M A T H E M A T I C A L   M E T H O D S   O F   S T A T I S T I C S

ASYMPTOTIC MINIMAX ESTIMATION IN NONPARAMETRIC AUTOREGRESSION

M. H. Neumann[1] and T. L. Thorarinsdottir[2]

[1]Friedrich-Schiller-Universität Jena, Institut für Stochastik
Ernst-Abbe-Platz 2, D-07743 Jena, Germany
E-mail: mneumann@mathematik.uni-jena.de

[2]University of Aarhus, Department of Mathematical Sciences
Ny Munkegade, DK–8000 Aarhus C, Denmark
E-mail: disa@imf.au.dk

We develop asymptotic theory for nonparametric estimators of the autoregression function. To deal with irregularities in the pattern of explanatory variables caused by their randomness, we propose a new estimator which is a modification of the Priestley–Chao kernel method. It is shown that this estimator has similar asymptotic properties to standard estimators of kernel type. We establish an asymptotic lower bound to the minimax risk in Sobolev classes and show that our modified Priestley–Chao estimator can get arbitrarily close to this efficiency bound.

Key words: exact asymptotics, minimax risk, nonparametric autoregression, nonparametric estimation.

2000 Mathematics Subject Classification: Primary 62G05; secondary 62M05.

1. Introduction

The purpose of this paper is, generally speaking, to develop some good estimator of the autoregression function. One possible criterion for the optimality of estimators is their maximum risk in certain classes of functions, leading to minimax or, more often, asymptotically minimax estimators.

Since nonparametric curve estimation based on a finite number of observations always contains some aspect of interpolation, it is most natural to impose some smoothness constraint on the class of functions in order to get a uniformly consistent method. Typical smoothness classes in which asymptotic minimax results have been derived for various statistical models (nonparametric regression, density estimation, spectral density estimation) are Hölder, Sobolev or Besov classes.

Numerous asymptotic minimax results concerning the rate of convergence can be found in the literature. It turns out that many of the commonly used estimators (kernel, spline, orthonormal series, and wavelet estimators) can be tuned to achieve such an optimal rate of convergence. However, even within these classes, there is a wide variety of asymptotically optimal methods. This suggests that the focus on rates of convergence is often too weak to find a method which deserves the term "best".

A new line of research by focussing both on the optimal rate of convergence and the optimal constant in the asymptotic minimax risk has been inspired by the seminal work of Pinsker (1980) who already solved all essential problems in the particular context of signal estimation in Gaussian white noise. Later, these results were carried over to spectral density estimation (Efromovich and Pinsker, 1981), density estimation (Efromovich and Pinsker, 1982), nonparametric regression with Gaussian errors (Nussbaum, 1985), which was extended to non-normal error distributions (Golubev and Nussbaum, 1990). Recently, results of this type were derived in the context of density estimation from discretely observed diffusion processes by Butucea and Neumann (2005). It has been shown that the asymptotic minimax risk can be obtained by optimally tuned Fourier series, kernel or spline estimators. If kernel estimators are concerned, for example, in the case of density estimation, the desired risk bound can be obtained by a unique kernel function which is connected with the smoothness class under consideration. Therefore, at a practical level, an objective criterion for the choice of a kernel function is given.

In this paper we intend to derive asymptotic minimax estimators of the autoregression function in Sobolev classes. We will show that the efficiency bound has the same structure as in the already studied models. In contrast to the related case of nonparametric regression, the actual difficulty of the problem depends on the distribution of the explanatory (lagged) variables, which in turn depends on the autoregression function m. In order to get a tight efficiency bound, we take this particular difficulty of the problem into account by the choice of the loss function which contains a term with the stationary density. We show that asymptotic efficiency can be obtained by kernel estimators with kernel functions converging (as $n \to \infty$) to that which is known to be optimal in the density case. While we give a complete proof of the lower asymptotic risk bound, our sequence of estimators attaining this bound is not fully adaptive since an optimal choice of the tuning parameter (bandwidth) is required; see equation (4.2). We believe, however, that standard techniques such as a plug-in method or leave-l_n-out cross-validation could be employed to achieve a fully data-driven method.

2. Main Assumptions and a Modified Priestley–Chao Estimator

Suppose we observe $X_0, \ldots, X_n$ which stem from a stationary and ergodic autoregressive process $(X_t)_{t \in \mathbb{Z}}$ obeying the equation

$$(2.1) \qquad X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t, \qquad t \in \mathbb{Z},$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ are independent and identically distributed random variables with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, and $E\varepsilon^{4+\kappa} < \infty$, for some $\kappa > 0$.

We intend to estimate the autoregression function m. Since the performance of any estimator necessarily deteriorates in regions with a low stationary density, we

restrict our attention to a fixed interval, without loss of generality $[0, 1]$, and ensure by appropriate conditions that the stationary density is bounded away from zero on this region.

We need two types of assumptions. Some of them ensure basic properties such as ergodicity of the process, while the other ones concern smoothness properties of the target function m and make certain rates of convergence of estimators possible. The following set of assumptions is mainly imposed to get ergodicity:

(A1)
   (i) $\limsup_{|x| \to \infty} |m(x)|/|x| < 1$,
   (ii) ε_t has an everywhere positive and Lipschitz continuous density p_ε,
   (iii) σ is a continuous function and there exist constants $0 < C_1 \le C_2 < \infty$ such that $C_1 \le \sigma(x) \le C_2$ for all x.

It is well known (see, for example, Doukhan, 1994, pp. 106–107) that a process satisfying assumption (A1) is geometrically ergodic. Moreover, for given $(\varepsilon_t)_{t \in \mathbb{Z}}$, there exists a unique stationary solution to (2.1) and the corresponding process is absolutely regular (β-mixing) with geometrically decaying coefficients. The stationary density π_m is everywhere positive and Lipschitz continuous.

In Section 4 below, where minimax results in certain classes of functions are concerned, we will require uniformity (in m) of these properties, which will be ensured by appropriate conditions on m. Regarding smoothness properties of m, we assume that

(A2) There exists some $\delta > 0$ such that m has β generalized derivatives on $[-\delta, 1 + \delta]$ with $\int_{-\delta}^{1+\delta} (m^{(\beta)}(z))^2 \, dz < \infty$.

The ultimate goal in this paper is to devise a sequence of estimators which is asymptotically minimax in Sobolev classes of functions. In the related problem of nonparametric regression, the standard estimator is still the one proposed by Nadaraya (1964) and Watson (1964). It is undoubtedly a good choice in the often studied cases of equidistant nonrandom design or random design with a sufficiently regular density. To diminish unfavorable bias effects in the case of nonrandom and nonequidistant design, an alternative weighting scheme has been proposed by Gasser and Müller (1979).

In the case of random design, local polynomial estimators first studied by Stone (1977, 1980), Cleveland (1979) and Katkovnik (1979) are considered to be the adequate tool; for a comparison of these approaches see, for example, Fan and Gijbels (1996).

In the context of nonparametric autoregression, local polynomial estimators seem to be the natural choice since they automatically adapt to irregularities in the pattern of explanatory (lagged) variables caused by their randomness. A pth order local polynomial estimator $\widetilde{m}_{n,LP}(x)$ is given as $\widehat{a}_0 = \widehat{a}_0(x; X_0, \dots, X_n)$, where $\widehat{a} = (\widehat{a}_0, \dots, \widehat{a}_{p-1})'$ minimizes

$$\sum_{t=1}^{n} w\left(\frac{x - X_{t-1}}{h_n}\right)\left(X_t - \sum_{j=0}^{p-1} a_j (x - X_{t-1})^j\right)^2.$$

It is in fact a local least squares estimator, with window function w and a sequence of bandwidths $(h_n)_{n \in \mathbb{N}}$. The sequence of bandwidths here and below is usually

assumed to satisfy $h_n \to 0$ and $nh_n \to \infty$, as $n \to \infty$. The estimator $\widetilde{m}_{n,LP}(x)$ can also be written in the form of a linear estimator as

$$\widetilde{m}_{n,LP}(x) = \sum_{t=1}^{n} \widetilde{K}_{h_n}\left(x, X_{t-1}, \{X_0, \dots, X_{n-1}\}\right) X_t,$$

which can then be approximated by an estimator with weights depending each on a single observation only,

$$\widetilde{\widetilde{m}}_{n,LP}(x) = \sum_{t=1}^{n} \widetilde{\widetilde{K}}_{h_n}(x, X_{t-1}) X_t;$$

see Neumann and Kreiss (1998). The latter approximation simplifies technical difficulties a lot and makes a mathematically rigorous asymptotic analysis of local polynomial estimators possible. However, for our ultimate goal of achieving the sharp asymptotic efficiency bound, it turns out that the "effective kernel function" (here $\widetilde{\widetilde{K}}_{h_n}(x, \cdot)$) has to be sufficiently close to some particular kernel which is known to be optimal in the related problems of filtering of a smooth signal from Gaussian white noise (Golubev, 1987) and nonparametric density estimation (Butucea and Neumann, 2005). Since we were not able to establish the connection between the effective kernel $\widetilde{\widetilde{K}}_{h_n}(x, \cdot)$ and the underlying window function w we could not find an asymptotically optimal local polynomial estimator explicitly.

To overcome this problem, we propose a modification of the kernel estimator of Priestley and Chao (1972). Their proposal, which was made in the context of nonparametric regression with equispaced design, amounts to estimating m by

$$\widetilde{m}_{n,PC}(x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - X_{t-1}}{h_n}\right) X_t.$$

It is well known that the bias of standard kernel regression estimators behaves favorably in correspondence with the smoothness of the target function if the design points show a regular pattern. In our case with random design points $X_0, \dots, X_{n-1}$, however, such a regularity cannot be guaranteed. To obtain nevertheless a favorable bias behavior, we will modify the kernel function in such a way that the empirical moments behave as the theoretical moments of a kernel of order β (see equation (2.4) below), which implies that enough terms in a Taylor series expansion of the bias cancel.

Now we define a modified version of the Priestley–Chao estimator as

$$(2.2) \qquad \widehat{m}_n(x) = \sum_{t=1}^{n} K_n(x, X_{t-1}) X_t,$$

where

$$(2.3) \qquad K_n(x, z) = \frac{1}{nh_n} K\left(\frac{x - z}{h_n}\right) \left[\widehat{c}_0(x) + \sum_{j=1}^{\beta-1} \widehat{c}_j(x) g_j\left(\frac{x - z}{h_n}\right)\right]$$

and $(h_n)_{n \in \mathbb{N}}$ is a sequence of nonrandom bandwidths. In order to derive asymptotic results, we will assume that this sequence of bandwidths satisfies

(A3) $h_n \underset{n \to \infty}{\longrightarrow} 0$ and $nh_n / \log n \underset{n \to \infty}{\longrightarrow} \infty$.

We assume that the kernel function K has compact support $[-C_K, C_K]$, is Lipschitz continuous and is of order β, that is,

$$\int K(u) u^l \, du = \delta_{l,0}, \qquad \text{for} \quad l = 0, \dots, \beta - 1,$$

and $\int |K(u)||u|^\beta \, du < \infty$. The functions $g_1, \dots, g_{\beta-1}$ are chosen so that $K(\cdot) g_j(\cdot)$ are Lipschitz continuous and that the matrix

$$M = \begin{pmatrix} \int K(u) \, du & \int K(u) \, g_1(u) \, du & \dots & \int K(u) \, g_{\beta-1}(u) \, du \\ \int K(u) \, u \, du & \int K(u) \, u \, g_1(u) \, du & \dots & \int K(u) \, u \, g_{\beta-1}(u) \, du \\ \vdots & \vdots & \ddots & \vdots \\ \int K(u) \, u^{\beta-1} \, du & \int K(u) \, u^{\beta-1} \, g_1(u) \, du & \dots & \int K(u) \, u^{\beta-1} \, g_{\beta-1}(u) \, du \end{pmatrix}$$

is nonsingular. And finally, the functions $\widehat{c}_0, \dots, \widehat{c}_{\beta-1}$ will be chosen so that, in spite of irregularities in the design, certain empirical moments behave as the theoretical ones of a βth order kernel, that is,

$$(2.4) \qquad \sum_{t=1}^{n} K_n(x, X_{t-1}) \left(\frac{x - X_{t-1}}{h_n} \right)^l = \delta_{l,0}, \qquad \text{for} \quad l = 0, \dots, \beta - 1.$$

The following lemma shows that, with a probability tending to one, the system of equations (2.4) has a unique solution. Before we proceed, we fix some notation. Denote by $\mathbb{X}_n = (X_0, \dots, X_{n-1})'$ the vector of explanatory (lagged) variables in the model (2.1). Furthermore, we denote by λ in terms of the form $O(n^{-\lambda})$ a constant which can be arbitrarily large, in some cases under the condition that a corresponding constant C_λ is appropriately chosen. These constants may be different at different places. We denote by π_m the stationary density of the process with autoregression function m.

Lemma 2.1. *Suppose that stationary observations obeying* (2.1) *are given, and that assumptions* (A1) *and* (A3) *are fulfilled. Then there exist sets* $\mathcal{X}_n \subseteq \mathbb{R}^n$ *with* $P(\mathbb{X}_n \notin \mathcal{X}_n) = O(n^{-\lambda})$ *such that, for* $\mathbb{X}_n \in \mathcal{X}_n$, (2.4) *possesses a unique solution* $(\widehat{c}_0(x), \dots, \widehat{c}_{\beta-1}(x))'$ *for all* $x \in [0, 1]$. *Furthermore, there exists a sequence* $(\delta_n)_{n \in \mathbb{N}}$ *with* $\delta_n = O(h_n + \sqrt{\log n / (nh_n)})$ *such that, with* $c_0(x) = 1/\pi_m(x)$ *and* $c_j(x) = 0$ *(*$j = 1, \dots, \beta - 1$*),*

$$\sup_{x \in [0,1]} |\widehat{c}_j(x) - c_j(x)| \leq \delta_n,$$

for $l = 0, \dots, \beta - 1$ *and* $\mathbb{X}_n \in \mathcal{X}_n$.

It will be shown in Theorem 3.1 below that the estimator $\widehat{m}_n$ has the usual asymptotic behavior on the favorable sets $\mathcal{X}_n$. Moreover, if we additionally take care that the correction factors $\widehat{c}_j(x)$ are bounded, then we can establish the asymptotic result for the usual L_2-loss; see Theorem 3.2 below.

Remark 2.1. So far, it is only required that the functions $g_1, \dots, g_{\beta-1}$ make the matrix M nonsingular; an appropriate choice of them is still left open. Since K is by assumption a kernel of order β, the first column of the matrix M is always

equal to $(1, 0, \ldots, 0)'$. Hence, it suffices to choose $g_1, \ldots, g_{\beta-1}$ in such a way that the submatrix

$$\widetilde{M} = \begin{pmatrix} \int K(u)\, u\, g_1(u)\, du & \ldots & \int K(u)\, u\, g_{\beta-1}(u)\, du \\ \vdots & \ddots & \vdots \\ \int K(u)\, u^{\beta-1}\, g_1(u)\, du & \ldots & \int K(u)\, u^{\beta-1}\, g_{\beta-1}(u)\, du \end{pmatrix}$$

is nonsingular. This would suggest, for example, the choice $g_i(u) = K(u)u^i$. Actually, we then had

$$c'\widetilde{M}c = \int (K(u))^2 \left(\sum_{i=1}^{\beta-1} c_i u^i \right)^2 du.$$

Since $\{u \colon K(u) \neq 0\}$ contains some interval, $c'\widetilde{M}c = 0$ if and only if $c = (0, \ldots, 0)'$, which means that $\widetilde{M}$, and hence M too, are nonsingular.

3. Asymptotic Properties of the Estimator

In contrast to the often studied case of nonparametric regression with regular nonrandom design, we face here the situation that the "random design" (that is, $X_0, \ldots, X_{n-1}$) can become, with a probability tending to zero as $n \to \infty$, extremely unfavorable. This is, for example, reflected by the fact that the correction factors $\widehat{c}_j(x)$ have a favorable behavior only up to a probability of $O(n^{-\lambda})$; see Lemma 2.1. These unfavorable events, where the design gets out of control, can possibly affect the asymptotic risk of the estimator. Therefore, we prove first the following result for a truncated loss function.

Theorem 3.1. *Suppose that stationary observations obeying* (2.1) *are given, and that assumptions* (A1) *to* (A3) *are fulfilled. Then, as $n \to \infty$,*

$$E_m \left[\int_0^1 \left(\widehat{m}_n(x) - m(x) \right)^2 dx \; \wedge \; 1 \right] = \frac{1}{nh_n} \int K^2(u)\, du \int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx$$

$$+ \int_0^1 \left(\int \frac{1}{h_n} K\!\left(\frac{x-y}{h_n} \right) m(y)\, dy - m(x) \right)^2 dx + o\!\left(\frac{1}{nh_n} + h_n^{2\beta} \right).$$

If m is β times continuously differentiable on $[-\delta, 1+\delta]$, for some $\delta > 0$, then the second term on the right-hand side is equal to

$$h_n^{2\beta} \left(\frac{\int K(u)u^\beta\, du}{\beta!} \right)^2 \int_0^1 \left(m^{(\beta)}(x) \right)^2 dx + o(h_n^{2\beta}).$$

In order to get rid of the truncation in the loss function, we have to avoid a too strong insistence on (2.4) which can possibly make the correction factors $\widehat{c}_0(x), \ldots, \widehat{c}_{\beta-1}(x)$ arbitrarily large. A simple remedy is to replace the $\widehat{c}_j$'s defined there by

$$\widehat{\widehat{c}}_j(x) = \begin{cases} \widehat{c}_j(x), & \text{if } |\widehat{c}_j(x)| \leq C, \\ 0, & \text{if } |\widehat{c}_j(x)| > C, \end{cases}$$

where the truncation point satisfies $C > \sup_{x \in [0,1]}\{1/\pi_m(x)\}$. Let $\widehat{\widehat{m}}_n$ be the estimator derived from $\widehat{m}_n$, with $\widehat{\widehat{c}}_0(x), \ldots, \widehat{\widehat{c}}_{\beta-1}(x)$ instead of $\widehat{c}_0(x), \ldots, \widehat{c}_{\beta-1}(x)$.

Theorem 3.2. *Suppose that stationary observations obeying* (2.1) *are given, and that assumptions* (A1) *to* (A3) *are fulfilled. Then, as* $n \to \infty$,

$$E_m\left[\int_0^1 \left(\widehat{\widehat{m}}_n(x) - m(x)\right)^2 dx\right] = \frac{1}{nh_n}\int K^2(u)\,du\,\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\,dx$$

$$+ \int_0^1 \left(\int \frac{1}{h_n}K\left(\frac{x-y}{h_n}\right)m(y)\,dy - m(x)\right)^2 dx + o\left(\frac{1}{nh_n} + h_n^{2\beta}\right).$$

4. Asymptotic minimax theory in Sobolev classes

In this section, we derive asymptotic minimax bounds in Sobolev classes and show that our modified Priestley–Chao estimators can get arbitrarily close to achieving these efficiency bounds. First, we have to impose uniform (in m) versions of the previously used assumptions (A1) and (A2). For reasons explained below, we restrict our consideration to the case that the ε_t's are standard normal distributed. Furthermore, σ is assumed to be a Lipschitz continuous function not depending on m with $C_1 \le \sigma(x) \le C_2 \; \forall x \in \mathbb{R}$ and $0 < C_1 \le C_2 < \infty$. To guarantee uniformity in the mixing properties, we assume that m is a member of the class of functions

$$\mathcal{M} = \big\{m \colon R \to \mathbb{R} \mid |m(x)| \le C_3|x| \text{ for all } |x| > C_4\big\},$$

where $C_3 < 1$ and $C_4 < \infty$ are fixed constants. To define an appropriate smoothness class, we fix $L, \delta_0 \in (0, \infty)$, and a nonnegative function M on $[0, \infty)$ with $M(\delta) \xrightarrow[\delta \to 0]{} 0$ and define

$$W_2^\beta(L) = \bigg\{m \colon m \text{ has } \beta \text{ generalized derivatives on } (-\delta_0, 1 + \delta_0)$$

$$\text{and } \int_{-\delta}^{1+\delta} (m^{(\beta)}(x))^2\,dx \le L + M(\delta) \quad \forall \delta \in [0, \delta_0)\bigg\}.$$

In the related case of nonparametric regression with homoscedastic errors and regular nonrandom design with density π, it is known from Theorem 1 in Golubev and Nussbaum (1990) that the minimax risk (with the usual L_2-loss) is proportional to $\left(\int_0^1 \pi^{-1}(x)\,dx\right)^{2\beta/(2\beta+1)}$. Therefore, in order to avoid a too conservative efficiency bound, we build the difficulty of estimation into our loss function and focus on the minimax risk

$$R_n = \inf_{\widetilde{m}_n} \sup_{m \in W_2^\beta(L) \cap \mathcal{M}} \left\{\left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\,dx\right)^{-\frac{2\beta}{2\beta+1}} E_m\|\widetilde{m}_n - m\|_{L_2([0,1])}^2\right\},$$

where the infimum is taken over all estimators of m based on $X_0, \ldots, X_n$. The idea of such a "self-normalizing" loss is not totally new. A similar idea appeared in Lepski, Mammen and Spokoiny (1997) and Lepski and Spokoiny (1997), in a different context.

To set a benchmark for our modified Priestley–Chao estimators, we state first an asymptotic lower bound to the minimax risk, which resembles the classical Pinsker

bound (see, for example, Pinsker (1980), Efromovich and Pinsker (1981, 1982), Nussbaum (1985), Golubev and Nussbaum (1990), Belitser and Levit (1996), Butucea and Neumann (2005), for such results in different settings). It follows from Corollary 4.1 below that this efficiency bound is sharp.

Theorem 4.1. *Suppose that stationary observations obeying* (2.1) *are given, and that assumption* (A3) *and the assumptions made in this section are fulfilled. Then*

$$\liminf_{n\to\infty} \inf_{\widetilde{m}_n} \sup_{m\in W_2^\beta(L)\cap\mathcal{M}} \left\{ n^{\frac{2\beta}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx \right)^{-\frac{2\beta}{2\beta+1}} E_m \|\widetilde{m}_n - m\|_{L_2([0,1])}^2 \right\}$$
$$\geq \gamma(\beta) L^{\frac{1}{2\beta+1}},$$

where $\gamma(\beta) = (2\beta+1)^{1/(2\beta+1)} \left[\beta/(\pi(\beta+1))\right]^{2\beta/(2\beta+1)}$ *is Pinsker's constant.*

The next theorem shows an upper bound for the uniform risk of our modified Priestley–Chao estimators $\widehat{\widehat{m}}_n$ from Section 2.

Theorem 4.2. *Suppose that stationary observations obeying* (2.1) *are given, and that assumption* (A3) *and the assumptions made in this section are fulfilled. Then, uniformly in* $m \in W_2^\beta(L)\cap\mathcal{M}$,

$$(4.1) \quad E_m \|\widehat{\widehat{m}}_n - m\|_{L_2([0,1])}^2 \leq \frac{1}{nh_n} \int K^2(u)\, du \int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx$$
$$+ h_n^{2\beta} \sup_\omega \left\{ (\widehat{K}(\omega) - 1)^2 \omega^{-2\beta} \right\} L + o\left(\frac{1}{nh_n} + h_n^{2\beta} \right).$$

Furthermore, if the sequence of bandwidths $(h_n)_{n\in\mathbb{N}}$ *is chosen in an asymptotically optimal way, that is,*

$$(4.2) \qquad h_n = n^{-\frac{1}{2\beta+1}} \left(\frac{\int_0^1 \sigma^2(x)/\pi_m(x)\, dx \int K^2(u)\, du}{2\beta \sup_\omega \{(\widehat{K}(\omega)-1)^2 \omega^{-2\beta}\} L} \right)^{\frac{1}{2\beta+1}} (1 + o(1)),$$

then

$$(4.3) \quad \limsup_{n\to\infty} \sup_{m\in W_2^\beta(L)\cap\mathcal{M}} \left\{ n^{\frac{2\beta}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx \right)^{-\frac{2\beta}{2\beta+1}} E_m \|\widehat{\widehat{m}}_n - m\|_{L_2([0,1])}^2 \right\}$$
$$\leq (2\beta+1) \left(\sup_\omega \left\{ (\widehat{K}(\omega) - 1)^2 \omega^{-2\beta} \right\} L \right)^{\frac{1}{2\beta+1}} \left(\frac{\int K^2(u)\, du}{2\beta} \right)^{\frac{2\beta}{2\beta+1}}.$$

Remark 4.1. So far, the asymptotically efficient estimator is not feasible since it still requires an appropriate choice of the bandwidths h_n. For known L and $\sigma^2(\cdot)$, one could simply use any consistent preliminary estimator of the stationary density π_m; for closely related results see Dalalyan and Kutoyants (2002) and Dalalyan (2005).

A more realistic approach should, however, not require prior knowledge of any quantity which is usually not known in advance. In this sense, a fully data-driven procedure for the choice of the bandwith is clearly preferable. Some cross-validation technique may be employed for this purpose; a thorough study of this is, however, beyond the scope of this paper.

Remark 4.2. Formally, the right-hand side of (4.3) is minimized by the choice $\widehat{K}_\beta(\omega) = (1 - |\omega|^\beta)_+$, which is the Fourier transform of

$$K_\beta(x) = -\frac{\beta!}{\pi} \sum_{j=1}^\beta \frac{\sin^{(j)}(x)}{(\beta - j)! x^{j+1}}.$$

If we plug K_β into (4.3), then we obtain by

$$\int K_\beta^2(u)\, du = \frac{1}{2\pi} \int (\widehat{K}_\beta(\omega))^2\, d\omega = \frac{2\beta^2}{\pi(\beta + 1)(2\beta + 1)}$$

that

$$\limsup_{n\to\infty} \sup_{m\in W_2^\beta(L)\cap\mathcal{M}} \left\{ n^{\frac{2\beta}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx \right)^{-\frac{2\beta}{2\beta+1}} E_m \|\widehat{\widehat{m}}_n - m\|_{L_2([0,1])}^2 \right\}$$
$$\leq \gamma(\beta)\, L^{\frac{1}{2\beta+1}}.$$

This does, however, not mean that we can reach the asymptotic efficiency bound by a modified Priestley–Chao estimator with kernel function K_β. This kernel is not compactly supported and is also not a kernel of order β in the usual sense since, for example, $\int |K_\beta(u)||u|^\beta\, du = \infty$ for $\beta \geq 1$.

Despite this negative result we can get arbitrarily close to the efficiency bound stated in Theorems 4.1 and 4.2 by using approximations of K_β by βth order kernels which do have compact support. This is made clear by the following lemma.

Lemma 4.1. *Let* $\widehat{K}_\beta(\omega) = (1 - |\omega|^\beta)_+$. *Then, for any* $\delta > 0$, *there exists a compactly supported and Lipschitz continuous kernel* $K_{\beta,\delta}$ *such that*

$$\left| |1 - \widehat{K}_\beta(\omega)|\, |\omega|^{-\beta} - |1 - \widehat{K}_{\beta,\delta}(\omega)|\, |\omega|^{-\beta} \right| \leq \delta \qquad \forall \omega$$

and

$$\|K_\beta - K_{\beta,\delta}\|_{L_2} \leq \delta.$$

Corollary 4.1. *Suppose that stationary observations obeying (2.1) are given, and that the assumptions made in this section are fulfilled. Then, for any* $\epsilon > 0$, *there exists a kernel function* $K_{\beta,\epsilon}$ *such that the estimator* $\widehat{\widehat{m}}_n$ *with bandwidths* $(h_n)_{n\in\mathbb{N}}$ *according to (4.2) satisfies*

$$\limsup_{n\to\infty} \sup_{m\in W_2^\beta(L)\cap\mathcal{M}} \left\{ n^{\frac{2\beta}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx \right)^{-\frac{2\beta}{2\beta+1}} E_m \|\widehat{\widehat{m}}_n - m\|_{L_2([0,1])}^2 \right\}$$
$$\leq \gamma(\beta)\, L^{\frac{1}{2\beta+1}}\, (1 + \epsilon).$$

Remark 4.3. Without the assumed normality of the innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$, one could still derive under appropriate regularity conditions an asymptotic lower bound result as in Theorem 4.1, where only $\sigma^2(x)$ has to be replaced by the inverse of the Fisher information of the family of densities $\{p_\varepsilon(\cdot - u)\colon u \in \mathbb{R}\}$. To achieve this asymptotic risk bound, one has to devise a likelihood-based method of estimation rather than taking our kernel-weighted mean of the observations. In the context of nonparametric regression with regular design, an asymptotically efficient estimator based on orthogonal series expansions and Pinsker's filter has been studied in the Diploma thesis of Sprünker (2005), written under the supervision of the first author. The purpose of the present paper is, however, to provide an adequate modification of the kernel method in the autoregressive case which includes, in particular, a remedy against irregularities in the pattern of explanatory variables. In order to preserve a clear presentation of the basic ideas we restrict ourselves to the Gaussian case, when sharp asymptotic minimaxity is concerned.

5. Proofs

Proof of Lemma 2.1. For any $x \in [0,1]$, let $M_n(x)$ be the $\beta \times \beta$ matrix with entries

$$(M_n(x))_{j,k} = \frac{1}{nh_n} \sum_t K\left(\frac{x - X_{t-1}}{h_n}\right) \left(\frac{x - X_{t-1}}{h_n}\right)^{j-1} g_{k-1}\left(\frac{x - X_{t-1}}{h_n}\right),$$

$j, k = 1, \ldots, \beta$. If $M_n(x)$ is nonsingular, then a solution to (2.4) exists and

$$M_n(x)\widehat{c}(x) = \mathbf{1}_\beta = (1, 0, \ldots, 0)'.$$

Furthermore, since $M\mathbf{1}_\beta = \mathbf{1}_\beta$, closeness of $\widehat{c}(x)$ to $(1/\pi_m(x), 0, \ldots, 0)'$ will follow from closeness of $M_n(x)$ to $\pi_m(x)M$; see (5.4) below.

We have that

$$\begin{aligned}
(5.1) \quad E_m[(M_n(x))_{j,k}] &= \int \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) \left(\frac{x-y}{h_n}\right)^{j-1} g_{k-1}\left(\frac{x-y}{h_n}\right) \pi_m(y)\, dy \\
&= \pi_m(x) \int K(u) u^{j-1} g_{k-1}(u)\, du + O(h_n) \\
&= \pi_m(x)\, M_{j,k} + O(h_n)
\end{aligned}$$

and

$$(5.2) \qquad \text{var}\left((M_n(x))_{j,k}\right) = \frac{1}{nh_n^2}\text{var}\left(K\left(\frac{x-X_0}{h_n}\right)\left(\frac{x-X_0}{h_n}\right)^{j-1}g_{k-1}\left(\frac{x-X_0}{h_n}\right)\right)$$

$$+ \frac{1}{n^2h_n^2}\sum_{t=1}^{n}\sum_{s:\,1\le|s-t|\le C_\lambda\log n}\text{cov}\left(K\left(\frac{x-X_{t-1}}{h_n}\right)\left(\frac{x-X_{t-1}}{h_n}\right)^{j-1}g_{k-1}\left(\frac{x-X_{t-1}}{h_n}\right),\right.$$

$$\left. K\left(\frac{x-X_{s-1}}{h_n}\right)\left(\frac{x-X_{s-1}}{h_n}\right)^{j-1}g_{k-1}\left(\frac{x-X_{s-1}}{h_n}\right)\right)$$

$$+ \frac{1}{n^2h_n^2}\sum_{t=1}^{n}\sum_{s:\,|s-t|>C_\lambda\log n}\text{cov}\left(K\left(\frac{x-X_{t-1}}{h_n}\right)\left(\frac{x-X_{t-1}}{h_n}\right)^{j-1}g_{k-1}\left(\frac{x-X_{t-1}}{h_n}\right),\right.$$

$$\left. K\left(\frac{x-X_{s-1}}{h_n}\right)\left(\frac{x-X_{s-1}}{h_n}\right)^{j-1}g_{k-1}\left(\frac{x-X_{s-1}}{h_n}\right)\right)$$

$$= O\left(\frac{1}{nh_n}\right) + O\left(\frac{\log n}{n}\right) + O(n^{-\lambda}).$$

Before we proceed, we recall the Bernstein-type inequality from Theorem 1.4.2.4 in Doukhan (1994, page 36). Assume that $Z_1,\dots,Z_n$ are β-mixing random variables. If

    (i) $EZ_t = 0$ for all t,
    (ii) there exists $\sigma^2 < \infty$ such that, for all n, m, $\frac{1}{m}E(Z_n + \cdots + Z_{n+m})^2 \le \sigma^2$,
    (iii) $|Z_t| \le M < \infty$ for all t,

then, for any $\varepsilon > 0$ and any $q \in (0, n/(1 + \varepsilon^2/4)]$, the following inequality holds:

$$P\left(\left|\sum_{t=1}^{n}Z_t\right| \ge x\right) \le 4\exp\left\{-\frac{(1-\varepsilon)x^2}{2(n\sigma^2 + qMx/3)}\right\} + 2\frac{n\beta_{[q\varepsilon^2/4]-1}}{q}.$$

In the case of geometrically mixing coefficients, this inequality can be simplified to

$$P\left(\left|\sum_{t=1}^{n}Z_t\right| \ge x\right) \le 4\exp\left\{-\frac{C_\lambda x^2}{n\sigma^2 + \log n\,x}\right\} + O(n^{-\lambda}),$$

where λ may be chosen arbitrarily large and $C_\lambda > 0$ appropriately. Therefore, we obtain from (5.1) and (5.2) that

$$P\left(|(M_n(x))_{j,k} - \pi_m(x)M_{j,k}| > C_\lambda\left(h_n + \left(\frac{\log n}{nh_n}\right)^{1/2}\right)\right) = O(n^{-\lambda}).$$

Since K and $K(\cdot)g_j(\cdot)$ $(j = 1,\dots,\beta-1)$ are Lipschitz continuous functions, we can deduce by an approximation on increasingly fine grids in conjunction with a simple continuity argument that also

$$(5.3) \quad P\left(|(M_n(x))_{j,k} - \pi_m(x)M_{j,k}| \le C_\lambda\left(h_n + \left(\frac{\log n}{nh_n}\right)^{1/2}\right) \quad \forall x \in [0,1]\right)$$

$$= 1 - O(n^{-\lambda})$$

holds true. Since the matrix M is by assumption nonsingular, we obtain that the matrices $M_n(x)$ are also nonsingular with $((M_n(x))^{-1}$ being bounded, with probability exceeding $1 - O(n^{-\lambda})$. This implies that

(5.4) $\widehat{c}(x) = (M_n(x))^{-1} M \mathbf{1}_\beta$

$$= \begin{pmatrix} 1/\pi_m(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + (M_n(x))^{-1} (\pi_m(x)M - M_n(x)) \begin{pmatrix} 1/\pi_m(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

which yields

$$P\left(|\widehat{c}_j(x) - c_j(x)| \le C_\lambda \left(h_n + \left(\frac{\log n}{n h_n} \right)^{1/2} \right) \quad \forall x \in [0,1] \right) = 1 - O(n^{-\lambda}). \qquad \square$$

Proof of Theorem 3.1. Recall that, according to Lemma 2.1, there exist sets $\mathcal{X}_n$ with

(5.5) $$P(\mathbb{X}_n \notin \mathcal{X}_n) = O(n^{-\lambda})$$

and, for $\mathbb{X}_n \in \mathcal{X}_n$, the system of equations (2.4) possesses a unique solution $(\widehat{c}_0(x), \dots, \widehat{c}_{\beta-1}(x))$ for all x with

(5.6) $$\sup_{x \in [0,1]} |\widehat{c}_j(x) - c_j(x)| \le \delta_n \xrightarrow[n \to \infty]{} 0.$$

From (5.5) we obtain that

(5.7) $E_m \left[\int_0^1 (\widehat{m}_n(x) - m(x))^2 \, dx \wedge 1 \right]$

$$= E_m \left[I(\mathbb{X}_n \in \mathcal{X}_n) \int_0^1 \left(\sum_{t=1}^n K_n(x, X_{t-1}) \sigma(X_{t-1}) \varepsilon_t \right)^2 dx \right]$$

$$+ E_m \left[I(\mathbb{X}_n \in \mathcal{X}_n) \int_0^1 \left(\sum_{t=1}^n K_n(x, X_{t-1}) m(X_{t-1}) - m(x) \right)^2 dx \right]$$

$$+ 2E_m \left[I(\mathbb{X}_n \in \mathcal{X}_n) \int_0^1 \left(\sum_{t=1}^n K_n(x, X_{t-1}) \sigma(X_{t-1}) \varepsilon_t \right) \right.$$

$$\left. \times \left(\sum_{t=1}^n K_n(x, X_{t-1}) m(X_{t-1}) - m(x) \right) dx \right] + O(n^{-1})$$

$$= T_1 + T_2 + T_3 + O(n^{-1}),$$

say.

First, we consider the stochastic term, T_1. We obtain by standard arguments that

$$E_m\left[\int_0^1\left(\frac{1}{nh_n}\sum_{t=1}^n K\left(\frac{x-X_{t-1}}{h_n}\right)\sigma(X_{t-1})\varepsilon_t\right)^2 dx\right]^2$$

$$= \frac{1}{n^4 h_n^2}\sum_{t_1,\ldots,t_4=1}^n E_m\left[\int_0^1 \frac{1}{h_n}K\left(\frac{x-X_{t_1-1}}{h_n}\right)K\left(\frac{x-X_{t_2-1}}{h_n}\right)dx\right.$$

$$\times\,\sigma(X_{t_1-1})\varepsilon_{t_1}\sigma(X_{t_2-1})\varepsilon_{t_2}$$

$$\left.\times\int_0^1 \frac{1}{h_n}K\left(\frac{x-X_{t_3-1}}{h_n}\right)K\left(\frac{x-X_{t_4-1}}{h_n}\right)dx\,\sigma(X_{t_3-1})\varepsilon_{t_3}\sigma(X_{t_4-1})\varepsilon_{t_4}\right]$$

$$= \frac{1}{(nh_n)^2}\left(\int \frac{1}{h_n}K^2\left(\frac{x-y}{h_n}\right)\sigma^2(x)\pi_m(x)\,dx\right)^2 + o\left(\frac{1}{(nh_n)^2}\right) = O\left(\frac{1}{(nh_n)^2}\right),$$

which implies by the Cauchy–Schwarz inequality that

$$(5.8)\quad E_m\left[I(\mathbb{X}_n\in\mathcal{X}_n)\int_0^1(\pi_m(x))^{-2}\left(\frac{1}{nh_n}\sum_{t=1}^n K\left(\frac{x-X_{t-1}}{h_n}\right)\sigma(X_{t-1})\varepsilon_t\right)^2 dx\right]$$

$$= E_m\left[\frac{1}{nh_n}\int_0^1(\pi_m(x))^{-2}\frac{1}{h_n}K\left(\frac{x-X_0}{h_n}\right)^2 dx\,\sigma^2(X_0)\right] + o\left((nh_n)^{-1}\right).$$

Moreover, it holds for $j=0,\ldots,\beta-1$ with $g_0(x)=1$ that

$$E_m\left[\int_0^1\left(\frac{1}{nh_n}\sum_{t=1}^n K\left(\frac{x-X_{t-1}}{h_n}\right)g_j\left(\frac{x-X_{t-1}}{h_n}\right)\sigma(X_{t-1})\varepsilon_t\right)^2 dx\right] = O\left((nh_n)^{-1}\right),$$

which implies, in conjunction with (5.6), that

$$(5.9)\qquad T_1 = \frac{1}{nh_n}\int K^2(u)\,du\int_0^1\frac{\sigma^2(x)}{\pi_m(x)}\,dx + o\left(\frac{1}{nh_n}\right).$$

Now we turn to the bias-type term, T_2. We use the Taylor expansion

$$m(y)-m(x) = \sum_{k=1}^{\beta-1}m^{(k)}(x)\frac{(y-x)^k}{k!} + \int_x^y\frac{(y-z)^{\beta-1}}{(\beta-1)!}m^{(\beta)}(z)\,dz,$$

with the usual convention that $\int_x^y\ldots = -\int_y^x\ldots$ if $y<x$. For $\mathbb{X}_n\in\mathcal{X}_n$, we have

that

$$(5.10) \quad \sum_{t=1}^{n} K_n(x, X_{t-1}) m(X_{t-1}) - m(x)$$

$$= \sum_{t=1}^{n} K_n(x, X_{t-1}) \int_x^{X_{t-1}} \frac{(X_{t-1} - z)^{\beta-1}}{(\beta - 1)!} m^{(\beta)}(z) \, dz$$

$$= \frac{1}{\pi_m(x)} \sum_{t=1}^{n} \frac{1}{nh_n} K\left(\frac{x - X_{t-1}}{h_n}\right) \int_x^{X_{t-1}} \frac{(X_{t-1} - z)^{\beta-1}}{(\beta - 1)!} m^{(\beta)}(z) \, dz$$

$$+ \sum_{j=0}^{\beta-1} \left(\widehat{c}_j(x) - c_j(x)\right) \sum_{t=1}^{n} \frac{1}{nh_n} K\left(\frac{x - X_{t-1}}{h_n}\right) g_j\left(\frac{x - X_{t-1}}{h_n}\right)$$

$$\times \int_x^{X_{t-1}} \frac{(X_{t-1} - z)^{\beta-1}}{(\beta - 1)!} m^{(\beta)}(z) \, dz.$$

Denote

$$f_n(u, v) = \frac{1}{(nh_n)^2} \int_0^1 \frac{1}{\pi_m^2(x)} \left[K\left(\frac{x - u}{h_n}\right) \int_x^u \frac{(u - z)^{\beta-1}}{(\beta - 1)!} m^{(\beta)}(z) \, dz \right.$$

$$\left. \times K\left(\frac{x - v}{h_n}\right) \int_x^v \frac{(v - z)^{\beta-1}}{(\beta - 1)!} m^{(\beta)}(z) \, dz \right] dx.$$

Since K is supported on $[-C_K, C_K]$, we obtain the following rough bound for $|f_n(u, v)|$:

$$|f_n(u, v)| \leq C \frac{1}{(nh_n)^2} h_n^{2\beta-2} \int_{\max\{u,v\}-C_K h_n}^{\min\{u,v\}+C_K h_n} dx$$

$$\times \int_{u-C_K h_n}^{u+C_K h_n} |m^{(\beta)}(z)| \, dz \int_{v-C_K h_n}^{v+C_K h_n} |m^{(\beta)}(z)| \, dz$$

$$\leq \frac{2 \, C_K \, C}{n^2} h_n^{2\beta-3} \left[(\min\{u, v\} - \max\{u, v\} + 2C_K h_n) \vee 0 \right]$$

$$\times \left\{ \int_{u-C_K h_n}^{u+C_K h_n} (m^{(\beta)}(z))^2 \, dz + \int_{v-C_K h_n}^{v+C_K h_n} (m^{(\beta)}(z))^2 \, dz \right\}.$$

Hence,

$$\sup_{u,v} |f_n(u, v)| = O\left(n^{-2} h_n^{2\beta-2}\right).$$

Moreover, since π_m is bounded,

$$E_m |f_n(X_0, X_0)| = O\left(n^{-2} h_n^{2\beta-1}\right),$$

and since all joint densities are also bounded,

$$E_m |f_n(X_s, X_t)| = O\left(n^{-2} h_n^{2\beta}\right),$$

uniformly in $s \neq t$. Denote by X_0' a random variable independent of X_0 and with the same distribution. Then we obtain the more precise result

$$
\begin{aligned}
E_m[f_n(X_0, X_0')] &= \int_0^1 \Bigg[\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) \Big(\int_x^y \frac{(y-z)^{\beta-1}}{(\beta-1)!} m^{(\beta)}(z)\, dz \Big) dy \\
&\quad + \int \Big(\frac{\pi_m(y)}{\pi_m(x)} - 1 \Big) \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) \Big(\int_x^y \frac{(y-z)^{\beta-1}}{(\beta-1)!} m^{(\beta)}(z)\, dz \Big) dy \Bigg]^2 dx \\
&= \int_0^1 \Bigg[\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) \Big(\int_x^y \frac{(y-z)^{\beta-1}}{(\beta-1)!} m^{(\beta)}(z)\, dz \Big) dy \Bigg]^2 dx + o(h_n^{2\beta}) \\
&= \int_0^1 \Big(\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) m(y)\, dy - m(x) \Big)^2 dx + o(h_n^{2\beta}).
\end{aligned}
$$

These estimates imply that

$$
\begin{aligned}
(5.11) \quad E_m\Bigg[I(\mathbb{X}_n \in \mathcal{X}_n) \int_0^1 \Big(\frac{1}{\pi_m(x)} \sum_{t=1}^n \frac{1}{nh_n} K\Big(\frac{x - X_{t-1}}{h_n}\Big) \\
\times \int_x^{X_{t-1}} \frac{(X_{t-1} - z)^{\beta-1}}{(\beta-1)!} m^{(\beta)}(z)\, dz \Big)^2 dx \Bigg] \\
= n^2 E_m[f_n(X_0, X_0')] + O\Big(h_n^{2\beta} \Big(\frac{1}{nh_n} + \frac{\log n}{n} \Big) \Big) \\
= \int_0^1 \Big(\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) m(y)\, dy - m(x) \Big)^2 dx + o(h_n^{2\beta}).
\end{aligned}
$$

Furthermore, we obtain by analogous considerations that

$$
\begin{aligned}
E_m\Bigg[\int_0^1 \Big(\sum_{t=1}^n \frac{1}{nh_n} K\Big(\frac{x - X_{t-1}}{h_n}\Big) g_j\Big(\frac{x - X_{t-1}}{h_n}\Big) \\
\times \int_x^{X_{t-1}} \frac{(X_{t-1} - z)^{\beta-1}}{(\beta-1)!} m^{(\beta)}(z)\, dz \Big)^2 dx \Bigg] = O(h_n^{2\beta}),
\end{aligned}
$$

which implies, in conjunction with (5.6) and (5.19), that

$$
(5.12) \qquad T_2 = \int_0^1 \Big(\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) m(y)\, dy - m(x) \Big)^2 dx + o(h_n^{2\beta}).
$$

Finally, it can be shown by straightforward calculations that

$$
(5.13) \qquad\qquad T_3 = o\Big(\frac{1}{nh_n} + h_n^{2\beta} \Big).
$$

Now (5.7), (5.9), (5.12), and (5.13) imply that

$$
\begin{aligned}
(5.14) \quad E_m\Bigg[\int_0^1 \big(\widehat{m}_n(x) - m(x)\big)^2 dx \wedge 1 \Bigg] &= \frac{1}{nh_n} \int K^2(u)\, du \int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx \\
&\quad + \int_0^1 \Big(\int \frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) m(y)\, dy - m(x) \Big)^2 dx + o\Big(\frac{1}{nh_n} + h_n^{2\beta} \Big).
\end{aligned}
$$

The second assertion is a well-known result. $\quad\square$

Proof of Theorem 4.1. This proof follows the pattern of the proof of the lower asymptotic risk bound in Butucea and Neumann (2005) and bears many similarities to the corresponding proof in Golubev and Nussbaum (1990). We include it since the dependence in the present context and the slightly modified loss function require certain modifications.

Let $\epsilon > 0$ be arbitrary. We will actually show that

$$(5.15) \quad R_n = \inf_{\widetilde{m}_n} \sup_{m \in \mathcal{U}_n} \left\{ n^{\frac{2\beta}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_m(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}} E_m \|\widetilde{m}_n - m\|_{L_2([0,1])}^2 \right\}$$

$$\geq \gamma(\beta) L^{\frac{1}{2\beta+1}} - \epsilon$$

holds for sufficiently large n, where

$$\mathcal{U}_n = \left\{ m_\theta^{(n)} : \theta \in \Theta_n \right\}$$

is an appropriate sequence of asymptotically least favorable parametric subclasses of $W_2^\beta(L) \cap \mathcal{M}$. The functions $m_\theta^{(n)}$ are of the form

$$(5.16) \quad m_\theta^{(n)}(x) = \sum_{j=1}^s \sum_{k=1}^{q_n} \theta_{j,k} \phi_{j,k}^{(n)}(x),$$

where $\phi_{j,k}^{(n)}$ is supported on the interval $[(k-1)/q_n, k/q_n]$ and will be derived from ϕ_j described below.

To define appropriate functions ϕ_j, $j \in \mathbb{N}$, we consider the eigenvalue problem

$$(-1)^\beta f^{(2\beta)}(x) = \lambda f(x), \qquad x \in [0,1],$$

with boundary conditions $f^{(k)}(0) = f^{(k)}(1) = 0$ for $k = 0, \ldots, \beta-1$. We arrange the solutions in such a way that the eigenvalues $(\lambda_j)_{j \in \mathbb{N}}$ are nondecreasing and choose the corresponding eigenfunctions $(f_j)_{j \in \mathbb{N}}$ so that they are orthonormal. (They are automatically orthogonal if they belong to different eigenvalues; to orthonormalize them we can use the Gram–Schmidt orthogonalization algorithm.) It is known that the eigenvalues satisfy the asymptotic relation

$$\lambda_j = (\pi j)^{2\beta}(1 + o(1)) \qquad \text{as} \quad j \to \infty,$$

see, e.g., Neumark (1960, Section II.4.9). From integration by parts we obtain that

$$\int_0^1 f_j^{(\beta)}(x) f_k^{(\beta)}(x) \, dx = \lambda_j \delta_{j,k} \quad \forall j, k \in \mathbb{N}.$$

Let ϕ_j be the continuation of f_j by zero outside the interval $[0,1]$. Then $\phi_j \in \widetilde{W}_2^\beta = \{ f \in L_2(\mathbb{R}) : f^{(\beta)} \in L_2(\mathbb{R}) \text{ and } f^{(k)}(0) = f^{(k)}(1) = 0 \ \forall k = 0, \ldots, \beta-1 \}$. We define

$$q_n = \left[s^{-1} \left(\frac{nL}{\int_0^1 \sigma^2(x)/\pi_0(x) \, dx \int_0^{1/\pi} (\pi x)^\beta (1 - (\pi x)^\beta) \, dx} \right)^{\frac{1}{2\beta+1}} \right],$$

where the choice of the integer s will be described below. (Here π_0 denotes the stationary density in the case $m \equiv 0$.) Furthermore, let

$$c_{k,n} = q_n \int_{(k-1)/q_n}^{k/q_n} \frac{\sigma^2(x)}{\pi_0(x)} \, dx.$$

We define the perturbations $\phi_{j,k}^{(n)}$ $(j = 1, \ldots, q_n; \; k = 1, \ldots, s)$ as

$$\phi_{j,k}^{(n)}(x) = n^{-1/2} \, q_n^{1/2} \, c_{k,n}^{1/2} \, \phi_j(q_n x - (k-1)).$$

Note that the function $m_\theta^{(n)}$ belongs to $W_2^\beta(L)$ if and only if the parameter $\theta = (\theta_{j,k})_{j=1,\ldots,q_n; k=1,\ldots,s}$ is contained in

$$(5.17) \qquad \Theta_n = \left\{ \theta \in \mathbb{R}^{sq_n} : \frac{1}{n} q_n^{2\beta} \sum_{j=1}^{s} \lambda_j \sum_{k=1}^{q_n} c_{k,n} \theta_{j,k}^2 \le L \right\}.$$

The risk R_n will be estimated by a certain Bayesian risk, which will enable us to calculate a lower efficiency bound explicitly. A sharp asymptotic risk bound can then be obtained by taking a sequence of asymptotically least favorable prior distributions. In view of available results in related settings, it could be anticipated that this can be achieved by sequences of asymptotically normal priors. Denote by $(\mu_N)_{N \in \mathbb{N}}$ a sequence of distributions with finite support, $\int x^2 \mu_N(dx) < 1$ and $\int x^4 \mu_N(dx) \le 3$ $\forall N$, and $\mu_N \Longrightarrow \mathcal{N}(0,1)$ as $N \to \infty$. Let $\mu_{N,j}$ be the distribution of a random variable $s_j Z_N$, where $Z_N \sim \mu_N$ and $s_j^2 = a(j/s)$, $j = 1, \ldots, s$, and $a(x) = (\pi x)^{-\beta} (1 - (\pi x)^\beta)_+$. As prior measure for the parameter vector θ, we take the product measure $\mu_N^{(n)} = \otimes_{j=1}^{s} \mu_{N,j}^{\otimes q_n}$, where $\theta_{j,k} \sim \mu_{N,j}$.

Now we obtain that

$$(5.18) \quad R_n \ge \inf_{\widetilde{m}_n} \left\{ n^{\frac{2\beta}{2\beta+1}} \int_{\Theta_n} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_{m_\theta^{(n)}}(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}} \right.$$

$$\left. \times \, E_{m_\theta^{(n)}} \| \widetilde{m}_n - m_\theta^{(n)} \|^2 \, \mu_N^{(n)}(d\theta) \right\} \ge R_{n,1}(1 + o(1)) - R_{n,2},$$

say, where

$$R_{n,1} = n^{\frac{2\beta}{2\beta+1}} \inf_{\widetilde{m}} \int_{\mathrm{supp}(\mu_N^{(n)})} E_{m_\theta^{(n)}} \| \widetilde{m} - m_\theta^{(n)} \|^2 \, \mu_N^{(n)}(d\theta) \left(\int_0^1 \frac{\sigma^2(x)}{\pi_0(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}},$$

$$R_{n,2} = n^{\frac{2\beta}{2\beta+1}} \sup_{\theta \in \mathrm{supp}(\mu_N^{(n)})} \left\{ \left(\int_0^1 \frac{\sigma^2(x)}{\pi_{m_\theta^{(n)}}(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}} \right\}$$

$$\times \sup_{\theta_1, \theta_2 \in \mathrm{supp}(\mu_N^{(n)})} \left\{ \| m_{\theta_1}^{(n)} - m_{\theta_2}^{(n)} \|^2 \right\} \mu_N^{(n)}(\Theta_n^c).$$

(The $o(1)$ term in (5.18) is due to the fact that $\pi_{m_\theta^{(n)}}(x)$ is replaced by $\pi_0(x)$ in the definition of $R_{n,1}$; the fact that $\sup_{\theta \in \Theta_n} \sup_{x \in [0,1]} |\pi_{m_\theta^{(n)}}(x) - \pi_0(x)| \xrightarrow[n \to \infty]{} 0$

follows along the lines of the proof of Theorem 3 in Franke, Kreiss, Mammen, and Neumann (2002). The second inequality in (5.18) follows from convexity of $\{m_\theta^{(n)} : \theta \in \Theta_n\}$, which implies that the Bayes estimator lies in this set.) Let $\Delta = \Delta(N) = 1 - \int x^2 \, \mu_N(dx)$. We choose s such that

$$\frac{1}{s^{\beta+1}} \sum_{j=1}^{s} \lambda_j s_j^2 \le \frac{1 - \Delta/2}{1 - \Delta} \int_0^{1/\pi} (\pi x)^\beta \big(1 - (\pi x)^\beta\big) \, dx.$$

Now we obtain that

$$\frac{1}{n} q_n^{2\beta} \sum_{j=1}^{s} \lambda_j \sum_{k=1}^{q_n} c_{k,n} E_{\mu_{N,j}}[\theta_{j,k}^2] = \frac{1}{n} q_n^{2\beta} \sum_{j=1}^{s} \lambda_j \int_0^1 \frac{\sigma^2(x)}{\pi_0(x)} \, dx \, (1 - \Delta) s_j^2 \le (1 - \Delta/2) L,$$

which implies by the weak law of large numbers that

$$\mu_N^{(n)}(\Theta_n^c) \xrightarrow[n\to\infty]{} 0.$$

Since

$$\sup_{\theta \in \mathrm{supp}(\mu_N^{(n)})} \left\{ \left(\int_0^1 \frac{\sigma^2(x)}{\pi_{m_\theta^{(n)}}(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}} \right\} = O(1)$$

and

$$\sup_{\theta_1, \theta_2 \in \mathrm{supp}(\mu_N^{(n)})} \left\{ \|m_{\theta_1}^{(n)} - m_{\theta_2}^{(n)}\|^2 \right\} = O(n^{-\frac{2\beta}{2\beta+1}}),$$

we obtain that

(5.19) $$R_{n,2} \xrightarrow[n\to\infty]{} 0.$$

Now we analyze the term $R_{n,1}$. Using the orthonormality of the perturbations we obtain that

(5.20) $$R_{n,1} = n^{-\frac{1}{2\beta+1}} \left(\int_0^1 \frac{\sigma^2(x)}{\pi_0(x)} \, dx \right)^{-\frac{2\beta}{2\beta+1}}$$

$$\times \sum_{j=1}^{s} \sum_{k=1}^{q_n} c_{k,n} \inf_{\widetilde{\theta}_{j,k}} \int_{\mathrm{supp}(\mu_N^{(n)})} E_\theta\big[|\widetilde{\theta}_{j,k} - \theta_{j,k}|^2\big] \, \mu_N^{(n)}(d\theta)$$

$$\ge n^{-\frac{1}{2\beta+1}} q_n \left(\int_0^1 \frac{\sigma^2(x)}{\pi_0(x)} \, dx \right)^{\frac{1}{2\beta+1}}$$

$$\times \sum_{j=1}^{s} \min_{1 \le k \le q_n} \inf_{\substack{\theta_{l,m} \in \mathrm{supp}(\mu_{N,j}), \\ (l,m) \ne (j,k)}} \inf_{\widetilde{\theta}_{j,k}} \int_{\mathrm{supp}(\mu_{N,j})} E_\theta\big[|\widetilde{\theta}_{j,k} - \theta_{j,k}|^2\big] \mu_{N,j}(d\theta_{j,k}),$$

that is, we can reduce our considerations to the separate analysis of certain one-dimensional Bayesian estimation problems.

To establish the link to Gaussian shift experiments whose analysis finally leads to explicit lower bounds, we are going to prove local asymptotic normality (LAN) for the family of one-dimensional subexperiments given by

$$\left\{ m_u^{(j,k_n)} = \sum_{(l,m)\colon\,(l,m)\neq(j,k_n)} \theta_{l,m}^{(j,n)}\phi_{l,m}^{(n)} + u\phi_{j,k_n}^{(n)} : u \in [cs_j, cs_j] \right\},$$

where k_n and $(\theta_{l,m})_{(l,m)\neq(j,k_n)}$ are sequences which afford the minimum on the right-hand side of (5.20).

We obtain, under $P_{m_0^{(j,k_n)}}$, that

$$
\begin{aligned}
\Lambda_u^{(j,k_n)} &:= \log \frac{p_{m_u^{(j,k_n)}}(X_0,\dots,X_n)}{p_{m_0^{(j,k_n)}}(X_0,\dots,X_n)} \\
&= \log \frac{p_{m_u^{(j,k_n)}}(X_0)\varphi\left(\frac{X_1-m_u^{(j,k_n)}(X_0)}{\sigma(X_0)}\right)\cdots\varphi\left(\frac{X_n-m_u^{(j,k_n)}(X_{n-1})}{\sigma(X_{n-1})}\right)}{p_{m_0^{(j,k_n)}}(X_0)\varphi\left(\frac{X_1-m_0^{(j,k_n)}(X_0)}{\sigma(X_0)}\right)\cdots\varphi\left(\frac{X_n-m_0^{(j,k_n)}(X_{n-1})}{\sigma(X_{n-1})}\right)} \\
&= \log \frac{p_{m_u^{(j,k_n)}}(X_0)}{p_{m_0^{(j,k_n)}}(X_0)} + \sum_{t=1}^n \log \frac{\varphi\left(\frac{\varepsilon_t+m_0^{(j,k_n)}(X_{t-1})-m_u^{(j,k_n)}(X_{t-1})}{\sigma(X_{t-1})}\right)}{\varphi\left(\frac{\varepsilon_t}{\sigma(X_{t-1})}\right)} \\
&= \log \frac{p_{m_u^{(j,k_n)}}(X_0)}{p_{m_0^{(j,k_n)}}(X_0)} + u\sum_{t=1}^n \phi_{j,k_n}^{(n)}(X_{t-1})\frac{\varepsilon_t}{\sigma^2(X_{t-1})} - \frac{u^2}{2}\sum_{t=1}^n \frac{(\phi_{j,k_n}^{(n)}(X_{t-1}))^2}{\sigma^2(X_{t-1})}.
\end{aligned}
$$

According to Lemma 6.1 in Grama and Neumann (2006) we have that

$$\log\left(\frac{p_{m_u^{(j,k_n)}}(X_0)}{p_{m_0^{(j,k_n)}}(X_0)}\right) \xrightarrow{P} 0.$$

Therefore, we obtain by a central limit theorem for triangular arrays of strongly mixing random variables (see Politis, Romano, and Wolf, 1997, Theorem A.1) that

$$(5.21) \qquad\qquad \Lambda_u^{(j,k_n)} \xrightarrow{d} \mathcal{N}\left(-\frac{u^2}{2}, u^2\right).$$

Now we proceed in the same way as Golubev and Nussbaum (1990) did in the proof of their Theorem A1. Because of the LAN property (5.21), we obtain, for any fixed N,

$$
\begin{aligned}
(5.22) \qquad \liminf_{n\to\infty} \inf_{\widetilde{\theta}_{j,k_n}} \int_{\mathrm{supp}(\mu_{N,j})} & E_{\theta^{(j)}}\left[|\widetilde{\theta}_{j,k_n} - \theta_{j,k_n}|^2\right]\mu_{N,j}(d\theta_{j,k_n}) \\
& \geq \inf_{\widetilde{\theta}_j} \int_{\mathrm{supp}(\mu_{N,j})} \widetilde{E}_{\theta_j}\left[|\widetilde{\theta}_j - \theta_j|^2\right]\mu_{N,j}(d\theta_j) + o(1),
\end{aligned}
$$

where $\theta^{(j)}$ is the parameter vector consisting of $(\theta_{l,m}^{(j,k_n)})_{(l,m)\neq(j,k_n)}$ and θ_{j,k_n}, and $\widetilde{E}_{\theta_j}$ is the expectation in the Gaussian shift experiment, where

$$(5.23) \qquad\qquad Y_j = \theta_j + \xi_j$$

is observed and $\xi_j \sim \mathcal{N}(0,1)$. Moreover, it follows from the arguments in the proof of Theorem A1 in Golubev and Nussbaum (1990) that the right-hand side of (5.22) converges to the Bayesian risk for experiment (5.23) with normal prior $\mu_{\infty,j} = \mathcal{N}(0, s_j^2)$ as the parameter N tends to infinity; see also Theorem 3.1 of Neumann and Spokoiny (1995). Therefore, we obtain that

$$(5.24) \qquad \lim_{N\to\infty} \liminf_{n\to\infty} \inf_{\widetilde{\theta}_{j,k_n}} \int_{\mathrm{supp}(\mu_{N,j})} E_{\theta^{(j)}}\big[|\widetilde{\theta}_{j,k_n} - \theta_{j,k_n}|^2\big]\, \mu_{N,j}(d\theta_{j,k_n})$$

$$\geq \inf_{\widetilde{\theta}_j} \int \widetilde{E}_{\theta_j}\big[|\widetilde{\theta}_j - \theta_j|^2\big] p_{\mathcal{N}(0,s_j^2)}(\theta_j)\, d\theta_j = \frac{s_j^2}{1+s_j^2}.$$

Choosing c sufficiently large we obtain from (5.20) and (5.22) that

$$\liminf_{n\to\infty} R_{n,1} \geq \Big(1 - \frac{1}{2}\epsilon\Big) L^{\frac{1}{2\beta+1}} \bigg(\int_0^{1/\pi} (\pi x)^\beta \big(1 - (\pi x)^\beta\big)\, dx\bigg)^{-\frac{1}{2\beta+1}} \frac{1}{s} \sum_{j=1}^{s} \frac{s_j^2}{1+s_j^2}.$$

This yields, for s large enough,

$$\liminf_{n\to\infty} R_{n,1} \geq (1-\epsilon) L^{\frac{1}{2\beta+1}} \bigg(\int_0^{1/\pi} (\pi x)^\beta (1 - (\pi x)^\beta)\, dx\bigg)^{-\frac{1}{2\beta+1}} \int_0^{1/\pi} (1 - (\pi x)^\beta)\, dx$$

$$= (1-\epsilon) L^{\frac{1}{2\beta+1}}\, \gamma(\beta).$$

This implies, in conjunction with (5.18) and (5.19), that inequality (5.15) is fulfilled and, hence, that the assertion holds true. □

Proof of Theorem 4.2. A close inspection of the computations in the proof of Theorem 3.1 (in particular, (5.14) in that proof) and the remark to the proof of Theorem 3.2 reveal that

$$E_m \|\widehat{\widehat{m}}_n - m\|_{L_2([0,1])}^2 \leq \frac{1}{nh_n} \int K^2(u)\, du \int_0^1 \frac{\sigma^2(x)}{\pi_m(x)}\, dx$$

$$+ \int_0^1 \bigg(\frac{1}{h_n} K\Big(\frac{x-y}{h_n}\Big) m(y)\, dy - m(x)\bigg)^2 dx + o\bigg(\frac{1}{nh_n} + h_n^{2\beta}\bigg)$$

holds *uniformly* in $m \in W_2^\beta(L) \cap \mathcal{M}$.

Let $\epsilon > 0$ be arbitrary. Now we can replace m by a function $\widetilde{m} \in L_1(\mathbb{R})$ with the property that, for sufficiently small $\delta > 0$,

$$\widetilde{m}(x) = m(x), \qquad \text{for all} \quad x \in [-\delta, 1+\delta]$$

and

$$\int_{-\infty}^{\infty} \big(\widetilde{m}^{(\beta)}(x)\big)^2 dx \leq \int_0^1 \big(m^{(\beta)}(x)\big)^2 dx + \epsilon.$$

Therefore, we obtain, for n large enough, that

$$\int_0^1 \left(\frac{1}{h_n}K\left(\frac{x-y}{h_n}\right)m(y)\,dy - m(x)\right)^2 dx$$

$$= \int_0^1 \left(\frac{1}{h_n}K\left(\frac{x-y}{h_n}\right)\widetilde{m}(y)\,dy - \widetilde{m}(x)\right)^2 dx$$

$$\leq \int_{-\infty}^{\infty} \left(\frac{1}{h_n}K\left(\frac{x-y}{h_n}\right)\widetilde{m}(y)\,dy - \widetilde{m}(x)\right)^2 dx$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left(\widehat{\frac{1}{h_n}K\left(\frac{\cdot}{h_n}\right)}(\omega) - 1\right)^2 \left(\widehat{\widetilde{m}}(\omega)\right)^2 d\omega$$

$$\leq h_n^{2\beta}\sup_{\omega}\left\{(\widehat{K}(\omega)-1)^2\omega^{-2\beta}\right\}\frac{1}{2\pi}\int_{-\infty}^{\infty}\omega^{2\beta}\left(\widehat{\widetilde{m}}(\omega)\right)^2 d\omega$$

$$\leq h^{2\beta}\sup_{\omega}\left\{(\widehat{K}(\omega)-1)^2\omega^{-2\beta}\right\}(L+\epsilon),$$

which yields the first assertion of the theorem. The proof of the second assertion is straightforward. $\square$

Proof of Lemma 4.1. We will derive the kernel $K_{\beta,\delta}$ in two steps. First, we approximate $\widehat{K}_\beta$ by a sufficiently often differentiable function whose inverse Fourier transform constitutes a kernel of order β. Then, in order to get a kernel with compact support, we taper this kernel (in time domain) and add appropriate correction terms which preserve the required moment properties.

First, we choose a $(\beta + 2)$-times continuously differentiable function g with bounded support and

$$(5.25) \qquad\qquad \widehat{K}_\beta(\omega) \leq g(\omega) \leq 1 \qquad \forall \omega \in \mathbb{R}$$

and $\|g - \widehat{K}_\beta\|_{L_2} \leq \sqrt{2\pi}\delta/2$. The latter property implies that

$$(5.26) \qquad\qquad \|G - K_\beta\|_{L_2} \leq \delta/2,$$

where G is the inverse Fourier transform of g, that is,

$$G(x) = F(g)(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ix\omega}g(\omega)\,d\omega.$$

According to Theorem 1.8 on page 4 in Stein and Weiss (1990), we have that

$$(5.27) \qquad x^k G(x) = (-i)^k F(g^{(k)})(x), \qquad \text{for} \quad k = 0,\dots,\beta + 2.$$

Since $g \in L_1$ and $g^{(\beta+2)} \in L_1$, it follows in particular that

$$(5.28) \quad \int (1+|x|)^\beta\,|G(x)|\,dx \leq \sup_{y}\left\{(1+|y|)^{\beta+2}|G(y)|\right\}\int (1+|x|)^{-2}\,dx < \infty.$$

Therefore,

$$\int_{-\infty}^{\infty} x^k G(x)\, dx = (-i)^k g^{(k)}(0) = \delta_{k,0}, \qquad \text{for} \quad k = 0, \dots, \beta - 1,$$

that is, G is a kernel of order β.

Let $r\colon \mathbb{R} \to [0,1]$ be a Lipschitz continuous function with $r(x) = 1$ if $|x| \le 1$ and $r(x) = 0$ if $|x| \ge 2$. Let $G_N(x) = G(x) r(x/N)$. We will add appropriate correction terms to G_N to preserve the properties of a kernel of order β. To this end, we choose compactly supported and Lipschitz continuous functions $H_0, \dots, H_{\beta-1}$ such that $\int x^l H_k(x)\, dx = \delta_{k,l}$ for $l, k = 0, \dots, \beta - 1$. Define

$$m_{N,k} = \int_{-\infty}^{\infty} x^k G(x)(1 - r(x/N))\, dx.$$

Then K_N with

$$K_N(x) = G_N(x) + \sum_{k=0}^{\beta-1} m_{N,k} H_k(x)$$

is by construction a Lipschitz continuous kernel of order β. By (5.28) we have that $m_{N,k} \xrightarrow[N\to\infty]{} 0$, which implies that

(5.29) $$\|K_N - G\|_{L_2} \xrightarrow[N\to\infty]{} 0.$$

By (5.25) and the construction of K_N we have that

$$g^{(k)}(0) = \widehat{K}_N^{(k)}(0) = \delta_{k,0}, \qquad \text{for} \quad k = 0, \dots, \beta - 1.$$

Furthermore, it holds that

$$\left| g^{(\beta)}(\omega) - \widehat{K}_N^{(\beta)}(\omega) \right| \le \left| g^{(\beta)}(\omega) - \widehat{G}_N^{(\beta)}(\omega) \right| + \sum_{k=0}^{\beta-1} |m_{N,k}|\, |\widehat{H}_k^{(\beta)}(\omega)|$$

$$\le M_{N,\beta} + \sum_{k=0}^{\beta-1} |m_{N,k}| \sup_\nu |\widehat{H}_k^{(\beta)}(\nu)|,$$

where $M_{N,\beta} = \int_N^{\infty} |x|^\beta \big(|G(x)| + |G(-x)| \big)\, dx \xrightarrow[N\to\infty]{} 0$. This implies by a Taylor expansion that

(5.30) $$|g(\omega) - \widehat{K}_N(\omega)| = \left| \int_0^\omega \left(g^{(\beta)}(\nu) - \widehat{K}_N^{(\beta)}(\nu) \right) \frac{(\omega - \nu)^{\beta-1}}{(\beta-1)!}\, d\nu \right|$$

$$\le \sup_\nu \left| g^{(\beta)}(\nu) - \widehat{K}_N^{(\beta)}(\nu) \right| \frac{|\omega|^\beta}{\beta!} \xrightarrow[N\to\infty]{} 0.$$

Now (5.29) and (5.30) imply that there exists an $N_\delta < \infty$ such that

$$\|K_{N_\delta} - G\|_{L_2} \le \delta/2$$

and

$$\left| g(\omega) - \widehat{K}_{N_\delta}(\omega) \right| |\omega|^{-\beta} \leq \delta.$$

These two properties and (5.25) and (5.26) imply that $K_{\beta,\delta} := K_{N_\delta}$ has the claimed properties. $\quad\square$

References

[1] E. N. Belitser and B. Y. Levit (1996), *Asymptotically minimax nonparametric regression in L_2*, Statistics, **28**, 105–122.

[2] C. Butucea and M. H. Neumann (2005), *Exact asymptotics for estimating the marginal density of discretely observed diffusion processes*, Bernoulli, **11**, 411–444.

[3] W. S. Cleveland (1979), *Robust locally weighted regression and smoothing scatterplots*, J. Amer. Statist. Assoc., **74**, 829–836.

[4] A. S. Dalalyan (2005), *Sharp adaptive estimation of the trend coefficient for ergodic diffusions*, Ann. Statist., **33**, 2507–2528.

[5] A. S. Dalalyan and Yu. A. Kutoyants (2002), *Asymptotically efficient trend coefficient estimation for ergodic diffusions*, Math. Methods Statist., **11**, 402–427.

[6] P. Doukhan (1994), *Mixing: Properties and Examples. Lecture Notes in Statistics*, Vol. 85, Springer, New York.

[7] S. Yu. Efromovich and M. S. Pinsker (1981), *Estimating a square integrable spectral density from a sequence of observations*, Problemy Peredachi Informatsii, **17**, 50–68 (in Russian). [English transl., Problems Inform. Transmission, **17** (1982), 182–196.]

[8] S. Yu. Efromovich and M. S. Pinsker (1982), *Estimation of square-integrable probability density of a random variable*, Problemy Peredachi Informatsii, **18**, 19–38 (in Russian). [English transl., Problems Inform. Transmission, **18** (1983), 175–189.]

[9] J. Fan and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

[10] J. Franke, J.-P. Kreiss, E. Mammen, and M. H. Neumann (2002), *Properties of the nonparametric autoregressive bootstrap*, J. Time Series Anal., **23**, 555–585.

[11] T. Gasser and H.-G. Müller (1979), *Kernel estimation of regression functions. In: Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, Vol. 757, Springer, New York, pp. 23–68.

[12] G. K. Golubev (1987), *Adaptive asymptotically minimax estimators of smooth signals*, Problemy Peredachi Informatsii, **23**, 57–67 (in Russian). [English transl., Problems Inform. Transmission, **23**, 47–55.]

[13] G. K. Golubev and M. Nussbaum (1990), *A risk bound in Sobolev class regression*, Ann. Statist., **18**, 758–778.

[14] I. G. Grama and M. H. Neumann (2006), *Asymptotic equivalence of nonparametric autoregression and nonparametric regression*, Ann. Statist., **34** (4), to appear.

[15] V. Ya. Katkovnik (1979), *Linear and nonlinear methods of nonparametric regression analysis*, Avtomatika, No. 5, 35–46.

[16] O. V. Lepski, E. Mammen, and V. G. Spokoiny (1997), *Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors*, Ann. Statist., **25**, 929–947.

[17] O. V. Lepski and V. G. Spokoiny (1997), *Optimal pointwise adaptive methods in nonparametric estimation*, Ann. Statist., **25**, 2512–2546.

[18] E. A. Nadaraya (1964), *On estimating regression*, Teor. Veroyatn. Primen., **9**, 141–142 (in Russian). [English transl., Theory Probab. Appl., **10**, 186–190.]

[19] M. H. Neumann and J.-P. Kreiss (1998), *Regression-type inference in nonparametric autoregression*, Ann. Statist., **26**, 1570–1613.

[20] M. H. Neumann and V. G. Spokoiny (1995), *On the efficiency of wavelet estimators under arbitrary error distributions*, Math. Methods Statist., **4**, 137–166.

[21] M. A. Neumark (1960), *Lineare Differentialoperatoren*, Akademie-Verlag, Berlin. (In German).

[22] M. Nussbaum (1985), *Spline smoothing in regression models and asymptotic efficiency in L_2*, Ann. Statist., **13**, 984–997.

[23] M. S. Pinsker (1980), *Optimal filtration of square-integrable signals in Gaussian noise*, Problemy Peredachi Informatsii, **16**, 52–68 (in Russian). [English transl., Problems Inform. Transmission, **16**, 120–133.]

[24] D. N. Politis, J. P. Romano, and M. Wolf (1997), *Subsampling for heteroskedastic time series*, J. Econometrics, **81**, 281–317.

[25] M. B. Priestley and M. T. Chao (1972), *Non-parametric function fitting*, J. Roy. Statist. Soc., Ser. B, **34**, 385–392.

[26] N. A. Sprünker (2005), *Efficient estimation in nonparametric regression*, Diploma thesis, TU Braunschweig, Germany.

[27] E. M. Stein and G. Weiss (1990), *Introduction to Fourier Analysis on Euclidean Spaces*, 6th ed., Princeton Univ. Press.

[28] C. J. Stone (1977), *Consistent nonparametric regression*, Ann. Statist., **5**, 595–645.

[29] C. J. Stone (1980), *Optimal rates of convergence for nonparametric estimators*, Ann. Statist., **8**, 1348–1360.

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