*The Annals of Statistics*
1996, Vol. 24, No. 3, 1106–1125

# ON OPTIMAL ADAPTIVE ESTIMATION OF A QUADRATIC FUNCTIONAL

BY SAM EFROMOVICH[1] AND MARK LOW[2]

*University of New Mexico and University of Pennsylvania*

Minimax mean-squared error estimates of quadratic functionals of smooth functions have been constructed for a variety of smoothness classes. In contrast to many nonparametric function estimation problems there are both regular and irregular cases. In the regular cases the minimax mean-squared error converges at a rate proportional to the inverse of the sample size, whereas in the irregular case much slower rates are the rule.

We investigate the problem of adaptive estimation of a quadratic functional of a smooth function when the degree of smoothness of the underlying function is not known. It is shown that estimators cannot achieve the minimax rates of convergence simultaneously over two parameter spaces when at least one of these spaces corresponds to the irregular case. A lower bound for the mean squared error is given which shows that any adaptive estimator which is rate optimal for the regular case must lose a logarithmic factor in the irregular case. On the other hand, we give a rather simple adaptive estimator which is sharp for the regular case and attains this lower bound in the irregular case. Moreover, we explicitly describe a subset of functions where our adaptive estimator loses the logarithmic factor and show that this subset is relatively small.

**1. Introduction.** The problem of estimating quadratic functionals such as $F(f) = \int_0^1 f^2(t)\,dt$ has been analyzed in density estimation, nonparametric regression and white noise models. See, for example, Ibragimov and Khas'minskii (1980), Ibragimov, Nemirovskii and Khas'minskii (1986), Donoho and Nussbaum (1990), Fan (1991) and Efromovich (1994). Since the equivalence results of Brown and Low (1992) and Nussbaum (1994) show that these statistical models are asymptotically equivalent whenever the underlying functions are sufficiently smooth, we shall restrict our attention to the classical filtering model

$$Y_n(t) = \int_0^t f(u)\,du + n^{-1/2}B(t), \qquad t \in [0,1], \tag{1.1}$$

where $B(t)$ is a standard Brownian motion on $[0,1]$ and we observe $Y_n = (Y_n(t),\ 0 \le t \le 1)$.

Received November 1992; revised September 1995.
[1]Work partially supported by NSF Grant DMS-91-23956.
[2]Work partially supported by an NSF Postdoctoral Research Fellowship.
*AMS* 1991 *subject classifications*. Primary 62C05; secondary 62E20, 62J02, 62G04, 62M99.
*Key words and phrases.* Functional estimation, adaptation, optimal risk convergence, filtering, nonparametric function.

Following Donoho, Liu and MacGibbon (1990), Donoho and Nussbaum (1990) and Fan (1991), we assume that the estimated function $f$ belongs to the hyperrectangle $H(\alpha, Q) = \{f\colon f(t) = \sum_{j=0}^{\infty} \theta_j \varphi_j(t), |\theta_0| < Q, |\theta_{2j-1}| \le Qj^{-\alpha}, |\theta_{2j}| \le Qj^{-\alpha}, \ j = 1, 2, \dots\}$, where $\{\varphi_0(x) = 1, \ \varphi_{2j-1}(t) = \sqrt{2}\sin(2\pi jt), \varphi_{2j}(t) = \sqrt{2}\cos(2\pi jt)\}$ is the classical trigonometric basis in $L_2(0,1)$ and $\theta_j = \langle f, \varphi_j \rangle$ is the corresponding Fourier coefficient, where $\langle f, \varphi \rangle = \int_0^1 f(t)\varphi(t)\,dt$ denotes the usual inner product in $L_2(0,1)$. The interested reader can easily extend the results in this paper to other familiar classes such as Lipschitz, Hölder and Sobolev classes.

Before we turn to the problem of adaptive estimation, let us recall the known results for minimax estimation. Fan (1991) shows that the maximum risk

$$R(\alpha, Q, \hat{F}_n) = \sup_{f \in H(\alpha, Q)} E_f\left\{\left(\hat{F}_n - F(f)\right)^2\right\} \tag{1.2}$$

satisfies the asymptotic equation

$$\begin{aligned} R(\alpha) &= \inf R(\alpha, Q, \hat{F}_n) \\ &\asymp \begin{cases} n^{-1}, & \text{if } \alpha \ge 3/4, \\ n^{-4(2\alpha-1)/(4\alpha-1)}, & \text{if } 1/2 < \alpha < 3/4, \end{cases} \end{aligned} \tag{1.3}$$

where the infimum is taken over all possible estimators $\hat{F}_n = \hat{F}_n(Y_n, \alpha, Q)$ and $a_n \asymp b_n$ means that there is a constant $C$ such that $C^{-1} < a_n/b_n < C$. If $\alpha \le 1/2$, then consistent estimation is impossible. Moreover, it is known from Ibragimov and Khas'minskii (1980) and Efromovich (1994) that $R(\alpha) = 4F(f)n^{-1}(1 + o(1))$ for $\alpha \ge 3/4$.

The line (1.3) exhibits the so-called elbow phenomenon in optimal rates and the existence of two natural cases for quadratic functional estimation: (i) a regular case, where the rate is proportional to the inverse of sample size, and (ii) an irregular case when the rate is significantly slower.

Many different rate optimal estimators are already known. We now present a projection estimator which is robust in $\alpha$ for the regular case:

$$\hat{F}(n, J(\alpha)) = \sum_{j=0}^{J(\alpha)} \left[\langle Y, \varphi_j \rangle^2 - n^{-1}\right]. \tag{1.4}$$

The window size (or so-called cutoff) $J(\alpha)$ is defined as

$$J(\alpha) = \begin{cases} n, & \text{if } \alpha \ge 3/4, \\ \lfloor n^{2/(4\alpha-1)} \rfloor, & \text{if } 1/2 < \alpha < 3/4, \end{cases} \tag{1.5}$$

where $\lfloor x \rfloor$ is the maximal integer which is not greater than $x$. From now on, so as to simplify notation, we drop the argument $n$ in functions such as $J(\alpha)$ whenever this does not lead to confusion.

Note that the estimator (1.4) achieves the optimal rate of convergence of $n^{-1}$ in the regular case when $\alpha \ge 3/4$. This permits us to restrict our attention only to the case $\alpha \in (1/2, 3/4]$.

In this paper we study the problem of estimating $F(f)$, $f \in H(\alpha, Q)$, using sequences of estimators which simultaneously minimize the maximum risk over a range of values $\alpha$, where $1/2 < \alpha \le 3/4$. Moreover, we explore the problem of adaptive sharp-optimal estimation when the error converges with optimal constant and rate whenever $\alpha > 3/4$. We shall see that in the irregular case our problem resembles the problem of adaptive estimation of linear functionals first studied by Lepskii (1990, 1992). See also Brown and Low (1992) and Efromovich and Low (1994, 1996).

In Section 2, by combining an inequality of Brown and Low (1992) and ideas of Ingster (1986), a lower bound is given to the maximum risk subject to an upper bound on the risk at a particular point. The bounds are developed for a large class of quadratic functionals over arbitrary orthosymmetric sets. A simple consequence of this inequality is that there do not exist sequences of estimators which attain the minimax rate of convergence given in (1.3) for two different values of $\alpha$ in the range $1/2 < \alpha \le 3/4$. More specifically, it is shown that attaining the minimax rate of convergence on the smaller parameter set results in at least an extra logarithmic factor in the maximum risk over the larger parameter space.

In Section 3 we introduce a rate-optimal estimator. In Section 4 we adopt a Bayes approach which shed light on the subset of functions where our estimator loses the extra logarithmic factor and it allows us to conclude that this subset is relatively small. Proofs are given in Section 5.

**2. Lower bounds for adaptive estimators.** It might be hoped that a sequence of estimators could be found which attains the minimax rate of convergence for mean-squared error over $H(\alpha, Q)$ for a range of values of $\alpha$. Unfortunately this is not generally the case even for two different values of $\alpha$, that is, $\alpha \in \{\beta_1, \beta_0\}$, $\beta_1 < \beta_0$, whenever $\beta_1 < 3/4$.

This result is a straightforward consequence of the following theorem, which gives a lower bound for the maximum risk subject to an upper bound for the risk at a particular parameter point.

Recall that the white noise observations (1.1) are equivalent to noisy observations of the Fourier coefficients given by

$$\eta_i = \theta_i + n^{-1/2} e_i, \qquad i = 0, 1, 2, \dots, \tag{2.1}$$

where $e_i$ are iid standard normal random variables.

Let $\Theta = (\theta_0, \theta_1, \dots)$. The set $\Phi$ is called orthosymmetric if whenever $\Theta \in \Phi$, then $(\pm\theta_0, \pm\theta_1, \dots) \in \Phi$ for all possible sequences of signs. Minimax theory over such sets was developed in Donoho and Nussbaum (1990).

Let $G(\Theta) = \sum_{i=0}^{\infty} q_i \theta_i^2$, where $q_i \ge 0$. In particular, if $q_i = 1$ for all $i$, then $G(\Theta)$ is equal to the quadratic functional $F(f)$ for $f = \sum_{i=0}^{\infty} \theta_j \varphi_j(t)$.

Define the function $w(\delta)$ by

$$w(\delta) = \sup\left\{G(\Theta) : \sum_{i=0}^{\infty} \theta_i^4 \le \delta^2, \Theta \in \Phi\right\}. \tag{2.2}$$

For any estimate $\hat{G}_n$ of $G(\Theta)$, based on (2.1), write

$$R(\Theta, \hat{G}_n) = E\Big\{\big(G(\Theta) - \hat{G}_n\big)^2\Big\}. \tag{2.3}$$

THEOREM 2.1. *Assume that* $\Phi$ *is orthosymmetric and set* $S(\delta) = \{\Theta: \sum_{i=0}^{\infty} \theta_i^4 \le \delta^2\}$. *If*

$$R(\Theta, \hat{G}_n) \le \varepsilon^2 \tag{2.4}$$

*for some positive* $\varepsilon$, *then*

$$\sup_{\Phi \cap S(\delta)} R(\Theta, \hat{G}_n) \ge w^2(\delta)\big[1 - 2\exp\{n^2\delta^2\}\varepsilon w^{-1}(\delta)\big]. \tag{2.5}$$

The proof of Theorem 2.1 is delayed until Section 4. In this section we use this theorem to show that it is not possible to construct a single sequence of estimators which achieve minimax rates of convergence over $H(\alpha, Q)$ for two different valeus of $\alpha \in \{\beta_0, \beta_1\}$ such that $1/2 < \beta_1 < \beta_0 \le 3/4$.

Set $a_n = c[\ln(n)n^{-2}]^{\beta_1/(4\beta_1 - 1)}$ for some positive constant $c$ and let

$$B(\beta) = [\ln(n)]^{(4\beta - 2)/(4\beta - 1)}, \tag{2.6}$$

$$J^*(\beta) = \Big\lfloor J(n, \beta)[B(\beta)]^{-1/(4\beta - 2)}\Big\rfloor, \tag{2.7}$$

where $J(\beta)$ is defined in(1.5).

For each $c$ which defines $a_n$ let $\Phi_n(c) = \{\Theta: \theta_i = \pm a_n, i = 0, 1, \dots, J^*(\beta_1), \theta_i = 0, i > J^*(\beta_1)\}$ and $\mathscr{R}_n^*(c) = \{f: f(t) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \Theta \in \Phi_n(c)\}$.

COROLLARY 2.1. *Let* $R(\alpha, Q, \hat{F}_n)$ *be defined by* (1.2), *where* $\hat{F}_n$ *is a sequence of estimators based on the white noise model* (1.1). *If* $1/2 < \beta_1 < \beta_0 \le 3/4$ *and*

$$R(\beta_0, Q, F_n) \asymp R(\beta_0), \tag{2.8}$$

*then there exists a positive constant* $c_0$ *such that for all* $0 < c \le c_0$,

$$\sup_{f \in H(\beta_1, Q) \cap \mathscr{R}_n^*(c)} E_f\Big\{\big(\hat{F}_n - F(f)\big)^2\Big\} \ge C^* R(\beta_1) B(\beta_1), \tag{2.9}$$

*where* $C^*$ *is a positive constant.*

To prove Corollary 2.1, set $G(\Theta) = \sum_{i=0}^{\infty} \theta_i^2$. Obviously $G(\Theta) = F(f)$ for $f(t) = \sum_{i=0}^{\infty} \theta_i \varphi_i(t)$. Then, for all $\Theta \in \Phi_n(c)$ simple algebra shows that $G(\Theta) \asymp c^2[\ln(n)n^{-2}]^{(2\beta_1 - 1)/(4\beta_1 - 1)}$ and $n^2 \sum_{i=0}^{\infty} \theta_i^4 \asymp c^4 \ln(n)$. Hence, if $R(\Theta, \hat{G}_n) \le dn^{-(8\beta_0 - 4)/(4\beta_0 - 1)} = \gamma_n^2$, then

$$\sup_{\Theta \in \Phi_n(c)} R(\Theta, \hat{G}_n) \ge \Big[c^2\big(\ln(n)n^{-2}\big)^{(2\beta_1 - 1)/(4\beta_1 - 1)}\Big]^2$$
$$\times\Big[1 - 4\exp\{2c^2\ln(n)\}\gamma_n c^{-2}\big(\ln(n)n^{-2}\big)^{(2\beta_1 - 1)/(4\beta_1 - 1)}\Big].$$

This inequality together with the assumption $\beta_0 > \beta_1 > 1/4$ means that for sufficiently small $c$,

$$\sup_{\Theta \in \Phi_n(c)} R\big(\Theta, \hat{G}_n\big) > C\big[\ln(n) n^{-2}\big]^{(4\beta_1 - 2)/(4\beta_1 - 1)} \asymp R(\beta_1) B(\beta_1). \tag{2.10}$$

Since $\Phi_n(c)$ is a subset of $H(\beta_1, Q) \cap \mathscr{R}_n^*(c)$ for all sufficiently small $c$, we obtain (2.9) and Corollary 2.1 is proved.

The following adaptive projection estimator shows that the lower bound given in (2.9) is sharp. See also the discussion in Section 4. Set for $1/2 < \beta_1 \le \min\{3/4, \beta_0\}$,

$$R^*(\beta_1) = 2J^*(\beta_1) n^{-2}, \tag{2.11}$$

$$\hat{I}(J_0, J_1) = \hat{F}(n, J_1) - \hat{F}(n, J_0), \tag{2.12}$$

$$I(J_0, J_1) = \sum_{j=J_0+1}^{J_1} \theta_j^2, \tag{2.13}$$

$$\begin{aligned} \mathscr{R}(\beta_0, \beta_1) &= \big\{f\colon I^2\big(J(\beta_0), J(\beta_1)\big) > C_0 R(\beta_1)\big\} \\ &\quad \cap \big\{f\colon I^2\big(J(\beta_0), J^*(\beta_1)\big) < 2\ln(n) R^*(\beta_1)(1 + \nu_n)\big\} \\ &\triangleq \mathscr{R}' \cap \mathscr{R}'', \end{aligned} \tag{2.14}$$

where $\nu_n = [\ln(n)]^{-1/(8\beta_1 - 1)}$ and $C_0 > 1$ is an arbitrarily large constant.

Assume that $\alpha \in \{\beta_0, \beta_1\}$ and define the statistic

$$\hat{k} = \begin{cases} 0, & \text{if } \hat{I}^2\big(J(\beta_0), J^*(\beta_1)\big) \le 2\ln(n) R^*(\beta_1), \\ 1, & \text{otherwise.} \end{cases} \tag{2.15}$$

THEOREM 2.2. *Let $\beta_0$ and $\beta_1$ be given such that $1/2 < \beta_1 < \min\{3/4, \beta_0\}$. Then the projection estimator* (1.4) *with adaptive window size $J(\beta_k)$ satisfies*,

$$\sup_{f \in H(\beta_0, Q)} E_f\Big\{\Big(\hat{F}\big(n, J(\beta_{\hat{k}})\big) - F(f)\Big)^2\Big\} \asymp R(\beta_0), \tag{2.16}$$

$$\sup_{f \in H(\beta_1, Q) \setminus \mathscr{R}(\beta_0, \beta_1)} E_f\Big\{\Big(\hat{F}\big(n, J(\beta_{\hat{k}})\big) - F(f)\Big)^2\Big\} \asymp R(\beta_1), \tag{2.17}$$

$$\sup_{f \in H(\beta_1, Q) \cap \mathscr{R}(\beta_0, \beta_1)} E_f\Big\{\Big(\hat{F}\big(n, J(\beta_{\hat{k}})\big) - F(f)\Big)^2\Big\} \asymp R(\beta_1) B(\beta_1). \tag{2.18}$$

Theorem 2.2 shows that the lower bound (2.9) is in fact rate sharp.

REMARK 2.1. Efromovich and Pinsker (1984) show that

$$\begin{aligned} &\sup_{f \in H(\beta_1, Q)} E_f\big\{\big|\hat{I}\big(J(\beta_0), J^*(\beta_1)\big) - I\big(J(\beta_0), J^*(\beta_1)\big)\big|^2\big\} \\ &\qquad = R^*(\beta_1)(1 + o(1)). \end{aligned} \tag{2.19}$$

Hence, $R^*(\beta_1)$is asymptotically equal to the maximum variance of the statistic $\hat{I}((J(\beta_0), J^*(\beta_1))$, which explains its central role in computing $\hat{k}$ in (2.15).

REMARK 2.2. The subset $H(\beta_1, Q) \setminus \mathscr{R}(\beta_0, \beta_1)$, where our estimator does not lose the logarithmic factor, has a simple interpretation. If $I^2(J(\beta_0), J(\beta_1)) \le C_0 R(\beta_1)$, then it does not matter whether $\hat{k}$ is equal to 0 or 1. On the other hand, if $I^2(J(\beta_0), J^*(\beta_1)) \ge 2\ln(n) R^*(\beta_1)(1+\nu_n)$, then the probability that $\hat{k} = 0$ (our estimator made the wrong decision) is sufficiently small.

REMARK 2.3. In the context of estimating linear functionals, Lepskii (1990) was the first to find optimal adaptive procedures which have maximum risk a logarithmic factor greater than the minimax risk. For two parameters $\{\beta_1, \beta_0\}$ Lepskii's algorithm performs as follows [see the Remark in Section 2 of Lepskii (1990)]: (i) optimal estimates $\tilde{F}(Y_n, \beta_j)$, $j = 0, 1$, are computed; (ii) if the difference $|\tilde{F}(Y_n, \beta_1) - \tilde{F}(Y_n, \beta_0)|$ is less than a specific threshold level, then we choose $\hat{F}(Y_n, \beta_1)$ and otherwise we use $\hat{F}(Y_n, \beta_0)$.

Our estimator resembles Lepskii's algorithm, but the difference is that we compare some specific statistics rather than optimal estimators. This approach allows us to describe the function subset $\mathscr{R}' \cap \mathscr{R}''$, where our estimator loses the logarithmic factor.

To shed light on how large this subset is, suppose that the function $f(t) = \sum_{j \ge 0} \theta_j \varphi_j(t)$ is generated by a Monte Carlo method, where $\theta_j$ are realizations of independent random variables $\Theta_j$ supported on $H(\alpha, Q)$. We make only one more assumption about these random variables, namely, that $E\{\sum_{j=J(\beta_0)+1}^{J(\beta_1)} \Theta_j^2\} \ge C_1 \max_{|\theta_j| \le Qj^{-\alpha}}[\sum_{j=J(\beta_0)+1}^{J(\beta_1)} \theta_j^2] n^{-\delta}$, where $0 \le \delta < \min\{(4\beta_0 - 1)^{-1}, 4(4\beta_1 - 2)(\beta_0 - \beta_1)/[(4\beta_0 - 1)(4\beta_1 - 1)]\}$ and $C_1$ is a positive constant. In particular, note that if $\Theta_j / j^{-\alpha}$ are iid and $E\{\Theta_1^2\} \ne 0$, then our assumption holds with $\delta = 0$. Thus, this assumption is extremely mild.

Straightforward calculation shows that

$$\Big[\mathrm{Var}\Big(\textstyle\sum_{j=J(\beta_0)+1}^{J(\beta_1)} \Theta_j^2\Big)\Big]^{1/2} \le C\Big[\max_{|\theta_j| \le Qj^{-\alpha}} \textstyle\sum_{j=J(\beta_0)+1}^{J(\beta_1)} \theta_j^2\Big] n^{-1/(4\beta_0 - 1)}$$

and

$$\ln(n) R^*(\beta_1) = \ln(n) n^{(-4\beta_1 + 2)/(4\beta_1 - 1)}.$$

Chebyshev's inequality then allows us to estimate the probability of the event $f \in \mathscr{R}''$:

$$\Pr\left(\left[\sum_{j=J(\beta_0)+1}^{J(\beta_1)} \Theta_j^2\right]^2 < 2\ln(n) R^*(\beta_1)(1+\nu_n)\right) < Cn^{-[(4\beta_0 - 1)^{-1} - \delta]}.$$

Hence although our adaptive estimator does lose a logarithmic factor over a subset of functions, the integrated risk over the priors described above converges at the same rate as the corresponding minimax risk. Further comparisons with a robust Bayes approach are given in Section 4.

**3. Adaptive sequence of estimators.** In this section the projection estimator (1.4) is investigated where the window size depends on the data. The choice of the window size can be most easily described for the case when we assume that $\alpha \in \{\beta_0, \beta_1, \dots, \beta_K\}$ and $3/4 \geq \beta_0 > \beta_1 > \cdots > \beta_K > 1/2$.

Set $J_k = J(\beta_k)$ and $J_k^* = J^*(\beta_k)$, where $J$ and $J^*$ are defined by (1.5) and (2.7). Let

$$\hat{k} = \min\{k: \hat{I}^2(J_k, J_l^*) \leq 3\ln(n) R^*(\beta_l),\ l > k,\ J_l^* > J_k,\ 0 \leq k \leq K\}, \tag{3.1}$$

$$\begin{aligned} \mathscr{R}(k) &= \{f: I^2(J_0, J_k) > CR(\beta_k)\} \\ &\quad \cap \{f: I^2(J_{k-1}, J_k^*) < 4\ln(n) R^*(\beta_k)\} \\ &\triangleq \mathscr{R}'(k) \cap \mathscr{R}''(k). \end{aligned} \tag{3.2}$$

The subset $\mathscr{R}(k)$ for $k = 1$ is similar to $\mathscr{R}(\beta_0, \beta_1)$ defined in (2.14), only here for the simplicity we set $\nu_n = 1$. Note, that if $l > k$, then for sufficiently large $n$ we get $J_l^* > J_k$.

The optimal properties of the adaptive estimator are then conveniently summarized by the following theorem.

THEOREM 3.1. *Let $\alpha \in \{\beta_0, \beta_1, \dots, \beta_K\}$, where $\{\beta_k\}$ are given and satisfy the relations $3/4 \geq \beta_0 > \beta_1 > \cdots > \beta_K > 1/2$. Then the projection estimator* (1.4) *with adaptive window size $J(\beta_{\hat{K}})$, where $\hat{k}$ is defined in* (3.1), *has the following statistical characteristics*:

(i) *For $\alpha = \beta_0$*,

$$\sup_{f \in H(\beta_0, Q)} E_f\left\{\left(\hat{F}(n, J(\beta_{\hat{k}})) - F(f)\right)^2\right\} \asymp R(\beta_0). \tag{3.3}$$

(ii) *For any $k = 2, 3, \dots, K$*,

$$\sup_{f \in H(\beta_k, Q) \setminus \mathscr{R}(k)} E_f\left\{\left(\hat{F}(n, J(\beta_{\hat{k}})) - F(f)\right)^2\right\} \asymp R(\beta_k), \tag{3.4}$$

$$\sup_{f \in H(\beta_k, Q) \cap \mathscr{R}(k)} E_f\left\{\left(\hat{F}(n, J(\beta_{\hat{k}})) - F(f)\right)^2\right\} \asymp R(\beta_k) B(\beta_k). \tag{3.5}$$

The desired extension to the case of unknown smoothness $\alpha \in (1/2, 3/4]$ is straightforward. First, we explicitly describe the procedure and then explain the underlying idea.

Let $N = N(n)$ be the maximum integer such that $d^{N-1} \leq n^{1-2/\ln\ln(n)}$, where $d > 2$ is a fixed integer. Set $J_k = d^{k-1} n$ and note that $\{J_1, \dots, J_N\}$ is a collection of window sizes. These $N$ window sizes essentially correspond to optimal window sizes for particular values $\alpha \in \{\beta_0, \beta_1, \dots, \beta_N\}$, where here $\beta_k = \beta_k(n)$ are defined as a solution to the equation $J(\beta_k) = J_k$.

Let $k^*(n)$ be the minimal integer $k$ such that $J_k^* \geq n$. Recall that $J_k^* = J_k[B(\beta_k)]^{-1/(4\beta_k - 2)}$ and define the statistics

$$\check{k} = \min\left\{k: \hat{I}^2(J_k^*, J_l^*) \leq 3\ln(n)R^*(\beta_l), l > k, k^*(n) \leq k \leq K\right\}$$

$$\tilde{k} = \begin{cases} \check{k}, & \text{if } \check{k} > k^*(n), \\ 0, & \text{if } \check{k} = k^*(n). \end{cases} \tag{3.6}$$

Define $\tilde{\mathscr{R}}(k) = \tilde{\mathscr{R}}'(k) \cap \tilde{\mathscr{R}}''(k)$, where $\tilde{\mathscr{R}}'(k) = \{f\colon I^2(n, J_k) \geq CR(\beta_k)\}$ and $\tilde{\mathscr{R}}''(k) = \{f\colon I^2(J^*_{k-t^*(k)}, J_k^*) \leq 4\ln(n)R^*(\beta_k)\}$, where $t^*(k)$ is any bounded sequence of integers.

THEOREM 3.2. *Let $\alpha > 1/2$ be unknown and fixed. Then the projection estimator* (1.4) *with adaptive window size $J(\beta_{\tilde{k}})$ satisfies*:

(i) *For the regular case, that is, for any $\alpha \geq 3/4$,*

$$\sup_{f \in H(\alpha, Q)} E_f\left\{\left(\hat{F}(n, J(\beta_{\tilde{k}})) - F(f)\right)^2\right\} \asymp R(\alpha). \tag{3.7}$$

(ii) *For the irregular case, that is, for $1/2 < \alpha < 3/4$,*

$$\sup_{f \in H(\alpha, Q) \setminus \tilde{R}(k(\alpha))} E_f\left\{\left(\hat{F}(n, J(\beta_{\tilde{k}})) - F(f)\right)^2\right\} \asymp R(\alpha), \tag{3.8}$$

$$\sup_{f \in H(\alpha, Q) \cap \tilde{\mathscr{R}}(k(\alpha))} E_f\left\{\left(\hat{F}(n, J(\beta_{\tilde{k}})) - F(f)\right)^2\right\} \asymp R(\alpha)B(\alpha). \tag{3.9}$$

Hence, the adaptive projection estimator has an optimal nonadaptive rate $R(\alpha)$ for risk convergence over all functions $f \in H(\alpha, Q)$ except for a relatively small (in the sense of the previous section) subset $\tilde{\mathscr{R}}$ and only for the irregular case $\alpha < 3/4$. Nevertheless, over this subset the recommended estimate has the optimal adaptive minimax rate $R(\alpha)B(\alpha)$ of risk convergence in the irregular case $\alpha < 3/4$ due to the lower bound of Corollary 2.1.

Now we are in a position to discuss the estimator. The underlying idea of our particular choice of window sizes is that for any $\alpha \in (1/2, 3/4]$ and sufficiently large $n$ there exists an integer $k(\alpha) = k(n, \alpha)$ such that $\alpha \in (\beta_{k(\alpha)+1}, \beta_{k(\alpha)}]$ and $R(\alpha)/R(\beta_{k(\alpha)}) \asymp 1$. Note also that $\beta_N = 1/2 + (1/4)[\ln\ln(n)]^{-1}(1 + o(1))$. Hence, in this way the set of choices $\alpha \in (1/2, 3/4]$ is artificially reduced for each $n$ to the case of $N = N(n) \asymp \ln(n)$ alternative values $\{\beta_k\}$ of $\alpha$, and this case was explored at the beginning of this section.

In the definition of $\tilde{k}$, we use only the set $\{J_l^*\}$ of window sizes because for the considered setting the inequalities $J_l^* > J_k$ are no longer valid simultaneously whenever $l > k$. Nevertheless, $J^*_{l+1} = dJ_l^*(1 + o(1))$ and therefore the set $\{J_l^*\}$ can be used for seeking an optimal window size similar to the

$\{J_k\}$. The latter leads only to one complication, namely, separate consideration for the cases $\alpha = 3/4$ and $\alpha < 3/4$, which leads to the two different cases on the right-hand side of (3.6). The asymptotic equality $n/J^*(\alpha) = o(1)$, which is valid for all $\alpha < 3/4$, explains the underlying idea of the second line in (3.6).

## 4. Discussion.

4.1. *Sharp adaptive estimation for the regular case*. Define a measure of efficiency of an adaptive estimator $\hat{F}(Y_n)$ as the ratio betwen the worst-case error of this estimator and the risk of a minimax estimator:

$$\Psi\big(\hat{F}(Y_n), \alpha, Q\big) = R\big(\alpha, Q, \hat{F}(Y_n)\big)/\inf R\big(\alpha, Q, \hat{F}(Y_n, \alpha, Q)\big), \tag{4.1}$$

where in this section the infimum is taken over all possible estimators $\hat{F}(Y_n, \alpha, Q)$. In other words, we compare the risk of adaptive estimators to the minimax risk over each $H(\alpha, Q)$.

It follows from the previous sections that

$$\Psi(\hat{F}(n, J(\beta_{\tilde{k}})), \alpha, Q) \asymp \begin{cases} 1, & \text{if}\, \alpha \geq 3/4, \\ (\ln(n))^{(4\alpha-2)/(4\alpha-1)}, & \text{if}\, 1/2 < \alpha < 3/4. \end{cases} \tag{4.2}$$

We have been interested until now only in optimal rates, and not in achieving the sharp-optimal constants. However, for the regular case $\alpha \geq 3/4$, estimation with both optimal constant and rate is possible. The following results of Efromovich (1994) make this issue more explicit.

Let $b_n$ be an arbitrary positive sequence such that $b_n$ tends to infinity slower than any power function of $n$. For the regular setting $\alpha \geq 3/4$, the sequence $b_n$ plays the same role as $B(\alpha)$ played in the irregular case $\alpha < 3/4$ considered earlier.

PROPOSITION 4.1. *Let $\alpha > 3/4$. Then*

$$\inf \sup_{f \in H(\alpha, Q)} \Big[ nE_f\Big\{\big(\hat{F}(Y_n, \alpha, Q) - F(f)\big)^2\Big\} - 4F(f)\Big] \geq o(1) \tag{4.3}$$

*and the estimator $\hat{F}(n, \lfloor n/b_n \rfloor)$ attains this lower bound.*

*Moreover, let $\tilde{F} = \tilde{F}(Y_n)$ be any estimator such that for some $\alpha > 3/4$,*

$$\sup_{f \in H(\alpha, Q)} \Big[ nE_f\Big\{\big(\tilde{F}(Y_n) - F(f)\big)^2\Big\} - 4F(f)\Big] = o(1). \tag{4.4}$$

*Then there exists no finite constant A such that for all $n$,*

$$\sup_{f \in H(3/4, Q)} nE_f\Big\{\big(\tilde{F}(Y_n) - F(f)\big)^2\Big\} \leq A. \tag{4.5}$$

Denote (only in this section)

$$I_0 = \sum_{j=\lfloor n/b_n \rfloor+1}^{n} \theta_j^2, \tag{4.6}$$

$$\hat{I}_0 = \sum_{j=\lfloor n/b_n \rfloor+1}^{n} \left(\hat{\theta}_j^2 - n^{-1}\right), \tag{4.7}$$

$$S = \left\{f: \sqrt{n^{-1}} < I_0 < 2\sqrt{b_n n^{-1}}\right\}, \tag{4.8}$$

$S^c$ is the complement of $S$. Define the estimator

$$\tilde{F}_n = \begin{cases} \hat{F}(n, \lfloor n/b_n \rfloor), & \text{if } \hat{I}_0 \le \sqrt{b_n n^{-1}}, \\ \hat{F}(n, \lfloor n/b_n \rfloor) + \hat{I}_0, & \text{otherwise.} \end{cases} \tag{4.9}$$

PROPOSITION 4.2. *The estimator $\tilde{F}_n$ is adaptively sharp-optimal, that is, it satisfies the relations*

$$\sup_{f \in H(\alpha, Q)} \left[ nE_f\left\{\left(\tilde{F}_n - F(f)\right)^2\right\} - 4F(f)\right] = o(1), \qquad \alpha > 3/4, \tag{4.10}$$

$$\sup_{f \in H(3/4, Q) \cap S^c} \left[ nE_f\left\{\left(\tilde{F}_n - F(f)\right)^2\right\}\right] < C, \tag{4.11}$$

$$\sup_{f \in H(3/4, Q) \cap S} \left[ nE_f\left\{\left(\tilde{F}_n - F(f)\right)^2\right\}\right] < Cb_n. \tag{4.12}$$

Hence, any estimator which is asymptotically sharp-optimal for $\alpha > 3/4$ cannot achieve the optimal rate $n^{-1}$ whenever $\alpha = 3/4$. However, this loss is not so drastic as for the irregular case.

The natural question is whether it is possible to combine these two cases and give an adaptive estimator which is sharp optimal for the regular case and rate-optimal for the irregular case in the sense of the minimax lower bounds.

The answer is that it is possible. In particular, it is not difficult to verify that the adaptive projection estimator

$$\tilde{F}(Y_n) = \begin{cases} \hat{F}(n, J(\beta_{\tilde{k}})), & \text{if } \tilde{k} \ge 1, \\ \tilde{F}_n, & \text{if } \tilde{k} = 0, \end{cases} \tag{4.13}$$

is (i) sharp-optimal for the case $\alpha > 3/4$, that is, (4.10) is valid for this estimator as well as for $\tilde{F}_n$; (ii) adaptively optimal for $\alpha = 3/4$, that is, (4.11) and (4.12) are valid for $\tilde{F}(Y_n)$ as well as for $\tilde{F}_n$; (iii) adaptively optimal for $1/2 < \alpha < 3/4$ in the sense of the lower bound of Section 2, that is, (3.8) and (3.9) are valid for $\tilde{F}(Y_n)$ as well as for $\hat{F}(n, J(\beta_{\tilde{k}}))$. The proof is straightforward and we leave it to the interested reader.

Hence, we have obtained the desired measure of efficiency:

$$\Psi\big(\tilde{F}(Y_n), \alpha, Q\big) \le \begin{cases} 1 + o(1), & \text{if } \alpha > 3/4, \\ Cb_n, & \text{if } \alpha = 3/4, \\ C[\ln(n)]^{(4\alpha-2)/(4\alpha-1)}, & \text{if } 1/2 < \alpha < 3/4. \end{cases} \tag{4.14}$$

Recall that $b_n$ tends to infinity as slow as desired as $n \to \infty$.

4.2. *Bayes approach.* A comparison of Theorem 3.2 and the lower bound of Corollary 2.1 shows that the suggested data dependent projection estimator attains the lower bound and, hence, in this sense, is optimal. Another adaptive optimality property can be given in terms of a minimax Bayes perspective. Such an approach has been taken by Heckman and Woodroofe (1991).

For any prior distribution $\mu$ supported on $H(\alpha, Q)$ and any estimator $\delta_n = \delta_n(Y_n, \alpha, Q)$ of $F(f)$ write

$$r_B(\mu, \delta_n) = E_\mu\Big\{\big(\delta_n - F(f)\big)^2\Big\}, \tag{4.15}$$

where $E_\mu$ denotes expectation with respect to the probability model (1.1) and the prior $\mu$.

The Bayes risk corresponding to $\mu$ can then be written as

$$R_B(n, \mu) = \inf_{\delta_n} r_B(\mu, \delta_n). \tag{4.16}$$

In particular, if $\mathscr{P}(\alpha, Q)$ is a collection of distributions $\mu$ such that $\Pr_\mu\{f \in H(\alpha, Q)\} = 1$, then

$$\sup_{\mu \in \mathscr{P}(\alpha, Q)} R_B(n, \mu) = R(\alpha), \tag{4.17}$$

where $R(\alpha)$ is the conventional minimax risk defined in (1.3).

Now for $\alpha < 3/4$ let $\mathscr{P}_n(\alpha, Q)$ be a subset of $\mathscr{P}(\alpha, Q)$ defined by

$$\mathscr{P}_n(\alpha, Q) = \mathscr{P}(\alpha, Q) \cap \Big\{\mu\colon \Pr_\mu\big\{f \in \tilde{\mathscr{R}}(k(\alpha))\big\} \le C/B(\alpha)\Big\}, \tag{4.18}$$

where the subset $\tilde{\mathscr{R}}(k(\alpha))$ is defined in Section 3. We showed in Section 2 how to construct a particular $\mu \in \mathscr{P}_n(\alpha, Q)$.

The maximum Bayes risk for the collection of priors $\mathscr{P}_n(\alpha, Q)$ has the same rate of risk convergence as the ordinary minimax risk, that is, the minimax risk

$$\sup_{\mu \in \mathscr{P}_n(\alpha, Q)} R_B(n, \mu) \asymp R(\alpha). \tag{4.19}$$

The following assertion is an obvious corollary to Theorem 3.2 which shows that the suggested data dependent projection estimator is Bayes minimax over the subset $\mathscr{P}_n(\alpha, Q)$.

COROLLARY 4.1. *Under the conditions of Theorem* 3.2, *a projection estimator* (1.4) *with adaptive window size* $J(\beta_{\tilde{k}})$ *is Bayes minimax, that is, for any* $\alpha > 1/2$,

$$\sup_{\mu \in \mathscr{P}_n(\alpha, Q)} E_\mu\Big\{\big(\hat{F}(n, J(\beta_{\tilde{k}})) - F(f)\big)^2\Big\} \asymp R(\alpha). \tag{4.20}$$

This result gives a new insight into the phenomena of minimax adaptive estimation.

## 5. Proofs.

PROOF OF THEOREM 2.1. The proof is based on three main ideas. First, by a convexity argument it is sufficient to consider estimators which are only functions of $Y_1^2, Y_2^2, \dots$. Equation (2.5) then follows from Theorem 1 of Brown and Low (1992) and the following lemma.

Recall, that if $X$ is a normal random variable with mean $\theta$ and variance $\sigma^2$, then $X^2$ has a noncentral $\chi^2$ distribution with density

$$\begin{aligned} f_{\theta,\sigma}(y) &= \big(1/2\sqrt{2\pi y\sigma^2}\big)\exp\big\{-(y+\theta^2)/2\sigma^2\big\} \\ &\quad\times\big[\exp\big\{\theta\sqrt{y}/\sigma^2\big\} + \exp\big\{-\theta\sqrt{y}/\sigma^2\big\}\big]. \end{aligned} \tag{5.1}$$

LEMMA 5.1. *Let* $f_{\theta,\sigma}$ *be the density of a noncentral* $\chi^2$ *distribution given by* (5.1). *Then*

$$\mathscr{I} = \int_0^\infty f_{\theta,\sigma}^2(y) f_{0,\sigma}^{-1}(y)\,dy \le \exp\{\theta^4/\sigma^4\}. \tag{5.2}$$

PROOF. A simple calculation yields that $\mathscr{I} = (1/2)(\exp\{\theta^2/\sigma^2\} + \exp\{-\theta^2/\sigma^2\})$ and then (5.2) follows from a comparison of the Taylor series for $\mathscr{I}$ and $\exp\{\theta^2/\sigma^2\}$. Lemma 5.1 is proved. □

We now continue the proof of Theorem 2.1. Let $\mathbf{Y} = \{Y_0, Y_1, \dots\}$. Define the $i$th sign change operator $\tau_i(\theta_0, \theta_1, \dots)$ by $\tau_i(\theta_0, \theta_1, \dots) = (\theta_0, \dots, \theta_{i-1}, -\theta_i, \theta_{i+1}, \dots)$. Since $\Phi$ is orthosymmetric if $\Theta \in \Phi$ and $\Theta \in S(\delta)$, then $\tau_i(\Theta) \in \Phi \cap S(\delta)$ and $E_{\tau_i,\Theta}\{(G(\tau_i\Theta) - \hat{G}_n(\mathbf{Y}))^2\} = E_\Theta\{(G(\Theta) - \hat{G}_n(\tau_i\mathbf{Y}))^2\}$. Hence, by convexity

$$\begin{aligned} &\tfrac{1}{2}\Big[E_\Theta\big\{G(\Theta) - \hat{G}_n(\mathbf{Y})\big)^2\big\} + E_{\tau_i\Theta}\Big\{\big(G(\tau_i\Theta) - \hat{G}_n(\tau_i\mathbf{Y})\big)^2\Big\}\Big] \\ &\quad\ge E_\Theta\Big\{G(\Theta) - (1/2)\big(\hat{G}_n(\mathbf{Y}) + \hat{G}_n(\tau_i\mathbf{Y})\big)^2\Big\}. \end{aligned}$$

Thus, to prove (2.5) attention may be restricted to estimators, which are functions of $Y_0^2, Y_1^2, \dots$. Let $\sigma_n = 1/n$ and set

$$\mathscr{I}_n(\Theta) = \prod_{i=1}^\infty \int_0^\infty f_{\theta_i,\sigma_n}^2(y) f_{0,\sigma_n}^{-1}(y)\,dy.$$

Then line (2.4) of Brown and Low (1992) yields that if $\hat{G}_n$ is an estimator which is only a function of $Y_0, Y_1, \ldots$ and for which (2.4) holds, then

$$R(\Theta, \delta_n) \geq G^2(\Theta)\left[1 - 2\mathscr{I}_n(\Theta)\,\varepsilon_n G^{-1}(\Theta)\right]. \tag{5.3}$$

Now by Lemma 5.1,

$$\mathscr{I}_n(\Theta) \leq \exp\left\{\sum_{i=0}^{\infty} n^2\theta_i^4\right\}. \tag{5.4}$$

Inequality (2.5) now follows from the definition of $w(\delta)$ and (5.3) and (5.4). Theorem 2.1 is proved. □

In the proof of Theorem 2.2 the following lemma plays a central role. Set $\sigma(k, k+r) = E_f^{1/2}\{(\hat{I}(k, k+r) - I(k, k+1))^2\}$ and note that $\sigma(k, k+r)$ is a functional of $f$ as well.

LEMMA 5.2. *For any natural $k, r$, any $0 \leq \gamma < 1$ and $f \in H(\alpha, Q)$,*

$$P_f\left(|\hat{I} - I| > u\sigma\right) \leq \exp\left\{-(1/2)u^2(1-\gamma)^2\right\}, \quad |u| \leq (1/2)\gamma n\sigma, \tag{5.5}$$

$$P_f\left(|\hat{I} - I| > u\sigma\right) \leq \exp\{-u\sigma n\gamma/4\}, \quad |u| > (1/2)\gamma n\sigma, \tag{5.6}$$

*where $\hat{I} = \hat{I}(k, k+r)$, $I = I(k, k+r)$ and $\sigma = \sigma(k, k+r)$.*

This lemma is a corollary of the following well-known exponential inequality [see Petrov (1987)].

PROPOSITION 5.1. *Suppose that $X_1, \ldots, X_r$ are independent random variables and $S = \sum_{j=1}^r X_j$. If there exist positive constants $g_1, \ldots, g_r$ and $T$ such that*

$$E\left\{\exp(tX_j)\right\} \leq \exp\left(g_j t^2/2\right), \quad j = 1, \ldots, r, \ |t| \leq T, \tag{5.7}$$

*then*

$$P(|S| > u) \leq \exp(-u^2/2G), \quad |u| \leq GT, \tag{5.8}$$

$$P(|S| > u) \leq \exp(-uT/2), \quad |u| > GT, \tag{5.9}$$

*where $G = \sum_{j=1}^r g_j$.*

REMARK 5.1. The factor $1/2$ in the exponent of the right-hand side of (5.9) is sharp as may be easily seen by considering a sequence of iid normal random variables $X_1, X_2, \ldots, X_r$. Therefore this factor is sharp in (5.5) as well.

PROOF OF LEMMA 5.2. Note that

$$\hat{I}(k, k+r) - I(k, k+r) = \sum_{j=k+1}^{k+r}\left[\langle Y, \varphi_j\rangle^2 - n^{-1} - \theta_j^2\right]$$

and $\langle Y, \varphi \rangle = \theta_j + n^{-1/2}\xi_j$, where $\xi_j$ are iid standard normal variables. Hence

$$\hat{I}(k, k+r) - I(k, k+r) = \sum_{j=k+1}^{k+r} \left[2n^{-1/2}\xi_j\theta_j + n^{-1}\left(\sigma_j^2 - 1\right)\right]$$

and it is possible to apply Proposition 5.1 to the independent random variables $X_j = 2n^{-1/2}\xi_j\theta_j + n^{-1}(\xi_j^2 - 1)$. Elementary algebra shows that for any $0 \le \gamma < 1$,

$$E\left\{\exp(tX_j)\right\} \le \exp\left(g_j t^2/2\right), \qquad |t| \le \gamma n/2,$$

where $g_j = 4(1-\gamma)^{-1/2}n^{-1}\theta_j^2 + 2n^{-2}(1-\gamma)^{-2}$. Hence

$$G = \sum_{j=k+1}^{k+r} g_j = 4(1-\gamma)^{-1/2}n^{-1}I(k, k+r) + 2rn^{-2}(1-\gamma)^{-2} \tag{5.10}$$

and using the following equality of Efromovich and Pinsker (1984),

$$E_f\left\{|\hat{I}(k, k+r) - I(k, k+r)|^2\right\} = 4n^{-1}I(k, k+r) + 2rn^{-2}, \tag{5.11}$$

(5.5) and (5.6) immediately follow. Lemma 5.2 is proved. □

Equality (5.11) and simple algebra show that the estimate (1.4) with nonadaptive window size (1.5) is rate-optimal. From Lemma 5.2 we also obtain the following corollary.

COROLLARY 5.1. *For any natural $k, r$, any $q > 0$ and $f \in H(\alpha, Q)$ there exists a function $C(q) < Cq$ such that*

$$E_f^{1/q}\left\{|\hat{I}(k, k+r) - I(k, k+r)|^q\right\} \le C(q)\sigma(k, k+r).$$

We are now ready to prove the upper bounds given in Theorem 2.2.

PROOF OF THEOREM 2.2. Here we use the notation $J_0 = J(\beta_0)$, $J_1 = J(\beta_1)$, $J^* = J^*(\beta_1)$, $I = I(J_0, J^*)$ and $\hat{I} = \hat{I}(J_0, J^*)$. In accordance with definition (2.15) of the adaptive window size $\hat{J} = J(\beta_{\hat{k}})$,

$$\begin{aligned} &\sup E_f\left\{\left[\hat{F}(n, \hat{J}) - F(f)\right]^2\right\} \\ &\quad \le \sum_{k=0}^{1} \sup E_f\left\{\chi(\hat{k} = k)\left[\hat{I}(0, J_k) - I(0, J_k) - \sum_{j > J_k} \theta_j^2\right]^2\right\} \\ &\quad \triangleq R_0 + R_1, \end{aligned} \tag{5.12}$$

where the supremum is taken over $f \in H(\alpha, Q)$ and $\chi(E)$ is the indicator of the event $E$.

First consider the case $\alpha = \beta_0$. Then a rate-optimal window size is $J_0$ and therefore $R_0 \le R(\alpha)$. For estimating $R_1$ we use the Hölder inequality

$$\begin{aligned} R_1 &\le \sup E_f^{1/p}\Big\{ \chi\big(\hat{I}^2 > 2\ln(n)R^*(\beta_1)\big)\Big\} \\ &\qquad \times E_f^{1/q}\left\{\left[\hat{I}(0, J_1) - I(0, J_1) - \sum_{j>J_1} \theta_j^2\right]^{2q}\right\} \\ &\triangleq \sup\{T_1 T_2\}, \end{aligned}$$

where $p^{-1} + q^{-1} = 1$, $q > 1$.

To estimate $T_1$ we note that $I \le CJ_0^{-2\alpha+1} = o(1)\sqrt{\ln^{-1/2}(n)R^*(\beta)}$ for all $f \in H(\alpha, Q)$ and that from (2.11) and (5.11) we get $\sigma^2(J_0, J_1^*) = R^*(\beta_1)(1 + o(1))$. Hence, for arbitrarily small $\gamma$ and $\kappa$ and sufficiently large $n$ it follows from Lemma 5.2 that

$$\begin{aligned} T_1 &\le \sup P_f^{1/p}\Big(\hat{I} - I > \sqrt{2\ln(n)}\,\sigma(J_0, J_1^*)(1-\gamma)\Big) \\ &\le \exp\big(-\ln(n)(1-\gamma)^2(1-\kappa)/p\big). \end{aligned} \tag{5.13}$$

To estimate $T_2$ use Corollary 5.1 to show

$$T_2 = E_f^{1/q}\left\{\left[\hat{I}(0, J_1) - I(0, J_1) - \sum_{j>J_0} \theta_j^2\right]^{2q}\right\} \le CqR(\beta_1). \tag{5.14}$$

Now recall that

$$R(\beta_1) \asymp n^{-4(2\beta_1 - 1)/(4\beta_1 - 1)}$$

and hence

$$T_2 \le Cq\, n^{-4(2\beta_1 - 1)/(4\beta_1 - 1)}.$$

Thus for sufficiently small positive $\gamma$, $\kappa$ and $q^{-1}$ it follows that $\sup(T_1 T_2) = o(1)n^{-1} = o(1)R(\beta_0)$. This together with $R_1 = o(1)R(\beta_0)$ and $R_0 \le R(\beta_0)$ shows that (2.16) holds.

Now consider the second case when $\alpha = \beta_1$. Then a rate-optimal window size is equal to $J_1$ and hence $R_1 \le R(\beta_1)$. Therefore for completing the proof it suffices to show that $R_0$ satisfies (2.17) and (2.18), where in definition (5.12) of $R_0$ the supremum is taken over the corresponding set of functions $f$.

First consider the case where $f \in H(\beta_1, Q)\setminus\mathscr{R}(\beta_0, \beta_1)$, that is, we are proving (2.17). There are two different subcases. First, we consider the case where $I^2(J_0, J_1) \le CR(\beta_1)$. Using (5.11) it follows that

$$\begin{aligned} \sup E_f\Big\{\big(\hat{F}(n, J_0) - F(f)\big)^2\Big\} &\asymp \sup\left[\left(\sum_{j>J_0} \theta_j^2\right)^2 + J_1 n^{-2}\right] \\ &\asymp \sup\big[I(J_0, J_1) + J_1^{-2\beta_1+1}\big]^2 \\ &\asymp R(\beta_1), \end{aligned} \tag{5.15}$$

where the supremum is taken over $f \in \{f\colon I^2(J_0, J_1) \le CR(\beta_1)\} \cap H(\beta_1, Q)$.

To complete the proof of (2.17) we consider the second case, where $I^2 = I^2(J_0, J^*) > 2(1+\nu_n)\ln(n)R^*(\beta)$. Taking the supremum over $H(\beta_1, Q) \cap (\mathscr{R}'')^c$ it follows that for sufficiently large $n$,

$$(5.16)\qquad \begin{aligned} R_0 &= \sup E_f\Big\{\chi\big(\hat{I}^2 < 2\ln(n)R^*(\beta_1)\big)\chi\big(I^2 > 2(1+\nu_n)\ln(n)R^*(\beta_1)\big) \\ &\qquad\qquad \times\big(\hat{F}(n, J_0) - F(f)\big)^2\Big\} \\ &\le \sup E_f\Big\{\chi\big(I - \hat{I} > (\nu_n/3)I\big)\big(\hat{F}_n(J_1) - F(f)\big)^2\Big\} \end{aligned}$$

and using the Chebyshev inequality we see that

$$(5.17)\qquad R_0 \le C\sup(\nu_n I)^{-2} E_f\Big\{(I-\hat{I})^2\big(\hat{F}(n, J_0) - F(f)\big)^2\Big\}.$$

To estimate the right-hand side of the last inequality we use the Cauchy-Schwarz inequality and Corollary 5.1. We obtain,

$$(5.18)\qquad \begin{aligned} R_0 &\le C\sup(\nu_n I)^{-2} E_f^{1/2}\Big\{(I-\hat{I})^4\Big\} E_f^{1/2}\Big\{\big(\hat{F}(n, J_0) - F(f)\big)^4\Big\} \\ &\le C(\nu_n I)^{-2} R^*(\beta_1)\big[I + (J^*)^{-2\beta_1+1}\big]^2. \end{aligned}$$

Note that

$$(J^*)^{-2\beta_1+1} \asymp R^{1/2}(\beta_1)[\ln(n)]^{(2\beta_1-1)/(4\beta_1-1)}$$

and

$$R^*(\beta_1) \asymp R(\beta_1)[\ln(n)]^{-1/(4\beta_1-1)}.$$

Hence it finally follows that for $f \in H(\beta_1, Q) \cap (\mathscr{R}'')^c$,

$$(5.19)\qquad R_0 \le C\nu^{-2}R(\beta_1)[\ln(n)]^{-1/(4\beta_1-1)} = CR(\beta_1).$$

Assertion (2.17) is thus proved. To finish the proof of Theorem 2.2 we have to show the validity of (2.18). In this case once again the optimal window size is proportional to $J_1$ and therefore $R_1 \asymp R(\beta_1)$. To estimate $R_0$ we note that

$$(5.20)\qquad \begin{aligned} R_0 &= \sup E_f\Big\{\big(\hat{F}(n, J_)) - F(f)\big)^2\Big\} \\ &\le C\sup\left[I^2(J_0, J^*) + J_0 n^{-1} + \left(\sum_{j>J^*}\theta^2\right)^2\right] \\ &\le C\Big[\ln(n)R^*(\beta_1) + (J^*)^{2(-2\beta_1+1)}\Big] \asymp R(\beta_1)B(\beta_1), \end{aligned}$$

where the supremum is over $f \in H(\beta_1, Q) \cap \mathscr{R}(\beta_0, \beta_1)$. Assertion (2.18) and, therefore, Theorem 2.2, are proved. □

PROOF OF THEOREM 3.1. Assume that $\alpha = \beta_k$ and, in the first place, consider the case $\hat{k} > k$. Due to the definition of $\hat{k}$, if $\hat{k} = k + t$, $t > 0$, then for some random integer $\hat{s}$, $0 \le \hat{s} \le K - \hat{k}$,

$$(5.21)\qquad \hat{I}^2(J_{\hat{k}-1}, J^*_{\hat{k}+\hat{s}}) > 3\ln(n)R^*(\beta_{\hat{k}+\hat{s}}).$$

Since for $f \in H(\beta_k, Q)$,

$$I^2\big(J_{\hat{k}-1}, J^*_{\hat{k}+\hat{s}}\big) \le CR(\beta_{\hat{k}-1}) \tag{5.22}$$

we obtain that

$$|\hat{I}(J_{\hat{k}-1}, J^*_{\hat{k}+\hat{s}}) - I(J_{\hat{k}-1}, J^*_{\hat{k}+\hat{s}})| \ge [3(1-\kappa_1)\ln(n)R^*(\beta_{\hat{k}+\hat{s}})]^{1/2}, \tag{5.23}$$

where hereafter constants $\kappa_i$ can be chosen arbitrary small as $n \to \infty$.

Thus, for any integer $0 < t \le K - k$, it follows from Lemma 5.2, (5.11) and the inequality $(1-\kappa_2)\sigma^2(J_{k+t}, J^*_{k+t+s}) \le R^*(\beta_{k+t+s})$ that

$$\begin{aligned} E_f\big\{\chi(\hat{k} = k+t)\big\} &= \sum_{s=0}^{K-k-t} E_f\big\{\chi(\hat{s} = s)\chi(\hat{k} = k+t)\big\} \\ &\le \sum_{s=0}^{K-k-t} \exp\{(-1/2)3(1-\kappa_3)\ln(n)\} \\ &\le K\exp\{-(3/2)(1-\kappa_3)\ln(n)\}. \end{aligned} \tag{5.24}$$

Also note that, by Corollary 5.1, if $q > 1$, then

$$E_f^{1/q}\Big\{\big[\hat{I}(0, J_{k+t+s}) - I(0, J_{k+t+s})\big]^{2q}\Big\} \le Cq^2R(\beta_{k+t+s}). \tag{5.25}$$

Thus, using (5.24) and (5.25) and the Hölder inequality we obtain for sufficiently large $n$ that

$$\begin{aligned} R_1(f) &\triangleq E_f\Big\{\chi(\hat{k} > k)\big(\hat{F}(J_{\hat{k}}) - F(f)\big)^2\Big\} \\ &= \sum_{t=1}^{K-k} E_f\Big\{\chi(\hat{k} = k+t)\big(\hat{F}(J_{k+1}) - F(f)\big)^2\Big\} \\ &\le \sum_{t=1}^{K-k} E_f^{1/p}\big\{\chi(\hat{k} = k+t)\big\} \\ &\quad \times E_f^{1/q}\Big\{\big[\hat{I}(0, J_{k+t}) - I(0, J_{k+t}) - I(J_{k+t}, \infty)\big]^{2q}\Big\} \\ &\le C\sum_{t=1}^{K-k}\big[K\exp\{-(3/2)(1-\kappa_3)\ln(n)/p\}\big]\big[q^2R(\beta_{k+t})\big] \\ &\le Cq^2Kn^{-3(1-\kappa_3)/2p}\sum_{t=1}^{K-k} n^{-4(2\beta_{k+t}-1)/(4\beta_{k+t}-1)q}. \end{aligned} \tag{5.26}$$

Setting $p = 5/4$, $\kappa_3 = 1/10$ and $q = 5$ we obtain that

$$\sup_{f \in H(\beta_k, Q)} R_1(f) < Cn^{-1-\kappa_4}, \tag{5.27}$$

where here $\kappa_4 > 8/100$. Hence the case $\hat{k} > k$ has no influence on risk convergence.

Now consider the case $\hat{k} < k$. Due to (3.1), for sufficiently large $n$,

$$\chi(\hat{k} < k) \le \chi\big(\hat{I}^2(J_{\hat{k}}, J_k^*) \le 3\ln(n)R^*(\beta_k)\big) \tag{5.28}$$

that yields

$$\begin{aligned} R_2(f) &\triangleq E_f\Big\{\chi(\hat{k} < k)\big(\hat{F}(J_{\hat{k}}) - F(f)\big)^2\Big\} \\ &= E_f\Big\{\chi(\hat{k} < k)\big[\big(\hat{F}(J_k^*) - F(f)\big) - \hat{I}(J_{\hat{k}}, J_k^*)\big]^2\Big\} \\ &\le 2E_f\Big\{\big(\hat{F}(J_k^*) - F(f)\big)^2\Big\} + 2(3\ln(n)R^*(\beta_k)) \\ &< CB(\beta_k)R(\beta_k). \end{aligned} \tag{5.29}$$

Hence, to complete the proof, it suffices to show that $R_2(f) < CR(\beta_k)$ for $f \in H(\beta_k, Q) \setminus (\mathscr{R}'(k) \cap \mathscr{R}''(k))$, that is, to prove (3.4). For $f \in (\mathscr{R}'(k))^c$ this is obvious. For $f \in (\mathscr{R}''(k))^c$ due to (5.20) we obtain that $\chi(\hat{k} < k) \le \chi(I(J_{\hat{k}}, J_k^*) - \hat{I}(J_{\hat{k}}, J_k^*) > cI(J_{\hat{k}}, J_k^*))$, where here $c = (\sqrt{4} - \sqrt{3})/2$. This inequality, together with (5.11) and the inequality

$$E_f\Big\{\big[\hat{I}(J_{k-t}, J_k^*) - I(J_{k-t}, J_k^*)\big]^2\Big\} \le CR^*(\beta_k), \qquad 0 < t \le k,$$

yields that for all $f \in (\mathscr{R}''(k))^c$,

$$\begin{aligned} R_2(f) \le \sum_{t=1}^{k} E_f\Big\{&\chi(\hat{k} = k - t)\xi\big(I^2(J_{k-t}, J_k^*) > 4\ln(n)R^*(\beta_k)\big) \\ &\times\big[\hat{F}(J_{k-t}) - F(f)\big]^2\Big\} \\ \le \sum_{t=1}^{k} E_f\Big\{&\chi(\hat{k} = k - t)\chi\big(I^2(J_{k-t}, J_k^*) > 4\ln(n)R^*(\beta_k)\big)\chi \\ &\times\Big(\big(I(J_{k-t}, J_k^*) - \hat{I}(J_{k-t}, J_k^*)\big)^2 \\ &> c\ln(n)R^*(\beta_k)\Big)\big[\hat{F}(J_{k-t}) - F(f) - \hat{I}(J_{k-t}, J_k^*)\big]^2\Big\}. \end{aligned}$$

To estimate the right-hand side of the last inequality we use the Cauchy–Schwarz inequality, Lemma 5.2 and Corollary 5.1. This yields

$$\begin{aligned} R_2(f) &\le C\sum_{t=1}^{k} \exp\{-(1/4)(1 - \kappa_4)c\ln(n)\} \\ &\quad\times\Big[E_f^{1/2}\Big\{\big[\hat{F}(J_k^*) - F(f)\big]^4\Big\} + E^{1/2}\big\{\chi(\hat{k} = k - t)\hat{I}^4(J_{k-t}, J_k^*\big\}\Big] \\ &\le CK\exp\{-(1/4)(1 - \kappa_4)c\ln(n)\}B(\beta_k)R(\beta_k) = o(1)R(\beta_k). \end{aligned} \tag{5.30}$$

Theorem 3.1 is proved. □

Proof of Theorem 3.2. In the first place, note that if $\alpha < 3/4$, then $k(\alpha) \to \infty$ as $n \to \infty$. Hence, discarding from the consideration in (3.6) the cases $1 \le k \le k^*$, where $k^*$ is the maximum integer such that $J_k^* \le n$, has no impact on the choice of the considered adaptive window size.

Let us use the shorthand $k = k(n, \alpha)$, that is, $\alpha \in (\beta_k, \beta_{k+1}]$ and consider the case $\tilde{k} > k + t_0$, where $t_0$ is a given positive integer. Similar to (5.21)–(5.23) and using the same notation we obtain that if $\tilde{k} > k + t_0$, then $\hat{I}^2(J_k^*, J_{\tilde{k}+\hat{s}}^*) > 3\ln(n)R^*(\beta_{\hat{k}+\hat{s}})$ and $I^2(J_k^*, J_{k+\hat{s}}^*) < C\ln(n)R^*(\beta_k)$. These two inequalities yield for sufficiently large $t_0$ that

$$(5.31)\quad |I(J_k^*, J_{\tilde{k}+\hat{s}}^*) - I(J_k^*, J_{\tilde{k}+\hat{s}}^*)| \ge \left[3(1-\kappa_1)\ln(n)R^*(\beta_{\tilde{k}+\hat{s}})\right]^{1/2}.$$

Using (5.31) instead of (5.23) and then straightforwardly repeating (5.24)–(5.26) we obtain that $E_f\{\chi(\tilde{k} > k + t_0)(\hat{F}(n, J_{\tilde{k}}) - F(f))^2\} < CN(n)n^{-1-\kappa_4}$ for sufficiently large $t_0$. Combining this with inequality $R(\beta_{k+t}) \le Cd^tR(\beta_k)$, which is valid for all $0 \le t \le t_0$, we get

$$(5.32)\qquad E_f\left\{\chi(\tilde{k} > k)\left(\hat{F}(n, J_{\hat{k}}) - F(f)\right)^2\right\} \le CR(\beta_k).$$

Now consider the case $\tilde{k} < k$. Similar to (5.28) and (5.29) we prove (3.9). The case $f \in \tilde{\mathscr{R}}'(k(\alpha))$ is obvious for proving (3.8). Hence, to complete the proof it suffices to verify (3.8) for $f \in \tilde{\mathscr{R}}''(k(\alpha))$. Using the Cauchy–Schwarz inequality, definition (3.6) and simple algebra we obtain that

$$\begin{aligned}
&\sum_{t=1}^{k} E_f\left\{\chi(\tilde{k} = k - t)\left(\hat{F}(n, J_{k-1}) - F(f)\right)^2\right\}\\
&\quad\le \sum_{t=1}^{k} E_f^{1/2}\left\{\chi(\tilde{k} = k - t)\right\}\\
(5.33)\qquad &\qquad\times E_f^{1/2}\left\{\hat{F}(n, J_{k-t}) + \hat{I}(J_{k-t}^*, J_k^*) - \hat{I}(J_{k-t}^*, J_k^*) - F(f)\right)^2\right\}\\
&\quad\le C\sum_{t=1}^{k} E_f^{1/2}\left\{\chi(\tilde{k} = k - t)\right\}R(\beta_k)\min\left\{d^{t(4\beta_k-2)}; B(\beta_k)\right\},
\end{aligned}$$

where the previously discussed inequalities $\hat{I}^2(J_{\tilde{k}}^*, J_k^*) \le CB(\beta_k)R(\beta_k)$ and $I^2(J_k^*, \infty) \le CB(\beta_k)R(\beta_k)$ were also used. Hence, (3.8) is obvious for $0 < t < t^*(k)$.

To estimate $E_f\{\chi(\tilde{k} = k - t)\}$ for $t \ge t^*(k)$ we note that

$$\begin{aligned}
\chi(\tilde{k} = k - t) &\le \chi\left(\tilde{I}^2(J_{k-t}^*, J_k^*)\right.\\
&\qquad\left.\le 3\ln(n)R^*(\beta_k)\right)\chi\left(I^2(J_{k-t}^*, J_k^*) \ge 4\ln(n)R^*(\beta_k)\right)\\
&\le \chi\Big(|I(J_{k-t}^*, J_k^*) - \hat{I}(J_{k-t}^*, J_k^*)|\\
&\qquad \ge \left(\sqrt{4} - \sqrt{3}\right)\left[\ln(n)R^*(\beta_k)\right]^{1/2}\Big).
\end{aligned}$$

Hence, using Lemma 5.2 and (5.12), we get for $t > t^*(k)$ that

$$E_f\left\{\chi(\hat{k} = k - t)\right\} \le \exp\{-\kappa_5\ln(n)\} = n^{-\kappa_5}.$$

Substituting these results into the right-hand side of (5.33) and noting that $k \le N < c\ln(n)$ we obtain that

$$\sum_{t=1}^{k} E_f\Big\{\chi(\hat{k} = k - t)\Big(\hat{F}(n, J_{k-t}) - F(f)\Big)^2\Big\} \le CR(\beta_k) \tag{5.34}$$

for the considered setting $f \in \mathscr{R}''(n,k)$ that means the validity of (3.8). Theorem 3.2 is proved. □

**Acknowledgments.** The authors are grateful for the comments of Larry Brown and referees.

## REFERENCES


BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* To appear.

DONOHO, D. L., LIU, R. C. and MACGIBBON, (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

EFROMOVICH, S. (1994). On adaptive estimation of nonlinear functionals. *Statist. Probab. Lett.* **19** 57–63.

EFROMOVICH, S. and LOW, M. (1994). Adaptive estimates of linear functionals. *Probab. Theory Related Fields* **98** 261–275.

EFROMOVICH, S. and LOW, M. (1996). On Bickel and Ritov's conjecture about adaptive estimation of some quadratic functionals. *Ann. Statist.* **24**. To appear.

EFROMOVICH, S. and PINSKER, M. S. (1984). Learning algorithm for nonparametric filtering. *Automat. Remote Control* **11** 1434–1440.

FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294.

HECKMAN, N. E. and WOODROOFE, M. (1991). Minimax Bayes estimation in nonparametric regression. *Ann. Statist.* **19** 2003–2014.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1980). Some estimation problems for stochastic differential equations. *Lecture Notes in Control and Inform. Sci. B* **25** 1–12. Springer, Berlin.

IBRAGIMOV, I. A., NEMIROVSKII, A. S. and KHAS'MINSKII, R. Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

INGSTER, YU. I. (1986). Minimax testing of nonparametric hypotheses on a distribution density in the $L_p$-metrics. *Theory Probab. Appl.* **31** 333–337.

LEPSKII, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LEPSKII, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. *Advances in Soviet Mathematics* **12** 87–106.

NUSSBAUM, M. (1994). Asymptotic equivalence of density estimation and white noise. Technical Report, IAAS, Berlin.

PETROV, V. V. (1987). *Limit Theorems for Sum of Independent Variables*. Nauka, Moscow (in Russian).



DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 17131
E-MAIL: efrom@math.unm.edu

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: lowm@stat.wharton.upenn.edu