Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Automatic bandwidth choice and confidence intervals in nonparametric regression

M.H. Neumann

submitted: 23th September 1993

Institut für Angewandte Analysis
und Stochastik
Mohrenstraße 39
D – 10117 Berlin
Germany

Preprint No. 61
Berlin 1993

1991 Mathematics Subject Classification. Primary 62 G 15; Secondary 62 G 07, 62 G 20.
Key words and phrases. Nonparametric regression, bandwidth choice, confidence intervals, Edgeworth expansions.

Herausgegeben vom
Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
D – 10117 Berlin

Fax: + 49 30 2004975
e-Mail (X.400): c=de;a=d400;p=iaas-berlin;s=preprint
e-Mail (Internet): preprint@iaas-berlin.d400.de

AUTOMATIC BANDWIDTH CHOICE AND CONFIDENCE INTERVALS IN NONPARAMETRIC REGRESSION

Michael H. Neumann
Institut für Angewandte Analysis und Stochastik,
Berlin

Abstract

In the present paper we combine the issues of bandwidth choice and construction of confidence intervals in nonparametric regression. We modify the $\sqrt{n}$-consistent bandwidth selector of Härdle, Hall and Marron (1991) such that it is appropriate for heteroscedastic data and show how one can adapt the bandwidth g of the pilot estimator $\widehat{m}_g$ in a reasonable data-dependent way. Then we compare the coverage accuracy of classical confidence intervals based on kernel estimators with data-driven bandwidths. We propose a method to put undersmoothing with a data-driven bandwidth into practice and show that this procedure outperforms explicit bias correction.

1991 Mathematics Subject Classification. Primary 62G15; Secondary 62G07, 62G20.
Key words and phrases. Nonparametric regression, bandwidth choice, confidence intervals, Edgeworth expansions.

1. Introduction

We assume observations

$$Y_i = m(x_i) + \varepsilon_i \ , \quad i = 1, ..., n \ ,$$

where the errors ε_i are independently, not necessarily identically distributed with zero mean and variance $v(x_i)$. The nonrandom design points x_i are assumed to be spaced on the unit interval $[0, 1]$, $x_1 < x_2 < \cdots < x_n$.
We aim at defining an interval estimator for the regression function m at some interior point x_0 of this interval. As a starting point we estimate $m(x_0)$ by a Gasser-Müller kernel estimator $\widehat{m}_h(x_0)$, see Gasser and Müller (1979) [3].
There exists already a very developed theory for confidence intervals based on kernel estimators with nonrandom bandwidths. Under assumptions on the decay of the bandwidths it is shown that these methods are consistent and, moreover, there are rates for the errors in coverage probability calculated. Hall (1991) [8] for confidence intervals for a density and Hall (1992) [9] for intervals in regression with i.i.d. errors found optimal rates for the bandwidths involved in the confidence interval procedure by optimizing the coverage accuracy. On the other hand, the majority of the available literature does not take into account the bandwidth choice that is necessary for practical applications. Some exceptions we are aware of are papers of Faraway and Jhun (1990) [2] for density estimation and Faraway (1990) [1] for the regression case, where the bandwidth as well as the quantile for confidence bands are obtained on the basis of the same bootstrap sample. However, the authors do not provide any rigorous result on the real coverage probability in comparison to the prescribed level.
Usually the first step in constructing asymptotic confidence intervals consists in the definition of an asymptotically normally distributed pivotal quantity. There are two commonly used methods to deal with the bias of the initial estimator $\widehat{m}_h(x_0)$, undersmoothing and explicit bias correction on the basis of yet another kernel estimator. In Hall (1991 and 1992) [8], [9] it is shown that the undersmoothing method leads to a better coverage accuracy. An analogous result is proved in Neumann (1992) [15] for the case of not necessarily identical error distributions, again for kernel estimators with nonrandom bandwidths. However, it does not seem to be clear at all, how we could choose an undersmoothed bandwidth in a reasonable way from the data. The problem is, that only rates but not the corresponding constants are known for bandwidths optimizing the coverage accuracy, so that we have no hint how to choose them for a fixed sample size n. So it seems that there is no reliable substitute for bandwidths chosen by some criterion connected to the risk behavior of the corresponding estimator.
The practical bandwidth choice will be the main goal of the present paper. Whereas we can apply the bias correction method with usual bandwidth selectors at all stages, we replace the pure undersmoothing method by a two-step procedure that yields the same rates for the coverage accuracy. As an estimator of the optimal global bandwidth we employ here with some minor modifications the $\sqrt{n}$-consistent bandwidth selector of Härdle, Hall and Marron (1991) [5], [HHM91] hereafter, based on plug-in estimates of the integrated variance and the integrated squared bias of $\widehat{m}_h$. To make this method fully data-driven, we propose a method how the bandwidth of the pilot estimate $\widehat{m}_g$ needed for the bias estimation can be optimally chosen from the data.
The second step in getting confidence intervals is the recognition of the distributions of the abovementioned pivotal quantities. A simple approach is given by a normal approximation, which provides rates of at best $O((nh)^{-1/2})$, as shown in several papers. In the present paper we restrict our considerations exclusively on a distribution recognition via

Edgeworth expansions for the pivotal quantities. Recently, the application of bootstrap techniques in general, and in the context of heteroscedastic error distributions the wild bootstrap proposed by Härdle and Mammen (1990) [7] in particular, has become quite popular, see also the paper of Härdle, Huet and Jolivet (1992) [6]. However, in Neumann (1992) [15] it is shown that we obtain via Edgeworth expansions the same rate for the recognition of the distribution. Since on the other hand the quantiles via Edgeworth expansions are explicitly given, there seems to be no need for the computationally more involved bootstrap.

We derive the quantiles on the basis of expansions for pivotal quantities with nonrandom bandwidths. The distributions of the pivotal quantities with data-dependent bandwidths are approximated on the basis of expansions for quantities with nonrandom bandwidths. We show that the resulting upper rate bound is the same up to a factor of order n^{δ} for arbitrarily small $\delta > 0$ as we would obtain by the (formal) Edgeworth expansion of the pivotal quantity.

2. A FULLY DATA-DRIVEN BANDWIDTH SELECTOR

As already mentioned, all kernel estimators included in the procedure should be equipped with data-driven bandwidths. Although the coverage accuracy is the main criterion for the performance of our methods, we choose the bandwidths according to the risk behavior of the corresponding estimator, since a specific choice for confidence interval purposes seems to be unrealistic from the practical point of view. Since only the behaviour of $\widehat{m}_h$ at x_0 influences the properties of the confidence interval, it seems to be on first sight reasonable to seek for an estimate of the locally optimal bandwidth. On the other hand, we agree with Härdle and Bowman (1988) [4] who claim that the potential advantages of local adaptive bandwidth selection in the context of confidence intervals are not clear. Roughly speaking, the initial estimator $\widehat{m}_h(x_0)$ will only serve as a vehicle to introduce a nondegenerate noise structure, whereas its bias will be corrected with the help of a second estimator $\widehat{m}_g$. On the other hand, since our methods are based on Edgeworth expansions of pivotal quantities with nonrandom bandwidths, we can expect a better coverage accuracy for intervals with random bandwidths that are very close to some fixed one.

For fixed h, the mean squared error of $\widehat{m}_h(x_0) = \sum W(x_0,h)_i Y_i$ can be written as $MSE(h) = V_h + B_h^2$, where $V_h = \sum_{i=1}^n W(x_0,h)_i^2 v(x_i)$ and $B_h = \sum_{i=1}^n W(x_0,h)_i\, m(x_i) - m(x_0)$.

We estimate V_h simply by

$$\widehat{V}_h = \sum_{i=1}^{n} W(x_0,h)_i^2\, \widehat{v}_i\,, \tag{2.1}$$

where

$$\widehat{v}_i = \widehat{\varepsilon}_i^2 = (Y_i - \widetilde{m}(x_i))^2\,, \tag{2.2}$$

and where $\widetilde{m}$ denotes yet another kernel estimator. Anticipating the following results we remark that the consistency of the bandwidth selector considered in this section as well as of the confidence intervals in the next section require a higher degree of smoothness for m than the basic kernel estimator $\widehat{m}_h$ can exploit. We assume throughout the present paper that m is $(r+s)$-times continuously differentiable and therefore we take $\widetilde{m}$ as a kernel estimator $\widehat{m}_{\widehat{f}}$ with an $(r+s)$-th order kernel and a bandwidth $\widehat{f}$, which is for simplicity chosen by cross-validation.

The bias B_h will be approximated by an estimator of the form

$$\widehat{B}_{h,g} = \sum_{i=1}^{n} W(x_0,h)_i\, \widehat{m}_g(x_i) - \widehat{m}_g(x_0)\,, \tag{2.3}$$

where $\widehat{m}_g(x)$ is another kernel estimator with weights $\widetilde{W}(x,g)_i$ based on an s-th order kernel $\widetilde{w}$ and a bandwidth g.
Now we intend to modify the bandwidth selector of [HHM91] such that it takes into account the possible heteroscedasticity of the data. An even more important issue is the data-dependent choice of the bandwidth of the auxiliary estimator $\widehat{m}_g$, which is used for the estimation of the bias.
We are going to estimate the bandwidth h_0 that is optimal with respect to the mean integrated squared error ($MISE$) of $\widehat{m}_h$, where the integration is because of boundary effects restricted to some interval $[c,d]$ with $0 < c < d < 1$ that should include for our purposes the point x_0. The $MISE$ splits up into an integrated variance part

$$IV(h) = \int_c^d \sum_{i=1}^{n} W(x,h)_i^2 v(x_i)\, dx = \sum_{i=1}^{n} \int_c^d W(x,h)_i^2 dx\; v(x_i)$$

and an integrated squared bias part

$$ISB(h) = \int_c^d \left(\sum_{i=1}^{n} W(x,h)_i m(x_i) - m(x) \right)^2 dx\,.$$

We estimate $IV(h)$ analogously to (2.1) by

$$\widehat{IV}(h) = \sum_{i=1}^{n} \int_c^d W(x,h)_i^2 dx\; \widehat{v}_i. \tag{2.4}$$

Provided an appropriate choice of g, the quantity

$$\overline{ISB}(h,g) = \int_c^d \left(\sum_{i=1}^{n} W(x,h)_i \widehat{m}_g(x_i) - \widehat{m}_g(x) \right)^2 dx = \sum_{i,j=1}^{n} A(h,g)_{ij} Y_i Y_j$$

with

$$A(h,g)_{ij} = \int_c^d \left(\sum_k W(x,h)_k \widetilde{W}(x_k,g)_i - \widetilde{W}(x,g)_i \right) \left(\sum_l W(x,h)_l \widetilde{W}(x_l,g)_j - \widetilde{W}(x,g)_j \right) dx\,,$$

could already serve as an estimator of $ISB(h)$. As remarked by [HHM91], such an estimator is biased due to the variance of the diagonal terms and therefore it is natural to estimate $ISB(h)$ by

$$\widehat{ISB}(h,g) = \overline{ISB}(h,g) - \sum_{i=1}^{n} A(h,g)_{ii}\, \widehat{v}_i\,. \tag{2.5}$$

An alternative approach in the framework of density estimation is proposed by Sheather and Jones (1991) [12], [16]. They recognize that the non-stochastic term bias has the opposite sign to the bias due to the smoothing, and they choose the auxiliary bandwidth such that these terms cancel. However, an appropriate choice of this bandwidth requires the estimation of higher derivatives, and if enough smoothness is present to do this reasonably well, then we could use it in a different way to improve on the whole procedure at other stages.

Now we have with $\widehat{IV}(h) + \widehat{ISB}(h,g)$ a pattern to estimate $MISE(h)$, but it remains to fix an appropriate value of g. This problem was not solved in an entirely satisfactory way in [HHM91], and the authors conjectured that there is no substitute for trying some number of different g's. Assuming slightly more regularity than in [HHM91], namely $m \in C^{2r+s}[0,1]$ instead of $m \in C^{2r\wedge r+s}[0,1]$, we obtain that the term of order $O(n^{-1})$ of the mean squared error of $\widehat{ISB}(h,g)$ as an estimator of $ISB(h)$ does not depend on g, whereas the smallest two of the remaining terms do. These terms can be used to get a reasonable, asymptotically optimal choice for g.
The assumptions needed for the following lemma as well as for the assumptions in the sequel are given in the Appendix.

Lemma 2.1. *Assume* (A_G), (A_{BW}). *Then*

(i) $$E\left(\widehat{ISB}(h,g) - ISB(h)\right)^2$$
$$\begin{aligned} = \; & h^{4r}C(h)n^{-1} \\ & + 4h^{4r}g^{2s}\kappa_r^4\lambda_s^2\left(\int_c^d m^{(r+s)}(x)m^{(r)}(x)\,dx\right)^2 \\ & + 2h^{4r}n^{-2}g^{-(4r+1)}\kappa_r^4\int_c^d (v(x)/d(x))^2\,dx\int\left(\int \tilde{w}^{(r)}(y)\tilde{w}^{(r)}(y+z)\,dy\right)^2 dz \\ & + o(h^{4r}(g^{2s} + n^{-2}g^{-(4r+1)})), \end{aligned}$$

where $C(h)$ *is bounded and* $\kappa_r = \int z^r w(z)\,dz$, $\lambda_s = \int z^s \tilde{w}(z)\,dz$,
(ii)

$$g_{opt} = \left(\frac{(4r+1)\int_c^d (v(x)/d(x))^2 \int(\int \tilde{w}^{(r)}(y)\tilde{w}^{(r)}(y+z)dy)^2 dz}{4s\lambda_s^2\left(\int_c^d m^{(r+s)}(x)m^{(r)}(x)\,dx\right)^2}\, n^{-2}\right)^{\frac{1}{2s+4r+1}} (1+o(1)).$$

Let $\overline{g}$ be any consistent estimator of g_{opt}, which satisfies at least

$$P\left(c \le n^{1/(s+2r+1/2)}\overline{g} \le C\right) = O(n^{-1})$$

for some positive constants c and C. Now we define $\widehat{h}$ as a measurable minimizer of $\widehat{MISE}(h,\overline{g}) = \widehat{IV}(h) + \widehat{ISB}(h,\overline{g})$, whose existence is ensured by a Lemma of Jennrich (1969) [11].

Remark. Analogously to Theorem 1 in [HHM91] one can prove that

$$\frac{\widehat{h} - h_0}{h_0} = O_p\left(n^{-\Delta}\right),$$

where $\Delta = 1/2 \wedge s/(s+2r+1/2)$.

The next point concerns the appropriate choice of the bandwidth g for the local bias estimator $\widehat{B}_{h,g}$. First we infer from Lemma 6.2 that

$$\widehat{B}_{h,g} - B_h = \int w(z)\int_0^{hz}\frac{(hz-y)^{r-1}}{(r-1)!}\left(\widehat{m}_g^{(r)}(x_0+y) - m^{(r)}(x_0+y)\right)dy\,dz + O_P\left(n^{-2}g^{-1}\right)$$

holds. Since $h \ll g$ holds for h and g of optimal orders, the task of estimating B_h is nearly equivalent to the estimation of $m^{(r)}(x_0)$ by $\widehat{m}_g^{(r)}(x_0)$. Because an optimal local bandwidth $g_{opt}(x_0)$ seems to be difficult to adapt, we intend to use an estimator of the optimal global bandwidth $g_0^{(r,s)}$ in this case. Müller and Stadtmüller (1987) [13] provided a nice idea to adapt such a quantity, which can be applied also in the case of heteroscedasticity. They observed that

$$g_0^{(r,s)} = C_{r,s}(w)C(m,v)n^{-\frac{1}{2(r+s)+1}}(1+o(1))$$

holds with some constant $C_{r,s}$ that depends on the kernel function w but not on the unknown functions m and v. On the other hand, the optimal global bandwidth for an estimator of m itself with an $(r+s)$-th order kernel $\overline{w}$ has the form

$$g_0^{(0,r+s)} = C_{0,r+s}(\overline{w})C(m,v)n^{-\frac{1}{2(r+s)+1}}(1+o(1)).$$

Now we can estimate $g_0^{(0,r+s)}$ by some asymptotically optimal bandwidth $\widehat{g_0^{(0,r+s)}}$, and then we obtain a consistent estimate of $g_0^{(r,s)}$ by

$$\widehat{g} = \frac{C_{r,s}(w)}{C_{0,r+s}(\overline{w})}\widehat{g_0^{(0,r+s)}}, \tag{2.6}$$

what spares us the more involved direct estimation of the constant $C(m,v)$.

3. Confidence intervals for $m(x_0)$

3.1. Construction principles for confidence intervals. All commonly used methods to establish confidence intervals are based on the principle to estimate first $m(x_0)$ by an initial estimator $\widehat{m}(x_0)$ and to estimate then the distribution of $\widehat{m}(x_0) - m(x_0)$. It will be distinguished between pivotal and nonpivotal methods. For the related problem of bootstrap confidence intervals in density estimation Hall (1992) [10] pointed out that pivotal methods, which are based on a quantity $\widehat{V}^{-1/2}(\widehat{\vartheta}-\vartheta)$ that contains an estimator $\widehat{V}$ of the variance of $\widehat{\vartheta}$, should be preferred to nonpivotal methods, which are simply based on an estimation of the distribution of $(\widehat{\vartheta}-\vartheta)$. Heuristically, this superiority seems to transfer to intervals based on Edgeworth expansions. If one reconstructs the unknown distribution, then one has to estimate the first cumulants of the quantity under consideration. In case of nonpivotal methods one estimates the cumulants of $(\widehat{\vartheta}-\vartheta)$ once, whereas in case of pivotal methods bias and variance of $\widehat{\vartheta}$ are already estimated in a first step and one can easily take the influence of these estimates into account in the distribution recognition step. In the present paper we restrict ourselves to pivotal methods.

The main problem with confidence intervals in nonparametric regression rests on the fact that a consistent estimator of $m(x_0)$ is necessarily biased. Strictly speaking, $MISE$-optimal estimators have bias and standard deviation of the same order. There are two common methods to deal with this problem, undersmoothing and subsequent bias correction. Hall (1991 and 1992) [8], [9] shows in situations closely related to ours that the first method leads to a better asymptotic coverage accuracy, at least in the case of nonrandom bandwidths.

An important goal of the present paper is, to provide methods, where all bandwidths are chosen by the data in a reasonable way. The only available guideline for a favourable choice seems to be the risk behaviour of the corresponding estimators and, hence, the bandwidths we deal with are of $MISE$-optimal order, which means that bias and standard deviation of the estimator can be expected to be of the same order. Therefore, we cannot apply the undersmoothing method in its pure form.

In contrast, it is possible to construct a bias corrected pivotal quantity on the basis of $MISE$-optimal kernel estimators by a normalization of the initial estimator with estimates of its bias and variance. Now it seems to be more natural to estimate the bias first and to divide then the corrected quantity by an estimator of its standard deviation.

It turns out that this method is equivalent to undersmoothing and, in accordance to the existing theory, we obtain a better coverage accuracy as by the first method.

3.2. Asymptotic distributions of the pivotal quantities. As already announced, we consider the simple bias-corrected pivotal quantity

$$(3.1) \qquad T_{h,g} = \frac{\widehat{m}_h(x_0) - \widehat{B}_{h,g} - m(x_0)}{\widehat{V}_h^{1/2}} = \frac{\sum \overline{W}_{h,g,i}\varepsilon_i + b_{h,g}}{\widehat{V}_h^{1/2}}$$

where $\widehat{V}_h$ and $\widehat{B}_{h,g}$ are defined by (2.1) and (2.3), respectively, and $b_{h,g} = \sum \overline{W}_{h,g,i}m(x_i) - m(x_0)$ denotes the remaining bias. Further, we obtain a method equivalent to undersmoothing by estimating the whole variance of the numerator of $T_{h,g}$ instead of that of $\widehat{m}_h(x_0)$. In this case the usual condition $h \ll g$, which is introduced to keep the variance of the bias estimator of smaller order than that of $\widehat{m}_h(x_0)$, is no longer necessary and we optimize g with respect to the asymptotic coverage accuracy by choosing it of the same oder as h. Since there is no other guideline for doing this in practice, we set simply $g = h$, where h will be chosen later by some data-dependent rule.
Note that the roles of $\widehat{B}_{h,g}$ and $\widehat{B}_{h,h}$ are very different. Whereas the quantity $\widehat{B}_{h,g}$ in $T_{h,g}$ *estimates* the bias, the term $\widehat{B}_{h,h}$ in U_h reduces only the non-stochastic part of $\widehat{m}_h(x_0)$, which provides a new estimator $\widehat{m}_h(x_0) - \widehat{B}_{h,h}$ with a squared bias of smaller order than its variance.
With $\overline{W}_{h,i} = \overline{W}_{h,h,i}$ and $b_h = b_{h,h}$ we get the pivotal quantity

$$(3.2) \qquad U_h = \frac{\sum \overline{W}_{h,i}\varepsilon_i + b_h}{\widehat{\widehat{V}}_h^{1/2}},$$

where $\widehat{\widehat{V}}_h = \sum \overline{W}_{h,i}^2 \widehat{v}_i$. To obtain knowledge about the asymptotic distributions of the pivotal quantities we intend to apply Edgeworth expansions as far as possible. For that we approximate the quantities of interest by certain smooth functions of random vectors. Using results of Skovgaard (1981 and 1986) [17], [18] we can prove then the validity of these (formal) expansions. To draw conclusions from the size of the difference of two random variables to the difference of their cumulative distribution functions in a convenient way we introduce the following notation.

Definition 3.1. Let $\{Y_n\}$ and $\{Z_n\}$ $(Z_n \geq 0$ *a.s.*$)$ be sequences of random variables, and let $\{\gamma_n\}$ be a sequence of positive reals. We write

$$Y_n = \widetilde{O}(Z_n, \gamma_n)$$

if

$$P(|Y_n| > CZ_n) \leq C\gamma_n$$

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

This notion differs obviously from the usual O_p, which would provide a similar property for an arbitrary constant γ instead of $C\gamma_n$ on the right-hand side. As a rule, for arbitrary $\delta,\ \lambda > 0$ we can conclude under sufficiently strong moment conditions on the distributions of the errors by Markov's and Whittle's inequalities that

$$(3.3) \qquad (a_n)'\underline{\varepsilon} = \widetilde{O}(n^{\delta}\,\|a_n\|,\, n^{-\lambda})$$

and

$$\underline{\varepsilon}' A_n \underline{\varepsilon} - E\underline{\varepsilon}' A_n \underline{\varepsilon} \;=\; \tilde{O}\left(n^{\delta}\sqrt{tr(A_n A_n')}\,,n^{-\lambda}\right) \tag{3.4}$$

hold uniformly over $a_n \in \mathbb{R}^n$ and arbitrary $(n \times n)$-matrices A_n, where $\underline{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)'$. Furthermore, we obtain similar assertions for random quantities a_n and A_n, which is made rigorous by Lemma 6.1 in the Appendix.
The following lemma shows that $\tilde{O}$ is an appropriate concept for the calculation of the cumulative distribution function of quantities that do not immediately admit an Edgeworth expansion.

Lemma 3.1. *Let $\{X_n\}$ be a sequence of random variables that admit the Edgeworth expansion*

$$P(X_n < t) \;=\; \Phi(t) \;+\; p_n(t)\phi(t) \;+\; O(u_n)$$

with some polynomials p_n of bounded order with bounded coefficients. Further, we assume $Y_n = \tilde{O}(\gamma_{n1}, \gamma_{n2})$. Then

$$P(X_n + Y_n < t) = P(X_n < t) + O(u_n + \gamma_{n1} + \gamma_{n2}).$$

The proof of this lemma follows immediately from the inequalities

$$P(X_n < t - C\gamma_{n1}) - P(|Y_n| < C\gamma_{n1}) \le P(X_n + Y_n < t) \le P(X_n < t + C\gamma_{n1}) + P(|Y_n| < C\gamma_{n1})$$

and the Lipschitz equicontinuity of the functions $\Phi(t) + p_n(t)\phi(t)$.

4. Coverage accuracy of the confidence intervals

4.1. Coverage accuracy in case of nonrandom bandwidths. First, we approximate the cumulative distribution functions of the pivotal quantities with *nonrandom* bandwidths via Edgeworth expansions. The following proposition serves then as a starting point to derive formulas for quantities with data-driven bandwidths.

Proposition 4.1. *Assume (A_G), (A_E) and $h = h(n)$ and $g = g(n)$ to be nonrandom.*

(i) *If $nh \to \infty$, $g \to 0$ and $h/g \to 0$ as $n \to \infty$, then*

$$\begin{aligned} P(T_{h,g} < t) \;&=\; \Phi(t) - \frac{b_{h,g}}{V_h^{1/2}}\,\phi(t) + \rho_n\,\frac{2t^2+1}{6}\,\phi(t) \\ &\quad - \frac{1}{2}\,\frac{\overline{V}_{h,g} - V_h}{V_h}\,t\phi(t) + O(g^{2s} + (h/g)^{2(r+1)} + (nh)^{-1}) \end{aligned}$$

holds uniformly over each compact set $t \in \mathcal{T}$, where
$\rho_n = V_h^{-3/2}\sum_i W(x_0,h)_i^3\, E\varepsilon_i^3$ and $\overline{V}_{h,g} = \sum_i \overline{W}^2_{h,g,i}\, v(x_i)$.

(ii) *If $nh \to \infty$ and $h \to 0$ as $n \to \infty$, then*

$$P(U_h < t) = \Phi(t) - \frac{b_h}{\overline{V}^{1/2}_{h,h}}\,\phi(t) + \overline{\rho}_n\,\frac{2t^2+1}{6}\,\phi(t) + O(h^{2s} + (nh)^{-1})$$

holds uniformly over each compact set $t \in \mathcal{T}$, where $\overline{\rho}_n = \overline{V}_{h,h}^{-3/2}\sum \overline{W}^3_{h,i}\, E\,\varepsilon_i^3$.

The proof of this proposition is essentially the same as that of Proposition 3.2 in [15] and may be sketched, w.l.o.g. for (i), as follows. First we approximate $T_{h,g}$ by

$$\tilde{T}_{h,g} = \left(\sum W(x_0,h)_i^2\,\varepsilon_i^2\right)^{-1/2}\left(\sum \overline{W}_{h,g,i}\varepsilon_i + b_{h,g}\right).$$

Be Lemma 3.2 in [15] we have

$$T_{h,g} - \widetilde{T}_{h,g} = \widetilde{O}\left((nh)^{-1}, n^{-1}\right). \tag{4.1}$$

The vector $S_n = \left(\sum_i \overline{W}_{h,g,i}\varepsilon_i, \sum_i W(x_0,h)_i^2\varepsilon_i^2\right)'$ is a sum of independent random vectors and admits in accordance to results of Skovgaard (1986) an Edgeworth expansion with a residual term of order $O\left((nh)^{-1-\delta}\right)$ for some $\delta > 0$. Since $\widetilde{T}_{h,g}$ is a smooth function of S_n, we infer from Theorem 3.2 and Remark 3.4 in Skovgaard (1981) the validity of the formal Edgeworth expansion of $\widetilde{T}_{h,g}$. To identify the expansion, we must calculate the cumulants of $\widetilde{T}_{h,g}$, which has already been done in [15]. By Lemma 3.1 we conclude from (4.1) that the expansions of $T_{h,g}$ and $\widetilde{T}_{h,g}$ are identical up to a term of order $O((nh)^{-1})$, which completes the proof.

For the rest of this subsection we assume that the nonrandom bandwidths h and g are chosen of the same order as $\widehat{h}$ and $\widehat{g}$ described above, namely $h \asymp n^{-1/(2r+1)}$ and $g \asymp n^{-1/(2(r+s)+1)}$. Now it is easy to see that

$$\begin{aligned} b_{h,g} &= O(h^r g^s),\ b_h = O(h^{r+s}), \\ V_h,\ \overline{V}_{h,h} &\asymp (nh)^{-1/2}, \\ \rho_n,\ \overline{\rho}_n &= O((nh)^{-1/2}) \\ \text{and} \quad \frac{\overline{V}_{h,g} - V_h}{V_h} &= O((h/g)^{r+1}). \end{aligned}$$

If $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution then we obtain

$$\begin{aligned} & P\left(m(x_0) \in (\widehat{m}_h(x_0) - \widehat{B}_{h,g} - u_{1-\alpha}\widehat{V}_h^{1/2}, \infty)\right) \\ &= P(T_{h,g} < u_{1-\alpha}) \\ &= 1 - \alpha + O(g^s + (nh)^{-1/2}) \\ &= 1 - \alpha + O\left(n^{-\frac{s}{2(r+s)+1}} + n^{-\frac{r}{2r+1}}\right) \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} & P\left(m(x_0) \in (\widehat{m}_h(x_0) - \widehat{B}_{h,h} - u_{1-\alpha}\widehat{\widehat{V}}_h^{1/2}, \infty)\right) \\ &= 1 - \alpha + O(h^s + (nh)^{-1/2}) \\ &= 1 - \alpha + O\left(n^{-\frac{s\wedge r}{2r+1}}\right) \end{aligned} \tag{4.3}$$

Estimating ρ_n and $\overline{\rho}_n$ by $\widehat{\rho}_n = \widehat{V}_h^{-3/2}\sum W(x_0,h)_i^3\widehat{v}_i^3$ and $\widehat{\overline{\rho}}_n = \widehat{\widehat{V}}_h^{-3/2}\sum \overline{W}_{h,h,i}^3\widehat{v}_i^3$, respectively, and inverting the expansions from Proposition 4.1 we obtain confidence intervals

$$I_{h,g} = \left(\widehat{m}_h(x_0) - \widehat{B}_{h,g} - (1 + \frac{2u_{1-\alpha}^2 + 1}{6}\widehat{\rho}_h)u_{1-\alpha}\widehat{V}_h^{1/2}, \infty\right)$$

and

$$\widetilde{I}_h = \left(\widehat{m}_h(x_0) - \widehat{B}_{h,h} - (1 + \frac{2u_{1-\alpha}^2 + 1}{6}\widehat{\overline{\rho}}_h)u_{1-\alpha}\widehat{\widehat{V}}_h^{1/2}, \infty\right)$$

with the coverage probabilities

$$\begin{aligned} P(m(x_0) \in I_{h,g}) &= 1 - \alpha + O(g^s + (h/g)^{r+1} + (nh)^{-1}) \\ &= 1 - \alpha + O\left(n^{-\frac{s}{2(r+s)+1}}\right) \end{aligned} \tag{4.4}$$

and

$$P\left(m(x_0) \in \tilde{I}_h\right) = 1 - \alpha + O(h^s + (nh)^{-1}) = 1 - \alpha + O\left(n^{-\frac{s \wedge 2r}{2r+1}}\right). \tag{4.5}$$

The equations (4.4) and (4.5) are also proved in [15].

4.2. The effect of the bandwidth choice to the coverage accuracy. Now we are going to consider the performance of confidence intervals in practical situations, i.e. intervals based on pivotal quantities $T_{\widehat{h},\widehat{g}}$ and $U_{\widehat{h}}$ involving estimators with data-driven bandwidths. These quantities do not immediately admit Edgeworth expansions, because they cannot be written as smooth functions of a sum of independent random vectors. We will use the approximations to $T_{h,g}$ and U_h given by Proposition 4.1 and treat the differences between $T_{\widehat{h},\widehat{g}}$ and $T_{h,g}$ as well as between $U_{\widehat{h}}$ and U_h by estimates based on $\tilde{O}$.
First, we consider the order of approximation of the optimal bandwidth by their estimates considered in Section 2. From here let $\delta > 0$ be an arbitrarily small quantity, whose occurrence is explained by the application of Lemma 6.1.

Lemma 4.1. *Under* (A_G) *and* (A_{BW}) *we have*

$$\frac{\widehat{h} - h_0}{h_0} = \tilde{O}\left(n^\delta n^{-\Delta}, n^{-1}\right),$$

where $\Delta = 1/2 \wedge s/(s + 2r + 1/2)$.

On the basis of Lemma 6.1 it is easy to see that

$$\begin{aligned} \frac{d}{dg}\{T_{\widehat{h},g}\} &= \tilde{O}(n^\delta g^{s-1}, n^{-1}) \\ \frac{d}{dh}\{T_{h,g}\} &= \tilde{O}(n^\delta h^{-1}, n^{-1}) \\ &\text{and} \\ \frac{d}{dh}\{\widehat{\rho}_h\} &= \tilde{O}(n^\delta h^{-1} (nh)^{-1/2}, n^{-1}). \end{aligned}$$

From the decomposition

$$\begin{aligned} T_{\widehat{h},\widehat{g}} - \frac{2u_{1-\alpha}^2 + 1}{6}\widehat{\rho}_{\widehat{h}} &= \left(T_{h,g} - \frac{2u_{1-\alpha}^2 + 1}{6}\widehat{\rho}_h\right) + \frac{2u_{1-\alpha}^2 + 1}{6}(\widehat{\rho}_h - \widehat{\rho}_{\widehat{h}}) \\ &\quad + (T_{\widehat{h},\widehat{g}} - T_{\widehat{h},g}) + (T_{\widehat{h},g} - T_{h,g}) \end{aligned}$$

we obtain by the Lemmas 3.1 and 4.1 and by (4.4) the following assertion.

Theorem 4.1. *: Assume* (A_G), (A_U), (A_{BW}), (A_E) *and* $|\widehat{g} - g_0|/g_0 = \tilde{O}(n^{-\gamma}, n^{-1})$ *for some* $\gamma > 0$. *Then*

$$\begin{aligned} P(m(x_0) \in I_{\widehat{h},\widehat{g}}) &= P(m(x_0) \in I_{h_0,g_0}) + O\left(n^\delta n^{-\Delta}\right) \\ &= 1 - \alpha + O(n^{-\frac{s}{2(r+s)+1}}). \end{aligned}$$

For confidence intervals based on the pivotal statistic $U_{\widehat{h}}$ we can derive in an analogous way estimates for the error in coverage probability. However, since U_h yields for nonrandom h better rates than $T_{h,g}$, the error due to the randomness of $\widehat{h}$ is not automatically majorized by the error in coverage probability of the confidence interval with nonrandom bandwidths. Hence, we look for a better approximation to $\widehat{h}$. The idea is quite simple. Neglecting the effect of the estimator $\widehat{m}_{\widehat{f}}$ involved in the $\widehat{v}_i$'s, the pivotal statistic U_h depends only on $O(nh)$ of the n observations, whereas the bandwidth selector uses all of them to a certain amount. We define another bandwidth $\widetilde{h}$ by a similar criterion, where only the observations in some neighborhood of x_0 of size $O(h_0)$ are excluded, such that the quantity $U_{\widetilde{h}}$ is based on a set of observations disjoint from that used for the choice of $\widetilde{h}$. Then the conditional distribution of $U_{\widetilde{h}}$ under $\widetilde{h}$ is the same as the unconditional distribution of U_h at the point $h = \widetilde{h}$. Thus Proposition 4.1 remains valid for $U_{\widetilde{h}}$ as well, and because $\widetilde{h}$ approximates $\widehat{h}$ better than h_0, we obtain a better estimate for the error in coverage probability as via an approximation to $U_{\widehat{h}}$ by U_{h_0}.
Let, for appropriate C,

$$\Delta_n = Cn^{-\frac{1}{2r+1}},$$

$$J_n = \{i \in \{1, ..., n\} \quad | \, |x_i - x_0| \leq \Delta_n\}$$

We replace $\widehat{MISE}(h, \overline{g})$ by

$$\widetilde{M}(h) = \widetilde{IV}(h) + \widetilde{ISB}(h), \tag{4.6}$$

where $\widetilde{IV}(h) = \sum_{i \notin J_n} \int_c^d W(x,h)_i^2 \, dx \, \varepsilon_i^2 + \sum_{i \in J_n} \int_c^d W(x,h)_i^2 \, dx \, v(x_i)$ and $\widetilde{ISB}(h) = \sum_{i,j \notin J_n} A(h, g_0)_{ij} Y_i Y_j + \sum_{(i,j): i \in J_n or j \in J_n} A(h, g_0)_{ij} E\, Y_i Y_j - \sum A(h, g_0)_{ii} v(x_i)$ and define $\widetilde{h}$ as a measurable function with

$$\widetilde{h} \in argmin_{h \in [h_0/2, 3h_0/2]} \widetilde{M}(h)$$

Let the constant C be chosen so large that $U_{\widetilde{h}}$ and $\widetilde{M}(h)$ are based on disjoint sets of observations.
Now we can prove analogously to Lemma 4.1 that

$$\frac{|\widehat{h} - \widetilde{h}|}{h_0} = \widetilde{O}\left(n^{\delta} n^{-\Delta'}, n^{-1}\right), \tag{4.7}$$

where $\Delta' = (1/2 + 1/(2(2r+1))) \wedge s/(s + 2r + 1/2)$.
The additional factor $n^{-1/2(2r+1)}$ comes into play, because, roughly speaking, the number of the Y_i's included in $\widetilde{M}(h) - \widehat{MISE}(h, g_0)$ is $O(n\Delta_n)$ rather than $O(n)$ as in $\widehat{MISE}(h, \overline{g})$.
Again by Lemma 6.1 we obtain

$$\frac{d}{dh}\{U_h\} = \widetilde{O}(n^{\delta} h^{-1}, n^{-1}) \quad \text{and} \quad \frac{d}{dh}\{\widehat{\overline{\rho}}_h\} = \widetilde{O}(n^{\delta} h^{-1} (nh)^{-1/2}, n^{-1}).$$

With the decomposition

$$U_{\widehat{h}} - \frac{2u_{1-\alpha}^2 + 1}{6} \widehat{\overline{\rho}}_{\widehat{h}} = (U_{\widetilde{h}} - \frac{2u_{1-\alpha}^2 + 1}{6} \widehat{\overline{\rho}}_{\widetilde{h}}) + \frac{2u_{1-\alpha}^2 + 1}{6} (\widehat{\overline{\rho}}_{\widetilde{h}} - \widehat{\overline{\rho}}_{\widehat{h}}) + (U_{\widehat{h}} - U_{\widetilde{h}})$$

we obtain by (ii) of Proposition 4.1, (4.7) and Lemma 3.1 the following theorem.

Theorem 4.2. *Under* (A_G), (A_U), (A_{BW}) *and* (A_E) *we have*

$$\begin{aligned} P(m(x_0) \in \widetilde{I}_{\widehat{h}}) &= P(m(x_0) \in \widetilde{I}_{\widetilde{h}}) + O\left(n^{\delta}\, n^{-\Delta'}\right) \\ &= 1 - \alpha + O\left(n^{-\frac{s\wedge 2r}{2r+1}} + n^{\delta}\, n^{-\Delta'}\right). \end{aligned}$$

5. Discussion

1) Comparing the results of the Theorems 4.1 and 4.2 we see that the undersmoothing method retains its superiority to the explicit bias-correction also in the case of data-dependent bandwidths chosen by the above criteria.
2) Our estimates via the $\widetilde{O}$-calculations seem to be on first sight somewhat rough and there arises the question whether we would obtain better estimates by formal Edgeworth expansions of the pivotal quantities $T_{\widehat{h},\widehat{g}}$ and $U_{\widehat{h}}$, respectively. Apart from the fact that the validity of these expansions is not immediately clear, it turns out that we would obtain the same rates as given by the Theorems 4.1 and 4.2 with exception of the factor n^{δ}. To see this, expand $U_{\widehat{h}}$ in the Taylor series
$$U_{\widehat{h}} = U_{\widetilde{h}} + (\widehat{h} - \widetilde{h})\, U_h'|_{h=\widetilde{h}} + \frac{(\widehat{h} - \widetilde{h})^2}{2}\, U_h''|_{h=h^*},$$
where h^* is between $\widetilde{h}$ and $\widehat{h}$. The third term on the right-hand side is of negligible order. All arguments can be conditioned on $\widetilde{h}$, since the conditional distribution of $U_{\widetilde{h}}$ is equal to the unconditional distribution of U_h at the point $h = \widetilde{h}$. If we follow the proof of (4.7), we see that the leading term of order $h_0 n^{-\Delta'}$ of $\widehat{h} - \widetilde{h}$ is given by
$$\frac{\sum_{i \in J_n} W_i(\varepsilon_i^2 - v(x_i)) + \sum_{(i,j): i \in J_n \text{ or } j \in J_n} A(h, g_0)_{i,j}(m(x_i)\varepsilon_j + \varepsilon_i m(x_j))}{-\widetilde{M}''(\widetilde{h})}.$$
On the other hand, we have
$$U_h' = \frac{\widehat{m}_h'(x_0)}{\widehat{V}_h^{1/2}} - \frac{\widehat{m}_h(x_0)\widehat{V}_h'}{2\widehat{V}_h^{3/2}},$$
which depends mainly on Y_i's with $i \in J_n$ and has an order of magnitude of $O(h_0^{-1})$. Therefore, the second term of the above Taylor series contributes a term of order $n^{-\Delta'}$ to the first cumulant of $U_{\widehat{h}}$, which leads to a difference of order at least $n^{-\Delta'}$ between the Edgeworth expansions of $U_{\widetilde{h}}$ and $U_{\widehat{h}}$.
3) One disappointing fact with confidence intervals in nonparametric regression is that we cannot obtain a size of the intervals that shrinks with the same rate as the standard deviation of optimal estimators. The reason is that we have actually two more or less separate problems, the estimation of $m(x_0)$ by some estimator $\widehat{m}(x_0)$ as well as the recognition of the distribution of $\widehat{m}(x_0) - m(x_0)$, which essentially consists in the estimation of the bias $E\,\widehat{m}(x_0) - m(x_0)$. To solve both problems satisfactorily, we have to apportion the smoothness assumption for both purposes, which requires the application of an suboptimal estimator $\widehat{m}(x_0)$.
4) The methods used in the present paper can obviously be applied to kernel estimators with bandwidths chosen by other selectors. In Neumann (1992) [14] it is shown that the cross-validation bandwidth $\widehat{h}_{CV}$ can be approximated by some random bandwidth $\widetilde{h}$, which is independent of those observations that enter into the estimator $\widehat{m}_{\widetilde{h}}(x_0)$, to an order of $\widetilde{O}\left(h_0 n^{-1/(2r+1)} n^{\delta}, n^{-\lambda}\right)$, which yields finally

an error in coverage probability of $O\left(n^{\delta} n^{-1/(2r+1)}\right)$. Another direction for an extension are alternative kernel estimators, as e.g. those of Nadaraya-Watson type.

5) Theoretically, it seems to be possible to take the variability of $U_{\widehat{h}}$ due to the randomness of the bandwidth $\widehat{h}$ into account by an Edgeworth expansion for the quantity

$$U_{\widetilde{h}} + (\widehat{h} - \widetilde{h})\, U_h'\big|_{h=\widetilde{h}} ,$$

which approximates $U_{\widehat{h}}$ better than $U_{\widetilde{h}}$. If we approximate $\widehat{h} - \widetilde{h}$ also by a Taylor series in linear and quadratic forms in $\underline{\varepsilon}$, then we can derive by results of Skovgaard (1981) cited above an higher order Edgeworth expansion of $U_{\widehat{h}}$ that can serve as a starting point for the choice of an appropriate quantile.
On the other hand, it is not clear at all whether this method does pay off for moderate sample sizes that can be expected to occur in practical situations. Moreover, the estimation of the terms of the abovementioned Edgeworth expansion involves the estimation of terms like $\widetilde{M}''(h)$, which seems to be challenging from the practical point of view.

6. Appendix

6.1. Assumptions. Here we list the assumptions needed for the assertions in the previous sections.

1) General assumptions (A_G)
   - The design points $x_i = x_i(n)$ are regularly spaced, i.e. $\int_0^{x_i} d(t)\, dt = (i-1/2)/n$, for some positive, continuous density d on $[0,1]$.
   - $w \in C^0[-1,1]$ is a kernel function of order $r \geq 2$
   - $\widetilde{w} \in C^r[-1,1]$ is a kernel function of order $s \geq 2$
   - $m \in C^{r+s}[0,1]$

2) Assumption for uniform approximations (A_U)
   - All moments of the ε_i's are uniformly bounded. (If we assume instead that only a finite number of moments are bounded, then we have to choose δ in dependence on this number and on the entropy of the families of vectors and matrices as indicated in the proof of Lemma 6.1.)

3) Assumptions especially for the choice of the optimal global bandwidth (A_{BW})
   - $m \in C^{2r+s}[0,1]$, $\int_c^d \left(m^{(r)}(x)\right)^2 dx \neq 0$,
   $\int_c^d \left[\int \frac{1}{h} w_{x,h}\left(\frac{z-x}{h}\right) m(z)\, dz \;-\; m(x)\right]^2 dx \neq 0$ for all $h > 0$
   - $v \in C^0[0,1]$ is bounded from zero

4) Assumptions for Edgeworth expansions (A_E)
   - A sufficiently large number of moments of the ε_i's are uniformly bounded.
   - Cramér's condition is uniformly satisfied by the random vectors $\alpha_i = (\varepsilon_i, \varepsilon_i^2, \varepsilon_i^3)'$ in some neighborhood of x_0, i.e.

$$\sup_{i:|x_i - x_0| \leq C} \; \sup_{\|t\| > b} \left| E \exp\{it'\alpha_i\}\right| < 1$$

   for some $C > 0$ and all $b > 0$.

6.2. Some technical lemmas.

Lemma 6.1. *(uniform $\widetilde{O}$-approximation)*
Let $\mathcal{A}^n = \{a_\theta^{(n)}\}_{\theta\in\Theta}$ and $\mathcal{A}^{n\times n} = \{A_\theta^{(n)}\}_{\theta\in\Theta}$ be families of n-vectors and $(n \times n)$-matrices, respectively. Further, define the ϵ-entropy $E_\epsilon(\mathcal{A}^{n\times n})$ of $\mathcal{A}^{n\times n}$, as the minimal

number of $(n\times n)$*-matrices* A_i *with the property that each* $A \in \mathcal{A}^{n\times n}$ *can be approximated by some* A_i *with* $\|A - A_i\| \le \epsilon$. *Analogously we define the* ϵ *-entropy* $E_\epsilon(\mathcal{A}^n)$ *of* $\mathcal{A}^n$. *Assume* $(A_U), E_{n^{-1/2-\beta}}(\mathcal{A}^n) = O(n^\lambda)$ *and* $E_{n^{-1-\beta}}(\mathcal{A}^{n\times n}) = O(n^\lambda)$ *for some* $\beta > 0, \lambda < \infty$. *Then*
(i) $\sup_{\theta\in\Theta}\{(\|a_\theta^{(n)}\| + n^{-\beta})^{-1}|a_\theta^{(n)'}\underline{\varepsilon}|\} = \tilde{O}(n^\delta, n^{-\gamma})$,
(ii) $\sup_{\theta\in\Theta}\{(\sqrt{tr(A_\theta^{(n)}A_\theta^{(n)'})} + n^{-\beta})^{-1}|\underline{\varepsilon}'A_\theta^{(n)}\underline{\varepsilon} - E\underline{\varepsilon}'A_\theta^{(n)}\underline{\varepsilon}|\} = \tilde{O}(n^\delta, n^{-\gamma})$
holds for appropriate $\delta > 0$ *and* $\gamma < \infty$, *which can be chosen arbitrarily small and large, respectively, if all moments of the* ε_i*'s are uniformly bounded.*

Proof. For a one-element set $\Theta = \{\theta_0\}$ we obtain (i) and (ii) by Markov's and Whittle's inequalities, see Whittle (1960) [19]. For general Θ we derive (i) and (ii) on the basis of that set of vectors and matrices, just given by the definition of the $n^{-1/2-\beta}$-entropy and $n^{-1-\beta}$- entropy, respectively. Let $\widehat{\theta}$ denote this parameter from the approximating grid with $\|a_\theta^{(n)} - a_{\widehat{\theta}}^{(n)}\| \le n^{-1/2-\beta}$. By Markov's, Whittle's and Bonferroni's inequalities we obtain that, for appropriate positive δ and γ,

$$\begin{aligned} \|(a_\theta^{(n)})'\underline{\varepsilon}\| &\le \|(a_{\widehat{\theta}}^{(n)})'\underline{\varepsilon}\| + \|a_\theta^{(n)} - a_{\widehat{\theta}}^{(n)}\|\,\|\underline{\varepsilon}\| \\ &= O\left(n^\delta\|a_{\widehat{\theta}}^{(n)}\| + n^{-1/2-\beta}n^{1/2+\delta}\right) \\ &= O\left(n^\delta\|a_\theta^{(n)}\| + n^\delta\, n^{-\beta}\right) \end{aligned}$$

holds uniformly over $\theta \in \Theta$ with a probability exceeding $1 - O(n^{-\gamma})$, which implies (i). (ii) can be proved analogously. □

The next lemma improves the residual term of order n^{-1} given in [3] for the expectation of Gasser-Müller kernel estimators.

Lemma 6.2. *Let* $w_{x,h}$ *be uniformly (in* x *and* h *) Lipschitz continuous of order 1 and let* $\{g_n\}$ *be a sequence of twice differentiable functions. Further assume that the design satisfies the condition given in* (A_G). *Then*

$$\begin{aligned} \sum_{j=1}^n W(x,h)_j g_n(x_j) &= \int_0^1 \frac{1}{h} w_{x,h}\left(\frac{z-x}{h}\right) g_n(z)dz \\ &\quad + O\left(n^{-2}h^{-1}\sup_{0\le z\le 1}\{|g_n'(z)|\} + n^{-2}\sup_{0\le z\le 1}\{|g_n''(z)|\}\right). \end{aligned}$$

The proof of this lemma is straightforward and therefore omitted.

6.3. Proofs.

Proof of Lemma 2.1. The calculations are very similar to these in the proof of Theorem 1 in [HHM91]. Therefore we indicate only the sources of the terms in (i). Some of these formulas will be used in the course of the proof of Lemma 4.1. First, we approximate the entries of the matrix $A(h,g)$ by

$$\begin{aligned} A_{ij} &= \frac{h^{2r}}{g^{2r+1}}(s_i - s_{i-1})(s_j - s_{j-1})\int \widetilde{w}^{(r)}(x)\widetilde{w}^{(r)}(x + \frac{x_i - x_j}{g})\,dx \\ &\quad + o\left(\frac{h^{2r}}{g^{2r+1}}n^{-2}\right) + O\left(n^{-4}h^{-4}g\right) \qquad \text{if } |x_i - x_j| \le Cg, \end{aligned} \tag{6.1}$$

whereas $A_{ij} = 0$ holds if $|x_i - x_j| > Cg$. Further, we have by $m \in C^{2r+s}[0,1]$

(6.2)
$$(A(h,g)m)_i = C(h,i)n^{-1}h^{2r} + O(n^{-1}h^{2r}g^s) + O(n^{-3}h^{r-1}g^{-r-1} + n^{-3}h^{r-3}).$$

Now we split up

$$(6.3)\, Var\left(\widetilde{ISB}(h,g)\right) = Var\,(\underline{\varepsilon}'A(h,g)\underline{\varepsilon}) + 4\,Var\,(\underline{m}'A(h,g)\underline{\varepsilon}) + 4\,Cov\,(\underline{\varepsilon}'A(h,g)\underline{\varepsilon}, \underline{m}'A(h,g)\underline{\varepsilon}),$$

where $\underline{m} = (m(x_1), \ldots, m(x_n))'$, and estimate the terms on the right-hand side separately. Those terms, which enter into the formula (i) are underlined. We have

$$\begin{aligned} Var\,(\underline{\varepsilon}'A(h,g)\underline{\varepsilon}) &= 2\sum_{i,j} A_{ij}^2 v(x_i)v(x_j)\,(1 + o(1)) \\ (6.4) \quad &= \underline{2\frac{h^{4r}}{g^{4r+1}}n^{-2}\kappa_r^4 \int_0^1 \left(\frac{v(x)}{d(x)}\right)^2 dx \int \left(\int \widetilde{w}^{(r)}(y)\,\widetilde{w}^{(r)}(y+z)\,dy\right)^2 dz}\,(1 + o(1)). \end{aligned}$$

Further, we obtain by (6.2)

$$\begin{aligned} Var\,(\underline{m}'A(h,g)\underline{\varepsilon}) &= \underline{m}'A(h,g)Diag[v(x_1), \ldots, v(x_n)]A(h,g)\underline{m} \\ (6.5) \quad &= \underline{h^{4r}C_3(h)n^{-1}} + O\left(h^{4r}n^{-1}g^s\right) \end{aligned}$$

and, by (6.1),

$$(6.6) \qquad Cov\,(\underline{\varepsilon}'A(h,g)\underline{\varepsilon}, \underline{m}'A(h,g)\underline{\varepsilon}) = O\left(h^{4r}n^{-2}g^{-2r-1}\right),$$

where the residual terms will be both majorized by the underlined term in (6.4). By

$$\begin{aligned} Var\left(\widehat{ISB}(h,g)\right) &= Var\left(\widetilde{ISB}(h,g)\right) \\ &\quad + O\left(\sqrt{Var(\widetilde{ISB}(h,g))}\sqrt{Var(\sum A_{ii}\widehat{v}_i)} + Var\left(\sum A_{ii}\widehat{v}_i\right)\right) \end{aligned}$$

and

$$Var\left(\sum A_{ii}\widehat{v}_i\right) \le E\left(\sum A_{ii}(\widehat{v}_i - v_i)\right)^2 = O\left(h^{4r}n^{-3}g^{-4r-2}\right)$$

we see that the remaining residual terms do not enter into the asymptotic formula. Finally, we have

$$\begin{aligned} E\widehat{ISB}(h,g) - ISB(h) &= \underline{m}'A(h,g)\underline{m} - ISB(h) + E\sum A_{ii}(v_i - \widehat{v}_i) \\ (6.7) \quad &= \underline{2h^{2r}g^s\kappa_r^2\lambda_s \int m^{(r+s)}(x)m^{(r)}(x)\,dx}\,(1 + o(1)) \\ &\quad + O(n^{-2}) + O\left(h^{2r}n^{-3/2}g^{-2r-1}\right), \end{aligned}$$

which completes the calculations needed for the proof of (i). □

Proof of Lemma 4.1. First, we investigate how well $MISE(h)$ is approximated by its estimate $\widehat{MISE}(h,\overline{g})$. Let $W_i = \int_c^d W(x,h)_i^2\,dx$. We split up

$$\begin{aligned}
\widehat{IV}(h) - IV(h) &= \sum_i W_i(\varepsilon_i^2 - v(x_i)) \\
&\quad + \sum_i W_i(m(x_i) - \widehat{m}_{\widehat{f}}(x_i))^2 \\
&\quad + 2\sum_i W_i\varepsilon_i(m(x_i) - \sum_j \overline{W}(x_i,\widehat{f})_j m(x_j)) \\
&\quad - 2\sum_i W_i\varepsilon_i \sum_j \overline{W}(x_i,\widehat{f})_j\varepsilon_j \\
&= T_1 + \cdots + T_4 .
\end{aligned}$$

By means of Lemma 6.1 we can easily estimate the terms T_1 through T_4. Let for convenience h be first restricted to the interval $[n^{-1}, 1/2]$. Using $W_i = O(n^{-2}h^{-1})$ we get

$$T_1 = \widetilde{O}\left((nh)^{-1}n^{\delta-1/2}, n^{-\lambda}\right).$$

By $n^{-1}\sum(m(x_i) - \widehat{m}_{\widehat{f}}(x_i))^2 = \widetilde{O}\left((n\widehat{f})^{-1} + \widehat{f}^{2(r+s)}, n^{-\lambda}\right)$ and $\widehat{f} = f_0 + \widetilde{O}\left(f_0^{3/2}n^{\delta}, n^{-\lambda}\right)$ with some $f_0 \asymp n^{-1/(2(r+s)+1)}$ we obtain

$$T_2 = \widetilde{O}\left((nh)^{-1}n^{-\frac{2(r+s)}{2(r+s)+1}}, n^{-\lambda}\right)$$

and

$$\begin{aligned}
T_3 &= \widetilde{O}\left(\sqrt{\sum_i (W_i[m(x_i) - \sum_j \overline{W}(x_i,\widehat{f})_j m(x_j)])^2}\, n^{\delta}, n^{-\lambda}\right) \\
&= \widetilde{O}\left(n^{-2}h^{-1}n^{1/2}\widehat{f}^{r+s}n^{\delta}, n^{-\lambda}\right) \\
&= \widetilde{O}\left((nh)^{-1}n^{-1/2}n^{-\frac{r+s}{2(r+s)+1}}n^{\delta}, n^{-\lambda}\right).
\end{aligned}$$

If we write T_4 in the form $\underline{\varepsilon}'M(h,\widehat{f})\underline{\varepsilon}$, we obtain by the relation $tr(M(h,f)'M(h,f)) = O(n^{-1}(nh)^{-2}(nf)^{-1})$ and Whittle's inequality for quadratic forms the following estimate

$$\begin{aligned}
T_4 &\leq \left|\underline{\varepsilon}'M(h,\widehat{f})\underline{\varepsilon} - E\,\underline{\varepsilon}'M(h,f)\underline{\varepsilon}|_{f=\widehat{f}}\right| + E\,\underline{\varepsilon}'M(h,f)\underline{\varepsilon}|_{f=\widehat{f}} \\
&= \widetilde{O}\left(\sqrt{tr(M(h,\widehat{f})'M(h,\widehat{f}))}n^{\delta}, n^{-\lambda}\right) + O\left(\sum_j W_j\overline{W}(x_j,\widehat{f})_j\right) \\
&= \widetilde{O}\left((nh)^{-1}n^{-1/2}n^{-\frac{r+s}{2(r+s)+1}}n^{\delta}, n^{-\lambda}\right) + \widetilde{O}\left((nh)^{-1}n^{-\frac{2(r+s)}{2(r+s)+1}}, n^{-\lambda}\right).
\end{aligned}$$

Next, we decompose

$$\begin{aligned}
\widehat{ISB}(h,\overline{g}) - ISB(h) &= \underline{\varepsilon}'A(h,\overline{g})\underline{\varepsilon} - \sum A(h,\overline{g})_{ii}v(x_i) \\
&\quad + \sum_i A(h,\overline{g})_{ii}(v(x_i) - \widehat{v}_i) \\
&\quad + 2\underline{m}'A(h,\overline{g})\underline{\varepsilon} \\
&\quad + \underline{m}'A(h,\overline{g})\underline{m} - ISB(h) \\
&= T_5 + \cdots + T_8.
\end{aligned}$$

By (6.1) we see

$$\begin{aligned} T_5 &= \tilde{O}\left(\sqrt{\sum A(h,\overline{g})_{ij}^2}\, n^{\delta}, n^{-\lambda}\right) \\ &= \tilde{O}\left(h^{2r} n^{-1} \overline{g}^{-2r-1/2} n^{\delta}, n^{-\lambda}\right). \end{aligned}$$

Analogously to the estimation of $\widehat{IV}(h) - IV(h)$ we obtain

$$\begin{aligned} T_6 &= \tilde{O}\left(h^{2r} n^{-2} \overline{g}^{-2r-1} n^{1/2} n^{\delta}, n^{-\lambda}\right) \\ &= \tilde{O}\left(h^{2r} n^{-1} \overline{g}^{-2r-1/2} (n\overline{g})^{-1/2} n^{\delta}, n^{-\lambda}\right), \\ T_7 &= \tilde{O}\left(\|A(h,\overline{g})\underline{m}\| n^{\delta}, n^{-\lambda}\right) \\ &= \tilde{O}\left(h^{2r} n^{-1/2} n^{\delta}, n^{-\lambda}\right) \\ \text{and} \qquad T_8 &= \tilde{O}\left(h^{2r} \overline{g}^{s}, n^{-\lambda}\right). \end{aligned}$$

Analogous estimates can be derived for $h \in [0, n^{-1})$, where $O(1)$-terms take the place of the $(nh)^{-1}$-terms. It is known that

$$MISE(h) \geq C\left(h^{2r} + ((nh)^{-1} \wedge 1)\right), \tag{6.8}$$

which implies in conjunction with the above calculations that

$$\frac{\widehat{MISE}(h,\overline{g}) - MISE(h)}{MISE(h)} = \tilde{O}\left(n^{-\Delta} n^{\delta}, n^{-\lambda}\right).$$

On the other hand, we have $MISE(h_0) = O(n^{-\frac{2r}{2r+1}})$, which implies by (6.8) that

$$P\left(\left|\widehat{h} - h_0\right| > h_0/2\right) = O\left(n^{-\lambda}\right). \tag{6.9}$$

For brevity we set $\widehat{M}(h) = \widehat{MISE}(h,\overline{g})$ and $M(h) = MISE(h)$. Because of $\widehat{M}'(h)\big|_{h=\widehat{h}} = M'(h)\big|_{h=h_0} = 0$ we obtain

$$\begin{aligned} 0 &= \left(\widehat{M} - M\right)'(\widehat{h}) + \left(M'(\widehat{h}) - M'(h_0)\right) \\ &= \left(\widehat{M} - M\right)'(\widehat{h}) + (\widehat{h} - h_0) M''(h^*) \end{aligned}$$

for some h^* between h_0 and $\widehat{h}$, which implies

$$\widehat{h} - h_0 = \frac{(\widehat{M} - M)'(\widehat{h})}{-M''(h^*)}.$$

By straightforward calculations one obtains that

$$M''(h) \asymp n^{-1} h^{-3} \tag{6.10}$$

holds for $h \asymp h_0$. The term $(\widehat{M} - M)'(h)$ can be splitted up in the same way as $\widehat{M}(h) - M(h)$. It turns out, that T_1' through T_8' are of the same order as T_1 through T_8, respectively, with an additional factor of order h_0^{-1}. Hence, we have

$$\left(\widehat{M} - M\right)'(\widehat{h}) = \tilde{O}\left(n^{-1} h_0^{-2} n^{-\Delta} n^{\delta}, n^{-\lambda}\right),$$

which implies

$$\widehat{h} - h_0 = \tilde{O}\left(h_0 n^{-\Delta} n^{\delta}, n^{-\lambda}\right).$$

□

Proof of (4.7). The proof of this equation is very similar to that of Lemma 4.1. First, we derive analogously to (6.9) that

$$P\left(\left|\tilde{h} - h_0\right| > h_0/2\right) = O(n^{-\lambda}) \tag{6.11}$$

holds. By $\widehat{M}'(h)\Big|_{h=\widehat{h}} = \widetilde{M}'(h)\Big|_{h=\widetilde{h}} = 0$ we obtain

$$\begin{aligned} 0 &= \left(\widehat{M} - \widetilde{M}\right)'(\widehat{h}) + \left(\widetilde{M}'(\widehat{h}) - \widetilde{M}'(\widetilde{h})\right) \\ &= \left(\widehat{M} - \widetilde{M}\right)'(\widehat{h}) + (\widehat{h} - \widetilde{h})\widetilde{M}''(h^{**}) \end{aligned}$$

for some h^{**} between $\widetilde{h}$ and $\widehat{h}$, which implies

$$\widehat{h} - \widetilde{h} = \frac{(\widehat{M} - \widetilde{M})'(\widehat{h})}{-\widetilde{M}''(h^{**})}.$$

By (6.10) we can prove due to Lemma 6.1 that $\widetilde{M}''(h^{**})^{-1} = \tilde{O}\left(nh_0^3, n^{-\lambda}\right)$. We have

$$\begin{aligned} \widehat{IV}(h) - \widetilde{IV}(h) &= \sum_{i \in J_n} W_i(\varepsilon_i^2 - v(x_i)) + T_2 + T_3 + T_4 \\ &= T_1^* + T_2 + T_3 + T_4 . \end{aligned}$$

Let $h \asymp h_0$. By $\# J_n = O(n\Delta_n)$ we obtain

$$(T_1^*)' = \tilde{O}\left((nh)^{-1}h^{-1}n^{-1/2}\Delta_n^{1/2}n^{\delta}, n^{-\lambda}\right),$$

which differs from T_1' by the factor $\Delta_n^{1/2}$. It can be seen that the terms T_2' through T_4' are all majorized by $(T_1^*)'$. Next, we decompose

$$\begin{aligned} &\widehat{ISB}(h,\overline{g}) - \widetilde{ISB}(h,g_0) \\ &= T_5 - \sum_{i,j \notin J_n} A(h,g_0)\varepsilon_i\varepsilon_j + \sum_{i \notin J_n} A(h,g_0)_{ii}v(x_i) \\ &\quad + T_6 \\ &\quad + \underline{m}'A(h,\overline{g})\underline{m} - ISB(h) - \underline{m}'A(h,g_0)\underline{m} + ISB(h) \\ &\quad + 2\underline{m}'\left(A(h,\overline{g}) - A(h,g_0)\right)\underline{\varepsilon} \\ &\quad + 2 \sum_{(i,j): i \in J_n \text{ or } j \in J_n} A(h,g_0)_{ij}m(x_i)\varepsilon_j \\ &= U_1 + \cdots + U_5. \end{aligned}$$

The terms U_1', U_2' and U_3' are of the same order as T_5', T_6' and T_8', respectively. By (6.2) we conclude that $\|(A(h,\overline{g}) - A(h,g_0))\underline{m}\| = \tilde{O}\left(h^{2r}n^{-1/2}g_0^s n^{\delta}, n^{-\lambda}\right)$ holds, which implies that

$$U_4' = \tilde{O}\left(h^{2r}n^{-1/2}g_0^s n^{\delta}, n^{-\lambda}\right).$$

Finally, we have

$$U_5' = \tilde{O}\left(h^{2r}n^{-1/2}h^{-1}\Delta_n^{1/2}n^{\delta}, n^{-\lambda}\right).$$

Collecting the upper estimates for $(T_1^*)'$, T_2' through T_4' and U_1' through U_5' we obtain

$$\left(\widehat{M}' - \widetilde{M}'\right)(\widehat{h}) = \tilde{O}\left(n^{-1}h_0^{-2}n^{-\Delta'}n^{\delta}, n^{-\lambda}\right),$$

which yields (4.7). □

Acknowledgment. Some part of the research has been carried out during a visit of the author to the INRA, Jouy-en-Josas, France. I would like to thank Emmanuel Jolivet for the suggestion of this interesting subject and for helpful discussions. Further, I thank Wolfgang Härdle and Steve Marron for their interest and valuable comments.

REFERENCES

1. Faraway, J. (1990). Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Statist. Comput. Simul.* **37**, 37-44.
2. Faraway, J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Amer. Statist. Assoc.* **85**, 1119-1122.
3. Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation. Lecture Notes in Math.* **757**, 23-68.
4. Härdle, W. and Bowman, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102-110.
5. Härdle, W., Hall, P. and Marron, J. S. (1991). Regression smoothing parameters that are not far from their optimum. *J. Amer. Statist. Assoc.* **87**, 227-233.
6. Härdle, W., Huet, S. and Jolivet, E. (1992). Better bootstrap confidence intervals for regression curve estimation. Discussion paper No. 6, Humboldt-Universität, Berlin.
7. Härdle, W. and Mammen, E. (1990) Bootstrap methods in nonparametric regression. CORE Discussion paper 9058, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
8. Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22**, 215-232.
9. Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20**, 675-694.
10. Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20**, 695-711.
11. Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.
12. Jones, M. C. and Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **11**, 511-514.
13. Müller, H. G. and Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15**, 182-201.
14. Neumann, M. H. (1992). On completely data-driven pointwise confidence intervals in nonparametric regression. Rapport technique 92-02, INRA, Départment de Biométrie, Jouy-en-Josas, France.
15. Neumann, M. H. (1992). Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure. Preprint No. 34, Institut für Angewandte Analysis und Stochastik, Berlin.
16. Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53**, 683-690.
17. Skovgaard, I. M. (1981). Transformations of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8**, 207-217.
18. Skovgaard, I. M. (1986). On multivariate Edgeworth expansions. *Int. Statist. Rev.* **54**, 169-186.
19. Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Prob. Appl.* **5**, 302-305.

M. H. NEUMANN
INSTITUT FÜR ANGEWANDTE ANALYSIS UND STOCHASTIK
HAUSVOGTEIPLATZ 5-7
D - 10117 BERLIN
GERMANY

Veröffentlichungen des Instituts für Angewandte Analysis und Stochastik

Preprints 1992

1. D.A. Dawson, J. Gärtner: Multilevel large deviations.
2. H. Gajewski: On uniqueness of solutions to the drift–diffusion–model of semiconductor devices.
3. J. Fuhrmann: On the convergence of algebraically defined multigrid methods.
4. A. Bovier, J.-M. Ghez: Spectral properties of one–dimensional Schrödinger operators with potentials generated by substitutions.
5. D.A. Dawson, K. Fleischmann: A super–Brownian motion with a single point catalyst.
6. A. Bovier, V. Gayrard: The thermodynamics of the Curie–Weiss model with random couplings.
7. W. Dahmen, S. Prößdorf, R. Schneider: Wavelet approximation methods for pseudodifferential equations I: stability and convergence.
8. A. Rathsfeld: Piecewise polynomial collocation for the double layer potential equation over polyhedral boundaries. Part I: The wedge, Part II: The cube.
9. G. Schmidt: Boundary element discretization of Poincaré–Steklov operators.
10. K. Fleischmann, I. Kaj: Large deviation probability for some rescaled superprocesses.
11. P. Mathé: Random approximation of finite sums.
12. C.J. van Duijn, P. Knabner: Flow and reactive transport in porous media induced by well injection: similarity solution.
13. G.B. Di Masi, E. Platen, W.J. Runggaldier: Hedging of options under discrete observation on assets with stochastic volatility.
14. J. Schmeling, R. Siegmund–Schultze: The singularity spectrum of self-affine fractals with a Bernoulli measure.
15. A. Koshelev: About some coercive inequalities for elementary elliptic and parabolic operators.
16. P.E. Kloeden, E. Platen, H. Schurz: Higher order approximate Markov chain filters.

17. H.M. Dietz, Y. Kutoyants: A minimum–distance estimator for diffusion processes with ergodic properties.

18. I. Schmelzer: Quantization and measurability in gauge theory and gravity.

19. A. Bovier, V. Gayrard: Rigorous results on the thermodynamics of the dilute Hopfield model.

20. K. Gröger: Free energy estimates and asymptotic behaviour of reaction–diffusion processes.

21. E. Platen (ed.): Proceedings of the 1^{st} workshop on stochastic numerics.

22. S. Prößdorf (ed.): International Symposium "Operator Equations and Numerical Analysis" September 28 – October 2, 1992 Gosen (nearby Berlin).

23. K. Fleischmann, A. Greven: Diffusive clustering in an infinite system of hierarchically interacting diffusions.

24. P. Knabner, I. Kögel–Knabner, K.U. Totsche: The modeling of reactive solute transport with sorption to mobile and immobile sorbents.

25. S. Seifarth: The discrete spectrum of the Dirac operators on certain symmetric spaces.

26. J. Schmeling: Hölder continuity of the holonomy maps for hyperbolic basic sets II.

27. P. Mathé: On optimal random nets.

28. W. Wagner: Stochastic systems of particles with weights and approximation of the Boltzmann equation. The Markov process in the spatially homogeneous case.

29. A. Glitzky, K. Gröger, R. Hünlich: Existence and uniqueness results for equations modelling transport of dopants in semiconductors.

30. J. Elschner: The h-p–version of spline approximation methods for Mellin convolution equations.

31. R. Schlundt: Iterative Verfahren für lineare Gleichungssysteme mit schwach besetzten Koeffizientenmatrizen.

32. G. Hebermehl: Zur direkten Lösung linearer Gleichungssysteme auf Shared und Distributed Memory Systemen.

33. G.N. Milstein, E. Platen, H. Schurz: Balanced implicit methods for stiff stochastic systems: An introduction and numerical experiments.

34. M.H. Neumann: Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure.

35. M. Nussbaum: Asymptotic equivalence of density estimation and white noise.

Preprints 1993

36. B. Kleemann, A. Rathsfeld: Nyström's method and iterative solvers for the solution of the double layer potential equation over polyhedral boundaries.

37. W. Dahmen, S. Prössdorf, R. Schneider: Wavelet approximation methods for pseudodifferential equations II: matrix compression and fast solution.

38. N. Hofmann, E. Platen, M. Schweizer: Option pricing under incompleteness and stochastic volatility.

39. N. Hofmann: Stability of numerical schemes for stochastic differential equations with multiplicative noise.

40. E. Platen, R. Rebolledo: On bond price dynamics.

41. E. Platen: An approach to bond pricing.

42. E. Platen, R. Rebolledo: Pricing via anticipative stochastic calculus.

43. P.E. Kloeden, E. Platen: Numerical methods for stochastic differential equations.

44. L. Brehmer, A. Liemant, I. Müller: Ladungstransport und Oberflächenpotentialkinetik in ungeordneten dünnen Schichten.

45. A. Bovier, C. Külske: A rigorous renormalization group method for interfaces in random media.

46. G. Bruckner: On the regularization of the ill–posed logarithmic kernel integral equation of the first kind.

47. H. Schurz: Asymptotical mean stability of numerical solutions with multiplicative noise.

48. J.W. Barrett, P. Knabner: Finite element approximation of transport of reactive solutes in porous media. Part I: Error estimates for non–equilibrium adsorption processes.

49. M. Pulvirenti, W. Wagner, M.B. Zavelani Rossi: Convergence of particle schemes for the Boltzmann equation.

50. J. Schmeling: Most β shifts have bad ergodic properties.

51. J. Schmeling: Self normal numbers.

52. D.A. Dawson, K. Fleischmann: Super–Brownian motions in higher dimensions with absolutely continuous measure states.

53. A. Koshelev: Regularity of solutions for some problems of mathematical physics.

54. J. Elschner, I.G. Graham: An optimal order collocation method for first kind boundary integral equations on polygons.

55. R. Schlundt: Iterative Verfahren für lineare Gleichungssysteme auf Distributed Memory Systemen.

56. D.A. Dawson, K. Fleischmann, Y. Li, C. Müller: Singularity of super–Brownian local time at a point catalyst.

57. N. Hofmann, E. Platen: Stability of weak numerical schemes for stochastic differential equations.

58. H.G. Bothe: The Hausdorff dimension of certain attractors.

59. I.P. Ivanova, G.A. Kamenskij: On the smoothness of the solution to a boundary value problem for a differential–difference equation.

60. A. Bovier, V. Gayrard: Rigorous results on the Hopfield model of neural networks.