quad + \frac{1}{2} \frac{b_n}{V_{2,n}^{1/2}} \frac{\sum_{i=1}^n d_n(x_0, x_i) \mu_3(x_i)}{V_{2,n}^{3/2}} t \phi(t) \\ &\quad - \left(\frac{\sum_{i=1}^n w_{k,h}^3(x_0, x_i) \mu_3(x_i)}{V_{2,n}^{3/2}} + \frac{3}{2} \frac{\sum_{i=1}^n d_n(x_0, x_i) \mu_3(x_i)}{V_{2,n}^{3/2}} \right. \\ &\quad \left. \left(1 - 3 \frac{V_{1,n}}{V_{2,n}} \right) \right) \frac{t^2 - 1}{6} \phi(t) + O((nh)^{-1+\gamma}) \end{aligned} \tag{2.6}$$

with

$$d_n(x_0, x_i) = w_{k,h}(x_0, x_i) \sum_{j=1}^{n} w_{k,h}^2(x_0, x_j) w_{r,\lambda_2}(x_j, x_i) = O((nh)^{-2} (n\lambda_2)^{-1}) \tag{2.7}$$

for $h \ll \lambda_2$ and

$$d_n(x_0, x_i) = O((nh)^{-3}).$$

for $h \gg \lambda_2$ or $h \asymp \lambda_2$.

For the corresponding bootstrap pivot the following lemma holds.

Lemma 2.2 *We assume (A1) to (A8). Then*

$$
\begin{aligned}
P_*(T^*_{1,nc} < t) &= \Phi(t) + \hat\rho_{n3}\frac{2t^2+1}{6}\phi(t) \\
&\quad + \frac{1}{2}\frac{\hat b_n}{\hat V_n^{1/2}}\hat\rho_{n3} t\phi(t) + O_P((nh)^{-1+\gamma})
\end{aligned}
\tag{2.8}
$$

where

$$
\hat\rho_{n3} = \hat V_n^{-3/2}\sum_{i=1}^n w^3_{k,h_*}(x_0,x_i)\hat\mu_{3,1}(x_i), \tag{2.9}
$$

$$
\hat b_n = \sum_{i=1}^n w_{k,h_*}(x_0,x_i)\hat m_g(x_i) - \hat m_g(x_0) \tag{2.10}
$$

and $\hat\mu_{3,1}(x_i) = \hat\epsilon_i^3$. *Furthermore, it holds true that*

$$
\begin{aligned}
P_*(T^*_{2,nc} < t) &= \Phi(t) + \frac{1}{2}\frac{\sum_{i=1}^n w^3_{k,h}(x_0,x_i)\hat{\mu_{3,2}}(x_i)}{\hat V_{2,n}^{3/2}}\phi(t) \\
&\quad + \frac{1}{2}\frac{\hat b_n}{\hat V_{2,n}^{1/2}}\frac{\sum_{i=1}^n d_n(x_0,x_i)\hat\mu_{3,2}(x_i)}{\hat V_{2,n}^{3/2}} t\phi(t) \\
&\quad - \Bigg(\frac{\sum_{i=1}^n w^3_{k,h}(x_0,x_i)\hat\mu_{3,2}(x_i)}{\hat V_{2,n}^{3/2}} + \frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\hat\mu_{3,2}(x_i)}{\hat V_{2,n}^{3/2}} \\
&\quad \left(1 - 3\frac{\hat V_{1,n}}{\hat V_{2,n}}\right)\Bigg)\frac{t^2-1}{6}\phi(t) + O((nh)^{-1+\gamma})
\end{aligned}
\tag{2.11}
$$

where the estimator $\hat\mu_{3,2}(x_i)$ *is defined in the introduction.*

Now we can substract the equations 2.4 and 2.6 from the corresponding bootstrapped ones in lemma 2.2. This gives the following upper bounds for the bootstrap approximation of the distribution of the pivot statistics $T_{\nu,n}$.

Lemma 2.3 *We assume (A1) to (A8). Then*

$$
\begin{aligned}
|P_*(T^*_{\nu,n} < t) - P(T_{\nu,n} < t)| &\leq \frac{1}{2}|b_n|\frac{|\hat V_{\nu,n} - V_{\nu,n}|}{V_{\nu,n}^{3/2}} \\
&\quad + O_P((h^k g^s (nh)^{1/2} + (h/g)^{k+1/2}) \\
&\quad + O((nh)^{-1}).
\end{aligned}
\tag{2.12}
$$

2.2 Undersmoothing case

When we choose $h \ll n^{-1/(2k+1)}$ then we have the following counterparts of the lemmas 2.1 and 2.2.

Lemma 2.4 *Assuming (A1) to (A8) and* $h \ll n^{-1/(2k+1)}$ *we have for arbitrarily small* $\gamma > 0$

$$\begin{aligned} P(T_{1,n} < t) &= \Phi(t) + \frac{b_n}{V_{1,n}^{1/2}}\phi(t) + \rho_{n3}\frac{2t^2+1}{6}\phi(t) \\ &\quad + \frac{1}{2}\frac{b_n}{V_n^{1/2}}\rho_{n3}t\phi(t) + O((nh)^{-1+\gamma}) \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} P(T_{2,n} < t) &= \Phi(t) + \frac{b_n}{V_{2,n}^{1/2}}\phi(t) + \frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\phi(t) \\ &\quad + \frac{1}{2}\frac{b_n}{V_{2,n}^{1/2}}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}t\phi(t) \\ &\quad - \left(\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} + \frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\right. \\ &\quad \left.\left(1 - 3\frac{V_{1,n}}{V_{2,n}}\right)\right)\frac{t^2-1}{6}\phi(t) + O((nh)^{-1+\gamma}) \end{aligned} \tag{2.14}$$

Lemma 2.5 *We assume (A1) to (A8) and* $h \ll n^{-1/(2k+1)}$*. Then*

$$\begin{aligned} P_*(T_{1,n}^* < t) &= \Phi(t) + \frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}}\phi(t) + \hat{\rho}_{n3}\frac{2t^2+1}{6}\phi(t) \\ &\quad + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_n^{1/2}}\hat{\rho}_{n3}t\phi(t) + O_P((nh)^{-1+\gamma}) \end{aligned} \tag{2.15}$$

and

$$\begin{aligned} P_*(T_{2,n}^* < t) &= \Phi(t) + \frac{\hat{b}_n}{\hat{V}_{2,n}^{1/2}}\phi(t) + \frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu\hat{_{3,2}}(x_i)}{\hat{V}_{2,n}^{3/2}} \\ &\quad + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_{2,n}^{1/2}}\frac{\sum_{i=1}^n d_n(x_0,x_i)\hat{\mu}_{3,2}(x_i)}{\hat{V}_{2,n}^{3/2}}t\phi(t) \\ &\quad - \left(\frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\hat{\mu}_{3,2}(x_i)}{\hat{V}_{2,n}^{3/2}} + \frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\hat{\mu}_{3,2}(x_i)}{\hat{V}_{2,n}^{3/2}}\right. \\ &\quad \left.\left(1 - 3\frac{\hat{V}_{1,n}}{\hat{V}_{2,n}}\right)\right)\frac{t^2-1}{6}\phi(t) + O((nh)^{-1+\gamma}) \end{aligned} \tag{2.16}$$

The proofs of the first parts of these lemmas are given in Neumann (1997), the proofs of the second parts are similarly to that given in lemma 2.1 and lemma 2.2.

Finally, with the calculations of the preceeding subsection it follows that relation 2.12 holds also true in the undersmoothing case.

In the following section we investigate the asymptotic behavior of the differences $|\hat{V}_{\nu,n} - V_{\nu,n}|$, $\nu = 1, 2$.

3 Asymptotic behavior of the bootstrap estimates of the variances $V_{\nu,n}$, $\nu = 1, 2$

In this section we calculate the asymptotic MSE of the bootstrap variance estimators $\hat{V}_{\nu,n}$, $\nu = 1, 2$.

3.1 Wild bootstrap

We assume (A1) and (A3) to (A5). Then

$$
\begin{aligned}
E(\hat{V}_{1,n} - V_{1,n})^2 &= E\left(\sum_{i=1}^{n} w_{k,h}^2(x_0, x_i)[\hat{\epsilon}_i^2 - \mu_2(x_i)]\right)^2 \\
&\sim \sum_{i=1}^{n} w_{k,h}^4(x_0, x_i) E[\epsilon_i^2 - \mu_2(x_i)]^2 \qquad (3.1) \\
&= \sum_{i=1}^{n} w_{k,h}^4(x_0, x_i)[\mu_4(x_i) - \mu_2^2(x_i)] \\
&\sim (nh)^{-3}[\mu_4(x_0) - \mu_2^2(x_0)] \int K_k^4(z)\,dz \qquad (3.2) \\
&= O((nh)^{-3}) \qquad (3.3)
\end{aligned}
$$

where 3.1 follows from lemma 8.1 (i) and 3.2 follows from lemma 8.2.

3.2 Moment-oriented bootstrap

We define $\nu_2(x_i) := Var\tilde{\mu}_2(x_i)$. We assume (A1) to (A6). Then, by lemma 8.3 it follows that

$$
\begin{aligned}
E(\hat{V}_{2,n} - V_{2,n})^2 &= E\left(\sum_{i,j=1}^{n} w_{k,h}^2(x_0, x_i) w_{r,\lambda_2}(x_i, x_j)[\hat{\mu}_2(x_j) - \mu_2(x_j)]\right)^2 \\
&= \sum_{i,j,k,l=1}^{n} w_{k,h}^2(x_0, x_i) w_{k,h}^2(x_0, x_j) w_{r,\lambda_2}(x_i, x_l) w_{r,\lambda_2}(x_j, x_k) \\
&\quad \times E[\hat{\mu}_2(x_l) - \mu_2(x_l)][\hat{\mu}_2(x_k) - \mu_2(x_k)] \\
&\sim \sum_{i,j,k,l=1}^{n} w_{k,h}^2(x_0, x_i) w_{k,h}^2(x_0, x_j) w_{r,\lambda_2}(x_i, x_l) w_{r,\lambda_2}(x_j, x_k) \\
&\quad \times \Bigg[3\frac{\nu_2(x_k)}{n\lambda_2} \int_{-\tau}^{\tau} K_r(z) K_r\left(\frac{x_l - x_k}{\lambda_2} + z\right) dz \\
&\quad + \left((-1)^r \frac{{\lambda_2}^r}{r!} \mu_2^{(r)}(x_k) \int z^r K_r(z)\,dz\right)\left((-1)^r \frac{{\lambda_2}^r}{r!} \mu_2^{(r)}(x_l) \int z^r K_r(z)\,dz\right)\Bigg] \\
&=: A_1. \qquad (3.4)
\end{aligned}
$$

Now, by lemma 8.2 and change in the integration variables we derive

$$
\begin{aligned}
& \sum_{l=1}^{n} w_{r,\lambda_2}(x_i, x_l) \int_{-1}^{1} K_r(z) K_r\left(\frac{x_l - x_k}{\lambda_2} + z\right) dz \\
\sim\ & \frac{1}{\lambda_2} \int_0^1 \int_{-1}^{1} K_r(z) K_r\left(\frac{u - x_k}{\lambda_2} + z\right) dz K_r\left(\frac{x_i - u}{\lambda_2}\right) du \\
\sim\ & \int_{(x_i-1)/\lambda_2}^{x_i/\lambda_1} \int_{-1}^{1} K_r(z) K_r\left(\frac{x_i - t\lambda_2 - x_k}{\lambda_2} + z\right) K_r(t)\, dz\, dt \\
\sim\ & \int_{(x_i-1)/\lambda_2}^{x_i/\lambda_1} \int_{-1}^{1} K_r(z) K_r\left(\frac{x_i - x_k}{\lambda_2} - t + z\right) K_r(t)\, dz\, dt \\
\sim\ & \int_{(x_i-1)/\lambda_2}^{x_i/\lambda_1} \int_{-1}^{1} K_r(z) K_r\left(\frac{x_i - x_k}{\lambda_2} + q\right) K_r(q)\, dz\, dq \\
\sim\ & \int_{-1}^{1} \int_{-1}^{1} K_r(z) K_r\left(\frac{x_i - x_k}{\lambda_2} + q\right) K_r(q)\, dz\, dq \\
\sim\ & \int_{-1}^{1} K_r\left(\frac{x_i - x_k}{\lambda_2} + q\right) K_r(q)\, dq. \qquad (3.5)
\end{aligned}
$$

Thus, treating x_k analogeously, we obtain

$$
\begin{aligned}
A_1 \ \sim\ & \sum_{i,j=1}^{n} w_{k,h}^2(x_0, x_i) w_{k,h}^2(x_0, x_j) \\
& \times \Bigg[3 \frac{\nu_2(x_j)}{n\lambda_2} \int_{-\tau}^{\tau} K_r(z) K_r\left(\frac{x_i - x_j}{\lambda_2} + z\right) dz \\
& + \left((-1)^r \frac{{\lambda_2}^r}{r!} \mu_2^{(r)}(x_j) \int z^r K_r(z)\, dz \right) \left((-1)^r \frac{{\lambda_2}^r}{r!} \mu_2^{(r)}(x_i) \int z^r K_r(z)\, dz \right) \Bigg] \\
=:\ & A_2. \qquad (3.6)
\end{aligned}
$$

On the other hand, from Müller & Stadtmüller (1987a), lemma 2.1, it follows that

$$
\nu_2(x_i) = Var \tilde{\mu}_2(x_i) \asymp \frac{1}{2}\left(\mu_4(x_i) + \mu_2^2(x_i)\right). \qquad (3.7)
$$

Furthermore, because of $supp(K_r) = [-1, 1]$ it holds that

$$
K_r\left(\frac{x_i - x_j}{\lambda_2} + z\right) = 0 \quad \text{if } |x_i - x_j| > \lambda_2. \qquad (3.8)
$$

We denote $\mathcal{I}_{j,\lambda_2} := \{i||x_i - x_j| \leq \lambda_2\}$ and remark that $\#\mathcal{I}_{j,\lambda_2} = O(n\lambda_2)$. Then, for $\lambda_2 \gg h$, it follows from 3.7, 3.8 and lemma 8.2 (analogeously to the calculations above) that

$$
A_2 \ \sim\ \sum_{j=1}^{n} \sum_{i \in \mathcal{I}_{j,\lambda_2}} w_{k,h}^2(x_0, x_i) w_{k,h}^2(x_0, x_j)
$$

$$\times\left[3\frac{\nu_2(x_j)}{n\lambda_2}\int_{-\tau}^{\tau} K_r(z)K_r\left(\frac{x_i-x_j}{\lambda_2}+z\right)\,dz\right.$$
$$\left.+\left((-1)^r\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_j)\int z^rK_r(z)\,dz\right)\left((-1)^r\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_i)\int z^rK_r(z)\,dz\right)\right]$$
$$\sim\ (nh)^{-2}\left((n\lambda_2)^{-1}\left(\int K_k^2(z)\,dz\right)^2\right.$$
$$\left.\times\left(\frac{3}{2}[\mu_4(x_0)+\mu_2^2(x_0)]\int K_r^2(z)\,dz+\left(\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_0)\int z^rK_r(z)\,dz\right)^2\right)\right)\tag{3.9}$$
$$=\ O((nh)^{-2}((n\lambda_2)^{-1}+\lambda_2^{2r}))\tag{3.10}$$

Analogeously, we deduce for $\lambda_2 \ll h$ or $\lambda_2 \asymp h$ that

$$A_1\ \sim\ (nh)^{-2}\left((nh)^{-1}\left(\int K_k^2(z)\,dz\right)^2\right.$$
$$\left.\times\left(\frac{3}{2}[\mu_4(x_0)+\mu_2^2(x_0)]\int K_r^2(z)\,dz+\left(\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_0)\int z^rK_r(z)\,dz\right)^2\right)\right)\tag{3.11}$$
$$=\ O((nh)^{-2}((nh)^{-1}+\lambda_2^{2r})).\tag{3.12}$$

From 3.12 we see that the moment-oriented bootstrap can always achieve the same rate of convergence as the wild bootstrap, namely by choosing $\lambda_2 \ll h$ small enough. Yet, by 3.10 it follows that, if the error variance is smooth enough, the moment-oriented bootstrap get a better rate for $\lambda_2 \asymp n^{-1/(2r+1)} \gg h$. In what follows we will restrict our attetion to this case, that is we assume $\lambda_2 \gg h$. Furthermore, recall that all calculations for the moment-oriented bootstrap are carried out under the assumption $n\lambda_2 \to \infty$. Therefore they can't generalized to the (unsmoothed) case $\lambda_2 = 0$ which is separately treated in the wild bootstrap subsection.

3.3 Local variance estimators based on second order differences

An alternative local estimator of the error variance (which has to be smoothed in a second step) could be defined by second order differences:

$$\tilde{\mu}_2(x_i) := \frac{(2Y_i - Y_{i+1} + Y_{i-1})^2}{6}.\tag{3.13}$$

For the asymptotic variance of this estimator we get by Müller & Stadtmüller (1987a), lemma 2.1 the same value as for the above defined first-order one, namely

$$\nu_2(x_i) = Var\tilde{\mu}_2(x_i) = \frac{1}{2}\left(\mu_4(x_i)+\mu_2^2(x_i)\right).$$

Therefore, we follow by lemma 8.3 and lemma 8.2 analogeously to 3.10 that for $\lambda_2 \gg h$

$$E(\hat{V}_{2,n}-V_{2,n})^2\ \sim\ \sum_{i,j,k,l=1}^{n} w_{k,h}^2(x_0,x_i)w_{k,h}^2(x_0,x_j)w_{r,\lambda_2}(x_i,x_l)w_{r,\lambda_2}(x_j,x_k)$$

$$
\begin{aligned}
&\times\left[5\frac{\nu_2(x_k)}{n\lambda_2}\int K_r(z)K_r\left(\frac{x_l-x_k}{\lambda_2}+z\right)\,dz\right.\\
&\left.+\left((-1)^r\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_k)\int z^2K_r(z)\,dz\right)^2\right]\\
\sim\;& (nh)^{-2}\Bigg((n\lambda_2)^{-1}\left(\int K_k^2(z)\,dz\right)^2\\
&\times\left(\frac{5}{2}[\mu_4(x_0)+\mu_2^2(x_0)]\int K_r^2(z)\,dz+\left(\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_0)\int z^2K_r(z)\,dz\right)^2\right)\Bigg)\\
=\;& O((nh)^{-2}((n\lambda_2)^{-1}+\lambda_2^{2r})) \qquad (3.14)
\end{aligned}
$$

3.4 Smoothed wild bootstrap

A third possibility is to smooth the classical wild bootstrap. First, note that

$$
\overline{\nu}_2(x_i) := Var\hat{\epsilon}_i^2 = \mu_4(x_i)+\mu_2^2(x_i). \qquad (3.15)
$$

Therefore, we deduce by 3.15, lemma 8.1, lemma 8.3 and lemma 8.2 that, analogeously to 3.10, it holds for $\lambda_2 \gg h$ that

$$
\begin{aligned}
E(\hat{V}_{2,n}-V_{2,n})^2 \;=\;& E\left(\sum_{i,j=1}^n w_{k,h}^2(x_0,x_i)w_{r,\lambda_2}(x_i,x_j)[\hat{\epsilon}_j^2-\mu_2(x_j)]\right)^2\\
=\;& \sum_{i,j,k,l=1}^n w_{k,h}^2(x_0,x_i)w_{k,h}^2(x_0,x_j)w_{r,\lambda_2}(x_i,x_l)w_{r,\lambda_2}(x_j,x_k)\\
&\times E[\hat{\epsilon}_l^2-\mu_2(x_l)][\hat{\epsilon}_k^2-\mu_2(x_k)]\\
\sim\;& \sum_{i,j,k,l=1}^n w_{k,h}^2(x_0,x_i)w_{k,h}^2(x_0,x_j)w_{r,\lambda_2}(x_i,x_l)w_{r,\lambda_2}(x_j,x_k)\\
&\times E[\epsilon_l^2-\mu_2(x_l)][\epsilon_k^2-\mu_2(x_k)]\\
=\;& \sum_{i,j,k=1}^n w_{k,h}^2(x_0,x_i)w_{k,h}^2(x_0,x_j)w_{r,\lambda_2}(x_i,x_k)w_{r,\lambda_2}(x_j,x_k)\\
&\times E[\epsilon_k^2-\mu_2(x_k)]^2\\
\sim\;& (nh)^{-2}\Bigg((n\lambda_2)^{-1}\left(\int K_k^2(z)\,dz\right)^2\\
&\times\left([\mu_4(x_0)+\mu_2^2(x_0)]\int K_r^2(z)\,dz+\left(\frac{{\lambda_2}^r}{r!}\mu_2^{(r)}(x_0)\int z^2K_r(z)\,dz\right)^2\right)\Bigg)\\
=\;& O((nh)^{-2}((n\lambda_2)^{-1}+\lambda_2^{2r})). \qquad (3.16)
\end{aligned}
$$

3.5 Comparison of constants

By 3.2, 3.11, 3.14 and 3.16 the constants are interesting in the case $k = r$ and MISE-optimal bandwidths. For example, for $k = r = 2$ and the Epanechnikov kernel we derive

$$\int_{-1}^{1} K_k^2 = \int_{-1}^{1} K_r^2 = 3/5, \; \int_{-1}^{1} K_k^4 = 9/35,$$

therefore

$$\left(\int_{-1}^{1} K_k^2\right)^2 \int_{-1}^{1} K_r^2 = 27/125 = 0.216 \leq \int_{-1}^{1} K_k^4 = 9/35 = 0.257.$$

For $k = r = 4$ and the Quartic kernel we derive

$$\int_{-1}^{1} K_k^2 = \int_{-1}^{1} K_r^2 = 5/7, \; \int_{-1}^{1} K_k^4 = 1125/2431 = 0.463,$$

therefore

$$\left(\int_{-1}^{1} K_k^2\right)^2 \int_{-1}^{1} K_r^2 = 125/329 = 0.380 \leq \int_{-1}^{1} K_k^4 = 0.463.$$

Hence, the constant for the smoothed wild bootstrap is better (in the variance part) whereas the constants of the moment-oriented procedures are worse than the constants in the wild bootstrap case. Yet, in all cases except the wild bootstrap case we have an additional positive bias that depend on the r-th derivative of the error variance.

4 Rates of convergence for the two bootstrap distributions

According to 3.3, 3.10 and lemma 2.3, we have the following asymptotic rates of the two different bootstrap approximations.

Theorem 4.1 *We assume (A1) to (A8). Then, for* $\nu = 1, 2$,

$$|P_*(T^*_{\nu,n} < t) - P(T_{\nu,n} < t)| \leq A_{\nu,n1} + O_P((h^k g^s (nh)^{1/2} + (h/g)^{k+1/2}) + O((nh)^{-1}). \tag{4.1}$$

where, for $h \ll \lambda_2 \asymp \lambda_3$,

$$A_{1,n1} = O_P(h^k) \quad \textit{(wild bootstrap)} \tag{4.2}$$

and

$$\begin{aligned} A_{2,n1} &= O_P(h^k (nh)^{1/2}((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2} \\ &\quad + (nh)^{-1/2}(((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2} + ((n\lambda_3)^{-1} + \lambda_3^{2r})^{1/2})) \end{aligned} \tag{4.3}$$

(moment-oriented bootstrap). Furthermore, for $h \gg \lambda_2 \asymp \lambda_3$ *or* $h \asymp \lambda_2 \asymp \lambda_3$ *we have* $A_{1,n1} = A_{2,n1} = O_P(h^k)$.

We will investigate this theorem for two different bandwidth choices. First, we can select the bandwidths with MISE-optimal rates. That is, we choose $h \asymp n^{-1/(2k+1)}$, $g \asymp n^{-1/(2(k+s)+1)}$ and $\lambda_2 \asymp \lambda_3 \asymp n^{-1/(2r+1)}$. Then $A_{1,n1} \asymp n^{-k/(2k+1)}$ and $A_{2,n1} \asymp n^{-r/(2r+1)}$ (the second and third term of 4.3 are of higher order), hence

$$\begin{aligned} |P_*(T^*_{1,n} < t) - P(T_{1,n} < t)| &= O_P\Bigg(n^{-\frac{k}{2k+1}} + n^{-\frac{s}{2(k+s)+1}} \\ &\quad + \left(n^{-\frac{1}{2k+1}} / n^{-\frac{1}{2(k+s)+1}} \right)^{k+1/2} + n^{-\frac{2k}{2k+1}} \Bigg) \\ &= O_P\left(n^{-\frac{k}{2k+1}} + n^{-\frac{s}{2(k+s)+1}} + n^{-\frac{2k}{2k+1}} \right) \\ &= O_P\left(n^{-\frac{k}{2k+1}} + n^{-\frac{s}{2(k+s)+1}} \right) \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} |P_*(T^*_{2,n} < t) - P(T_{2,n} < t)| &= O_P\left(n^{-\frac{r}{2r+1}} + n^{-\frac{s}{2(k+s)+1}} + n^{-\frac{2k}{2k+1}} \right) \\ &= O_P\left(n^{-\frac{r}{2r+1}} + n^{-\frac{s}{2(k+s)+1}} \right). \end{aligned} \tag{4.5}$$

because of $n^{-s/(2(k+s)+1)} \gg n^{-2k/(2k+1)}$ for $s \geq 0$. Let us first consider 4.4. For fixed smoothness $l := k + s$ of the regression function $m(\bullet)$, the first term of 4.4 is monotonically decreasing in k whereas the second one is monotonically decreasing in s. Hence, we have to weight these two terms in order to get the best possible rate of convergence for $|P_*(T^*_{1,n} < t) - P(T_{1,n} < t)|$. Now,

$$\begin{aligned} n^{-\frac{k}{2k+1}} = n^{-\frac{s}{2(k+s)+1}} \quad &\Leftrightarrow \quad \frac{k}{2k+1} = \frac{s}{2(k+s)+1} \\ &\Leftrightarrow \quad s = k(2k+1). \end{aligned} \tag{4.6}$$

Hence,

$$l = k + s = k + k(2k+1) \Leftrightarrow k^2 + k - l = 0 \tag{4.7}$$

for which $k = \sqrt{1 + l/2} - 1$ is the positive solution. From 4.4, 4.6 and 4.7 it follows that

$$\min_{k,s} |P_*(T^*_{1,n} < t) - P(T_{1,n} < t)| = O_P\left(n^{-\frac{k}{2k+1}} \right) \tag{4.8}$$

where $k = \sqrt{1 + l/2} - 1$ and $s = k(2k+1)$.

Let us now consider 4.5. Weighting the two terms of the right side of 4.5 we get

$$\begin{aligned} n^{-\frac{r}{r+1}} = n^{-\frac{s}{2(k+s)+1}} \quad &\Leftrightarrow \quad \frac{r}{r+1} = \frac{s}{2(k+s)+1} \\ &\Leftrightarrow \quad s = r(2k+1). \end{aligned} \tag{4.9}$$

For this value of s it follows from 4.5 that

$$\min_{k,s} |P_*(T^*_{2,n} < t) - P(T_{2,n} < t)| = O_P\left(n^{-\frac{r}{2r+1}} \right). \tag{4.10}$$

Summing up 4.5 to 4.10 we conclude that the moment-oriented bootstrap gives a better rate of convergence if $n^{-r/(2r+1)} \ll n^{-k/(2k+1)}$, that is if

$$k = \sqrt{1 + l/2} - 1 < r \Leftrightarrow l < 2r(r+2).$$

This holds especially true if the regression function and the error variance have the same amount of smoothness (that is l=r).

Second, we can choose the bandwidths $h, g, \lambda_2, \lambda_3$ such that the rates of convergence in theorem 4.1 are as fast as possible. To begin this, we weight the two terms containing the bandwidth g because they are monotonically inreasing and decreasing in g, respectively. Doing this, we have

$$(nh)^{1/2} h^k g^s \asymp (h/g)^{k+1/2} \Leftrightarrow g \asymp n^{-\frac{1}{2(k+s)+1}}. \tag{4.11}$$

For this value of g we get from theorem 4.1

$$\begin{aligned} |P_*(T^*_{1,n} < t) - P(T_{1,n} < t)| &= O_P\left(h^k + h^k n^{-\frac{s}{2(k+s)+1}} (nh)^{1/2} + (nh)^{-1}\right) \\ &= O_P\left(h^k \left(1 + \left(h n^{\frac{2k+1}{2(k+s)+1}}\right)^{1/2}\right) + (nh)^{-1}\right) \end{aligned} \tag{4.12}$$

Note that the first and the second term of 4.12 is monotonically increasing in h whereas the third term is decreasing in h. Hence, in order to minimize the maximum of these three terms over the bandwidth h we have to weight the maximal term of the two first ones with the third one. Let us first consider the case when the first term in 4.12 is bigger, that is

$$h n^{\frac{2k+1}{2(k+s)+1}} \ll 1 \Leftrightarrow h \ll n^{-\frac{2k+1}{2(k+s)+1}}. \tag{4.13}$$

Then we obtain for the optimal bandwidth h

$$h^k \asymp (nh)^{-1} \Leftrightarrow h \asymp n^{-\frac{1}{k+1}}. \tag{4.14}$$

On the other hand it holds that

$$\begin{aligned} n^{-\frac{1}{k+1}} \ll n^{-\frac{2k+1}{2(k+s)+1}} \quad &\Leftrightarrow \quad \frac{1}{k+1} > \frac{2k+1}{2(k+s)+1} \\ &\Leftrightarrow \quad 2(k+s) + 1 > (2k+1)(k+1) \\ &\Leftrightarrow \quad s > k(k+1/2). \end{aligned} \tag{4.15}$$

Thus we have proved that for $s > k(k+1/2)$ the optimal choices of the bandwidth are $g \asymp n^{-1/(2(k+s)+1)}$ and $h \asymp n^{-1/(k+1)}$. Additionally, it follows from theorem 4.1 that in this case the moment-oriented bootstrap gives a better rate of convergence because the dominant first term of the moment-oriented bootstrap is of order $O_P(h^k (nh)^{1/2} ((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2})$, see 4.3 (instead of $O_P(h^k)$ for the wild bootstrap, see 4.2) and can be improved by the choice of $\lambda_2 = n^{-1/(2r+1)}$ if the smoothness r of the error variance and the third order error moment is high enough. That is, we choose $\lambda_2 \asymp \lambda_3 \asymp n^{-1/2r+1}$ and we weight (recalling that we have assumed $\lambda_2 \asymp \lambda_3 \gg h$, see the comments at the end of subsection 3.2)

$$h^k (nh)^{1/2} ((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2} \asymp (nh)^{-1} \quad \Leftrightarrow \quad h^k (nh)^{1/2} n^{\frac{r}{2r+1}} \asymp (nh)^{-1}$$

$$
\begin{aligned}
&\Leftrightarrow \quad h^{2k+3} \asymp n^{-3} n^{\frac{2r}{2r+1}} \\
&\Leftrightarrow \quad h^{2k+3} \asymp n^{-\frac{4r+3}{2r+1}} \\
&\Leftrightarrow \quad h \asymp n^{-\frac{4r+3}{(2r+1)(2k+3)}}. \qquad (4.16)
\end{aligned}
$$

The corresponding rates of convergence are

$$
|P_*(T_{1,n}^* < t) - P(T_{1,n} < t)| = O_P\left(n^{-\frac{k}{k+1}}\right) \qquad (4.17)
$$

(according to 4.12) and

$$
\begin{aligned}
|P_*(T_{2,n}^* < t) - P(T_{2,n} < t)| &= O_P((nh)^{-1}) \\
&= O_P\left(\left(nn^{-\frac{4r+3}{(2r+1)(2k+3)}}\right)^{-1}\right) \\
&= O_P\left(n^{-\frac{2(2kr+k+r)}{(2r+1)(2k+3)}}\right) \\
&= O_P\left(n^{-\frac{2k}{2k+3}} n^{-\frac{2r}{(2r+1)(2k+3)}}\right) \qquad (4.18)
\end{aligned}
$$

(according to 4.16). From 4.17 and 4.18 it follows that the moment-oriented bootstrap performs better if

$$
\begin{aligned}
n^{-\frac{k}{k+1}} \gg n^{-\frac{2k)}{2k+3}} n^{-\frac{2r}{(2r+1)(2k+3)}} \quad &\Leftrightarrow \quad \frac{k}{k+1} < \frac{2k)}{2k+3} n^{-\frac{2r}{(2r+1)(2k+3)}} \\
&\Leftrightarrow \quad (2r+1)k < 2r(k+1) \\
&\Leftrightarrow \quad k < 2r. \qquad (4.19)
\end{aligned}
$$

Let us now consider the case when the second term in 4.12 is bigger, that is

$$
h \gg n^{-\frac{2k+1}{2(k+s)+1}}. \qquad (4.20)
$$

Because of $\lambda_2 \asymp \lambda_3 \gg h$ we obtain from 4.12 the optimal bandwidth h by

$$
\begin{aligned}
h^k n^{-\frac{s}{2(k+s)+1}} (nh)^{1/2} \asymp (nh)^{-1} \quad &\Leftrightarrow \quad h^{k+1/2+1} \asymp n^{-1} n^{-1/2} n^{\frac{s}{2(k+s)+1}} \\
&\Leftrightarrow \quad h^{2k+3} \asymp n^{-3} n^{\frac{2s}{2(k+s)+1}} \\
&\Leftrightarrow \quad h^{2k+3} \asymp n^{-\frac{6k+4s+3s}{2(k+s)+1}} \\
&\Leftrightarrow \quad h \asymp n^{-\frac{6k+4s+3s}{(2(k+s)+1)(2h+3)}}. \qquad (4.21)
\end{aligned}
$$

On the other hand it holds that

$$
n^{-\frac{6k+4s+3s}{(2(k+s)+1)(2h+3)}} \gg n^{-\frac{2k+1}{2(k+s)+1}} \Leftrightarrow s < k(k+1/2). \qquad (4.22)
$$

Hence, for $s < k(k+1/2)$ the choice of the bootstrap method does not influence the leading term in 4.1 but only a term of higher order. The rates of convergence are for $\nu = 1, 2$ (according to 4.12 and 4.21)

$$
\begin{aligned}
|P_*(T_{\nu,n}^* < t) - P(T_{\nu,n} < t)| &= O_P\left(\left(nn^{-\frac{6k+4s+3s}{(2(k+s)+1)(2k+3)}}\right)^{-1}\right) \\
&= O_P\left(n^{-\frac{2k(2k+2s+1)+6s}{(2(k+s)+1)(2k+3)}}\right) \\
&= O_P\left(n^{-\frac{2k}{2k+3}} n^{-\frac{6s}{(2(k+s)+1)(2k+3)}}\right) \qquad (4.23)
\end{aligned}
$$

Furthermore, recall that for $\lambda_2 \asymp \lambda_3 \ll h$ the two considered bootstrap methods have equal asymptotic performance, too. That is, in the second case the two bootstrap methods have the same asymptotic rate.

From 4.18 and 4.23 it is obvious that weighting the second and the third term of 4.12 gives a better coverage probability than weighting the first and third term if $r < 3s$ (that is, for $r < 3s$ the two bootstrap methods have the same rates of convergence). For $r > 3s$ the moment-oriented bootstrap performs better.

The following corollary sums up the results of the preceeding calculations.

Corollary 4.1 *We assume (A1) to (A8). Then, for MISE-optimal bandwidths h, g, λ_2 the moment-oriented bootstrap has a better rate of convergence if*

$$l < 2r(r+2).$$

This is especially fulfilled for equal amount of smoothness for the regression function, the second and third order error moment. For a bandwidth choice that gives the optimal rate for the bootstrap approximation the moment-oriented bootstrap has a better rate of convergence if

$$r > 3s \quad \textit{and} \quad s > k(k+1/2). \tag{4.24}$$

A sufficient condition for the latter case is

$$r > 3l.$$

If condition 4.24 is not fulfilled, the two bootstrap methods have the same asymptotic rate. This is especially the case for equal amount of smoothness of the regression function, the second and third order error moment.

5 Bootstrap confidence intervals

In this section we will investigate asymptotic rates of the coverage probabilities for confidence intervals obtained by the two bootstrap methods. We will consider one-sided, rather than two-sided confidence intervals although the latter ones are probably of greater practical interest. The reason is that the coverage error of two-sided intervals is not so sensitive to the position of critical points as in the case of one-sided intervals (for a discussion of this problem see Hall, 1991, section 3.7). Thus, one-sided intervals give a more rigorous assessment of the behavior of the considered bootstrap methods. Furthermore, results for two-sided intervals are easily deduced from those for one-sided intervals.

The technique of using Edgeworth expansions in order to obtain bootstrap confidence intervals was largely developed by Hall (see e.g. Hall, 1992b). The idea is to invert the Edgeworth expansions and then to deal with the different terms separately by the delta-method. We define for $\nu = 1, 2$ and $\alpha \in (0, 1)$ the bootstrap crital values $\hat{t}_{\nu,\alpha}$ by

$$P_*(T^*_{\nu,n} \leq \hat{t}_{\nu,\alpha}) = 1 - \alpha.$$

The following theorem holds for $h \ll n^{-1/(2k+1)}$ or $h \asymp n^{-1/(2k+1)}$.

Theorem 5.1 *We assume (A1) to (A8). Then, for $\nu = 1, 2$, it holds that*

$$P(T_{\nu,n} < \hat{t}_{\nu,\alpha}) = 1 - \alpha + \delta_{\nu,n} + O((nh)^{-1+\gamma}) \tag{5.1}$$

where

$$\delta_{1,\alpha} = O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k)$$

and

$$\delta_{2,\alpha} = O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k (nh)^{1/2} \lambda_2^r + h^k h/\lambda_2).$$

We will investigate this theorem in the same way as theorem 4.1. Hereby note that the terms in theorem 5.1 are only different from those of theorem 4.1 by the term $(h/g)^{k+1}$ instead of $(h/g)^{k+1/2}$ in theorem 4.1 and in the terms containing the bandwidth λ_2. Thus, the following calculations will be similar to those leading to corollary 4.1. For MISE-optimal bandwidths we derive

$$\begin{aligned} \delta_{1,\alpha} &= O\Bigg(n^{-\frac{s}{2(k+s)+1}} + \left(n^{-\frac{1}{2k+1}} / n^{-\frac{1}{2(k+s)+1}} \right)^{k+1} \\ &\qquad + n^{-\frac{k}{2k+1}} + n^{-\frac{2k}{2k+1}} \Bigg) \\ &= O\left(n^{-\frac{k}{2k+1}} + n^{-\frac{s}{2(k+s)+1}} \right) \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} \delta_{2,\alpha} &= O\Bigg(n^{-\frac{k}{2k+1}} + n^{-\frac{s}{2(k+s)+1}} \\ &\qquad + n^{-\frac{r}{2r+1}} + n^{-\frac{k+1}{2k+1}} / n^{-\frac{1}{2r+1}} \Bigg) \\ &= O\left(n^{-\frac{s}{2(k+s)+1}} + n^{-\frac{r}{2r+1}} \right) \end{aligned} \tag{5.3}$$

because of

$$n^{-\frac{k+1}{2k+1}} / n^{-\frac{1}{2r+1}} \ll n^{-\frac{r}{2r+1}}$$

for $r > k$, which holds true in the MISE-optimal case with $h \ll \lambda_2 \asymp \lambda_3$. Note that the asymptotic rates in the equations 5.2 and 5.3 are the same as the asymptotic rates in the equations 4.4 and 4.5 concerning the speed of convergence of the two bootstrap distributions. Hence, the conclusions of corollary 4.1 for MISE-optimal bandwidth hold also true in the case of coverage probabilities for confidence intervals.

Now we analyze the choice of the bandwidths $h, g, \lambda_2, \lambda_3$ such that the rates of convergence in theorem 5.1 are as fast as possible. At first, we weight again the two terms containing the bandwidth g because they are monotonically inreasing and decreasing in g, respectively. Thus we obtain

$$(nh)^{1/2} h^k g^s \asymp (h/g)^{k+1} \Leftrightarrow g \asymp (h/n)^{-\frac{1}{2(k+s+1)}}. \tag{5.4}$$

For this value of g we get from theorem 4.1

$$\delta_{1,\alpha} = O\left(h^k \left(1 + n^{\frac{k+1}{2(k+s+1)}} h^{\frac{k+2s+1}{2(k+s+1)}} \right) + (nh)^{-1} \right) \tag{5.5}$$

where the first and the second term of 5.5 is monotonically increasing in h whereas the third term is decreasing in h.

We begin our investigations with the case when the first term in 5.5 is bigger, that is

$$n^{\frac{k+1}{2(k+s+1)}} h^{\frac{k+2s+1}{2(k+s+1)}} \ll 1 \Leftrightarrow h \ll n^{-\frac{k+1}{k+2s+1}}. \tag{5.6}$$

Then we obtain for the optimal bandwidth h

$$h^k \asymp (nh)^{-1} \Leftrightarrow h \asymp n^{-\frac{1}{k+1}}. \tag{5.7}$$

On the other hand it holds that

$$\begin{aligned} n^{-\frac{1}{k+1}} \ll n^{-\frac{k+1}{k+2s+1}} \quad &\Leftrightarrow \quad \frac{1}{k+1} > \frac{k+1}{k+2s+1} \\ &\Leftrightarrow \quad k+2s+1 > (k+1)^2 \\ &\Leftrightarrow \quad s > k(k+1)/2. \end{aligned} \tag{5.8}$$

Hence, for $s > k(k+1)/2$ the moment-oriented bootstrap has a better rate of convergence whereas for $s \leq k(k+1)/2$ the two bootstrap distributions have the same rate of convergence.

To obtain the optimal rate for the moment-oriented bootstrap, we weight, according to theorem 5.1,

$$(nh)^{1/2}\lambda_2^r \asymp h/\lambda_2 \Leftrightarrow \lambda_2 \asymp (h/n)^{\frac{1}{2(r+1)}}. \tag{5.9}$$

The corresponding rates of convergence are (see theorem 5.1)

$$\delta_{1,\alpha} = O\left(n^{-\frac{k}{k+1}}\right) \tag{5.10}$$

and

$$\begin{aligned} \delta_{2,\alpha} &= O(h^k(nh)^{1/2}(\lambda_2^r + h/\lambda_2)) \\ &= O\left(h^k(nh)^{1/2}(h/n)^{\frac{r}{2(r+1)}}\right) \\ &= O\left(h^{k+1/2+\frac{r}{2(r+1)}} n^{-\frac{r}{2(r+1)}}\right) \\ &= O\left(n^{-\frac{2k+1}{2(k+1)}} n^{-\frac{k+2}{2(k+1)}\frac{r}{r+1}}\right). \end{aligned} \tag{5.11}$$

Let us now consider the case when the second term in 5.5 is bigger, that is

$$n^{\frac{k+1}{2(k+s+1)}} h^{\frac{k+2s+1}{2(k+s+1)}} \gg 1 \Leftrightarrow h \gg n^{-\frac{k+1}{k+2s+1}}. \tag{5.12}$$

Then we weight the second and the third term of 5.5, as follows.

$$\begin{aligned} h^k n^{\frac{k+1}{2(k+s+1)}} h^{\frac{k+2s+1}{2(k+s+1)}} \asymp (nh)^{-1} \quad &\Leftrightarrow \quad n^{-2(k+s+1)+k+1} \asymp h^{2(k+s+1)(k+1)+k+2s+1} \\ &\Leftrightarrow \quad h \asymp n^{-\frac{3k+2s+3s}{5k+4s+3+2k^2+2ks}}. \end{aligned} \tag{5.13}$$

The rates of convergence for the coverage probabilities are in that case for $\nu = 1, 2$ (according to 5.5 and 5.13)

$$\begin{aligned} \delta_{\nu,\alpha} &= O((nh)^{-1} \\ &= O\left(\left(nn^{-\frac{3k+2s+3}{5k+4s+3+2k^2+2ks}}\right)^{-1}\right) \\ &= O\left(n^{-\frac{2(k+s+k^2+ks}{5k+4s+3+2k^2+2ks}})\right). \end{aligned} \tag{5.14}$$

Now, by 5.11 and 5.14, after some algebraic calculations it is easily seen that the moment-oriented bootstrap performs better if

$$\begin{aligned} & n^{-\frac{2k+1}{2(k+1)}} n^{-\frac{k+2}{2(k+1)}\frac{r}{r+1}} \ll n^{-\frac{2(k+s+k^2+ks}{5k+4s+3+2k^2+2ks)}} \\ \Leftrightarrow \quad & r > \frac{4k^2+7k+2ks+3}{2k^3+13k^2+20k+2k^2s+10ks+8s+9}. \end{aligned} \tag{5.15}$$

That is, when the smoothness r of the error variance is higher than the right hand side of 5.15 then we can improve the wild bootstrap by the choice of $s > k(k+1)/2$. Note that the right hand side of 5.15 is for $k, s \geq 1$ obviously smaller than 1. Hence, we have derived the following corollary.

Corollary 5.1 *We assume (A1) to (A8). Then, for MISE-optimal bandwidths h, g, λ_2 the moment-oriented bootstrap has a better rate of convergence for the coverage probability of confidence intervals if*

$$l < 2r(r+2).$$

This is especially fulfilled for equal amount of smoothness for the regression function, the second and third order error moment. For a bandwidth choice that gives the optimal rate for the coverage probablities, the moment-oriented bootstrap has a better rate of convergence for any smoothness $r \geq 2$ if we choose

$$s > k(k+1)/2. \tag{5.16}$$

6 Discussion

1. From theorem 4.1 and corollary 4.1 it follows that the (conditional) bootstrap distribution while using the moment-oriented bootstrap method better approximates the true distribution of the pivotal statistic $T_{\nu,n}$ if the error moments $\mu_2(\bullet)$ and $\mu_3(\bullet)$ are sufficiently smooth in comparison to the regression function $m(\bullet)$. If we choose the bandwidths of MISE-optimal order then a sufficient condition for a superior behavior of the moment-oriented bootstrap is equal order of smoothness of the regression function $m(\bullet)$ and the error moments $\mu_2(\bullet)$ and $\mu_3(\bullet)$. The reason of this property is that the estimator $\hat{m}_h(x_0)$ can only use smoothness of order k instead the full smoothness of order $k + s$ of the regression function because the "rest" of the smoothness (of order s) is needed to estimate the bias of $\hat{m}_h(x_0)$.

2. If we estimate the error variance by estimators based on higher order differences, then, by 3.9 and 3.14, the constant becomes worse whereas the rate of convergence remains the same. The reason is that there are more covariance terms in the asymptotic expansion (see lemma 8.3). The best asymptotic constant gives a smoothed classical wild bootstrap (see 3.16). Yet, simulation results in Bunke (1997) indicate that bootstrap methods based on higher order differences have a better behavior for small and moderate sample sizes. In this sense, we do not recommend a smoothed classical wild bootstrap.

3. By 3.9, the important constant in the asymptotic expansion is determined by the r-th order derivative $\mu_2^{(r)}(x_0)$ of the error variance and by the kurtosis $\mu_4(x_0)/\mu_2^2(x_0)$ of the error distribution at x_0.

4. Error moments higher than third order do not influence first and second order asymptotics. Yet, Bunke (1997) indicates that their estimation can be important for the small and moderate sample bevavior.

5. By corollary 5.1, the moment-oriented bootstrap achieves better rates for the coverage probability of studentized confidence in the case of equal smoothness of the regression function and the second and third order error moment.

6. The results of this paper are derived for nonrandom bandwidths. Yet, they can easily be generalized to random, data-driven bandwidths by use of a full-crossvalidation bandwidths choice criterion (see Bunke, Droge & Polzehl, 1995 and Sommerfeld, 1997) and techniques of proving of Neumann (1992, 1995) and Sommerfeld (1997).

7 Proofs

Proof of lemma 2.1: The proof of relation 2.4 is essentially that of proposition 3.1 in Neumann (1997). The only difference to his paper is that he derived Edgeworth expansions for $T_{1,n} - \hat{b}_n/\hat{V}_{1,n}^{1/2}$ instead of $T_{1,nc} = b_n/V_{1,n}^{1/2}$. Note that we can correct by the unknown term $b_n/V_{1,n}^{1/2}$ because it is implicitely estimated by the bootstrap.

Here we will only give a sketch of the proof containing the (very few) differences to that of Neumann. In part 1 of his proof he shows by results of Skovgaard (1986) the validity of an expansion of arbitrary length of the random vector

$$S_{1,n} := B_{1,n}^{-1/2} \sum_{j=1}^{n} \alpha_{1,j} \tag{7.1}$$

where $\alpha_{1,j} := (n\,h\,w_{k,h}(x_0,x_j)\epsilon_j, (nh)^2 w_{k,h}^2(x_0,x_j)(\epsilon_j^2 - \mu_2(x_j)))'$ and $B_{1,n} := Cov(\alpha_{1,j})$. Now, analogeously to part 2 of that proof we can derive from results of Skovgaard (1981) the validity of the expansion of a sufficiently regular sequence of functions $f_n(S_{1,n})$. In order to do that, we approximate $T_{1,nc}$ by

$$\tilde{T}_{1,nc} \;\; := \;\; \frac{\sum_{j=1}^{n} w_{k,h}(x_0,x_j)Y_j - m(x_0)}{\tilde{V}_{1,n}^{1/2}} - \frac{b_n}{V_{1,n}^{1/2}}$$

$$= \frac{\sum_{j=1}^n w_{k,h}(x_0,x_j)\epsilon_j + b_n}{\sqrt{\sum_{j=1}^n w_{k,h}^2(x_0,x_j)(\epsilon_j^2 - \mu_2(x_j)) + V_{1,n}}} - \frac{b_n}{V_{1,n}^{1/2}} \tag{7.2}$$

$$= (\alpha_{n1} + b_n)(\alpha_{1,n2} + V_{1,n})^{-1/2} - \frac{b_n}{V_{1,n}^{1/2}} \tag{7.3}$$

where $\tilde{V}_{1,n} := \sum_{j=1}^n w_{k,h}^2(x_0,x_j)\epsilon_j^2$. By a Taylor expansion of $(\alpha_{n2} + V_{1,n})^{-1/2}$ at $V_{1,n}$ we derive

$$\tilde{T}_{1,nc} = T'_{1,nc} + \frac{5}{32}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}^3}{{V'_{1,n}}^{7/2}} \tag{7.4}$$

where

$$\begin{aligned} T'_{1,nc} &= \frac{\alpha_{n1} + b_n}{V_{1,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}} + \frac{3}{8}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}^2}{V_{1,n}^{5/2}} - \frac{b_n}{V_{1,n}^{1/2}} \\ &= \frac{\alpha_{n1}}{V_{1,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}} + \frac{3}{8}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}^2}{V_{1,n}^{5/2}} \end{aligned} \tag{7.5}$$

and $V'_{1,n}$ is between $V_{1,n}$ and $\alpha_{1,n2} + V_{1,n}$.

Recall that Neumann defined for a sequence of random variables $\{Y_n\}$ and for sequences of constants $\{\gamma_{n1}\}$ and $\{\gamma_{n2}\}$ the notation

$$Y_n := \tilde{O}(\gamma_{n1}, \gamma_{n2})$$

if

$$P(|Y_n| > C\gamma_{n1}) \leq C\gamma_{n2}$$

holds for $n \geq 1$ and some $C < \infty$.

Analogeously to his proof we deduce that

$$\tilde{T}_{1,nc} - T'_{1,nc} = \tilde{O}((nh)^{-3/2+4\gamma}, n^{-1}) \tag{7.6}$$

for arbitrary $\gamma > 0$. Here, it does not matter that the bias b_n is of order $O((nh)^{-1/2})$ instead of $o((nh)^{-1/2})$ as in Neumann's derivation. Hence we derive, according to lemma 3.1 in Neumann (1992b), that the Edgeworth expansions of $\tilde{T}_{1nc}$ and T'_{nc} coincide up to a term of order $O((nh)^{-3/2+4\gamma} + n^{-1})$ and, hence, it suffices to state this expansion for $T'_{1,nc}$.

The rest of the proof goes as that of proposition 3.1 in Neumann (1997). The only differences are that the first order cumulant of $T'_{1,nc}$ is

$$\tilde{\kappa}_{1,n} = -\frac{1}{2}\rho_{n3} + O((nh)^{-1})$$

and that (with Neumann's notations) $\overline{V}_{1,n} = V_{1,n}$ because of $\overline{\overline{W}}_{nj} = \overline{W}_{nj}$. The last relation holds because in this paper there is no subsequent bias correction by an estimator $\hat{b}_n/\hat{V}_{1,n}^{1/2}$ but by $b_n/V_{1,n}^{1/2}$.

Hence, we have the following expansion for $\tilde{T}_{1,nc}$:

$$\begin{aligned} P(\tilde{T}_{1,nc} < t) &= \Phi(t) + \rho_{n3}\frac{2t^2+1}{6}\phi(t) \\ &\quad + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\phi(t) + O((nh)^{-1}) \end{aligned} \tag{7.7}$$

To complete the proof of this lemma, by lemma 3.1 of Neumann (1997) it suffices to show that for arbitrarily small $\gamma > 0$ it holds that

$$T_{1,nc} - \tilde{T}_{1,nc} = \tilde{O}((nh)^{-1+\gamma}, n^{-1}).$$

In order to do that, we write

$$\begin{aligned} T_{1,nc} - \tilde{T}_{1,nc} &= (\alpha_{n1} + b_n)\left(\frac{1}{\sqrt{\hat{V}_{1,n}}} - \frac{1}{\sqrt{\tilde{V}_{1,n}}}\right) \\ &= (\alpha_{n1} + b_n)\frac{\sqrt{\tilde{V}_{1,n}} - \sqrt{\hat{V}_{1,n}}}{\sqrt{\hat{V}_{1,n}\tilde{V}_{1,n}}} \\ &= (\alpha_{n1} + b_n)\frac{\tilde{V}_{1,n} - \hat{V}_{1,n}}{\sqrt{\hat{V}_{1,n}\tilde{V}_{1,n}}(\sqrt{\tilde{V}_{1,n}} + \sqrt{\hat{V}_{1,n}})}. \end{aligned} \tag{7.8}$$

Hence, from

$$\hat{\epsilon}_i^2 - \epsilon_i^2 = (\hat{m}_h(x_i) - m(x_i))^2 - 2\epsilon_i(\hat{m}_h(x_i) - m(x_i))$$

(see 8.2) we derive that

$$\begin{aligned} \tilde{V}_{1,n} - \hat{V}_{1,n} &= \sum_{i=1}^n w_{k,h}^2(x_0, x_i)(\hat{\epsilon}_i^2 - \epsilon_i^2) \\ &= \sum_{i=1}^n w_{k,h}^2(x_0, x_i)(\hat{m}_h(x_i) - m(x_i))^2 + 2\sum_{i=1}^n w_{k,h}^2(x_0, x_i)(m(x_i) - \hat{m}_h(x_i))\epsilon_i \\ &=: T_{n1} + 2T_{n2}. \end{aligned} \tag{7.9}$$

By lemma 8.1 in Neumann (1992) it holds for arbitrary small $\delta > 0$ that

$$\begin{aligned} \hat{m}_h(x_i) - m(x_i) &= \sum_{j=1}^n w_{k,h}(x_i, x_j)\epsilon_j + \sum_{j=1}^n w_{k,h}(x_i, x_j)m(x_j) - m(x_i) \\ &= \sum_{j=1}^n w_{k,h}(x_i, x_j)\epsilon_j + O(h^k) \\ &= \tilde{O}\left(\left(\sum_{j=1}^n w_{k,h}^2(x_i, x_j)\right)^{1/2} n^\delta, n^{-1}\right) + O(h^k) \\ &= \tilde{O}((nh)^{-1/2}n^\delta + h^k, n^{-1}) \\ &= \tilde{O}((nh)^{-1/2}n^\delta, n^{-1}) \end{aligned} \tag{7.10}$$

where the last relation holds because of $h^k \asymp (nh)^{-1/2}$. Hence,

$$T_{n1} = \tilde{O}((nh)^{-2}n^{2\delta}, n^{-1}). \tag{7.11}$$

To deal with T_{n2} we decompose

$$\begin{aligned} T_{n2} &= \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\left(m(x_i) - \sum_{j=1}^{n} w_{k,h}(x_i,x_j)m(x_j)\right)\epsilon_i \\ &\quad + \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)\left(\sum_{j=1}^{n} w_{k,h}(x_i,x_j)\epsilon_j\right)\epsilon_i \\ &=: T_{n3} + T_{n4}. \end{aligned} \tag{7.12}$$

According to lemma 8.1 in Neumann (1992) we derive

$$\begin{aligned} T_{n3} &= \tilde{O}\left(\left(\sum_{i=1}^{n} w_{k,h}^4(x_0,x_i)\left(m(x_i) - \sum_{j=1}^{n} w_{k,h}(x_i,x_j)m(x_j)\right)^2\right)^{1/2} n^{\delta}, n^{-1}\right) \\ &= \tilde{O}((nh)^{-3/2}h^k n^{\delta}, n^{-1}) \\ &= \tilde{O}((nh)^{-2}n^{\delta}, n^{-1}). \end{aligned} \tag{7.13}$$

Furthermore, note that

$$\begin{aligned} T_{n4} &= \sum_{i,j=1}^{n} w_{k,h}^2(x_0,x_i)w_{k,h}(x_i,x_j)\epsilon_j\epsilon_i \\ &= \epsilon' M_h \epsilon \\ &= E\epsilon' M_h \epsilon + (\epsilon' M_h \epsilon - E\epsilon' M_h \epsilon) \\ &=: T_{n5} + T_{n6} \end{aligned} \tag{7.14}$$

where M_h is a $n \times n$ - matrix with (i,j)-th element $w_{k,h}^2(x_0,x_i)w_{k,h}(x_i,x_j)$. Obviously,

$$\begin{aligned} T_{n5} &= \sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)w_{k,h}(x_i,x_i)\mu_2(x_i) \\ &= O((nh)^{-2}). \end{aligned} \tag{7.15}$$

On the other hand, again by lemma 8.1 of Neumann (1992) it follows that

$$\begin{aligned} T_{n6} &= \tilde{O}(tr(M_h V M_h' V)^{1/2} n^{\delta}, n^{-1}) \\ &= \tilde{O}(tr(M_h M_h')^{1/2} n^{\delta}, n^{-1}) \\ &= \tilde{O}\left(\left(\sum_{i,j=1}^{n} w_{k,h}^2(x_0,x_i)w_{k,h}(x_i,x_j)w_{k,h}^2(x_0,x_j)w_{k,h}(x_j,x_i)\right)^{1/2} n^{\delta}, n^{-1}\right) \\ &= \tilde{O}((nh)^{-2}n^{\delta}, n^{-1}). \end{aligned} \tag{7.16}$$

because $V = diag(\mu_2(x_1), \ldots, \mu_2(x_n))$ is bounded. Finally, we get by 7.8, 7.11, 7.13, 7.15 and 7.16 that

$$T_{1,nc} - \tilde{T}_{1,nc} = \tilde{O}((nh)^{-1+\gamma}, n^{-1}).$$

Thus, 2.4 is proven.

The validity of the Edgeworth expansion 2.6 follows analogeously to the first part of this proof. To identify this expansion, we calculate the corresponding cumulants in the following. In order to do that, we use again the paper of Skovgaard (1981) to identify an expansion of $f_n(S_{2,n})$ where

$$S_{2,n} := B_{2,n}^{-1/2} \sum_{j=1}^{n} \alpha_{2,j}, \tag{7.17}$$

$\alpha_{2,j} := (n\, h\, w_{k,h}(x_0, x_j)\epsilon_j, (nh)^2 w_{k,h}^2(x_0, x_j) \sum_{i=1}^n w_{r,\lambda_2}(x_j, x_i)(\tilde{\tilde{\mu}}_2(x_i) - \mu_2(x_i)))'$, $\tilde{\tilde{\mu}}_2(x_i) := (\epsilon_i - \epsilon_{i-1})^2/2$ and $B_{2,n} := Cov(\alpha_{2,j})$. As in the proof of 2.4 we approximate $T_{2,nc}$ by

$$\begin{aligned} \tilde{T}_{2,nc} &:= \frac{\sum_{j=1}^n w_{k,h}(x_0, x_j)Y_j - m(x_0)}{\tilde{V}_{2,n}^{1/2}} - \frac{b_n}{V_{1,n}^{1/2}} \\ &= \frac{\sum_{j=1}^n w_{k,h}(x_0, x_j)\epsilon_j + b_n}{\sqrt{\sum_{i,j=1}^n w_{k,h}^2(x_0, x_i) w_{r,\lambda_2}(x_i, x_j)(\epsilon_j^2 - \mu_2(x_j)) + V_{2,n}}} - \frac{b_n}{V_{2,n}^{1/2}} \qquad (7.18) \\ &= (\alpha_{n1} + b_n)(\alpha_{2,n2} + V_{2,n})^{-1/2} - \frac{b_n}{V_{2,n}^{1/2}} \qquad (7.19) \end{aligned}$$

where

$$\tilde{V}_{2,n} := \sum_{i,j=1}^n w_{k,h}^2(x_0, x_i) w_{r,\lambda_2}(x_i, x_j) \tilde{\tilde{\mu}}_2(x_j).$$

By a Taylor expansion of $(\alpha_{2,n2} + V_{2,n})^{-1/2}$ at $V_{2,n}$ we derive

$$\tilde{T}_{2,nc} = T'_{2,nc} + \frac{5}{32} \frac{(\alpha_{n1} + b_n)\alpha_{2,n2}^3}{{V'_{2,n}}^{7/2}} \tag{7.20}$$

where

$$\begin{aligned} T'_{2,nc} &= \frac{\alpha_{n1} + b_n}{V_{2,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{2,n2}}{V_{2,n}^{3/2}} + \frac{3}{8}\frac{(\alpha_{n1} + b_n)\alpha_{2,n2}^2}{V_{2,n}^{5/2}} - \frac{b_n}{V_{2,n}^{1/2}} \\ &= \frac{\alpha_{n1}}{V_{2,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{2,n2}}{V_{2,n}^{3/2}} + \frac{3}{8}\frac{(\alpha_{n1} + b_n)\alpha_{2,n2}^2}{V_{2,n}^{5/2}} \qquad (7.21) \end{aligned}$$

and $V'_{2,n}$ is between $V_{2,n}$ and $\alpha_{2,n2} + V_{2,n}$.

Now, analogeously to the proof of proposition 3.1 in Neumann (1997) we deduce that it is sufficient to state the expansion for

$$T''_{2,nc} := \frac{\alpha_{n1}}{V_{2,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{2,n2}}{V_{2,n}^{3/2}}. \tag{7.22}$$

In the following we calculate the cumulants $\kappa_{\nu,n}$ $(\nu = 1, 2, 3, \ldots)$ of $T''_{2,nc}$. We have

$$\kappa_{1,n} = ET''_{2,nc} = -\frac{1}{2}\frac{E\alpha_{n1}\alpha_{2,n2}}{V_{2,n}^{3/2}} + O((nh)^{-1}) \tag{7.23}$$

where, because of (A6) and 8.29,

$$\begin{aligned} E\alpha_{n1}\alpha_{2,n2} &= \sum_{i=1}^{n}\left(w_{k,h}(x_0,x_i)\sum_{j=1}^{n} w_{k,h}^2(x_0,x_j)w_{r,\lambda_2}(x_j,x_i)\right)\mu_3(x_j) \\ &= \sum_{i\in\mathcal{I}}^{n} d_n(x_0,x_i)\mu_3(x_i) \end{aligned} \tag{7.24}$$

where $\#\mathcal{I} = \min\{nh, n\lambda_2\}$ and

$$d_n(x_0,x_i) = \begin{cases} O((nh)^{-3}, & if\ h \gg \lambda_2\ or\ h \asymp \lambda_2 \\ O((n\lambda_2)^{-1}(nh)^{-2}, & if\ h \ll \lambda_2. \end{cases}$$

Hence,

$$E\alpha_{n1}\alpha_{2,n2} = \begin{cases} O((nh)^{-2}, & if\ h \gg \lambda_2\ or\ h \asymp \lambda_2 \\ O((n\lambda_2)^{-1}(nh)^{-1}, & if\ h \ll \lambda_2. \end{cases} \tag{7.25}$$

Furthermore,

$$E(T''_{2,nc})^2 = 1 - \frac{b_n}{V_{2,n}^{1/2}}\frac{E\alpha_{n1}\alpha_{2,n2}}{V_{2,n}^{3/2}} + O((nh)^{-1}) \tag{7.26}$$

$$\begin{aligned} E(T''_{2,n})^3 &= \frac{E\alpha_{n1}^3}{V_{2,n}^{3/2}} - \frac{3}{2}\frac{E\alpha_{n1}^2(\alpha_{n1}+b_n)\alpha_{2,n2}}{V_{2,n}^{5/2}} + O((nh)^{-1}) \\ &= \frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \\ &\quad -\frac{9}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\frac{V_{1,n}}{V_{2,n}} + O((nh)^{-1}). \end{aligned} \tag{7.27}$$

Hence, $\kappa_{2,n} = E(T''_{2,nc}) + O((nh)^{-1})$ and

$$\begin{aligned} \kappa_{3,n} &= E(T''_{2,nc})^3 - 3E(T''_{2,nc})^2 ET''_{2,nc} + (ET''_{2,nc})^3 \\ &= E(T''_{2,nc})^3 - 3E(T''_{2,nc})^2 ET''_{2,nc} + O((nh)^{-1}) \\ &= \frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \\ &\quad +\frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\left(1 - 3\frac{V_{1,n}}{V_{2,n}}\right) + O((nh)^{-1}). \end{aligned} \tag{7.28}$$

Thus, according to the proof of proposition 3.1 in Neumann (1997), we have for the characteristic function $\hat{\eta}_n(\bullet)$ of $\tilde{T}_{2,nc}$ the relation

$$\begin{aligned} \hat{\eta}_n(t) &= \exp\left(-\frac{t^2}{2}\right)\Bigg(1 + it\kappa_{1,n} + \frac{(it)^2}{2!}(\kappa_{2,n}-1) \\ &\quad + \frac{(it)^3}{3!}\kappa_{3,n} + O((nh)^{-1}(t^2+t^8))\Bigg), \end{aligned} \tag{7.29}$$

which implies

$$
\begin{aligned}
P(\tilde{T}_{2,nc} < t) &= \Phi(t) + \frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\phi(t) \\
&+ \frac{1}{2}\frac{b_n}{V_{2,n}^{1/2}}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} t\phi(t) \\
&- \left(\frac{\sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} + \frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \right. \\
&\left. \left(1 - 3\frac{V_{1,n}}{V_{2,n}}\right)\right)\frac{t^2-1}{6}\phi(t) + O((nh)^{-1+\gamma}). \qquad (7.30)
\end{aligned}
$$

Accordind to lemma 3.1 in Neumann (1997), it remains only to show that

$$\tilde{T}_{2,nc} - T_{2,nc} = \tilde{O}((nh)^{-1+\gamma}, n^{-1}). \qquad (7.31)$$

We will show it in what follows. Note that analogeously to 7.8 it holds that

$$T_{2,nc} - \tilde{T}_{2,nc} = (\alpha_{n1} + b_n)\frac{\tilde{V}_{2,n} - \hat{V}_{2,n}}{\sqrt{\hat{V}_{2,n}\tilde{V}_{2,n}}(\sqrt{\tilde{V}_{2,n}} + \sqrt{\hat{V}_{2,n}})}. \qquad (7.32)$$

Furthermore, recall that

$$\tilde{V}_{2,n} - \hat{V}_{2,n} = \sum_{i,j=1}^n w_{k,h}^2(x_0,x_i)w_{r,\lambda_2}(x_i,x_j)(\tilde{\tilde{\mu}}_2(x_j) - \tilde{\mu}_2(x_j)) \qquad (7.33)$$

and, because of the Lipschitz continuity of the regression function $m(\bullet)$,

$$
\begin{aligned}
|\tilde{\tilde{\mu}}_2(x_j) - \tilde{\mu}_2(x_j)| &= |-(m(x_j) - m(x_{j-1})(m(x_j) - m(x_{j-1} + 2(\epsilon_j - \epsilon_{j-1})| \\
&\leq |m(x_j) - m(x_{j-1})||(m(x_j) - m(x_{j-1} + 2(\epsilon_j - \epsilon_{j-1})| \\
&\leq C|m(x_j) - m(x_{j-1})| \\
&\leq C'|x_j - x_{j-1}| \\
&= O(n^{-1}). \qquad (7.34)
\end{aligned}
$$

Finally, summing up 7.32, 7.33 and 7.34 completes the proof of this lemma.

□

Proof of lemma 2.2: In the case of a moment-oriented bootstrap which is based on a continous distribution the validity of this expansion follows by the same arguments as given in the proof of lemma 2.1. The validity of this expansion in a discrete distribution based bootstrap case is proven in Neumann (1997), proposition 4.1. He showed that in this case the ϵ_i^* do not fulfill Cramér's condition. However, we are not in the case of a sum of lattice distributions. Hence, he is able to prove some version of Petrov's condition. The rest of the proof goes completely analogeous to the parts of the proof of lemma 2.1 where we identified the Edgeworth expansions.

$\square$

Proof of lemma 2.3: Recall that, according to lemma 2.1, it holds for $\nu = 1, 2$ that

$$\begin{aligned} P(T_{\nu,n} < t) &= P(T_{\nu,nc} < t - b_n/V_{\nu,n}^{1/2}) \\ &= \Phi(t - b_n/V_{\nu,n}^{1/2}) + R_{\nu,n0} + O((nh)^{-1+\gamma}) \end{aligned} \tag{7.35}$$

with

$$\begin{aligned} R_{1,n0} &= \rho_{n3}\frac{2(t - b_n/V_{\nu,n}^{1/2})^2 + 1}{6}\phi(t - b_n/V_{\nu,n}^{1/2}) \\ &\quad + \frac{1}{2}\frac{b_n}{V_{\nu,n}^{1/2}}\rho_{n3}(t - b_n/V_{\nu,n}^{1/2})\phi(t - b_n/V_{\nu,n}^{1/2}), \end{aligned} \tag{7.36}$$

$$\begin{aligned} R_{2,n0} &= \frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0, x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \\ &\quad + \frac{1}{2}\frac{b_n}{V_{2,n}^{1/2}}\frac{\sum_{i=1}^n d_n(x_0, x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}(t - b_n/V_n^{1/2})\phi(t - b_n/V_n^{1/2}) \\ &\quad - \left(\frac{1}{2}\frac{\sum_{i=1}^n w_{k,h}^3(x_0, x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} + \frac{3}{2}\frac{\sum_{i=1}^n d_n(x_0, x_i)\mu_3(x_i)}{V_{2,n}^{3/2}}\right. \\ &\quad \left.\left(1 - 3\frac{V_{1,n}}{V_{2,n}}\right)\right)\frac{(t - b_n/V_n^{1/2})^2 - 1}{6}\phi(t - b_n/V_n^{1/2}) \end{aligned} \tag{7.37}$$

and

$$P_*(T_{\nu,n}^* < t) = \Phi(t - \hat{b}_n/\hat{V}_{\nu,n}^{1/2}) + \hat{R}_{\nu,n0} + O_P((nh)^{-1+\gamma}) \tag{7.38}$$

where $\hat{R}_{\nu,n0}$ are the bootstrap counterparts of $R_{\nu,n0}$ for $\nu = 1, 2$. Hence,

$$\begin{aligned} |P(T_{\nu,n} < t) - P_*(T_{\nu,n}^* < t)| &\leq \int_{t - \max(b_n/V_{\nu,n}^{1/2}, \hat{b}_n/\hat{V}_{\nu,n}^{1/2})}^{t - \min(b_n/V_{\nu,n}^{1/2}, \hat{b}_n/\hat{V}_{\nu,n}^{1/2})} \phi(t)\,dt + |R_{\nu,n0} - \hat{R}_{\nu,n0}| + O_P((nh)^{-1+\gamma}) \\ &\leq \left|\frac{b_n}{V_{\nu,n}^{1/2}} - \frac{\hat{b}_n}{\hat{V}_{\nu,n}^{1/2}}\right| + |R_{\nu,n0} - \hat{R}_{\nu,n0}| + O_P((nh)^{-1+\gamma}) \end{aligned} \tag{7.39}$$

Now, note that we have by lemma 8.5 that

$$\hat{\rho}_{n3} - \rho_{n3} = O_P((nh)^{-1}).$$

From this relation we obtain

$$|R_{1,n0} - \hat{R}_{1,n0}| = O_P((nh)^{-1}). \tag{7.40}$$

Analogeously, we deduce by lemma 8.5

$$\begin{aligned} |R_{2,n0} - \hat{R}_{2,n0}| &= O_P(((nh)^{-1/2}[(n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2}) \\ &\quad + (n\lambda_3)^{-1} + \lambda_3^{2r})^{1/2})])(1 + h^k(nh)^{1/2})) \\ &= O_P((nh)^{-1}(1 + h^k(nh)^{1/2})) \end{aligned} \tag{7.41}$$

because of $h \ll \lambda_2$.

Now we consider the first difference in relation 7.39. By a Taylor expansion of $\hat{V}_{\nu,n}^{-1/2}$ at $V_{\nu,n}$ we get

$$\hat{V}_{\nu,n}^{-1/2} = V_{\nu,n}^{-1/2} - \frac{1}{2}V_{\nu,n}^{-3/2}(\hat{V}_{\nu,n} - V_{\nu,n}) + \frac{3}{8}\tilde{V}_{\nu,n}^{-5/2}(\hat{V}_{\nu,n} - V_{\nu,n})^2 \tag{7.42}$$

where $\tilde{V}_{\nu,n}$ is between $V_{\nu,n}$ and $\hat{V}_{\nu,n}$. Therefore,

$$\begin{aligned}\frac{\hat{b}_n}{\hat{V}_{\nu,n}^{1/2}} - \frac{b_n}{V_{\nu,n}^{1/2}} &= \hat{b}_n\left(\frac{1}{V_{\nu,n}^{1/2}} - \frac{1}{2}\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} + \frac{3}{8}\frac{(\hat{V}_{\nu,n} - V_{\nu,n})^2}{\tilde{V}_{\nu,n}^{5/2}}\right) - \frac{b_n}{V_{\nu,n}^{1/2}} \\ &= \frac{\hat{b}_n - b_n}{V_{\nu,n}^{1/2}} - \frac{1}{2}b_n\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} \\ &\quad -\frac{1}{2}(\hat{b}_n - b_n)\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} + \frac{3}{8}\hat{b}_n\frac{(\hat{V}_{\nu,n} - V_{\nu,n})^2}{\tilde{V}_{\nu,n}^{5/2}}.\end{aligned} \tag{7.43}$$

Now, recall that $V_{\nu,n} = O((nh)^{-1})$, $\tilde{V}_{\nu,n} = O_P((nh)^{-1})$ $(\nu = 1,2)$, $\hat{b}_n = O_P(h^k)$, furthermore $\hat{b}_n - b_n = O_P(h^k g^s + (ng)^{-1/2}(h/g)^k))$ by lemma 8.6 and $\hat{V}_{1,n} - V_{1,n} = O_P((nh)^{-3/2})$, $\hat{V}_{2,n} - V_{2,n} = O_P((nh)^{-1}(\lambda_2^{2r} + (n\lambda_2)^{-1})^{1/2})$ by 3.3 and 3.10. Hence, for $\nu = 1,2$ it holds that

$$(\hat{b}_n - b_n)\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} = O_P((h^k g^s + (ng)^{-1/2}(h/g)^k)((nh)^{-3/2}) + (nh)^{-1}(\lambda_2^{2r} + (n\lambda_2)^{-1})^{1/2})) \tag{7.44}$$

and

$$\hat{b}_n\frac{(\hat{V}_{\nu,n} - V_{\nu,n})^2}{\tilde{V}_{\nu,n}^{5/2}} = O_P(h^k((nh)^{-1/2} + (\lambda_2^{2r} + (n\lambda_2))^{-1/2})) \tag{7.45}$$

are of higher asymptotic order than

$$\frac{\hat{b}_n - b_n}{V_{\nu,n}^{1/2}} = O_P((h^k g^s + (ng)^{-1/2}(h/g)^k)(nh)^{1/2}) = O_P((h^k g^s (nh)^{1/2} + (h/g)^{k+1/2})$$

and

$$\begin{aligned}b_n\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} &= O_P(h^k(nh)^{3/2}((nh)^{-3/2} + (nh)^{-1}(\lambda_2^2 r + (n\lambda_2))^{-1/2})) \\ &= O_P((nh)^{-1/2}h^k)\end{aligned}$$

because of $h \ll \lambda_2$. Additionally,

$$\frac{\hat{b}_n - b_n}{V_{\nu,n}^{1/2}}$$

does obviously not depend on estimators of the error moments. Therefore we obtain

$$\frac{\hat{b}_n}{\hat{V}_{\nu,n}^{1/2}} - \frac{b_n}{V_{\nu,n}^{1/2}} = -\frac{1}{2}b_n\frac{\hat{V}_{\nu,n} - V_{\nu,n}}{V_{\nu,n}^{3/2}} + O_P((h^k g^s(nh)^{1/2} + (h/g)^{k+1/2}) + O((nh)^{-1}). \tag{7.46}$$

Finally, from 7.39, 7.40 and 7.46 it follows the validity of equation 2.12.

$\square$

Proof of theorem 5.1: *Part I* ($h \ll n^{-1/(2k+1)}$):
At first, we consider the wild bootstrap case. We invert the Edgeworth expansions 2.13 and 2.15. That is, from 2.13 it follows by Taylor expansions of $\Phi(\bullet)$ and $\phi(\bullet)$ at t that

$$\begin{aligned}
& P\left(T_{1,n} < t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right) \\
= \;& \Phi\left(t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right) \\
& +\rho_{n3}\frac{2\left(t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right)^2 + 1}{6}\phi\left(t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right) \\
& +\frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}\left(t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right)\phi\left(t + \frac{b_n}{V_n^{1/2}} + \rho_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}t\right) \\
& +O((nh)^{-1+\gamma}) \\
= \;& \Phi(t) + O((nh)^{-1+\gamma}). \qquad (7.47)
\end{aligned}$$

The corresponding bootstrap version is given by

$$\begin{aligned}
& P_*\left(T_{1,n}^* < t + \frac{\hat{b}_n}{\hat{V}_n^{1/2}} + \hat{\rho}_{n3}\frac{2t^2+1}{6} + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}}\hat{\rho}_{n3}t\right) \\
= \;& \Phi(t) + O_P((nh)^{-1+\gamma}). \qquad (7.48)
\end{aligned}$$

Now we define for $\alpha \in (0,1)$ the crital values t_α, $\hat{t}_\alpha$ and z_α by

$$P_*(T_{1,n}^* \le \hat{t}_\alpha) = P(T_{1,n} \le t_\alpha) = \Phi(z_\alpha) = 1 - \alpha. \qquad (7.49)$$

Then it follows from 7.47 and 7.48 that

$$\begin{aligned}
& P\left(T_{1,n} < z_\alpha + \frac{b_n}{V_{1,n}^{1/2}} + \rho_{n3}\frac{2z_\alpha^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}z_\alpha\right) \\
= \;& \Phi(z_\alpha) + O((nh)^{-1+\gamma}) \\
= \;& \alpha + O((nh)^{-1+\gamma}) \\
= \;& P(T_{1,n} < t_\alpha) + O((nh)^{-1+\gamma}) \qquad (7.50)
\end{aligned}$$

and

$$\begin{aligned}
& P_*\left(T_{1,n}^* < z_\alpha + \frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}} + \hat{\rho}_{n3}\frac{2z_\alpha^2+1}{6} + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}}\hat{\rho}_{n3}z_\alpha\right) \\
= \;& \Phi(z_\alpha) + O_P((nh)^{-1+\gamma}) \\
= \;& \alpha + O_P((nh)^{-1+\gamma}) \\
= \;& P_*(T_{1,n}^* < \hat{t}_\alpha) + O_P((nh)^{-1+\gamma}), \qquad (7.51)
\end{aligned}$$

respectively. From 7.50 and 7.51 we obtain

$$t_\alpha = z_\alpha + \frac{b_n}{V_{1,n}^{1/2}} + \rho_{n3}\frac{2z_\alpha^2+1}{6} + \frac{1}{2}\frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}z_\alpha + O_P((nh)^{-1+\gamma})$$

and

$$\hat{t}_\alpha = z_\alpha + \frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}} + \hat{\rho}_{n3}\frac{2z_\alpha^2+1}{6} + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}}\hat{\rho}_{n3}z_\alpha + O_P((nh)^{-1+\gamma}),$$

hence

$$\begin{aligned}\hat{t}_\alpha &= t_\alpha + \left(\frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}} - \frac{b_n}{V_{1,n}^{1/2}}\right) + (\hat{\rho}_{n3} - \rho_{n3})\frac{2z_\alpha^2+1}{6} \\ &\quad + \frac{1}{2}z_\alpha\left(\frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}}\hat{\rho}_{n3} - \frac{b_n}{V_{1,n}^{1/2}}\rho_{n3}\right) + O_P((nh)^{-1+\gamma}). \end{aligned} \tag{7.52}$$

For $\nu = 1, 2$ we denote the coverage error we seek by

$$\delta_{\nu,\alpha} := P(T_{\nu,n} < \hat{t}_\alpha) - (1-\alpha).$$

and we claim that for arbitrary small $\gamma > 0$ it holds that

$$\delta_{1,\alpha} = O((nh)^{1/2}h^k g^s + (h/g)^{k+1} + h^k + (nh)^{-1+\gamma}). \tag{7.53}$$

Result 7.53 may be proved by the so-called "delta-method" (see e.g. Hall, 1992), as follows. We write

$$\Delta_1 := \frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}$$

and

$$\begin{aligned}\Delta_2 &:= (\hat{\rho}_{n3} - \rho_{n3})\frac{2z_\alpha^2+1}{6} - \frac{1}{2}z_\alpha\left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}}\hat{\rho}_{n3} - \frac{b_n}{V_n^{1/2}}\rho_{n3}\right) \\ &\quad + O_P((nh)^{-1+\gamma})\end{aligned}$$

and remark that we can rewrite 7.52 as

$$\hat{t}_\alpha = t_\alpha + \Delta_1 + \Delta_2 + O_P((nh)^{-1+\gamma}) \tag{7.54}$$

$$=: t_\alpha + \Delta_1 + \tilde{\Delta}_2. \tag{7.55}$$

Now note that by Markov's and Whittle's inequalities and by the results of the sections 2 and 3 we have

$$P(|\Delta_2| > (nh)^{-1+\gamma}) = O(((nh)^{-1+\gamma}) =: \eta_n. \tag{7.56}$$

Furthermore,

$$P(T_{1,n} < \hat{t}_\alpha) = P(T_{1,n} < t_\alpha + \Delta_1 + \tilde{\Delta}_2) \tag{7.57}$$

$$\begin{aligned} &\le P(T_{1,n} < t_\alpha + \Delta_1 + \eta_n) + P(|\tilde{\Delta}_2| > \eta_n) \\ &\le P(T_{1,n} < t_\alpha + \Delta_1 + \eta_n) + O(\eta_n) \end{aligned}$$

$$= P(T_{1,n} - \Delta_1 - \eta_n < t_\alpha) + O(\eta_n) \tag{7.58}$$

where 7.57 follows by 7.55 and 7.58 follows by 7.56. Analogeously,

$$P(T_{1,n} < \hat{t}_\alpha) \geq P(T_{1,n} - \Delta_1 + \eta_n < t_\alpha) + O(\eta_n). \tag{7.59}$$

As in the proof of lemma 2.1 we expand $T_{1,n} - \Delta_1$ in an Edgeworth series. In order to do that, we approximate

$$\begin{aligned} T_{1,n,\Delta} &:= T_{1,n} - \Delta_1 \\ &= \frac{\hat{m}(x_0) - m(x_0)}{\hat{V}_{1,n}^{1/2}} - \frac{\hat{b}_n}{\hat{V}_{1,n}^{1/2}} + \frac{b_n}{V_{1,n}^{1/2}} \\ &= \frac{\sum_{i=1}^n \overline{w}(x_0,x_i)m(x_i) - m(x_0) + \sum_{i=1}^n \overline{w}(x_0,x_i)\epsilon_i}{\hat{V}_{1,n}^{1/2}} + \frac{b_n}{V_{1,n}^{1/2}} \end{aligned} \tag{7.60}$$

with

$$\overline{w}(x_0,x_i) := w_{k,h}(x_0,x_i) - \sum_{j=1}^n w_{k,h}(x_0,x_j)w_{(k+s,k),g}(x_j,x_i) - w_{(k+s,k),g}(x_0,x_i) \tag{7.61}$$

by

$$\tilde{T}_{1,n,\Delta} := \frac{\sum_{i=1}^n \overline{w}(x_0,x_i)m(x_i) - m(x_0) + \sum_{i=1}^n \overline{w}(x_0,x_i)\epsilon_i}{\tilde{V}_{1,n}^{1/2}} + \frac{b_n}{V_{1,n}^{1/2}}. \tag{7.62}$$

Then, by Skovgaard (1986) the validity of an Edgeworth expansion of $\tilde{T}_{1,n,\Delta}$ follows.

We calculate the cumulants of $\tilde{T}_{1,n,\Delta}$, as follows. At first, recall that as in relation 7.43 we have by lemma 8.6 and 7.3

$$\begin{aligned} \Delta_1 &= \frac{\hat{b}_n - b_n}{V_{1,n}^{1/2}} - \frac{1}{2}b_n\frac{\tilde{V}_{1,n} - V_{1,n}}{V_{1,n}^{3/2}} + R_n \\ &= \frac{O(h^k g^s) + \sum_{i=1}^n u_n(x_i)\epsilon_i}{V_{1,n}^{1/2}} - \frac{1}{2}b_n\frac{\alpha_{1,n2}}{V_{1,n}^{3/2}} + R_n \\ &=: \Delta_1' + R_n \end{aligned} \tag{7.63}$$

with $u_n(x_i) = O((ng)^{-1}(h/g)^k)$ defined in lemma 8.6 and a remainder term R_n of higher order. On the other hand, we remark that it is analogeously to 7.5 sufficient to approximate T_n by

$$T_{1,n}' := \frac{\alpha_{n1} + b_n}{V_{1,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}}. \tag{7.64}$$

That is, we appoximate $\tilde{T}_{1,n,\Delta}$ by

$$\begin{aligned} T_{1,n,\Delta}' := &\frac{\alpha_{n1} + b_n}{V_{1,n}^{1/2}} - \frac{1}{2}\frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}} \\ &- \frac{O(h^k g^s) + \sum_{i=1}^n u_n(x_i)\epsilon_i}{V_{1,n}^{1/2}} - \frac{1}{2}b_n\frac{\alpha_{1,n2}}{V_{1,n}^{3/2}}. \end{aligned} \tag{7.65}$$

Recall that we defined α_{n1} and $\alpha_{1,n2}$ in 7.3. Because of

$$E\alpha_{1,n2} = \sum_{i=1}^{n} w_{k,h}(x_0, x_i) E(\epsilon_i^2 - \mu_2(x_i)) = 0$$

it holds that

$$E\Delta_1 = O((nh)^{1/2} h^k g^s),$$

hence

$$\tilde{\kappa}_{1,n} := ET'_{1,n,\Delta} = \kappa_{1,n} + O((nh)^{1/2} h^k g^s) + O((nh)^{-1+\gamma}). \tag{7.66}$$

where $\kappa_{i,n}$ denotes the i-th cumulant of $T_{1,n}$. Now, denote

$$T''_{1,n,\Delta} := T'_{1,n,\Delta} - \frac{b_n}{V_n^{1/2}}.$$

and

$$T''_{1,n} := T'_{1,n} - b_n / V_n^{1/2}.$$

Then

$$\begin{aligned}
E(T''_{1,n,\Delta})^2 &= E(T''_{1,n})^2 + 2E\Bigg(\frac{\alpha_{n1}}{V_{1,n}^{1/2}} \frac{1}{2} b_n \frac{\alpha_{1,n2}}{V_{1,n}^{3/2}} \\
&\quad + \frac{1}{2} \frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}} \frac{O(h^k g^s) + \sum_{i=1}^{n} u_n(x_i)\epsilon_i}{V_{1,n}^{1/2}} \\
&\quad - \frac{1}{2} \frac{(\alpha_{n1} + b_n)\alpha_{1,n2}}{V_{1,n}^{3/2}} \frac{1}{2} b_n \frac{\alpha_{1,n2}}{V_{1,n}^{3/2}} \\
&\quad - \frac{\alpha_{n1}}{V_n^{1/2}} \frac{O(h^k g^s) + \sum_{i=1}^{n} u_n(x_i)\epsilon_i}{V_n^{1/2}} \Bigg) + O((nh)^{-1+\gamma}) \\
&=: ET_{1,n}^2 + T_1 + T_2 + T_3 + T_4 + O((nh)^{-1+\gamma}).
\end{aligned} \tag{7.67}$$

It holds that

$$\begin{aligned}
E\alpha_{n1}\alpha_{1,n2} &= E \sum_{i=1}^{n} w_{k,h}(x_0, x_i)\epsilon_i \sum_{j=1}^{n} w_{k,h}^2(x_0, x_i)(\epsilon_j^2 - \mu_2(x_j)) \\
&= \sum_{i=1}^{n} w_{k,h}^3(x_0, x_i)\mu_3(x_i) \\
&= O((nh)^{-2}).
\end{aligned} \tag{7.68}$$

Therefore,

$$T_1 = \frac{\sum_{i=1}^{n} w_{k,h}^3(x_0, x_i)\mu_3(x_i)}{V_{1,n}^2} b_n = O(h^k). \tag{7.69}$$

Furthermore,

$$T_2 = V_{1,n}^{-2} E\Bigg(\left(\sum_{i=1}^{n} w_{k,h}(x_0, x_i)\epsilon_i + O(h^k) \right) \left(\sum_{j=1}^{n} w_{k,h}^2(x_0, x_j)(\epsilon_j^2 - \mu_2(x_j)) \right)$$

$$
\begin{aligned}
& \times\left(O(h^k g^s)+\sum_{l=1}^{n} u_n(x_l)\epsilon_l\right)\Bigg) \\
= \; & V_{1,n}^{-2}\Bigg(\sum_{i,j=1}^{n} w_{k,h}^2(x_0,x_i)u_n(x_i)w_{k,h}(x_0,x_j)\mu_2(x_j)\mu_2(x_i) \\
& +O(h^k)\sum_{i=1}^{n} w_{k,h}^2(x_0,x_i)u_n(x_i)\mu_3(x_i) \\
& +O(h^k g^s)\sum_{i=1}^{n} w_{k,h}^3(x_0,x_i)\mu_3(x_i)\Bigg) \\
= \; & O((h/g)^{k+1}+h^k(h/g)^{k+1}+h^k g^s) \\
= \; & O((h/g)^{k+1}+h^k g^s) \qquad (7.70)
\end{aligned}
$$

It holds

$$T_3=-\frac{b_n}{2V_{1,n}^3}E((\alpha_{n1}+b_n)\alpha_{1,n2}^2).$$

Now,

$$E\alpha_{n1}\alpha_{1,n2}^2=\sum_{i=1}^{n} w_{k,h}^5(x_0,x_i)\mu_5(x_i)=O((nh)^{-4})$$

and

$$b_nE\alpha_{1,n2}^2=\sum_{i=1}^{n} w_{k,h}^4(x_0,x_i)\mu_4(x_i)=O(h^k(nh)^{-3}).$$

Therefore,

$$
\begin{aligned}
T_3 &= O((nh)^3h^k((nh)^{-4}+h^k(nh)^{-3})) \\
&= O(h^k((nh)^{-1}+h^k). \qquad (7.71)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
T_4 &= -2V_{1,n}^{-1}\sum_{i=1}^{n} w_{k,h}(x_0,x_i)u_n(x_i)\mu_2(x_i) \\
&= O((h/g)^{k+1}). \qquad (7.72)
\end{aligned}
$$

From 7.69, 7.70, 7.71 and 7.72 it follows that

$$\tilde{\kappa}_{2,n}=E(T''_{1,n,\Delta})^2-(ET''_{1,n,\Delta})^2=\kappa_{2,n}+O((h/g)^{k+1}+h^k+(nh)^{-1}). \qquad (7.73)$$

Furthermore,

$$
\begin{aligned}
E(T''_{1,n,\Delta})^3 &= E(T''_{1,n})^3+E(T''_{1,n})^2\Delta'_1+ET''_{1,n}(\Delta'_1)^2+E(\Delta'_1)^3 \\
&= O((nh)^{1/2}h^kg^s+(nh)^{-1}) \qquad (7.74)
\end{aligned}
$$

because of

$$
\begin{aligned}
E(T''_{1,n})^2\Delta'_1 &= \frac{O(h^kg^s)}{V_{1,n}^{3/2}}E\alpha_{n1}^2\sum_{i=1}^{n} u_n(x_i)\epsilon_i+O((nh)^{-1}) \\
&= O((nh)^{1/2}h^kg^s+(nh)^{-1}),
\end{aligned}
$$

$ET''_{1,n}(\Delta_1')^2 = O((nh)^{-1})$ and $E(\Delta_1')^3 = O((nh)^{-1})$. It is easily deduced that

$$E(T''_{1,n,\Delta})^l = E(T''_{1,n})^l + O((nh)^{-1}) \tag{7.75}$$

for $l \geq 4$. Hence,

$$\begin{aligned} \tilde{\kappa}_{3,n} &= E(T''_{1,n,\Delta})^3 - 3E(T''_{1,n,\Delta})^2 ET''_{1,n,\Delta} + 2(ET''_{1,n,\Delta})^3 \\ &= \kappa_{3,n} + O((nh)^{1/2} h^k g^s + (nh)^{-1}) \end{aligned} \tag{7.76}$$

and

$$\tilde{\kappa}_{l,n} = \kappa_{l,n} + O((nh)^{-1}) = O((nh)^{-1}) \tag{7.77}$$

for $l \geq 4$.

From 7.66, 7.73, 7.76, 7.77 and

$$\tilde{T}_{1,n} - T_{1,n} = \tilde{O}((nh)^{-1+\gamma}, n^{-1})$$

for arbitrarily small $\gamma > 0$ we deduce that

$$P(T_{1,n} - \Delta_1 < t) = P(T_{1,n} < t) + O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k + (nh)^{-1+\gamma}). \tag{7.78}$$

Again by an Edgeworth expansion of $T_{1,n} + \eta_n$ we easily derive that

$$P(T_n < t_\alpha + \eta_n) = P(T_n < t_\alpha) + O(\eta_n). \tag{7.79}$$

Finally, by 7.58, 7.59, 7.78 and 7.79 we obtain that

$$\begin{aligned} P(T_{1,n} < \hat{t}_\alpha) &= P(T_{1,n} < t_\alpha) + O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k + (nh)^{-1+\gamma}) \\ &= \alpha + O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k + (nh)^{-1+\gamma}). \end{aligned} \tag{7.80}$$

Thus, we have shown 7.53.

For the moment-oriented bootstrap we claim that

$$\delta_{2,\alpha} = O((nh)^{1/2} h^k g^s + (h/g)^{k+1} + h^k (nh)^{1/2} \lambda_2^r + h^k h/\lambda_2 + (nh)^{-1+\gamma}). \tag{7.81}$$

The proof of relation 7.81 follows the lines of that of 7.53. In the following we only deal with the differences to the proof of 7.53.

As we will see below, there are only differences to the wild bootstrap case when $h \ll \lambda_2$ what is assumed in the following. According to lemma 8.3 it holds that

$$\begin{aligned} E\tilde{V}_{2,n} &= \sum_{i=1}^n w_{k,h}^2(x_0, x_i) E\hat{\mu}_2(x_i) \\ &\asymp \sum_{i=1}^n w_{k,h}^2(x_0, x_i) E\left(\mu_2(x_i) + \frac{\lambda_s^r}{r!} \mu_2^{(r)}(x_i) \int_{-\tau}^{\tau} u^r K_r(u)\, du\right) \\ &\asymp V_{2,n} + \sum_{i=1}^n w_{k,h}^2(x_0, x_i) \left(\frac{\lambda_2^r}{r!} \mu_2^{(r)}(x_i) \int_{-\tau}^{\tau} u^r K_r(u)\, du\right) \\ &\asymp V_{2,n} + (nh)^{-1} \left(\frac{\lambda_2^r}{r!} \mu_2^{(r)}(x_0) \int_{-\tau}^{\tau} u^r K_r(u)\, du\right) \int K_k^2(z)\, dz \\ &= V_{2,n} + O((nh)^{-1} \lambda_2^r). \end{aligned} \tag{7.82}$$

Hence,

$$\tilde{\tilde{\kappa}}_{1,n} := ET'_{2,n,\Delta} = \kappa_{1,n} + O(h^k (nh)^{1/2}\lambda_2^r + (nh)^{1/2}h^k g^s) + O((nh)^{-1+\gamma}). \tag{7.83}$$

In higher order cumulants there are only differences in the term T_1 and in terms of higher order. These differences arise (according to 7.25) only if $h \ll \lambda_2$. Thus, from 7.25 it follows that

$$\begin{aligned} T_1 &= \frac{b_n}{V_n^2} E\alpha_{n1}\alpha_{2,n2} \\ &= O(h^k (nh)(n\lambda_2)^{-1}) \\ &= O(h^k h/\lambda_2). \end{aligned} \tag{7.84}$$

Therefore,

$$\tilde{\tilde{\kappa}}_{2,n} = \kappa_{2,n} + O(h^k\lambda_2/h + (h/g)^{k+1} + (nh)^{-1}). \tag{7.85}$$

Part II ($h \ll n^{-1/(2k+1)}$):
The proof in the MISE-optimal case $h \asymp n^{-1/(2k+1)}$ is similar to part I of the proof of this theorem. For $t := \tilde{t} + b_n/V_{1,n}^{1/2}$ we deduce that, by 2.4 and 2.8,

$$\begin{aligned} &\Phi(\tilde{t}) + \rho_{n3}\frac{2\tilde{t}^2+1}{6}\phi(\tilde{t}) + \frac{1}{2}\frac{b_n}{V_n^{1/2}}\rho_{n3}\tilde{t}\phi(\tilde{t}) + O((nh)^{-1}) \\ = \ & P\left(T_{1,n} - \frac{b_n}{V_n^{1/2}} < \tilde{t}\right) \\ = \ & P\left(T_{1,n} < \tilde{t} + \frac{b_n}{V_n^{1/2}}\right) \\ = \ & P(T_{1,n} < t) + O((nh)^{-1}) \end{aligned} \tag{7.86}$$

and

$$\begin{aligned} &\Phi(\tilde{t}) + \hat{\rho}_{n3}\frac{2\tilde{t}^2+1}{6}\phi(\tilde{t}) + \frac{1}{2}\frac{\hat{b}_n}{\hat{V}_n^{1/2}}\hat{\rho}_{n3}\tilde{t}\phi(\tilde{t}) + O_P((nh)^{-1}) \\ = \ & P_*\left(T_{1,n}^* - \frac{\hat{b}_n}{\hat{V}_n^{1/2}} < \tilde{t}\right) \\ = \ & P_*\left(T_{1,n}^* < \tilde{t} + \frac{\hat{b}_n}{\hat{V}_n^{1/2}}\right) \\ = \ & P_*\left(T_{1,n}^* < \tilde{t} + \frac{b_n}{V_n^{1/2}} + \left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}\right)\right) \\ = \ & P_*\left(T_{1,n}^* < t + \left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}\right)\right) + O_P((nh)^{-1}). \end{aligned} \tag{7.87}$$

Note that

$$\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}} = o_p(1)$$

vanishes asymptotically and is nonstochastic in the bootstrap world. Therefore, we obtain by inversion of 7.87 (see part I of this proof) that

$$P_*\left(T^*_{1,n} < t + \left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}\right)\right) = P_*(T^*_{1,n} < t) + \left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}\right)\phi(t) + O_P((nh)^{-1}). \tag{7.88}$$

Hence, we have derived

$$\begin{aligned} & P_*(T^*_{1,n} < t) - P(T_{1,n} < t) \\ = \; & \left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}} - \frac{b_n}{V_n^{1/2}}\right)\phi(t) + (\hat{\rho}_{n3} - \rho_{n3})\frac{2\tilde{t}^2+1}{6}\phi(\tilde{t}) + \frac{1}{2}\left(\frac{\hat{b}_n}{\hat{V}_n^{1/2}}\hat{\rho}_{n3} - \frac{b_n}{V_n^{1/2}}\rho_{n3}\right)\tilde{t}\phi(\tilde{t}) \\ & + O_P((nh)^{-1}). \end{aligned} \tag{7.89}$$

The rest of the proof is analogeous to part I.

□

8 Some technical lemmas

Lemma 8.1 *Assume (A1), (A3) and (A4). Then*

$$\begin{aligned} (i) \quad E[\hat{\epsilon}_i^2 - \mu_2(x_i))(\hat{\epsilon}_j^2 - \mu_2(x_j)] &= E[\epsilon_i^2 - \mu_2(x_i)][\epsilon_j^2 - \mu_2(x_j)] \\ &\quad + O(h^k\delta_{ij} + h^{4k} + (nh)^{-1}), \\ (ii) \quad E[\hat{\epsilon}_i^3 - \mu_3(x_i))(\hat{\epsilon}_j^3 - \mu_3(x_j)] &= E[\epsilon_i^3 - \mu_3(x_i)][\epsilon_j^3 - \mu_3(x_j)] \\ &\quad + O(h^k\delta_{ij} + (nh)^{-1}) \end{aligned}$$

where δ_{ij} denotes the Kronecker delta.

Proof: (i)

$$\begin{aligned} & E[\hat{\epsilon}_i^2 - \mu_2(x_i))(\hat{\epsilon}_j^2 - \mu_2(x_j)] \\ = \; & E[(\hat{\epsilon}_i^2 - \epsilon_i^2) + (\epsilon_i^2 - \mu_2(x_i))][(\hat{\epsilon}_j^2 - \epsilon_j^2) + (\epsilon_j^2 - \mu_2(x_j))] \\ = \; & E[\epsilon_i^2 - \mu_2(x_i)][\epsilon_j^2 - \mu_2(x_j)] + E[\hat{\epsilon}_i^2 - \epsilon_i^2][\hat{\epsilon}_j^2 - \epsilon_j^2] \\ & + E[\hat{\epsilon}_i^2 - \epsilon_i^2)][\epsilon_j^2 - \mu_2(x_j)] + E[\hat{\epsilon}_j^2 - \epsilon_j^2)][\epsilon_i^2 - \mu_2(x_i)] \\ = \; & E[\epsilon_i^2 - \mu_2(x_i)][\epsilon_j^2 - \mu_2(x_j)] + R_1 + R_2 + R_3 \end{aligned} \tag{8.1}$$

Recall that $\epsilon_i = Y_i - m(x_i)$ and $\hat{\epsilon}_i = Y_i - \hat{m}_h(x_i)$. Therefore

$$\begin{aligned} \hat{\epsilon}_i^2 - \epsilon_i^2 &= 2Y_i(m(x_i) - \hat{m}_h(x_i)) + \hat{m}_h^2(x_i) - m^2(x_i) \\ &= (\hat{m}_h(x_i) - m(x_i))^2 - 2\epsilon_i(\hat{m}_h(x_i) - m(x_i)) \end{aligned} \tag{8.2}$$

Furthermore we have

$$\hat{m}_h(x_i) - m(x_i) = b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l \tag{8.3}$$

with the bias

$$b_n(x_i) = \sum_{l=1}^{n} w_{k,h}(x_i, x_l)m(x_l) - m(x_i)$$

defined in lemma 2.1. Because of $E\epsilon_j^2 = \mu_2(x_j)$, $b_n(x_i) = O(h^k)$ and $w_{k,h}(x_i, x_j) = O((nh)^{-1})$ we derive from 8.1, 8.2 and 8.3

$$\begin{aligned}
R_2 &= E\left(\left(b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)^2 - 2\epsilon_i\left(b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)\left(\epsilon_j^2 - \mu_2(x_j)\right)\right) \\
&= 2b_n(x_i)w_{k,h}(x_i, x_j)E(\epsilon_j(\epsilon_j^2 - \mu_2(x_j))) + w_{k,h}^2(x_i, x_j)E(\epsilon_j^2(\epsilon_j^2 - \mu_2(x_j))) \\
&\quad -2E\left((\epsilon_i b_n(x_i) + w_{k,h}(x_i, x_i)\epsilon_i^2)(\epsilon_j^2 - \mu_2(x_j))\right) \\
&= O(h^k(nh)^{-1} + (nh)^{-2} + \delta_{ij}(h^k + (nh)^{-1})). \qquad (8.4)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
R_1 &= E\left(\hat{\epsilon}_i^2 - \epsilon_i^2\right)\left(\hat{\epsilon}_j^2 - \epsilon_j^2\right) \\
&= E\left(\left(b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)^2 - 2\epsilon_i\left(b_n(x_i)\sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)\right) \\
&\quad \times\left(\left(b_n(x_j) + \sum_{l=1}^{n} w_{k,h}(x_j, x_l)\epsilon_l\right)^2 - 2\epsilon_j\left(b_n(x_j)\sum_{l=1}^{n} w_{k,h}(x_j, x_l)\epsilon_l\right)\right). (8.5)
\end{aligned}$$

We calculate the different terms of 8.5. First,

$$\begin{aligned}
&\left(b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)^2 \\
&= b_n^2(x_i) + \sum_{l,m=1}^{n} w_{k,h}(x_i, x_l)w_{k,h}(x_j, x_m)\epsilon_l\epsilon_m + 2b_n(x_i)\sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l.
\end{aligned}$$

Hence,

$$\begin{aligned}
&E\left(\left(b_n(x_i) + \sum_{l=1}^{n} w_{k,h}(x_i, x_l)\epsilon_l\right)^2\left(b_n(x_j) + \sum_{l=1}^{n} w_{k,h}(x_j, x_l)\epsilon_l\right)^2\right) \\
&= b_n^2(x_i)b_n^2(x_j) + 4b_n(x_i)b_n(x_j)\sum_{l=1}^{n} w_{k,h}(x_i, x_l)w_{k,h}(x_j, x_l)E\epsilon_l^2 \\
&\quad + \sum_{l,m=1}^{n}\left(w_{k,h}^2(x_i, x_l)w_{k,h}^2(x_j, x_m) + 2w_{k,h}(x_i, x_l)w_{k,h}(x_i, x_m)w_{k,h}(x_j, x_l)w_{k,h}(x_j, x_m)\right)E\epsilon_l^2\epsilon_m^2 \\
&\quad +2b_n(x_i)\sum_{l=1}^{n} w_{k,h}(x_i, x_l)w_{k,h}^2(x_j, x_l)E\epsilon_l^3 + 2b_n(x_j)\sum_{l=1}^{n} w_{k,h}(x_j, x_l)w_{k,h}^2(x_i, x_l)E\epsilon_l^3 \\
&= O(h^{4k} + h^{2k}(nh)^{-1} + (nh)^{-2} + h^k(nh)^{-2}). \qquad (8.6)
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
& E\left(\epsilon_i\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)\left(b_n(x_j)+\sum_{l=1}^n w_{k,h}(x_j,x_l)\epsilon_l\right)^2\right)\\
= {} & b_n(x_i)\left(w_{k,h}^2(x_j,x_i)E\epsilon_i^3+2b_n(x_j)w_{k,h}(x_j,x_i)E\epsilon_i^2\right)+b_n^2(x_j)w_{k,h}^2(x_i,x_i)E\epsilon_i^2\\
& +w_{k,h}(x_i,x_i)\sum_{l=1}^n w_{k,h}^2(x_j,x_l)E\epsilon_i^2\epsilon_l^2+w_{k,h}(x_j,x_i)\sum_{l=1}^n w_{k,h}(x_i,x_l)w_{k,h}(x_j,x_l)E\epsilon_i^2\epsilon_l^2\\
& +b_n(x_i)w_{k,h}(x_i,x_i)w_{k,h}(x_j,x_i)E\epsilon_i^3\\
= {} & O(h^k((nh)^{-2}+(nh)^{-1})+h^{2k}(nh)^{-1}+(nh)^{-2}+h^k(nh)^{-2})\\
= {} & O(h^{2k}(nh)^{-1}+(nh)^{-2}). \qquad (8.7)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
& E\left(\epsilon_i\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)\epsilon_j\left(b_n(x_j)+\sum_{l=1}^n w_{k,h}(x_j,x_l)\epsilon_l\right)\right)\\
= {} & \delta_{ij}b_n(x_i)b_n(x_j)E\epsilon_i^2-b_n(x_i)w_{k,h}(x_j,x_i)E\epsilon_i^2\\
& -b_n(x_j)w_{k,h}(x_i,x_j)E\epsilon_j^2+\sum_{l=1}^n w_{k,h}(x_i,x_l)w_{k,h}(x_j,x_l)E\epsilon_l^2\\
= {} & O(h^{2k}\delta_{ij}+h^k(nh)^{-1}+(nh)^{-1})\\
= {} & O(h^{2k}\delta_{ij}+(nh)^{-1}). \qquad (8.8)
\end{aligned}
$$

From 8.6, 8.7 and 8.8 it follows that

$$R_1=O(h^{2k}\delta_{ij}+h^{4k}+(nh)^{-1}). \qquad (8.9)$$

Obviously, R_3 is of the same asymptotic order as R_2. Hence, we obtain from 8.1, 8.4 and 8.9

$$E[\hat{\epsilon}_i^2-\mu_2(x_i))(\hat{\epsilon}_j^2-\mu_2(x_j)]=E[\epsilon_i^2-\mu_2(x_i)][\epsilon_j^2-\mu_2(x_j)]+O(h^{2k}\delta_{ij}+h^{4k}+(nh)^{-1}),$$

which completes the proof of the first part of the lemma.

(ii) The proof of the second part is in spirit the same as that of the first part. Therefore we present only the main steps. First, we write

$$
\begin{aligned}
& E[\hat{\epsilon}_i^3-\mu_3(x_i))(\hat{\epsilon}_j^3-\mu_3(x_j)]\\
= {} & E[(\hat{\epsilon}_i^3-\epsilon_i^3)+(\epsilon_i^3-\mu_3(x_i))][(\hat{\epsilon}_j^3-\epsilon_j^3)+(\epsilon_j^3-\mu_3(x_j))]\\
= {} & E[\epsilon_i^3-\mu_3(x_i)][\epsilon_j^3-\mu_3(x_j)]+E[\hat{\epsilon}_i^3-\epsilon_i^3][\hat{\epsilon}_j^3-\epsilon_j^3]\\
& +E[\hat{\epsilon}_i^3-\epsilon_i^3)][\epsilon_j^3-\mu_3(x_j)]+E[\hat{\epsilon}_j^3-\epsilon_j^3)][\epsilon_i^3-\mu_3(x_i)]\\
= {} & E[\epsilon_i^3-\mu_3(x_i)][\epsilon_j^3-\mu_3(x_j)]+R_1+R_2+R_3. \qquad (8.10)
\end{aligned}
$$

Now, the difference $\hat{\epsilon}_i^3-\epsilon_i^3$ can be written as (see 8.2 and 8.3)

$$
\begin{aligned}
\hat{\epsilon}_i^3-\epsilon_i^3 &= -3Y_i^2(\hat{m}_h(x_i)-m(x_i))+3Y_i(\hat{m}_h^2(x_i)-m^2(x_i))-(\hat{m}_h^3(x_i)-m^3(x_i))\\
&= -3\epsilon_i^2(\hat{m}_h(x_i)-m(x_i))+3\epsilon_i(\hat{m}_h(x_i)-m(x_i))^2-(\hat{m}_h(x_i)-m(x_i))^3
\end{aligned}
$$

$$
\begin{aligned}
&= -3\epsilon_i^2\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)+3\epsilon_i\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)^2\\
&\quad -\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)^3 \qquad (8.11)
\end{aligned}
$$

with $b_n(x_i)=O(h^k)$ defined in lemma 2.1. From 8.10 and 8.11 we derive

$$
\begin{aligned}
R_2 &= E\Bigg(\Bigg(-3\epsilon_i^2\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)+3\epsilon_i\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)^2\\
&\quad -\left(b_n(x_i)+\sum_{l=1}^n w_{k,h}(x_i,x_l)\epsilon_l\right)^3\Bigg)(\epsilon_j^3-\mu_3(x_j))\Bigg)\\
&= -3\left(b_n(x_i)(\mu_5(x_i)-\mu_2(x_i)\mu_3(x_i))\delta_{ij}+w_{k,h}(x_i,x_j)\mu_2(x_i)\mu_4(x_j)\right)+R_5\\
&= O(h^k\delta_{ij}+(nh)^{-1}) \qquad (8.12)
\end{aligned}
$$

because

$$
\begin{aligned}
R_5 &= 3\Bigg(b_n^2(x_i)\mu_4(x_i)\delta_{ij}+2b_n(x_i)w_{k,h}(x_i,x_i)(\mu_5(x_i)-\mu_2(x_i)\mu_3(x_i))\delta_{ij}\\
&\quad +(2w_{k,h}(x_i,x_i)w_{k,h}(x_i,x_j)+w_{k,h}^2(x_i,x_i))\mu_2(x_i)\mu_4(x_j))\\
&\quad -b_n^2(x_i)w_{k,h}(x_i,x_j)\mu_4(x_j)-b_n(x_i)w_{k,h}^2(x_i,x_j)(\mu_5(x_j)-\mu_2(x_j)\mu_3(x_j))\\
&\quad -\sum_{l=1}^n w_{k,h}^2(x_i,x_l)w_{k,h}(x_i,x_j)\mu_2(x_l)\mu_4(x_j)\Bigg)\\
&= O((h^{2k}+h^k(nh)^{-1})\delta_{ij}+(nh)^{-2}+h^{2k}(nh)^{-1}+h^k(nh)^{-2}+(nh)^{-2})\\
&= O((h^{2k}+h^k(nh)^{-1})\delta_{ij}+(nh)^{-2}+h^{2k}(nh)^{-1}) \qquad (8.13)
\end{aligned}
$$

is of higher asymptotic order.

Furthermore, we easily deduce that

$$
\begin{aligned}
R_1 &= E(-3b_n(x_i)\epsilon_i^2)(-3b_n(x_j)\epsilon_j^2)+E(-3\epsilon_i^2)(-3\epsilon_j^2)\sum_{l=1} w_{k,h}^2(x_i,x_l)\epsilon_l^2+R_6\\
&= O(\delta_{ij}h^{2k}+(nh)^{-1}) \qquad (8.14)
\end{aligned}
$$

because R_6 includes only terms of higher order. Finally, we obtain from 8.10, 8.12 and 8.14

$$
E[\hat\epsilon_i^2-\mu_2(x_i))(\hat\epsilon_j^2-\mu_2(x_j)]=E[\epsilon_i^3-\mu_3(x_i)][\epsilon_j^3-\mu_3(x_j)]+O(h^k\delta_{ij}+(nh)^{-1})
$$

which completes the proof of the lemma.

□

The following two lemmas are generalizations of assertions contained in appendix 2 of Gasser & Müller (1979).

Lemma 8.2 *We assume (A1), (A4), (A5) and that $g(\cdot)$ is some continous function on the interval $[0,1]$. Then for any $r \geq 1$ it holds that*

$$\begin{aligned}
\sum_{i=1}^{n} w_{k,h}^{r}(x_0, x_i) g(x_i) &= \frac{1}{(nh)^{r-1}} \frac{1}{h} \int_0^1 g(u) K_k^r\left(\frac{x_0 - u}{h}\right) du \\
&\quad + O((nh)^{-r-(1-\gamma_{K_k})}) && (8.15)\\
&= \frac{g(x_0)}{(nh)^{r-1}} \int_{-\tau}^{\tau} K_k^r(z)\, dz + R_n && (8.16)
\end{aligned}$$

where

$$R_n = \begin{cases} o(1), & g \text{ continous} \\ O(h^k + ((nh)^{-r-(1-\gamma_{K_k})}), & \text{if } r = 1 \text{ and } g \text{ } k\text{-times continously differentiable.} \end{cases}$$

Proof: We have

$$\begin{aligned}
&\left| \sum_{i=1}^{n} w_{k,h}^{r}(x_0, x_i) g(x_i) - \frac{1}{(nh)^{r-1}} \frac{1}{h} \int_0^1 g(u) K_k^r\left(\frac{x_0 - u}{h}\right) du \right| \\
= &\left| \sum_{i \in \mathcal{I}} \frac{1}{h^r} \left(\int_{s_{i-1}}^{s_i} K_k^r\left(\frac{x_0 - u}{h}\right) du \right)^r g(x_i) - \frac{1}{(nh)^{r-1}} \frac{1}{h} \int_0^1 g(u) K_k^r\left(\frac{x_0 - u}{h}\right) du \right| \\
= &\ A_1 && (8.17)
\end{aligned}$$

where $\mathcal{I}$ is the index set of design points x_i which the weight function $w_{k,h}^r(x_0, x_i)$ does not weight to zero. Then it holds obviously $\#\mathcal{I} = O(nh)$ and $x_i \in \mathcal{I}$ iff $|x_0 - x_i| < h$. From 8.17, the mean value theorem, the continuity of the function g and the Lipschitz continuity of the kernel function K_k we derive for $\xi_i, \theta_i \in [s_{i-1}, s_i]$

$$\begin{aligned}
A_1 &= \left| \sum_{i \in \mathcal{I}} \left[\frac{1}{h^r} (s_i - s_{i-1})^r K_k^r\left(\frac{x_0 - \xi_i}{h}\right) g(x_i) \right.\right. \\
&\quad \left.\left. - \frac{1}{(nh)^{r-1}} \frac{1}{h} (s_i - s_{i-1}) g(\theta_i) K_k^r\left(\frac{x_0 - \theta_i}{h}\right) \right] \right| \\
&= \frac{1}{(nh)^r} \left| \sum_{i \in \mathcal{I}} \left(K_k^r\left(\frac{x_0 - \xi_i}{h}\right) - K_k^r\left(\frac{x_0 - \theta_i}{h}\right) \right) g(x_i) \right| \\
&\leq \frac{1}{(nh)^r} \sum_{i \in \mathcal{I}} \left| K_k^r\left(\frac{x_0 - \xi_i}{h}\right) - K_k^r\left(\frac{x_0 - \theta_i}{h}\right) \right| |g(x_i)| \\
&\leq \frac{1}{(nh)^r} \sum_{i \in \mathcal{I}} \left| \frac{\theta_i - \xi_i}{h} \right|^{\gamma_{K_k}} |g(x_i)| \\
&= O((nh)^{-r-(1-\gamma_{K_k})}). && (8.18)
\end{aligned}$$

Note that the last relation holds according to $|\theta_i - \xi_i| = O(n^{-1})$ and $\#\mathcal{I} = O(nh)$. Thus, 8.15 is proven. For continous funtion g, the equation 8.16 follows by $h \to 0$, the continuity of g and

$$\int_0^1 g(u) K_k^r\left(\frac{x_0 - u}{h}\right) du = h \int_{-\tau}^{\tau} g(x_0 - zh) K_k^r(z)\, dz$$

$$\asymp \quad hg(x_0)\int_{-\tau}^{\tau} K_k^r(z)\,dz. \tag{8.19}$$

When $r=1$ and g is k-times continously differentiable, a Taylor expansion of g up to to the kernel order k gives for $\tilde{h}\in[0,h]$

$$\begin{aligned}
&\int_0^1 g(u)K_k\left(\frac{x_0-u}{h}\right)\,du \\
= \;& h\int_{-\tau}^{\tau} g(x_0-zh)K_k(z)\,dz \\
= \;& h\Bigg(g(x_0)\int_{-\tau}^{\tau} K_k(z)\,dz + \sum_{\kappa=1}^{k-1}(-1)^{\kappa}\frac{h^{\kappa}}{\kappa!}g^{(\kappa)}(x_0)\int_{-\tau}^{\tau} z^{\kappa}K_k(z)\,dz \\
&+(-1)^k\frac{h^k}{k!}\int_{-\tau}^{\tau} g^{(k)}(x_0-z\tilde{h})z^kK_k(z)\,dz\Bigg) \\
= \;& h\left(g(x_0)\int_{-\tau}^{\tau} K_k(z)\,dz + (-1)^k\frac{h^k}{k!}\int_{-\tau}^{\tau} g^{(k)}(x_0-z\tilde{h})z^kK_k(z)\,dz\right) \\
= \;& h\left(g(x_0)\int_{-\tau}^{\tau} K_k(z)\,dz + O(h^k)\right) \qquad (8.20)
\end{aligned}$$

which completes the proof of this lemma.

□

We denote

$$\hat{\mu}_{s,m_1,m_2}(x_i) := \sum_{l=1}^{n} w_{r,\lambda_s}(x_i,x_l)\tilde{\mu}_{s,m_1,m_2}(x_l) \tag{8.21}$$

where

$$\tilde{\mu}_{s,m_1,m_2}(x_i) := \left(\sum_{\kappa=m_1}^{m_2} a_{\kappa}Y_{l-\kappa}\right)^s \tag{8.22}$$

with constants a_κ that fulfill the inequalities $a_{m_1}\neq 0$, $a_{m_2}\neq 0$. For example, for $s=2$, $m_1=0$ and $m_2=1$ we have

$$\hat{\mu}_{2,0,1}(x_l) = \sum_{l=1}^{n} w_{r,\lambda_2}(x_i,x_l)\tilde{\mu}_{2,0,1}(x_l)$$

with

$$\tilde{\mu}_{2,0,1}(x_l) = (a_0Y_l + a_1Y_{l-1})^2.$$

Furthermore, we define $\nu_s(x_k,x_l) := cov(\tilde{\mu}_s(x_k),\tilde{\mu}_s(x_l))$ and $\nu_s(x_k) := \nu_s(x_k,x_k)$. Then, recall that (under (A1) and (A6)) by lemma 2.1 in Müller & Stadtmüller (1987a) and lemma 8.4 it follows that $\nu_s(\cdot,\cdot)$ is continous for $s=2,3$.

The following lemma generalizes theorems 1 and 2 of Gasser & Müller (1979) for the moments of the above kernel smoother based on such s - dependent errors $\tilde{\mu}_s$ which appear because of the higher order differences.

Lemma 8.3 *Under (A1) to (A6) is*

$$
\begin{aligned}
E\hat{\mu}_{s,m_1,m_2}(x_0) &= \mu_s(x_0) + \frac{\lambda_s^r}{r!}\mu_s^{(r)}(x_0)\int_{-\tau}^{\tau} u^r K_r(u)\,du \\
&\quad + O(n^{-1}) + o(\lambda_s^r), \qquad (8.23) \\
cov(\hat{\mu}_{s,m_1,m_2}(x_i), \hat{\mu}_{s,m_1,m_2}(x_j)) &\asymp (2(m_2 - m_1) + 1)\frac{1}{n\lambda_s^2}\int_0^1 \nu_s(u,u) \\
&\quad \times K_r\left(\frac{x_i - u}{\lambda_s}\right) K_r\left(\frac{x_j - u}{\lambda_s}\right) du \qquad (8.24) \\
&\asymp (2(m_2 - m_1) + 1)\frac{\nu_s(x_i, x_i)}{n\lambda_s} \\
&\quad \times \int_{-\tau}^{\tau} K_r(z) K_r\left(\frac{x_j - x_i}{\lambda_s} + z\right) dz \qquad (8.25)
\end{aligned}
$$

Proof: The equation 8.23 is proven in Gasser & Müller (1979), appendix 1. Without restriction of generality we prove 8.24 and 8.25 for the case $s = 2$, $m_1 = 0$, $m_2 = 1$, the proofs for local estimators with other parameters s, m_1 and m_2 are essentially the same.

At first, recall that $\nu_2(x_k, x_l) := cov(\tilde{\mu}_2(x_k), \tilde{\mu}_2(x_l)) = 0$ for $|k - l| \geq 2$. Hence,

$$
\begin{aligned}
& cov(\hat{\mu}_2(x_i), \hat{\mu}_2(x_j)) \\
= & \sum_{k\in\mathcal{I}_i, l\in\mathcal{I}_j, |k-l|\leq 1} w_{r,\lambda_2}(x_i, x_k) w_{r,\lambda_2}(x_j, x_l) \nu_2(x_k, x_l) \\
= & \frac{1}{\lambda_2^2}\sum_{k\in\mathcal{I}_i, l\in\mathcal{I}_j, |k-l|\leq 1} \nu_2(x_k, x_l)\left(\int_{s_{k-1}}^{s_k} K_r\left(\frac{x_i - u}{\lambda_2}\right) du\right)\left(\int_{s_{l-1}}^{s_l} K_r\left(\frac{x_j - u}{\lambda_2}\right) du\right).
\end{aligned}
$$

Therefore we derive by the mean value theorem

$$
\begin{aligned}
& \left| cov(\hat{\mu}_2(x_i), \hat{\mu}_2(x_j)) - 3\frac{1}{n\lambda_2^2}\int_0^1 \nu_2(u,u) K_r\left(\frac{x_i - u}{\lambda_2}\right) K_r\left(\frac{x_j - u}{\lambda_2}\right) du \right| \\
= & \left| \frac{1}{\lambda_2^2}\sum_{k\in\mathcal{I}_i, l\in\mathcal{I}_j, |k-l|\leq 1} \nu_2(x_k, x_l)\left(\int_{s_{k-1}}^{s_k} K_r\left(\frac{x_i - u}{\lambda_2}\right) du\right)\left(\int_{s_{l-1}}^{s_l} K_r\left(\frac{x_j - u}{\lambda_2}\right) du\right) \right. \\
& \left. -3\frac{1}{n\lambda_2^2}\sum_{l\in\mathcal{I}_i\cap\mathcal{I}_j}\int_{s_{l-1}}^{s_l} \nu_2(u,u) K_r\left(\frac{x_i - u}{\lambda_2}\right) K_r\left(\frac{x_j - u}{\lambda_2}\right) du \right| \\
= & \frac{1}{(n\lambda_2)^2}\left| \sum_{k\in\mathcal{I}_i, l\in\mathcal{I}_j, |k-l|\leq 1} \nu_2(x_k, x_l) K_r\left(\frac{x_i - \xi_k}{\lambda_2}\right) K_r\left(\frac{x_j - \theta_l}{\lambda_2}\right) \right. \\
& \left. -3\sum_{l\in\mathcal{I}_i\cap\mathcal{I}_j} \nu_2(\eta_l, \eta_l) K_r\left(\frac{x_i - \eta_l}{\lambda_2}\right) K_r\left(\frac{x_j - \eta_l}{\lambda_2}\right) \right| \\
= & A_2 \qquad (8.26)
\end{aligned}
$$

Now, note that $|k-l| \leq 1$ implies $|x_k - x_l| = O(n^{-1})$. Thus we derive by the Lipschitz continuity of the kernel K_r, by the continuity of $\nu_2(\cdot,\cdot)$ and by $\#\{k,l||k-l| \leq 1\} = 3$

that

$$
\begin{aligned}
A_2 &\asymp \frac{3}{(n\lambda_2)^2}\left|\sum_{l\in\mathcal{I}_i\cap\mathcal{I}_j}\nu_2(x_l,x_l)\Bigg(K_r\left(\frac{x_i-\xi_k}{\lambda_2}\right)K_r\left(\frac{x_j-\theta_l}{\lambda_2}\right)\right. \\
&\quad \left.-K_r\left(\frac{x_i-\eta_l}{\lambda_2}\right)K_r\left(\frac{x_j-\eta_l}{\lambda_2}\right)\Bigg)\right| \\
&\le \frac{3}{(n\lambda_2)^2}\sum_{l\in\mathcal{I}_i\cap\mathcal{I}_j}\nu_2(x_l,x_l)\Bigg(K_r\left(\frac{x_i-\theta_l}{\lambda_2}\right)\left|K_r\left(\frac{x_j-\xi_l}{\lambda_2}\right)K_r\left(\frac{x_j-\eta_l}{\lambda_2}\right)\right| \\
&\quad +K_r\left(\frac{x_j-\eta_l}{\lambda_2}\right)\left|K_r\left(\frac{x_i-\theta_l}{\lambda_2}\right)K_r\left(\frac{x_i-\eta_l}{\lambda_2}\right)\right|\Bigg) \\
&\le \frac{3}{(n\lambda_2)^2}\sum_{l\in\mathcal{I}_i\cap\mathcal{I}_j}\nu_2(x_l,x_l)\Bigg(K_r\left(\frac{x_i-\theta_l}{\lambda_2}\right)\left|\frac{\eta_l-\xi_l}{\lambda_2}\right|^{\gamma_{K_r}} \\
&\quad +K_r\left(\frac{x_j-\eta_l}{\lambda_2}\right)\left|\frac{\eta_l-\theta_l}{\lambda_2}\right|^{\gamma_{K_r}}\Bigg) \\
&= o((n\lambda_2)^{-1}) \qquad (8.27)
\end{aligned}
$$

which proves 8.24. For the proof of 8.25 note that by the continuity of $\nu(\cdot,\cdot)$ it follows that

$$
\begin{aligned}
&\int_0^1 \nu_2(u,u)K_r\left(\frac{x_i-u}{\lambda_2}\right)K_r\left(\frac{x_j-u}{\lambda_2}\right)du \\
&= \lambda_2\int_{-\tau}^{\tau}\nu_2(x_i+z\lambda_2,x_i+z\lambda_2)K_r(z)K_r\left(\frac{x_j-x_i}{\lambda_2}+z\right)dz \\
&\asymp \lambda_2\nu_2(x_i,x_i)\int_{-\tau}^{\tau}K_r(z)K_r\left(\frac{x_j-x_i}{\lambda_2}+z\right)dz \qquad (8.28)
\end{aligned}
$$

which completes the proof of the lemma.

□

Recall that we defined in 8.22 a local estimate

$$
\tilde{\mu}_{3,m_1,m_2}(x_i) := \left(\sum_{\kappa=m_1}^{m_2} a_\kappa Y_{i-\kappa}\right)^3
$$

for the third order error moment $\mu_3(x_i)$. Further note that Bunke (1997) shows that a necessary condition for the unbiasedness of $\tilde{\mu}_{3,m_1,m_2}(x_i)$ under local linear regression models is that the equalities $\sum_{j=m_1}^{m_2} a_j = 0$ and $\sum_{j=m_1}^{m_2} a_j^3 = 1$ hold which we assume in what follows. The following lemma generalizes lemma 2.1 from Müller & Stadtmüller (1987a) for such estimators of third order error moments.

Lemma 8.4 *We assume (A1) and (A6). Then*

$$
E\tilde{\mu}_{3,m_1,m_2}(x_i) \asymp \mu_3(x_i) \qquad (8.29)
$$

and

$$
\begin{aligned}
Var\tilde{\mu}_{3,m_1,m_2}(x_i) &\asymp \mu_6(x_i)\sum_{j=m_1}^{m_2} a_j^6 + 15\mu_2(x_i)\mu_4(x_i)\sum_{j\neq l} a_j^2 a_l^4 \\
&\quad + 9\mu_3^2(x_i) + 15\mu_2^3(x_i)\sum_{j\neq l,\, l\neq k,\, j\neq k} a_j^2 a_l^2 a_k^2. \qquad (8.30)
\end{aligned}
$$

Proof: At first, because of the Lipschitz continuity of the regression function $m(x)$ and because of $\sum_{\kappa=m_1}^{m_2} a_\kappa = 0$ we have

$$
\begin{aligned}
\sum_{\kappa=m_1}^{m_2} a_\kappa Y_{i-\kappa} &= \sum_{\kappa=m_1}^{m_2} a_\kappa \epsilon_{i-\kappa} + \sum_{\kappa=m_1}^{m_2} a_\kappa m(x_{i-\kappa}) \\
&= \sum_{\kappa=m_1}^{m_2} a_\kappa \epsilon_{i-\kappa} + O(n^{-\alpha}). \qquad (8.31)
\end{aligned}
$$

Then, by 8.31 and the Lipschitz continuity of $\mu_3(\cdot)$ we derive

$$
\begin{aligned}
E\tilde{\mu}_{3,m_1,m_2}(x_i) &= \sum_{\kappa=m_1}^{m_2} a_\kappa^3 E\epsilon_{i-\kappa}^3 + O(n^{-\alpha}) \\
&= \sum_{\kappa=m_1}^{m_2} a_\kappa^3 \mu_3(x_{i-\kappa}) + O(n^{-\alpha}) \\
&= \mu_3(x_i)\sum_{\kappa=m_1}^{m_2} a_\kappa^3 + O(n^{-\alpha} + n^{-\zeta_3}) \\
&= \mu_3(x_i) + O(n^{-\alpha} + n^{-\zeta_3}). \qquad (8.32)
\end{aligned}
$$

Now, using similar arguments and the independence of the error terms ϵ_i we show that

$$
\begin{aligned}
E\tilde{\mu}_{3,m_1,m_2}^2(x_i) &= E\left(\sum_{\kappa=m_1}^{m_2} a_\kappa \epsilon_{i-\kappa}\right)^2 + O(n^{-\alpha}) \\
&= \mu_6(x_i)\sum_{j=m_1}^{m_2} a_j^6 + \binom{6}{2}\mu_2(x_i)\mu_4(x_i)\sum_{j\neq l} a_j^2 a_l^4 \\
&\quad + \frac{1}{2}\binom{6}{3}\sum_{j\neq l} a_j^3 a_l^3 \mu_3^2(x_i) + 15\mu_2^3(x_i)\sum_{j\neq l,\, l\neq k,\, j\neq k} a_j^2 a_l^2 a_k^2 \\
&\quad + O\left(n^{-\alpha} + n^{-(\zeta_2+\zeta_3+\zeta_4+\zeta_6)}\right) \qquad (8.33)
\end{aligned}
$$

which completes the proof of the lemma because of

$$
Var\tilde{\mu}_{3,m_1,m_2}(x_i) = E\tilde{\mu}_{3,m_1,m_2}^2(x_i) + \left(E\tilde{\mu}_{3,m_1,m_2}(x_i)\right)^2
$$

and

$$
\sum_{j=m_1}^{m_2} a_j^3 = 1.
$$

□

The following lemmas give asymptotic rates for differences of coefficients of the Edgeworth expansions in lemma 2.1 and lemma 2.2.

Lemma 8.5 *Assume (A1) to (A6). Then, for* $h \ll \lambda_2 \asymp \lambda_3$,

$$\hat{\rho}_{n3} - \rho_{n3} = O_P((nh)^{-1}), \tag{8.34}$$

$$\begin{aligned} & \frac{\sum_{i=1}^n w_{k,h}(x_0,x_i)\hat{\mu}_3(x_i)}{\hat{V}_{2,n}^{3/2}} - \frac{\sum_{i=1}^n w_{k,h}(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \\ = \;& O_P((nh)^{-1/2}((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2}) + ((n\lambda_3)^{-1} + \lambda_3^{2r})^{1/2})) \end{aligned} \tag{8.35}$$

and

$$\begin{aligned} & \frac{\sum_{i=1}^n d_n(x_0,x_i)\hat{\mu}_3(x_i)}{\hat{V}_{2,n}^{3/2}} - \frac{\sum_{i=1}^n d_n(x_0,x_i)\mu_3(x_i)}{V_{2,n}^{3/2}} \\ = \;& O_P((nh)^{-1/2}((n\lambda_2)^{-1} + \lambda_2^{2r})^{1/2}) + ((n\lambda_3)^{-1} + \lambda_3^{2r})^{1/2})) \end{aligned} \tag{8.36}$$

where $d_n(x_0, x_i)$ *is defined in 2.7. For* $h \gg \lambda_2 \asymp \lambda_3$ *all expressions are of order*$O_P((nh)^{-1})$.

Proof: The proof goes along the lines fo that of 3.12. We show this lemma only for $h \ll \lambda_2 \asymp \lambda_3$, the other case is analogeous. Recall that because of $h = h_*$

$$\begin{aligned} \hat{\rho}_{n3} - \rho_{n3} &= \hat{V}_{1,n}^{-3/2} \sum_{i=1}^n w_{k,h}^3(x_0,x_i)\hat{\mu}_3(x_i) - V_{1,n}^{-3/2} \sum_{i=1}^n w_{k,h}^3(x_0,x_i)\mu_3(x_i) \\ &= V_{1,n}^{-3/2} \sum_{i=1}^n w_{k,h}^3(x_0,x_i)\left(\hat{\mu}_3(x_i) - \mu_3(x_i)\right) \\ &\quad + \left(\hat{V}_{1,n}^{-3/2} - V_{1,n}^{-3/2}\right) \sum_{i=1}^n w_{k,h}^3(x_0,x_i)\hat{\mu}_3(x_i) \end{aligned} \tag{8.37}$$

By Taylor expansion of $\hat{V}_{1,n}^{-3/2}$ at $V_{1,n}$ we get for some $\tilde{V}_{1,n}$ between $V_{1,n}$ and $\hat{V}_{1,n}$:

$$\hat{V}_{1,n}^{-3/2} - V_{1,n}^{-3/2} = -\frac{3}{2}\tilde{V}_{1,n}^{-5/2}(\hat{V}_{1,n} - V_{1,n}) = O_P(nh) \tag{8.38}$$

because of $\tilde{V}_{1,n} = O_P((nh)^{-1})$ and $\hat{V}_{1,n} - V_{1,n} = O_P(nh^{-3/2})$ which holds due to 3.3. Hence,

$$\left(\hat{V}_{1,n}^{-3/2} - V_{1,n}^{-3/2}\right) \sum_{i=1}^n w_{k,h}^3(x_0,x_i)\hat{\mu}_{3,1}(x_i) = O_P((nh)^{-1}). \tag{8.39}$$

Now, we have

$$E\left(\sum_{i=1}^n w_{k,h}^3(x_0,x_i)[\hat{\epsilon}_i^3 - \mu_3(x_i)]\right)^2$$

$$\begin{aligned}
&\sim \sum_{i=1}^{n} w_{k,h}^6(x_0,x_i)E[\epsilon_i^3-\mu_3(x_i)]^2 && (8.40)\\
&= \sum_{i=1}^{n} w_{k,h}^6(x_0,x_i)[\mu_4(x_i)-\mu_3^2(x_i)] \\
&\sim (nh)^{-5}[\mu_6(x_0)-\mu_3^2(x_0]\int K_k^6(z)\,dz && (8.41)\\
&= O((nh)^{-5}) && (8.42)
\end{aligned}$$

where 8.40 follows from lemma 8.1 (ii) and 8.41 follows from lemma 8.2. Hence, from 8.42 and $V_{1,n}=O((nh)^{-1})$ it follows that

$$V_{1,n}^{-3/2}\sum_{i=1}^{n} w_{k,h}^3(x_0,x_i)\,(\hat{\mu}_{3,1}(x_i)-\mu_3(x_i))=O_P((nh)^{-1}). \qquad (8.43)$$

Summing up 8.39 and 8.43 proves 8.34.

For the proof of 8.35 recall that according to 3.10 it holds that $\hat{V}_{2,n}-V_{2,n}=O_P((nh)^{-1}((n\lambda_2)^{-1}+\lambda_2^{2r})^{1/2}))$. Therefore we derive analogeously to 8.38

$$\hat{V}_{2,n}^{-3/2}-V_{2,n}^{-3/2}=O_P((nh)^{3/2}((n\lambda_2)^{-1}+\lambda_2^{2r})^{1/2}). \qquad (8.44)$$

Thus,

$$\left(\hat{V}_{2,n}^{-3/2}-V_{2,n}^{-3/2}\right)\sum_{i=1}^{n} w_{k,h}^3(x_0,x_i)\hat{\mu}_{3,2}(x_i)=O_P((nh)^{-1/2}((n\lambda_2)^{-1}+\lambda_2^{2r})^{1/2}). \quad (8.45)$$

The following rest of the proof is analogeous to that of 3.10. By lemma 8.3 and lemma 8.2 we have

$$\begin{aligned}
& E\left(\sum_{i=1}^{n} w_{k,h}^3(x_0,x_i)[\hat{\mu}_{3,2}(x_i)-\mu_3(x_i)]\right)^2 \\
= & \sum_{i,j=1}^{n} w_{k,h}^3(x_0,x_i)w_{k,h}^3(x_0,x_j)E[\hat{\mu}_{3,2}(x_i)-\mu_3(x_i)][\hat{\mu}_{3,2}(x_j)-\mu_3(x_j)] \\
\sim & \sum_{i,j=1}^{n} w_{k,h}^3(x_0,x_i)w_{k,h}^3(x_0,x_j) \\
& \times\left[5\frac{\nu_3(x_i)}{n\lambda_3}\int_{-\tau}^{\tau}K_r(z)K_r\left(\frac{x_j-x_i}{\lambda_3}+z\right)\,dz\right. \\
& \left.+\left((-1)^r\frac{{\lambda_3}^r}{r!}\mu_3^{(r)}(x_i)\int z^rK_r(z)\,dz\right)\left((-1)^r\frac{{\lambda_3}^r}{r!}\mu_3^{(r)}(x_j)\int z^rK_r(z)\,dz\right)\right] \\
=: & A_1. && (8.46)
\end{aligned}$$

Now, from lemma 8.4 it follows that

$$\begin{aligned}
\nu_3(x_i) &= Var\tilde{\mu}_3(x_i) \\
&\sim \frac{16}{9}\mu_6(x_i)+45\mu_2(x_i)\mu_4(x_i)+9\mu_3^2(x_i)+\frac{135}{2}\mu_2^3(x_i). && (8.47)
\end{aligned}$$

Finally, from 8.47 and lemma 8.2 we obtain

$$
\begin{aligned}
A_1 \sim\; & (nh)^{-4}\Bigg(\left(\int K_k^3(z)\,dz\right)^2 \\
& \times\Bigg[5(n\lambda_3)^{-1}\left(\frac{16}{9}\mu_6(x_0)+45\mu_2(x_0)\mu_4(x_0)+9\mu_3^2(x_0)+\frac{135}{2}\mu_2^3(x_0)\right)\int K_r^2(z)\,dz \\
& +\left(\frac{{\lambda_3}^r}{r!}\mu_3^{(r)}(x_0)\int z^r K_r(z)\,dz\right)^2\Bigg] \qquad (8.48) \\
=\; & O((nh)^{-4}((n\lambda_3)^{-1}+\lambda_3^{2r})) \qquad (8.49)
\end{aligned}
$$

which completes the proof of 8.35. The proof of 8.36 is analogeous.

□

Lemma 8.6 *Under (A1) to (A5) it holds that*

$$\hat{b}_n - b_n = O_P\left(h^k g^s + (h/g)^k (ng)^{-1/2}\right).$$

Proof: Recall that because of $h_* = h$ we have

$$
\begin{aligned}
\hat{b}_n - b_n &= \sum_{i=1}^n w_{k,h_*}(x_0,x_i)\hat{m}_g(x_i) - \hat{m}_g(x_0) \\
&\quad - \sum_{i=1}^n w_{k,h}(x_0,x_i)m(x_i) - m(x_0) \\
&= \sum_{i=1}^n w_{k,h}(x_0,x_i)\left[\hat{m}_g(x_i)-m(x_i)\right] - \left[\hat{m}_g(x_0)-m(x_0)\right] \\
&\quad + \sum_{i=1}^n \left[w_{k,h_*}(x_0,x_i)-w_{k,h}(x_0,x_i)\right]\hat{m}_g(x_i) \\
&= \sum_{i=1}^n w_{k,h}(x_0,x_i)\left[\hat{m}_g(x_i)-m(x_i)\right] - \left[\hat{m}_g(x_0)-m(x_0)\right] \\
&= \sum_{i=1}^n w_{k,h}(x_0,x_i)b_{n,g}(x_i) + \sum_{i,j=1}^n w_{k,h}(x_0,x_i)w_{(k+s,k),g}(x_i,x_j)\epsilon_j \\
&\quad - \left(b_{n,g}(x_0) + \sum_{j=1}^n w_{(k+s,k),g}(x_0,x_j)\epsilon_j\right) \\
&= \sum_{i=1}^n w_{k,h}(x_0,x_i)b_{n,g}(x_i) - b_{n,g}(x_0) \\
&\quad + \sum_{j=1}^n \left(\sum_{i=1}^n w_{k,h}(x_0,x_i)w_{(k+s,k),g}(x_i,x_j)\right.
\end{aligned}
$$

$$-w_{(k+s,k),g}(x_0, x_j)\Bigg)\epsilon_j \tag{8.50}$$

where

$$\begin{aligned} b_{n,g}(x_0) &:= \sum_{j=1}^{n} w_{(k+s,k),g}(x_0, x_j)m(x_j) - m(x_0) && (8.51)\\ &\asymp h^k \, Bias\, \hat{m}_g^{(k)}(x_0). && (8.52) \end{aligned}$$

The last relation follows from

$$\hat{m}_g^{(k)}(x_0) = h^{-k}\sum_{j=1}^{n} w_{(k+s,k),g}(x_0, x_j)m(x_j).$$

Now, according Gasser & Müller (1984) it holds that

$$\begin{aligned} Bias\, \hat{m}_g^{(k)}(x_0) &\asymp (-1)^{k+s}\frac{g^s}{(k+s)!}m^{(k+s)}(x_0)\int_{-\tau}^{\tau} K_{(k+s,k)}(z)z^{k+s}\, dz\\ &= O(g^s). \end{aligned}$$

Thus,

$$b_{n,g}(x_0) \asymp (-1)^{k+s}\frac{h^k g^s}{(k+s)!}m^{(k+s)}(x_0)\int_{-\tau}^{\tau} K_{(k+s,k)}(z)z^{k+s}\, dz = O(h^k g^s) \tag{8.53}$$

from which follows that

$$c_n(x_0) = O(h^k g^s). \tag{8.54}$$

On the other hand, we have

$$\begin{aligned} u_n(x_j) &:= \sum_{i=1}^{n} w_{k,h}(x_0, x_i)w_{(k+s,k),g}(x_i, x_j) - w_{(k+s,k),g}(x_0, x_j)\\ &= \frac{1}{hg}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K_k\left(\frac{x_0-u}{h}\right)\, du \int_{s_{j-1}}^{s_j} K_{(k+s),k}\left(\frac{x_i-u}{g}\right)\, du\\ &\quad -\frac{1}{g}\int_{s_{j-1}}^{s_j} K_{(k+s),k}\left(\frac{x_0-u}{g}\right)\, du\\ &= \frac{1}{hg}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K_k\left(\frac{x_0-u}{h}\right)\, du \int_{s_{j-1}}^{s_j}\Bigg(K_{(k+s),k}\left(\frac{x_i-u}{g}\right)\, du\\ &\quad -K_{(k+s),k}\left(\frac{x_0-u}{g}\right)\, du\Bigg)\\ &=: A_1. && (8.55) \end{aligned}$$

Now, according to (A6) we derive by a Taylor expansion of $K_{(k+s),k}\left(\frac{x_i-u}{g}\right)$ at $\left(\frac{x_0-u}{g}\right)$ that

$$\int_{s_{j-1}}^{s_j}\left(K_{(k+s),k}\left(\frac{x_i-u}{g}\right)\, du - K_{(k+s),k}\left(\frac{x_0-u}{g}\right)\, du\right)$$

$$
\begin{aligned}
&= \sum_{\kappa=1}^{k} \left(\frac{x_i - x_0}{g} \right)^{\kappa} \frac{1}{g} \int_{s_{j-1}}^{s_j} K^{(\kappa)}_{(k+s),k} \left(\frac{x_0 - u}{g} \right) du + R_n \\
&= \sum_{\kappa=1}^{k} \left(\frac{x_i - x_0}{g} \right)^{\kappa} \int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} K^{(\kappa)}_{(k+s),k}(z)\, dz + R_n
\end{aligned}
$$

with a higher order remainder term R_n. By integration by parts we have for $\kappa = 1$

$$
\begin{aligned}
&\int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} K^{(\kappa)}_{(k+s),k}(z)\, dz \\
&= K'_{(k+s),k} \left(\frac{x_0 - s_{j-1}}{g} \right) du - K'_{(k+s),k} \left(\frac{x_0 - s_j}{g} \right) du - \int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} z K_{(k+s),k}(z)\, dz \\
&= - \int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} z K_{(k+s),k}(z)\, dz \\
&\leq - \int_{-\tau}^{\tau} z K_{(k+s),k}(z)\, dz
\end{aligned}
$$

for sufficiently large n, that is for sufficiently small bandwidth g. For general $\kappa \geq 1$ we derive analogeously that

$$
\begin{aligned}
\int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} K^{(\kappa)}_{(k+s),k}(z)\, dz &= (-1)^{\kappa} \int_{(x_0-s_j)/g}^{(x_0-s_{j-1})/g} z^{\kappa} K_{(k+s),k}(z)\, dz \\
&\leq (-1)^{\kappa} \int_{-\tau}^{\tau} z^{\kappa} K_{(k+s),k}(z)\, dz
\end{aligned}
$$

where the last relation holds for sufficiently large n. Recall that for $\kappa \leq k-1$ it holds $\int_{-\tau}^{\tau} z^{\kappa} K_{(k+s),k}(z)\, dz = 0$ Hence,

$$
\begin{aligned}
A_1 &\leq (-1)^k \frac{1}{h} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_k \left(\frac{x_0 - u}{h} \right) du \left(\frac{x_i - x_0}{g} \right)^k \int_{(x_0-s_{j-1})/g}^{(x_0-s_j)/g} z^k K^{(k)}_{(k+s),k}(z)\, dz + R_n \\
&= (-1)^k \frac{1}{h} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_k \left(\frac{x_0 - u}{h} \right) du \left(\frac{x_i - x_0}{g} \right)^k \\
&\quad \times \frac{1}{g} \int_{s_{j-1}}^{s_j} \left(\frac{x_0 - u}{g} \right)^k K^{(k)}_{(k+s),k} \left(\frac{x_0 - u}{g} \right) du + R_n \\
&\leq (-1)^k \frac{1}{h} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_k \left(\frac{x_0 - u}{h} \right) du \left(\frac{x_i - x_0}{g} \right)^k \max(K)(ng)^{-1} + R_n \\
&= \sum_{i=1}^{n} w_{k,h}(x_0, x_i) O((ng)^{-1}(h/g)^k) \\
&= O((ng)^{-1}(h/g)^k).
\end{aligned}
$$

Finally,

$$
E \left(\sum_{j=1}^{n} u_n(x_j) \epsilon_j \right)^2 = \sum_{j \in \mathcal{I}_j} u_n^2(x_j) E \epsilon_j^2
$$

$$
\begin{aligned}
&= \sum_{j \in \mathcal{I}_j} u_n^2(x_j)\mu_2(x_j) \\
&= O((ng)^{-1}(h/g)^{2k}). \qquad (8.56)
\end{aligned}
$$

Note that the last relation follows because (due to the weights in $u_n(x_j)$) the sum is taken over a set $\mathcal{I}_j$ with cardinality ng. Thus, the lemma is proven.

□

References

Bunke, O. (1997). Bootstrapping in Heteroscedastic Regression Situations. *Discussion Paper*, Sonderforschungsbereich 373, Humboldt University, Berlin, to appear.

Bunke, O., Droge, B. & Polzehl, J. (1995). Model Selection, Transformations and Variance Estimation in Nonlinear Regression. *Discussion Paper* **52**, Sonderforschungsbereich 373, Humboldt-Universität, Berlin.

Gasser, T. & Müller, H.G. (1979). Kernel Estimation of Regression Functions, in Smoothing Techniques in Curve Estimation. *Lecture Notes in Mathematics* **757**, 23-68.

Gasser, T. & Müller, H.G. (1984). Estimating Regression Functions and Their Derivatives by the Kernel Method. *Scand. J. Statist.* **11**, 171-185.

Gasser, T. , Seifert, B. & Wolf, A. (1993). Nonparametric Estimation of Residual Variance Revisited. *Biometrika* **80**, 375-383.

Hall, P. (1983). Inverting an Edgeworth Expansion. *Ann. Statist.* **11**, 569-576.

Hall, P. (1991). Edgeworth Expansions for Nonparametric Density Estimators, with Applications. *Statistics* **22**, 215-232.

Hall, P. (1992a). On Bootstrap Confidence Intervals in Nonparametric Regression. *Ann. Statist.* **20**, 695-711.

Hall, P. (1992b). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

Härdle, W., Huet, S. & Jolivet, E. (1995). Better Bootstrap Confidence Intervals for Regression Curve Estimation. *Statistics* **26**, 287-306.

Härdle, W. & Mammen, E. (1993). Comparing Nonparametric Versus Parametric Regression Fits. *Ann. Statist.* **21**, 1926-1947.

Müller, H.-G. & Stadtmüller, W. (1987a). Estimation of Heteroscedasticity in Regression Analysis. *Ann. Statist.* **15**, 610-625

Müller, H.-G. & Stadtmüller, W. (1987b). Variable Bandwidth Kernel Estimators of Regression Curves. *Ann. Statist.* **15**, 182-201.

Neumann, M. H. (1992). On Completely Data-driven Pointwise Confidence Intervals in Nonparametric Regression. *Rapport Technique* **92-02**, INRA, Dépt. de Biométrie, Jouy-en-Josas, France.

Neumann, M. H. (1995). Automatic Bandwidth Choice and Confidence Intervals in Nonparametric Regression. *Ann. Statist.* **23**, 1937-1959.

Neumann, M. H. (1997). Pointwise Confidence Intervals in Nonparametric Regression with Heteroscedastic Error Structure. *Statistics* **29**, 1-36.

Sacks, J. & Ylvisaker, D. (1970). Designs for Regression Problems with Correlated error. *Ann. Math. Statist.* **41**, 2057-2074.

Skovgaard, I. M. (1981). Transformations of an Edgeworth Expansion by a Sequence of Smooth Functions. *Scand. J. Statist.* **8**, 207-217.

Skovgaard, I. M. (1986). On Multivariate Edgeworth Expansions. *Internat. Statist. Rev.* **54**, 169-186.

Sommerfeld, V. (1997). Construction of Automatic Confidence Intervals in Nonparametric Heteroscedastic Regression by a Moment-Oriented Bootstrap. *Discussion Paper* **22**, Sonderforschungsbereich 373, Humboldt-University, Berlin.

Wu, C. F. J. (1986). Jackknife, Bootstrap and other Resampling Methods in Regression Analysis (with Discussion). *Ann. Statist.* **14** 1261-1295.