Sharp Adaptive Estimation of Quadratic Functionals

Jussi Klemelä*
Department of Statistics, Economics Faculty
University of Mannheim
L 7 3-5 Verfügungsgebäude
68131 Mannheim, Germany
Email: klemela@rumms.uni-mannheim.de
Fax +49 621 1811931

May 29, 2006

Abstract

Estimation of a quadratic functional of a function observed in the Gaussian white noise model is considered. A data-dependent method for choosing the amount of smoothing is given. The method is based on comparing certain quadratic estimators with each other. It is shown that the method is asymptotically sharp or nearly sharp adaptive simultaneously for the "regular" and "irregular" region. We consider l_p bodies and construct bounds for the risk of the estimator which show that for $p = 4$ the estimator is exactly optimal and for example when $p \in [3, 100]$, then the upper bound is at most 1.055 times larger than the lower bound. We show the connection of the estimator to the theory of optimal recovery. The estimator is a calibration of an estimator which is nearly minimax optimal among quadratic estimators.

Mathematics Subject Classifications: primary 62G07; secondary 62G20
Key Words and phrases: Adaptive curve estimation, exact constants in nonparametric smoothing, minimax risk, optimal recovery, quadratic functionals.
Short title: Estimation of Quadratic Functionals

*Writing of this article was financed by Deutsche Forschungsgemeinschaft under project MA1026/6-2, CIES, France, and Jenny and Antti Wihuri Foundation.

1 Introduction

A data-dependent choice of the right amount of smoothing is one of the central issues in nonparametric function estimation. We study this problem when estimating quadratic functionals, for example the squared integral of a function. We consider the case where the smoothness of the function of interest is unknown and we want to construct estimators which adapt to the unknown smoothness of this underlying function.

Estimation of quadratic functionals has been studied by Ibragimov and Hasminskii (1980) and Bickel and Ritov (1988). They showed that when the unknown function is sufficiently smooth, quadratic functionals can be estimated with parametric $\sqrt{n}$-rate, otherwise the rate is slower. We will call the cases with $\sqrt{n}$-rate of convergence *regular* cases and the cases with slower than $\sqrt{n}$-rate of convergence *irregular* cases. Ibragimov, Nemirovskii, and Hasminskii (1986) and Fan (1991) derived minimax rates under various geometric constraints on the parameter space. Donoho and Nussbaum (1990) constructed minimax estimators among quadratic estimators.

The above mentioned results were non-adaptive in the sense that no theory for the data-dependent choice of the amount of smoothing was given. Efroimovich (1994) gave a sharp adaptive procedure for the regular case. Efroimovich and Low (1996) gave a rate optimal adaptive procedure for the irregular case which was simultaneously sharp adaptive for the regular case. They considered l_2 body and Hölder body ($p = \infty$). Laurent and Massart (2000) gave an adaptive estimator based on model selection. They considered l_p bodies also for the case $p < 2$, and Besov bodies. Johnstone (2001) and Gayraud and Tribouley (1999) have studied threshold estimators. Other references include Huang and Fan (1999) who consider estimators based on local polynomial regression and Cai and Low (2005) who consider parameter spaces that are not quadratically convex.

The existence of several rate optimal adaptive procedures for the irregular case gives motivation to study sharp asymptotics also in the irregular region. We consider l_p bodies when $2 < p < \infty$. We consider also the estimation of integrals of squared derivatives, unlike previously mentioned adaptive results, which were restricted to the estimation of integrals of squared functions.

In the connection of the estimation of linear functionals, for example estimating the value of the function at one point, sharp adaptive estimation procedures were given by Lepski and Spokoiny (1997), Tsybakov (1998), Klemelä and Tsybakov (2001). On the other hand, Donoho and Nussbaum (1990) showed how the estimation of quadratic functionals can be reduced to the estimation of linear functionals. This article combines the theory of Donoho and Nussbaum (1990) and the theory of Klemelä and Tsybakov

(2001), and gives an estimation procedure for the estimation of quadratic functionals, which is sharp or nearly sharp adaptive simultaneously for the regular and the irregular case.

The sharp adaptive procedure given in this article is of similar type as the method of Efroimovich and Low (1996), that is, Lepski method is applied to choose between diagonal quadratic estimators. A difference to Efroimovich and Low (1996) is that the parameters of the procedure: the threshold and the coefficients of the quadratic estimators, are chosen in a more careful way. The more careful choice of the parameters leads to sharper optimality of the procedure. The filters of the quadratic estimators, from which we choose one, have similar shape as the filters of the quadratic estimators considered by Donoho and Nussbaum (1990), and they showed these estimators to be minimax optimal among quadratic estimators. These quadratic estimators are also solutions to a certain optimal recovery problem.

It has become standard in nonparametric function estimation literature to first present results for the simplest observation model, namely the Gaussian white noise model, and later extend results to more complicated and realistic observation models. For smooth functions the extension can be done by the theory of asymptotic equivalence, see for example Brown and Low (1996) and Nussbaum (1996). When the parameter space contains nonsmooth functions, then the asymptotic equivalence in Le Cam's sense with, say, density estimation does not hold, but there are several examples where asymptotic properties of the types of estimators we consider in this article are still similar in both models, see for example Korostelev and Nussbaum (1999) and Butucea (2001).

Let the observation be

$$y_i = \theta_i + \varepsilon\, z_i, \ \ i = 1, 2, \ldots \tag{1}$$

where z_i are i.i.d. standard Gaussian random variables and $\varepsilon > 0$. This observation arises also as the Fourier coefficients of the continuous Gaussian white noise model $dY(t) = f(t)dt + \varepsilon dW(t)$, $t \in [0, 1]$, where W is the standard Brownian motion. We consider asymptotics when $\varepsilon \to 0$. For the case when the observation consists of n discrete observations, the noise calibration $\varepsilon = n^{-1/2}$ often leads to the asymptotic equivalence with the Gaussian white noise model. We want to estimate the value of the quadratic functional

$$Q_\kappa(\theta) = Q(\theta) = \sum_{i=1}^{\infty} i^{2\kappa}\theta_i^2 \tag{2}$$

where $\kappa \geq 0$.

We consider l_p bodies $\Theta_{s,L}$ in the sequence space with s as the smoothness index and L as a function of the radius. We define l_p bodies as

$$\Theta_{s,L} = \left\{ \theta \in \mathbf{R}^\infty : \sum_{i=1}^{\infty} i^q |\theta_i|^p \leq L^{p/2} \right\}, \quad 1 \leq p < \infty \tag{3}$$

where

$$q = p(s - 1/p + 1/2). \tag{4}$$

Note that with this notation Hölder body is defined as $\{\theta : |\theta_i| \leq L^{1/2} i^{-s-1/2}\}$. As follows from Ibragimov and Hasminskii (1980), Bickel and Ritov (1988), Ibragimov, Nemirovskii, and Hasminskii (1986), and Fan (1991), the minimax rate for the parameter space $\Theta_{s,L}$ is ε^{2r} with the exponent

$$r = \frac{4(s-\kappa)}{4s+1} < 1/2,$$

when $\kappa < s < 2\kappa + 1/4$, but when $s \geq 2\kappa + 1/4$, then the minimax rate equals the parametric rate ε. We assume, however, that s and L are unknown and that we only know that $s \in [s_*, \infty)$ and $L \in [L_*, L^*]$ for given $s_* < 2\kappa + 1/4$ and $0 < L_* < L^* < \infty$. This means that we are sure that the unknown sequence $\theta \in \Theta_{s,L}$ for some $(s, L) \in [s_*, \infty) \times [L_*, L^*]$. Efroimovich and Low (1996) showed that, similar to the case of the estimation of linear functionals, the optimal rate for the case of unknown s is larger than the minimax rate by a logarithmic factor, in the irregular region $\kappa < s < 2\kappa + 1/4$. Namely, the optimal adaptive rate in the irregular case is

$$\widetilde{\varphi}_s = (\varepsilon^2 \log \varepsilon^{-1})^r.$$

Efroimovich (1994) showed that at the point $s = 2\kappa + 1/4$ the optimal adaptive rate is $\varepsilon b_\varepsilon$ where $b_\varepsilon \to \infty$ slower than any power function of ε^{-1}, and for $s > 2\kappa + 1/4$, the optimal adaptive rate is ε.

We want to find estimators which are sharp in the asymptotic adaptive minimax sense simultaneously for the irregular and regular case. That is, we will prove the following results.

1. We will first construct estimator Q_ε^* and find constants $\widetilde{C}_{s,L,p}$ such that

$$\limsup_{\varepsilon \to 0} \sup_{(s,L) \in B} (\widetilde{C}_{s,L,p} \widetilde{\varphi}_s)^{-\rho} \sup_{\theta \in \Theta_{s,L}} E_\theta \, |Q_\varepsilon^* - Q(\theta)|^\rho \leq 1, \tag{5}$$

   where $B = [s_*, s'] \times [L_*, L^*]$, $s' < 2\kappa + 1/4$, and $\rho \geq 1$. To exclude the boundary case, we took the supremum over s only over the set $[s_*, s']$ but we can let $s' \to 2\kappa + 1/4$ sufficiently slowly as $\varepsilon \to 0$.

2. Secondly, we find constants $\widetilde{c}_{s,L,p}$ such that

$$\liminf_{\varepsilon\to 0} \inf_{\hat{Q}} \sup_{(s,L)\in B} (\widetilde{c}_{s,L,p}\widetilde{\varphi}_s)^{-\rho} \sup_{\theta\in\Theta_{s,L}} E_\theta \left|\hat{Q} - Q(\theta)\right|^\rho \geq 1, \tag{6}$$

   where the infimum is taken over all estimators.

3. Thirdly, we construct estimator Q^*_ε in such a way that it is sharp adaptive for the regular region $s > 2\kappa + 1/4$.

For the case $p = 4$ we have that $\widetilde{C}_{s,L,p} = \widetilde{c}_{s,L,p}$. When $\kappa = 0$, then the ratio of $\widetilde{C}_{s,L,p}$ to $\widetilde{c}_{s,L,p}$ is bounded by 1.055, uniformly over $s \in (0, 1/4)$ and $p \in [3, 100]$. When $2 < p < 3$, then the bound is still good but the bound will become slowly worse when $p \downarrow 2$.

The proofs of the results will reveal that calibrations of the algorithms which are optimal for a certain optimal recovery problem, are optimal estimators for the statistical problem. We give in Theorem 3 the exact solution for the optimal recovery problem when $p = 4$ and nearly exact solutions when $2 < p < \infty$, which explains why we reach optimal and nearly optimal constants for the statistical problem.

The definition of the estimator is given in Section 2.1. Results are presented as Theorem 1 and Theorem 2 in Section 2.2. Discussion of the results is given in Section 2.3. Proofs are given in Section 3, Section 4, and in the Appendix of the technical report (Klemelä, 2002).

The following notation is used. Denote $\alpha_\varepsilon \sim \beta_\varepsilon$ if $\lim_{\varepsilon\to 0}(\alpha_\varepsilon/\beta_\varepsilon) = 1$. Denote $\alpha_\varepsilon \asymp \beta_\varepsilon$ if $0 < \liminf_{\varepsilon\to 0}(\alpha_\varepsilon/\beta_\varepsilon) \leq \limsup_{\varepsilon\to 0}(\alpha_\varepsilon/\beta_\varepsilon) < \infty$. The l_2-norm is denoted by $\|\cdot\|_2$.

2 Definition of the estimator and results

We consider the observation (1) and the estimation of the quadratic functional defined in (2).

Scale of classes. Consider a collection of l_p bodies $\Theta_{s,L} \subset \mathbf{R}^\infty$ defined in (3). We assume that $(s, L) \in A$, where

$$A = [s_*, \infty) \times [L_*, L^*],$$
$$\max\{\kappa, (3 - p + 4\kappa)/2p\} < s_* < 2\kappa + 1/4, \tag{7}$$

$p > 2$, and $0 < L_* < L^* < \infty$. Here A is the range of adaptation. The parameter p is assumed to be known and fixed. We do not need to know the values of L_*, L^* to construct our estimators. Since $s_* < 2\kappa + 1/4$, we are partly in the irregular region and partly in the regular region. We need to know s_*, see (13).

We have assumed that $s_* < 2\kappa + 1/4$. When $s_* \geq 2\kappa + 1/4$, then the smoothness parameter is always in the regular region. Adaptive estimation in this regular case was considered by Efroimovich (1994). We will assume that the irregular region cannot be excluded. The appropriate estimator for the regular case is in fact given in (20).

2.1 Definition of the estimator

The estimator to be defined is obtained by filtering the sequence of shifted squared observations and choosing the filter sequence in a data-dependent way.

Filter sequence. We will filter the shifted squared observations with the sequence

$$g_i = g_i(s, L, \gamma) = i^{2\kappa}(1 + bi^q)^{-1}, \ \ i = 1, 2, \ldots, \tag{8}$$

where q is defined in (4),

$$b = b(s, L, \gamma) = \left(\frac{(4\kappa + 1) I_2^{1/2} \gamma}{4(s - \kappa) I_1^{(p-2)/p} L} \right)^{2q/(4s+1)}, \tag{9}$$

$$I_1 = \int_0^\infty t^q \phi_{s,p}^{p/(p-2)}(t)\, dt, \tag{10}$$

$$I_2 = \int_0^\infty \phi_{s,p}^2(t)\, dt, \tag{11}$$

and

$$\phi_{s,p}(t) = t^{2\kappa}(1 + t^q)^{-1}. \tag{12}$$

The finiteness of the integral I_2 is guaranteed by (7). Parameter γ will be chosen later as $\gamma \asymp \varepsilon^2 \log \varepsilon^{-1}$, see (18).

Grid. We will construct a finite grid on $[s_*, 2\kappa + 1/4]$. Corresponding to each grid-point there will be defined a quadratic estimator. Then a statistic $\hat{s}$ will be defined which has values on the grid and thus will choose one of the estimators.

Let the grid be

$$B_{grid} = \{s_1, \ldots, s_m\},$$

where the values s_i and the integer m are chosen such that:

$$s_* = s_1 < \cdots < s_m = 2\kappa + 1/4, \tag{13}$$

with

$$k_1(\log \delta^{-1})^{-c_1} \le s_{i+1} - s_i \le k_2(\log \delta^{-1})^{-c_2}, \quad i = 1, \ldots, m-2 \tag{14}$$

where $k_2 > k_1 > 0$, $c_1 \ge c_2 > 1$, and where

$$\delta = \sqrt{2}\varepsilon^2. \tag{15}$$

Let also

$$s_m - s_{m-1} = k_3(\log\log\log \delta^{-1})^{-c_3} \tag{16}$$

where $k_3, c_3 > 0$.

Scale of estimators. We will define quadratic estimators $Q_{s,L,\gamma}$ for $(s, L) \in A$ by

$$Q_{s,L,\gamma} = \sum_{i=1}^{\infty} g_i(s, L, \gamma)(y_i^2 - \varepsilon^2) \tag{17}$$

where observation y_i is defined in (1) and filter sequence g_i is defined in (8).

Without loosing generality, we may assume that $1 \in [L_*, L^*]$. Let us define a quadratic estimator corresponding to each grid-point by calibrating the estimator $Q_{s,1,\gamma}$, defined in (17). First, “adaptive factor” is defined, denoting $r(s) = 4(s - \kappa)/(4s + 1)$, as

$$d_\varepsilon(s, s') = 2\rho[r(s') - r(s)] \log \delta^{-1}$$

where δ was defined in (15) and $\rho \ge 1$ will be the power of the loss function. ”Noise under adaptation” is defined as

$$\tilde{\delta} = \tilde{\delta}(s) = \delta \, d_\varepsilon(s, 2\kappa + 1/4). \tag{18}$$

Define

$$\widehat{Q}_s = \begin{cases} Q_{s,1,\tilde{\delta}} & \text{for } s = s_1, \ldots, s_{m-1}, \\ \bar{Q}_\varepsilon & \text{for } s = s_m = 2\kappa + 1/4. \end{cases} \tag{19}$$

Estimator $\bar{Q}_\varepsilon$ is defined by

$$\bar{Q}_\varepsilon = \begin{cases} \sum_{i=1}^{\lfloor J/b_\varepsilon \rfloor} i^{2\kappa}(y_i^2 - \varepsilon^2), & \text{when } I \le \varepsilon\sqrt{b_\varepsilon}, \\ \sum_{i=1}^{\lfloor J/b_\varepsilon \rfloor} i^{2\kappa}(y_i^2 - \varepsilon^2) + I, & \text{when } I > \varepsilon\sqrt{b_\varepsilon} \end{cases} \tag{20}$$

where $J = \varepsilon^{2/(4\kappa+1)}$, b_ε is such that

$$\lim_{\varepsilon\to 0} b_\varepsilon = \infty, \ \ \lim_{\varepsilon\to 0}(b_\varepsilon \varepsilon^a) = 0 \text{ for all } a > 0, \tag{21}$$

and

$$I = \sum_{i=\lfloor J/b_\varepsilon \rfloor+1}^{\lfloor J \rfloor} i^{2\kappa}\left(y_i^2 - \varepsilon^2\right).$$

Estimator $\bar{Q}_\varepsilon$ was proposed in Efroimovich (1994).

Adaptive estimator. Adaptive estimator has the form $\widehat{Q}_{\hat{s}}$ where $\hat{s}$ is suitably chosen statistic. To define $\hat{s}$ we follow the approach of Lepski (1990, 1991, 1992a,b). The statistic $\hat{s}$ is defined as the largest of those s-values in the grid for which the corresponding estimator does not differ significantly from the estimators corresponding to the smaller s-values. That is, define

$$\hat{s} = \max\left\{ s \in B_{grid} : \left|\widehat{Q}_s - \widehat{Q}_{s'}\right| \le \eta(s') \text{ for all } s' \in B_{grid},\ s' \le s \right\}$$

where the threshold η is defined as

$$\eta(s) = \sigma_s d_\varepsilon(s, 2\kappa + 1/4) \tag{22}$$

where

$$\sigma_s = \delta \left\| g(s, 1, \tilde{\delta}(s)) \right\|_2 \tag{23}$$

and δ is defined in (15). Finally, the estimator of $Q(\theta)$ is defined as

$$Q_\varepsilon^* = \widehat{Q}_{\hat{s}}. \tag{24}$$

2.2 Results

The supremum of the risk of an estimator Q_ε over the parameter space $\Theta_{s,L}$, with respect to normalising factor $\psi > 0$, is denoted

$$\mathcal{R}_{\varepsilon,s,L}(Q_\varepsilon, \psi) = \psi^{-\rho} \sup_{\theta\in\Theta_{s,L}} E_\theta\left(|Q_\varepsilon - Q(\theta)|^\rho\right)$$

where $\rho \ge 1$. The normalising factor to be defined depends on smoothness index s. Recall that $\tilde{\delta}(s)$ was defined in (18) so that $\tilde{\delta}(s) \asymp \varepsilon^2 \log \varepsilon^{-1}$. Define

$$\varphi_{s,L} = \tilde{\delta}(s)^{4(s-\kappa)/(4s+1)} L^{(4\kappa+1)/(4s+1)}. \tag{25}$$

Define the sharp constant for the upper bound by

$$\begin{aligned} C_{s,p} &= I_1^{(p-2)(4\kappa+1)/[p(4s+1)]} I_2^{2(s-\kappa)/(4s+1)} \\ &\quad \times \left(\frac{4\kappa+1}{4(s-\kappa)}\right)^{4(s-\kappa)/(4s+1)} \frac{4s+1}{4\kappa+1}\,, \end{aligned} \tag{26}$$

where I_1 is defined in (10) and I_2 is defined in (11). Define the sharp constant for the lower bound by

$$c_{s,p} = I_1^{-2(4\kappa+1)/[p(4s+1)]} I_3^{-2(s-\kappa)/(4s+1)} I_4 \tag{27}$$

where

$$I_3 = \int_0^\infty \phi_{s,p}^{4/(p-2)}, \tag{28}$$

$$I_4 = \int_0^\infty t^{2\kappa} \phi_{s,p}^{2/(p-2)}(t)dt, \tag{29}$$

and $\phi_{s,p}$ is defined in (12).

Theorem 1 gives sharp bounds when $s < 2\kappa + 1/4$. In fact, we consider case $s \le s_{m-1}$, so that smoothness index s is bounded away from $2\kappa + 1/4$. However, $s_{m-1} \to 2\kappa + 1/4$ as $\varepsilon \to 0$.

Theorem 1 *Let $p > 2$. Denote $B = [s_*, s_{m-1}] \times [L_*, L^*]$. Then, for the estimator Q_ε^* defined in (19) and in (24),*

$$\limsup_{\varepsilon\to 0} \sup_{(s,L)\in B} \mathcal{R}_{\varepsilon,s,L}(Q_\varepsilon^*, C_{s,p}\varphi_{s,L}) \le 1. \tag{30}$$

We have the lower bound

$$\liminf_{\varepsilon\to 0} \inf_{Q_\varepsilon} \sup_{(s,L)\in B} \mathcal{R}_{\varepsilon,s,L}(Q_\varepsilon, c_{s,p}\varphi_{s,L}) \ge 1 \tag{31}$$

where the infimum is taken with respect to all estimators.

In Theorem 1 we considered case $s < 2\kappa + 1/4$. Let us next consider the regular area $[2\kappa + 1/4, \infty)$. We will will prove the rate optimality on the boundary $s = 2\kappa + 1/4$ and sharp optimality for cases $s > 2\kappa + 1/4$.

Theorem 2 *Let estimator Q_ε^* be defined in (19) and in (24). Then, for the case $s = 2\kappa + 1/4$,*

$$\limsup_{\varepsilon\to 0} \left(\varepsilon^2 b_\varepsilon\right)^{-1} \sup_{L\in[L_*,L^*]} \sup_{\theta\in\Theta_{2\kappa+1/4,L}} E_\theta\left(|Q_\varepsilon^* - Q(\theta)|^2\right) < \infty \tag{32}$$

where b_ε was defined in (21). For $s > 2\kappa + 1/4$,

$$\lim_{\varepsilon\to 0} \sup_{L\in[L_*,L^*]} \sup_{\theta\in\Theta_{s,L}} \left[\varepsilon^{-2} E_\theta\left(|Q_\varepsilon^* - Q(\theta)|^2\right) - 4Q_{2\kappa}(\theta)\right] = 0, \tag{33}$$

where $Q_{2\kappa}(\theta)$ is defined in (2).

A proof of Theorem 1 is given in Section 4.5 and a proof of Theorem 2 is given in the Appendix A of the technical report.

2.3 Discussion

2.3.1 Sharpness of the bound of Theorem 1

When $p = 4$, then the upper and lower bounds of Theorem 1 are equal, so that we have

$$\begin{aligned} C_{s,4} = c_{s,4} &= \left[\frac{1}{4s+1} B \left(1 - \frac{4\kappa+1}{4s+1}, 1 + \frac{4\kappa+1}{4s+1} \right) \right]^{1/2} \\ &\times \left(\frac{4\kappa+1}{4(s-\kappa)} \right)^{2(s-\kappa)/(4s+1)} \frac{4s+1}{4\kappa+1} \end{aligned} \tag{34}$$

where $B(c,d) = \int_0^1 x^{c-1}(1-x)^{d-1}\,dx$, $c,d > 0$, is the beta function. To see this we apply the properties $B(c,d) = B(c-1,d+1)(c-1)/d$ and $B(c,d) = B(c,d+1)(c+d)/d$.

When $p \neq 4$ the constants $C_{s,p}$ and $c_{s,p}$ are close to each other. Figure 1 shows the ratio $C_{s,p}/c_{s,p}$ when $(s,p) \in (0,1/4) \times [3,100]$ and $\kappa = 0$. From Figure 1 and related figures one may conclude that

$$\sup_{s\in(0,1/4),p\in[3,4]} \frac{C_{s,p}}{c_{s,p}} \leq 1.055, \qquad \sup_{s\in(0,1/4),p\in[4,100]} \frac{C_{s,p}}{c_{s,p}} \leq 1.04$$

when $\kappa = 0$. That is, the bound is better when $p \geq 4$, than in the case $3 \leq p \leq 4$. When $p < 3$, then the bound is still good but it starts deteriorating when $p \downarrow 2$. Remember, that we have also to restrict ourselves to the case $s > (3-p)/2p$, to guarantee the finiteness of (11).

When $\kappa = 2$, then

$$\sup_{s\in(0,1/4),p\in[3,4]} \frac{C_{s,p}}{c_{s,p}} \leq 1.019, \qquad \sup_{s\in(0,1/4),p\in[4,100]} \frac{C_{s,p}}{c_{s,p}} \leq 1.27.$$

That is, the bound is better when $3 \leq p \leq 4$. To reproduce and modify the Figures, see the Appendix L of the technical report.

2.3.2 Elbow in rates

The optimal adaptive rates for the estimation of the quadratic functional are

$$\begin{array}{ll} (\varepsilon^2 \log \varepsilon^{-1})^{4(s-\kappa)/(4s+1)}, & s \leq s_{m-1} \\ \varepsilon\sqrt{b_\varepsilon}, & s = 2\kappa + 1/4 \\ \varepsilon, & s > 2\kappa + 1/4 \end{array}$$

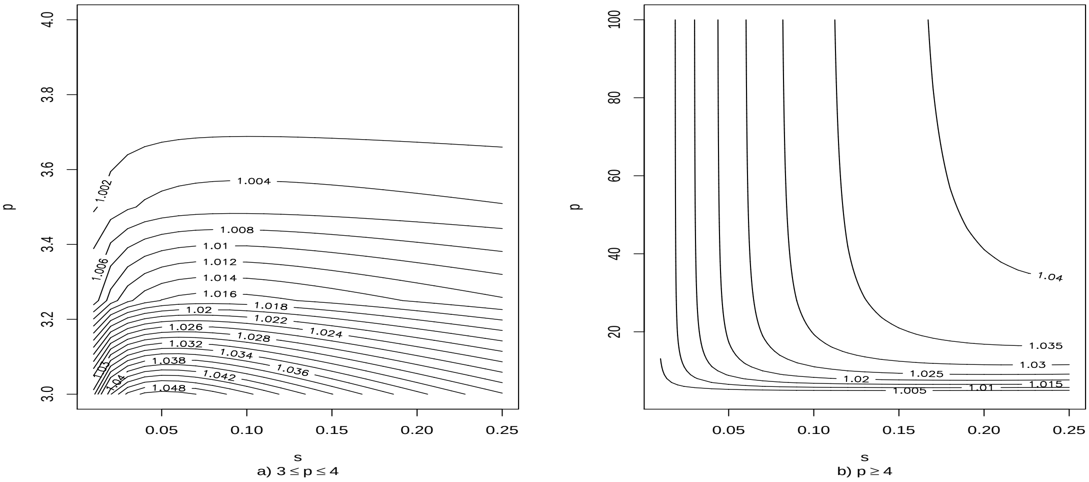

Figure 1: Contour of $C_{s,p}/c_{s,p}$ as a function of (s,p). That is, we plot the ratio of the constant in the upper bound to the constant in the lower bound. In a) we have $(s,p) \in (0,1/4)\times[3,4]$. In b) we have $(s,p) \in (0,1/4)\times[4,100]$.

where s_{m-1} satisfies (16), so that $s_{m-1} \to 2\kappa + 1/4$, and where b_ε satisfies (21). We have considered scale $[s_*,\infty)$ where $\max\{\kappa, (3-p+4\kappa)/2p\} < s_* < 2\kappa + 1/4$. We have constructed an estimator which is sharp adaptive in region $[s_*, s_{m-1}] \cup (2\kappa+1/4,\infty)$. For the region $(s_{m-1}, 2\kappa+1/4)$ we have not made claims on the rate of convergence of the estimator. The estimator is proved to be adaptive rate optimal when $s = 2\kappa + 1/4$ and sharp adaptive optimal when $s > 2\kappa + 1/4$.

2.3.3 Normalizing factor and the modulus of continuity

Following Donoho and Nussbaum (1990), let us define the *modulus of continuity* of the functional $Q(\theta)$ defined in (2) by

$$\Omega_{s,L}(\gamma) = \sup\left\{ Q(\theta) : \theta \in \Theta_{s,L},\ \sum_{i=1}^{\infty} \theta_i^4 \le \gamma^2 \right\}.$$

It was shown by Donoho and Nussbaum (1990) that the minimax rate of convergence with the parameter space $\Theta_{s,L}$ is given under certain conditions by the modulus of continuity $\Omega_{s,L}(\varepsilon^2)$. One may show that for the case $p = 4$ the optimal normalizing factor of Theorem 1 is equal to the modulus of continuity at the effective noise:

$$\Omega_{s,L}\left(\tilde{\delta}(s)\right) = c_{s,p}\varphi_{s,L} = C_{s,p}\varphi_{s,L}.$$

This follows from the next theorem and (34).

Theorem 3 *Let $\Theta_{s,L}$ be the l_p body defined in (3). Let $p > 2$ and $s > \max\{\kappa, (3-p+4\kappa)/(2p)\}$. We have that*

$$
\begin{aligned}
c_{s,p} &\leq \liminf_{\gamma\to 0} \frac{\Omega_{s,L}(\gamma)}{\gamma^{4(s-\kappa)/(4s+1)} L^{(4\kappa+1)/(4s+1)}} \\
&\leq \limsup_{\gamma\to 0} \frac{\Omega_{s,L}(\gamma)}{\gamma^{4(s-\kappa)/(4s+1)} L^{(4\kappa+1)/(4s+1)}} \leq C_{s,p}
\end{aligned}
$$

where $c_{s,p}$ and $C_{s,p}$ are the constants of Theorem 1 defined in (26) and (27).

Theorem 3 is proved in Section 3. Theorem 3 has independent interest since according to Donoho and Nussbaum (1990), it implies bounds for the risks of quadratic estimators.

2.3.4 Optimal recovery

The weigths g_i of the estimator $Q_{s,L,\gamma}$ are related to a certain optimal recovery problem. Namely, we have for the case $p = 4$ the asymptotic equality

$$
\begin{aligned}
&\sup_{\theta\in\Theta_{s,L}} \sup_{\|(\theta_i^2)-(\nu_i^2)\|_2\leq\gamma} \left| \sum_{i=1}^\infty g_i\nu_i^2 - \sum_{i=1}^\infty i^{2\kappa}\theta_i^2 \right| \\
&\sim \inf_{(q_i)} \sup_{\theta\in\Theta_{s,L}} \sup_{\|(\theta_i^2)-(\nu_i^2)\|_2\leq\gamma} \left| \sum_{i=1}^\infty q_i\nu_i^2 - \sum_{i=1}^\infty i^{2\kappa}\theta_i^2 \right|
\end{aligned}
$$

where g_i are defined in (8) and $\gamma \to 0$. One can find the general formula for the optimal recovery filter from Donoho and Nussbaum (1990). The general formula for the optimal recovery filter is given by calibrating the extremal sequence θ_i which achieves the supremum in the definition of the modulus of continuity.

2.3.5 Comparison with the estimation of linear functionals

Reduction to the linear case. An idea behind the proof of Theorem 1 is to reduce the quadratic case to the linear case. The observation was defined as $y_i = \theta_i + \varepsilon z_i$ where z_i are iid standard Gaussian. We have that $E_\theta Q_{s,L,\gamma} = \sum_{i=1}^\infty g_i(s,L,\gamma)\theta_i^2$. Let us define the bias term as

$$
B(s,L,\gamma) = \sup_{\theta\in\Theta_{s,L}} \left| \sum_{i=1}^\infty g_i(s,L,\gamma)\theta_i^2 - Q(\theta) \right|. \tag{35}
$$

We can write $Q_{s,L,\gamma} - E_\theta Q_{s,L,\gamma} = Z_{s,L,\gamma} + U_{s,L,\gamma}(\theta)$ where

$$Z_{s,L,\gamma} = \varepsilon^2 \sum_{i=1}^{\infty} g_i (z_i^2 - 1) \tag{36}$$

and

$$U_{s,L,\gamma}(\theta) = 2\varepsilon \sum_{i=1}^{\infty} g_i \theta_i z_i. \tag{37}$$

Let us define the standard deviation term as

$$R(s, L, \gamma) = \gamma \, \|g(s, L, \gamma)\|_2 \tag{38}$$

where $g = (g_i)$. Note that $\mathrm{Var}(Z_{s,L,\gamma}) = 2\varepsilon^4 \sum_{i=1}^{\infty} g_i^2$ and thus $\mathrm{std}(Z_{s,L,\delta}) = R(s, L, \delta)$ where $\delta = \sqrt{2}\varepsilon^2$. It will turn out that the heteroscedastic term $U_{s,L,\gamma}(\theta)$ is in a certain sense asymptotically negligible when $s < 2\kappa + 1/4$. Variance term $Z_{s,L,\gamma}$ is linear in $(z_i^2 - 1)$ and bias term $B(s, L, \gamma)$ is linear in (θ_i^2).

Normalizing factor. In this article it is shown that for $p = 4$ the adaptive minimax risk is asymptotically equal to

$$\Omega_{s,L}(\tilde{\delta}(s)), \qquad \tilde{\delta}(s) = \delta 2\rho(r^{\#} - r) \log \delta^{-1},$$

where $\tilde{\delta}(s)$ is the "effective noise", $r = 4(s - \kappa)/(4s + 1)$, $r^{\#} = 4(s^{\#} - \kappa)/(4s^{\#} + 1)$, $\delta = \sqrt{2}\varepsilon^2$, and $s^{\#} = 2\kappa + 1/4$. Let us compare this to the case of the estimation of linear functionals. It was shown by Ibragimov and Hasminskii (1984) and Donoho and Liu (1991) that the minimax rate for the estimation of a linear functional $T(\theta)$ is given by $\omega_{s,L}(\varepsilon)$ where

$$\omega_{s,L}(\gamma) = \sup \left\{ T(\theta) : \theta \in \Theta_{s,L}, \sum_{i=1}^{\infty} \theta_i^2 \le \gamma^2 \right\}.$$

It was shown by Klemelä and Tsybakov (2001) that under certain conditions the adaptive minimax risk is asymptotically equal to

$$\omega_{s,L}(\tilde{\varepsilon}(s)), \qquad \tilde{\varepsilon}(s) = \varepsilon(2\rho(r^* - r)\log \varepsilon^{-1})^{1/2},$$

where $r = 2(s - \kappa)/(2s + 1)$, $r^* = 2(s^* - \kappa)/(2s^* + 1)$, and s^* is an upper bound for the smoothness index. Klemelä and Tsybakov (2004) consider the case $s^* \to \infty$ and in this case the exact constant is get from the previous by setting $s^* = \infty$.

3 Proof of Theorem 3

We will prove Theorem 3 before proving Theorems 1 and 2 because the proof of this theorem will contain facts to be used later.

Upper bound. Let (g_i) be the filter sequence defined in (8). We have that

$$Q(\theta) = \sum_{i=1}^{\infty} (i^{2\kappa} - g_i)\theta_i^2 + \sum_{i=1}^{\infty} g_i \theta_i^2.$$

Let θ be such that

$$\sum_{i=1}^{\infty} \theta_i^4 \le \gamma^2, \qquad \sum_{i=1}^{\infty} i^q |\theta_i|^p \le L^{p/2}.$$

Then,

$$\sum_{i=1}^{\infty} g_i\theta_i^2 \le \left(\sum_{i=1}^{\infty} g_i^2\right)^{1/2} \left(\sum_{i=1}^{\infty} \theta_i^4\right)^{1/2} \le \gamma \left(\sum_{i=1}^{\infty} g_i^2\right)^{1/2}. \tag{39}$$

We have

$$\sum_{i=1}^{\infty} g_i^2 = \sum_{i=1}^{\infty} i^{4\kappa}(1+bi^q)^{-2} \sim b^{-(4\kappa+1)/q} \int_0^{\infty} \phi_{s,p}^2 \tag{40}$$

where $\phi_{s,p}$ is defined in (12). Also, denoting $r = p/2$, $r' = p/(p-2)$, because $i^{2\kappa} - g_i = bi^{q+2\kappa}/(1+bi^q) = bi^q g_i$,

$$\begin{aligned} \sum_{i=1}^{\infty}(i^{2\kappa} - g_i)\theta_i^2 &\le \left(\sum_{i=1}^{\infty} i^q|\theta_i|^p\right)^{2/p} \left(\sum_{i=1}^{\infty}[(i^{2\kappa} - g_i)i^{-q/r}]^{r'}\right)^{1/r'} \\ &\le bL\left(\sum_{i=1}^{\infty} i^q g_i^{p/(p-2)}\right)^{(p-2)/p}. \end{aligned} \tag{41}$$

We have

$$\begin{aligned} \sum_{i=1}^{\infty} i^q g_i^{p/(p-2)} &= \sum_{i=1}^{\infty} i^q [i^{2\kappa}(1+bi^q)^{-1}]^{p/(p-2)} \\ &\sim b^{-1-1/q-2\kappa p/[q(p-2)]} \int_0^{\infty} t^q \phi_{s,p}^{p/(p-2)}(t)\, dt. \end{aligned} \tag{42}$$

Finally, by the choice of b in (9),

$$\begin{aligned} Q(\theta) &\le b^{\alpha} L I_1^{(p-2)/p} + b^{-\beta}\gamma I_2^{1/2} \\ &= (I_2^{1/2}\gamma)^{\alpha/(\alpha+\beta)} (I_1^{(p-2)/p} L)^{\beta/(\alpha+\beta)} \left(\frac{\beta}{\alpha}\right)^{\alpha/(\alpha+\beta)} \left(1 + \frac{\alpha}{\beta}\right) \\ &= \gamma^{4(s-\kappa)/(4\kappa+1)} L^{(4\kappa+1)/(4s+1)} C_{s,p}, \end{aligned} \tag{43}$$

where $\alpha = 1 - (p-2)(q+1)/(pq) - 2\kappa/q = 2(s-\kappa)/q$, $\beta = (4\kappa+1)/(2q)$, I_1 is defined in (10), and I_2 is defined in (11). The upper bound is proved.

Lower bound. Let

$$\bar{\theta}_i = \left[a_1 i^{2\kappa}(1+b_1 i^q)^{-1}\right]^{1/(p-2)} \tag{44}$$

where a_1 and b_1 are such that

$$\sum_{i=1}^{\infty} \bar{\theta}_i^4 = \gamma^2, \qquad \sum_{i=1}^{\infty} i^q |\bar{\theta}_i|^p = L^{p/2}. \tag{45}$$

We have that

$$\sum_{i=1}^{\infty} \bar{\theta}_i^4 \sim a_1^{4/(p-2)} b_1^{-1/q - 8\kappa/[q(p-2)]} \int_0^\infty \phi_{s,p}^{4/(p-2)}. \tag{46}$$

Also,

$$\sum_{i=1}^{\infty} i^q |\bar{\theta}_i|^p \sim a_1^{p/(p-2)} b_1^{-1-1/q-2\kappa p/[q(p-2)]} \int_0^\infty t^q \phi_{s,p}^{p/(p-2)}(t)\, dt. \tag{47}$$

From (46),

$$a_1^{2/(p-2)} \sim \gamma b_1^{(p-2+8\kappa)/[2q(p-2)]} I_3^{-1/2} \tag{48}$$

and from (47),

$$a_1^{2/(p-2)} \sim L b_1^{[2(p-2)(q+1)+4\kappa p]/[qp(p-2)]} I_1^{-2/p} \tag{49}$$

where I_3 is defined in (28) and I_1 is defined in (10). Combining (48) and (49),

$$b_1 \sim L^{-1}\gamma \left(I_1^{2/p} I_2^{-1/2}\right)^{2q/(4s+1)} \tag{50}$$

where we used $q+1 = p(s+1/2)$. Now

$$\begin{aligned} Q(\bar{\theta}) &= \sum_{i=1}^{\infty} i^{2\kappa} \bar{\theta}_i^2 \\ &= a_1^{2/(p-2)} \sum_{i=1}^{\infty} i^{2\kappa} [i^{2\kappa}(1+b_1 i^q)^{-1}]^{2/(p-2)} \\ &\sim a_1^{2/(p-2)} b_1^{-1/q-2\kappa/q-4\kappa/[q(p-2)]} \int_0^\infty t^{2\kappa} \phi_{s,p}^{2/(p-2)}(t)\, dt. \end{aligned} \tag{51}$$

The lower bound follows from (48), (50), and (51).

4 Proof of Theorem 1

In the following, let $C, C', C_1, C_2, \ldots$ denote generic positive constants, possibly different in different occurrences, which can depend only on s_*, L_*, L^*, κ, p. We denote for shortness

$$s^{\#} = 2\kappa + 1/4,$$
$$\psi_{s,L} = C_{s,p}\varphi_{s,L}, \tag{52}$$

where $C_{s,p}$ is defined in (26) and $\varphi_{s,L}$ is defined in (25). Define also

$$r(s) = \frac{4(s-\kappa)}{4s+1}\,. \tag{53}$$

4.1 Preliminary lemmas

Define,

$$\tilde{b} = \left(\frac{\beta I_2^{1/2}}{\alpha I_1^{(p-2)/p}}\right)^{1/(\alpha+\beta)},$$

$\alpha = 2(s-\kappa)/q$, $\beta = (4\kappa+1)/(2q)$, $q = p(s-1/p+1/2)$, I_1 is defined in (10), and I_2 is defined in (11). Define, for $s \in [s_*, s^{\#}]$,

$$b(s) = L\tilde{b}^{\alpha} I_1^{(p-2)/p}$$

and

$$v(s) = \gamma \tilde{b}^{-\beta} I_2^{1/2}.$$

We call $b(s)$ the bias component and $v(s)$ the standard deviation component.

Let us state a result concerning the bias-standard deviation composition.

Lemma 4 *(i) For $C_{s,p}$ defined in (26),*

$$C_{s,p} = b(s) + v(s),$$

(ii) $0 < \inf_{s\in[s_*,s^{\#}]} b(s)$, $0 < \inf_{s\in[s_*,s^{\#}]} v(s)$,

(iii) b and v are continuous as a function of $s \in [s_*, s^{\#}]$,

(iv) for $B(s, L, \gamma)$ as in (35), $r(s)$ as in (53),

$$\limsup_{\gamma\to 0} \sup_{(s,L)\in A_0} \frac{B(s,L,\gamma)}{\gamma^{r(s)} L^{1-r(s)} b(s)} \le 1,$$

where $A_0 = [s_*, s^{\#}] \times [L_*, L^*]$, *(v) for $R(s,L,\gamma)$ as in (38), $r(s)$ as in (53),*

$$\limsup_{\gamma\to 0} \sup_{(s,L)\in A_0} \frac{R(s,L,\gamma)}{\gamma^{r(s)} L^{1-r(s)} v(s)} \le 1,$$

where $A_0 = [s_*, s^\#] \times [L_*, L^*]$.

Proof. Item (i) follows from (43). Items (ii) and (iii) follow from the definitions of $b(s)$ and $v(s)$. Item (iv) follows from (41) and (42). Item (v) follows from (39) and (40). □

We formulate one additional lemma on the bias of the estimator.

Lemma 5 *Let*

$$\mathcal{B}_{s,L}(s',L',\gamma) = L\,(\gamma/L')^{4(\tilde{s}-\kappa)/(4s'+1)}\,\tilde{b}(s,s')$$

where $(s,L),(s',L') \in [s_*, s^\#] \times [L_*, L^*]$,

$$\tilde{s} = \tilde{s}(s,s') = \min\{s'+c, s\},$$

$c = p/4 - 1/2$,

$$\tilde{b}(s,s') = \left(\frac{\beta' I_2(s')^{1/2}}{\alpha' I_1(s')^{(p-2)/p}}\right)^{2/(4s'+1)} \left[\int_0^\infty \left(\frac{t^{2\kappa+q'-\tilde{q}/r}}{1+t^{q'}}\right)^{r'} dt\right]^{1/r'}$$

where $\alpha' = 2(s'-\kappa)/q'$, $\beta' = (4\kappa+1)/(2q')$, $q' = p(s'-1/p+1/2)$, $\tilde{q} = \tilde{q}(s,s') = p(\tilde{s}-1/p+1/2)$, $I_1(s)$ is defined in (10), $I_2(s)$ is defined in (11), $r = p/2$, and $r' = p/(p-2)$. Then,

(i) for $A_0 = [s_*, s^\#] \times [L_*, L^*]$,

$$\limsup_{\gamma\to 0} \sup_{(s,L),(s',L')\in A_0} \frac{\sup_{\theta\in\Theta_{s,L}} \left|\sum_{i=1}^\infty g_i(s',L',\gamma)\theta_i^2 - Q(\theta)\right|}{\mathcal{B}_{s,L}(s',L',\gamma)} \le 1,$$

(ii) $0 < \inf_{s,s'\in[s_*,s^\#]} \tilde{b}(s,s')$, $\tilde{b}$ is continuous, and $\tilde{b}(s,s) = b(s)$.

Lemma 5 is proved in the Appendix C of the technical report.

Lemma 5 (i) says that $\mathcal{B}_{s,L}(s',L',\gamma)$ is an upper bound for the bias when the parameter space is $\Theta_{s,L}$ but the estimator is constructed for the parameter space $\Theta_{s',L'}$. It says also that when $s' < s$, then $\mathcal{B}_{s,L}(s',L',\gamma)$ is smaller than $B(s',L',\gamma)$ satisfying Lemma 4 (iv). We will apply Lemma 5 to guarantee that when we use pessimistically the estimator corresponding to smoothness index s', when the true smoothness index is $s > s'$, then we still profit from the true smoothness in the sense that the bias will be smaller than $B(s',L',\gamma)$. The Appendix M of the technical report explains why Lemma 5 does not hold for the case $p = 2$.

We will need the nestedness of the l_p bodies.

Lemma 6 *For $s_* \le s' < s < \infty$, $L \in [L_*, L^*]$, $\Theta_{s,L} \subset \Theta_{s',L}$.*

This lemma follows from the definition of the l_p body.

4.2 Technical lemmas

We will need the following technical lemma.

Lemma 7 *Let $\tilde{\delta}$ be defined in (18). For $s' < s < s^\#$ and $s''' \le s'' < s^\#$, denoting $r(s) = 4(s-\kappa)/(4s+1)$,*

$$\frac{\tilde{\delta}^{r(s)}(s'')}{\tilde{\delta}^{r(s')}(s''')} \le \exp\left\{-\frac{1}{2}[r(s) - r(s')]\log \delta^{-1}\right\}, \tag{54}$$

for $s' < s \le s_{m-1}$,

$$\frac{\tilde{\delta}^{r(s')}(s')}{\tilde{\delta}^{r(s)}(s)} \le C_1(s^\# - s_{m-1})^{-C_2} \exp\left\{\frac{1}{2\rho}\, d_\varepsilon(s', s)\right\}, \tag{55}$$

and for $s' < s \le s^\#$,

$$\frac{\tilde{\delta}^{r(s')}(s')}{\delta^{r(s)}} \le C_3(\log \delta^{-1})^{C_4} \exp\left\{\frac{1}{2\rho}\, d_\varepsilon(s', s)\right\}. \tag{56}$$

Also, for $s < s^\#$,

$$C_5 \tilde{\delta}^{r(s)}(s) \le \psi_{s,L}, \eta(s), B(s, L, \tilde{\delta}(s)) \le C_6 \tilde{\delta}^{r(s)}(s). \tag{57}$$

Lemma 7 is proved in the Appendix E of the technical report.

We will need certain exponential bounds for the stochastic parts of the estimator. Denote for shortness

$$Z_s = Z_{s,1,\tilde{\delta}(s)} \qquad U_s(\theta) = U_{s,1,\tilde{\delta}(s)}(\theta)$$

where $Z_{s,1,\tilde{\delta}(s)}$ and $U_{s,1,\tilde{\delta}(s)}(\theta)$ are defined in (36) and (37).

Lemma 8 *For $\rho = 0$ or $\rho \ge 1$, with σ_s defined in (23),*

(i) *when $x_\varepsilon > 0$,*

$$x_\varepsilon = o\left(\frac{\sigma_s^2}{\varepsilon^2 |g|_\infty}\right)$$

where $|g|_\infty = \sup_{i=1,2,\dots} |g_i(s, L, \tilde{\delta}(s))|$, then for sufficiently small ε,

$$E\left(|Z_s|^\rho I(|Z_s| \ge x_\varepsilon)\right) \le C \exp\left\{-\frac{1}{2}\frac{x_\varepsilon}{\sigma_s}\right\} \sum_{i=0}^{\lceil \rho \rceil} x_\varepsilon^{\rho - i} \sigma_s^i.$$

(ii) *for $x > 0$,*

$$E\left(|U_s(\theta)|^\rho I(|U_s(\theta)| \geq x)\right) \\ \leq C \exp\left\{-\frac{1}{2}\frac{x^2}{\mathrm{Var}(U_s(\theta))}\right\}\left(x^\rho + (\mathrm{Var}(U_s(\theta)))^{\rho/2}\right).$$

Lemma 8 is proved in the Appendix F of the technical report.

In the application of Lemma 8 we will use the fact that by Lemma 13 (iii) (given in the Appendix B of the technical report),

$$\frac{\sigma_s^2}{\varepsilon^2|g|_\infty} \geq C\,\frac{\tilde{\delta}(s)^{2r+c-1}}{d_\varepsilon(s, s^\#)} \tag{58}$$

where $c = 4\kappa/(4s+1)$ so that $2r + c - 1 = 4(s-\kappa)/(4s+1) - 1/(4s+1)$.

Let us prove that the probability of under-estimating largely the value of s by the statistic $\hat{s}$ is small, uniformly over $\theta \in \Theta_\nu$. Let us denote

$$s^- = s^-(s) = s - \frac{\log\log\log\delta^{-1}}{\log\delta^{-1}}. \tag{59}$$

Lemma 9 *Let $s \in [s_*, \infty)$ and $L \in [L_*, L^*]$. Let us denote $\nu = (s, L)$. Then, for $s' \in B_{grid} = \{s_1, \dots, s_m = s^\#\}$, $s' < s^-(\min\{s, s^\#\})$, for sufficiently small ε,*

$$\sup_{\theta\in\Theta_\nu} P_\theta\left(\hat{s} = s'\right) \leq C\mathrm{Card}(B_{grid}) \exp\left\{-\frac{1}{2}\,d_\varepsilon(s', s^\#)(1-\gamma_\varepsilon)\right\}$$

where $\gamma_\varepsilon = (\log\log\delta^{-1})^{-C'}$.

Lemma 9 is proved in the Appendix G of the technical report.

4.3 Lemmas to prove the upper bound of Theorem 1

Let us prove that the risk vanishes asymptotically when we are in the case where statistic $\hat{s}$ under estimates smoothness index s.

Lemma 10 *For ψ_ν defined in (52), Q_ε^* defined in (24) and in (19), and $s^- = s^-(\min\{s, s^\#\})$ defined in (59),*

$$\lim_{\varepsilon\to 0}\sup_{\nu\in B'}\sup_{\theta\in\Theta_\nu} E_\theta\left(\psi_\nu^{-\rho}\left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s} < s^-\right)\right) = 0$$

where either $B' = [s_*, s_{m-1}] \times [L_*, L^*]$, or $B' = [s_m, \infty) \times [L_*, L^*]$ where $s_m = s^\# = 2\kappa + 1/4$.

Proof. Let $s \in [s_*, s_{m-1}] \cup [s_m, \infty)$, $L \in [L_*, L^*]$ and denote $\nu = (s, L)$. Let $s' \in B_{grid}$, $s' < s^-$. By Lemma 6,

$$\sup_{\theta \in \Theta_{s,L}} \left| E_\theta \widehat{Q}_{s'} - Q(\theta) \right| \leq \sup_{\theta \in \Theta_{s_{\min},L}} \left| E_\theta \widehat{Q}_{s'} - Q(\theta) \right|$$

where $s_{min} = \min\{s, s^\#\}$. Then by Lemma 5 and Lemma 4 (iv), because $s' < s_{min}$, and by (57), for sufficiently small ε,

$$\sup_{\theta \in \Theta_{s_{min},L}} \left| E_\theta \widehat{Q}_{s'} - Q(\theta) \right| \leq C_1 \mathcal{B}_{s,L}(s', 1, \tilde{\delta}(s')) \leq C_2 B(s', 1, \tilde{\delta}(s')) \leq C_3 \psi_{s',L}.$$

Thus, for $\theta \in \Theta_{s,L}$, for sufficiently small ε,

$$\begin{aligned} \left| \widehat{Q}_{s'} - Q(\theta) \right| &\leq \left| E_\theta \widehat{Q}_{s'} - Q(\theta) \right| + |Z_{s'}| + |U_{s'}(\theta)| \\ &\leq C_1 \psi_{s',L} + |Z_{s'}| + |U_{s'}(\theta)| \,. \end{aligned}$$

Thus, for sufficiently small ε,

$$\sup_{\theta \in \Theta_\nu} E_\theta \left(\psi_\nu^{-\rho} \left| Q_\varepsilon^* - Q(\theta) \right|^\rho I\left(\hat{s} < s^- \right) \right) \leq \rho_1(\nu) + \rho_2(\nu) \tag{60}$$

where

$$\rho_1(\nu) = C \sum_{s' \in B_{grid}, s' < s^-} \sup_{\theta \in \Theta_\nu} P_\theta \left(\hat{s} = s' \right) \psi_\nu^{-\rho} \left(\psi_{s',L} + \tau(s') \right)^\rho , \tag{61}$$

$$\begin{aligned} \rho_2(\nu) = C &\sum_{s' \in B_{grid}, s' < s^-} \psi_\nu^{-\rho} \sup_{\theta \in \Theta_\nu} E \left[\left(\psi_{s',L} + |Z_{s'}| + |U_{s'}(\theta)| \right)^\rho \right. \\ &\left. I \left(|Z_{s'}| + |U_{s'}(\theta)| > \tau(s') \right) \right] , \end{aligned} \tag{62}$$

and

$$\tau(s') = 2\sigma_{s'} \left[d_\varepsilon(s', \min\{s, s^\#\}) + \left(\log \delta^{-1} \right)^{1/2} \right]$$

where σ_s is defined in (23). We have that

$$\lim_{\varepsilon \to 0} \sup_{\nu \in A} \rho_1(\nu) = 0 \tag{63}$$

and

$$\lim_{\varepsilon \to 0} \sup_{\nu \in A} \rho_2(\nu) = 0. \tag{64}$$

Equations (63) and (64) are proved in the Appendix H and Appendix I of the technical report. Lemma 10 follows from (60), (63), and (64). □

To prove the upper bound for Theorem 1 we need also to consider case when statistic $\hat{s}$ is not underestimating smoothness index s. We need to consider only the case $s \leq s_{m-1}$.

Lemma 11 *For $B = [s_*, s_{m-1}] \times [L_*, L^*]$,*

$$\limsup_{\varepsilon\to 0} \sup_{\nu\in B} \sup_{\theta\in\Theta_\nu} E_\theta \left(\psi_\nu^{-\rho} \left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s} \geq s^-\right)\right) \leq 1 \qquad (65)$$

where $s^- = s^-(s)$ is defined in (59).

Proof. Let $\nu = (s, L) \in B$. Denote by $\gamma_{\varepsilon i}$, $i = 1, 2, \ldots$, such generic positive numbers for which $\lim_{\varepsilon\to 0} \gamma_{\varepsilon i} = 0$ and which do not depend on s, L. Let $\bar{s}$ be such that

$$\left(\delta \log \delta^{-1}\right)^{1/(4\bar{s}+1)} = \left(\frac{\delta \log \delta^{-1}}{L}\right)^{1/(4s+1)}.$$

That is,

$$\bar{s} = s - \frac{(s + 1/4)\log L}{\log \delta^{-1} - \log\log \delta^{-1} + \log L}.$$

Denote $\tilde{s} = \tilde{s}(s, \bar{s})$ where $\tilde{s}$ is as in Lemma 5. Then, by Lemma 5 and Lemma 4 (iv), denoting $r(s) = 4(s - \kappa)/(4s + 1)$,

$$\begin{aligned}
\mathcal{B}_{s,L}(\bar{s}, 1, \tilde{\delta}(\bar{s})) &\leq L\tilde{\delta}^{4(\tilde{s}-\kappa)/(4\bar{s}+1)}(\bar{s})\tilde{b}(s, \bar{s})(1 + \gamma_{\varepsilon 1}) \\
&= L\left(2\rho(r(s^{\#}) - r(\bar{s}))\delta \log \delta^{-1}\right)^{4(\tilde{s}-\kappa)/(4\bar{s}+1)} \tilde{b}(s, \bar{s})(1 + \gamma_{\varepsilon 1}) \\
&= L\left(2\rho(r(s^{\#}) - r(\bar{s}))\,\frac{\delta \log \delta^{-1}}{L}\right)^{4(\tilde{s}-\kappa)/(4s+1)} \tilde{b}(s, \bar{s})(1 + \gamma_{\varepsilon 1}) \\
&\leq L\left(2\rho(r(s^{\#}) - r(s))\,\frac{\delta \log \delta^{-1}}{L}\right)^{4(s-\kappa)/(4s+1)} b(s)(1 + \gamma_{\varepsilon 2}) \\
&= L\left(\frac{\tilde{\delta}(s)}{L}\right)^{r(s)} b(s)(1 + \gamma_{\varepsilon 2}) \\
&\leq B(s, L, \tilde{\delta}(s))(1 + \gamma_{\varepsilon 3}).
\end{aligned} \qquad (66)$$

Also, by Lemma 4 (v),

$$\begin{aligned}
R(\bar{s}, 1, \tilde{\delta}(\bar{s})) &\leq \tilde{\delta}^{4(\bar{s}-\kappa)/(4\bar{s}+1)}(\bar{s})v(\bar{s})(1 + \gamma_{\varepsilon 4}) \\
&= \tilde{\delta}(\bar{s})\tilde{\delta}^{-(4\kappa+1)/(4\bar{s}+1)}(\bar{s})v(\bar{s})(1 + \gamma_{\varepsilon 4}) \\
&= \tilde{\delta}(\bar{s})\left(2\rho(r(s^{\#}) - r(\bar{s}))\delta \log \delta^{-1}\right)^{-(4\kappa+1)/(4\bar{s}+1)} v(\bar{s})(1 + \gamma_{\varepsilon 4}) \\
&= \tilde{\delta}(\bar{s})\left(2\rho(r(s^{\#}) - r(\bar{s}))\frac{\delta \log \delta^{-1}}{L}\right)^{-(4\kappa+1)/(4s+1)} v(\bar{s})(1 + \gamma_{\varepsilon 4}) \\
&\leq \tilde{\delta}(s)\left(2\rho(r(s^{\#}) - r(s))\frac{\delta \log \delta^{-1}}{L}\right)^{-(4\kappa+1)/(4s+1)} v(s)(1 + \gamma_{\varepsilon 5})
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{\delta}(s) \left(\frac{\tilde{\delta}(s)}{L} \right)^{-(4\kappa+1)/(4s+1)} v(s)(1+\gamma_{\varepsilon 5}) \\
&\leq R(s, L, \tilde{\delta}(s))(1+\gamma_{\varepsilon 6}). \qquad (67)
\end{aligned}
$$

Let $s^+ \in B_-$ be the largest grid point $\leq \bar{s}$. Let us denote $\mathcal{S}_1 = \{s' \in B_- : s^- \leq s' \leq s^+\}$ and $\mathcal{S}_2 = \{s' \in B_- : s^+ < s' \leq s_{m-1}\}$. Let us write $\mathcal{R}^+_{\varepsilon,\nu} = \mathcal{R}^{(1)}_{\varepsilon,\nu} + \mathcal{R}^{(2)}_{\varepsilon,\nu}$ where

$$
\mathcal{R}^{(1)}_{\varepsilon,\nu} = \sup_{\theta \in \Theta_\nu} E_\theta \left(\psi_\nu^{-\rho} \left| Q^*_\varepsilon - Q(\theta) \right|^\rho I\left(\hat{s} \in \mathcal{S}_1 \right) \right)
$$

and

$$
\mathcal{R}^{(2)}_{\varepsilon,\nu} = \sup_{\theta \in \Theta_\nu} E_\theta \left(\psi_\nu^{-\rho} \left| Q^*_\varepsilon - Q(\theta) \right|^\rho I\left(\hat{s} \in \mathcal{S}_2 \right) \right).
$$

Let us start with $\mathcal{R}^{(1)}_{\varepsilon,\nu}$. For $s' \in \mathcal{S}_1$, by Lemma 5, where $\tilde{s} = \tilde{s}(s, s')$ is as in Lemma 5,

$$
\begin{aligned}
\mathcal{B}_{s,L}(s', 1, \tilde{\delta}(s')) &\leq L\tilde{\delta}^{4(\tilde{s}-\kappa)/(4s'+1)}(s')\tilde{b}(s,s')(1+\gamma_{\varepsilon 7}) \\
&\leq L\tilde{\delta}^{4(\tilde{s}-\kappa)/(4\bar{s}+1)}(s')\tilde{b}(s,s')(1+\gamma_{\varepsilon 7}) \\
&\leq L\tilde{\delta}^{4(s-\kappa)/(4\bar{s}+1)}(\bar{s})\tilde{b}(s,\bar{s})(1+\gamma_{\varepsilon 8}) \\
&\leq \mathcal{B}_{s,L}(\bar{s}, 1, \tilde{\delta}(\bar{s}))(1+\gamma_{\varepsilon 9}).
\end{aligned}
$$

Thus, for $s' \in \mathcal{S}_1$, $\theta \in \Theta_\nu$, by (66),

$$
\begin{aligned}
\left| E_\theta \hat{Q}_{s'} - Q(\theta) \right| &\leq \mathcal{B}_{s,L}(s', 1, \tilde{\delta}(s'))(1+\gamma_{\varepsilon 10}) \\
&\leq \mathcal{B}_{s,L}(\bar{s}, 1, \tilde{\delta}(\bar{s}))(1+\gamma_{\varepsilon 11}) \\
&\leq B(s, L, \tilde{\delta}(s))(1+\gamma_{\varepsilon 12}) \\
&\leq B(s, L, \tilde{\delta}(s))(1+\gamma_{\varepsilon 12}) + R(s, L, \tilde{\delta}(s)) \\
&\leq \psi_\nu(1+\gamma_{\varepsilon 13}).
\end{aligned}
$$

Thus, when $s' \in \mathcal{S}_1$, $\theta \in \Theta_\nu$,

$$
\begin{aligned}
\left| \hat{Q}_{s'} - Q(\theta) \right| &\leq \left| E_\theta \hat{Q}_{s'} - Q(\theta) \right| + |Z_{s'}| + |U_{s'}(\theta)| \\
&\leq \psi_\nu(1+\gamma_{\varepsilon 13}) + |Z_{s'}| + |U_{s'}(\theta)|.
\end{aligned}
$$

Let us define

$$
\xi = (\sigma_s \psi_\nu)^{1/2}.
$$

Then, for $\theta \in \Theta_\nu$, applying similar inequality as in (82),

$$\begin{aligned}
& E_\theta\left(\psi_\nu^{-\rho}\left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s} \in \mathcal{S}_1\right)\right) \\
= & \sum_{s' \in \mathcal{S}_1} E_\theta\left(\psi_\nu^{-\rho}\left|\hat{Q}_{s'} - Q(\theta)\right|^\rho I\left(\hat{s} = s'\right)\right) \\
\leq & \sum_{s' \in \mathcal{S}_1} E_\theta\left(\left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-1}|Z_{s'}| + \psi_\nu^{-1}|U_{s'}(\theta)|\right)^\rho I\left(\hat{s} = s'\right)\right) \\
\leq & \left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-1}\xi\right)^\rho P_\theta\left(\hat{s} \in \mathcal{S}_1\right) + C_1 \sum_{s' \in \mathcal{S}_1} \\
& E\left(\left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-\rho}|Z_{s'}|^\rho + \psi_\nu^{-\rho}|U_{s'}(\theta)|^\rho\right) I\left(|Z_{s'}| + |U_{s'}(\theta)| > \xi\right)\right) \\
\leq & \left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-1}\xi\right)^\rho P_\theta\left(\hat{s} \in \mathcal{S}_1\right) + C_2 \sum_{s' \in \mathcal{S}_1} \\
& \left[\left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-\rho}\xi^\rho\right) P\left(|Z_{s'}| > \xi/2\right) + \psi_\nu^{-\rho} E\left(|Z_{s'}|^\rho I\left(|Z_{s'}| > \xi/2\right)\right)\right. \\
& + \left(1 + \gamma_{\varepsilon 13} + \psi_\nu^{-\rho}\xi^\rho\right) P\left(|U_{s'}(\theta)| > \xi/2\right) \\
& \left. + \psi_\nu^{-\rho} E\left(|U_{s'}(\theta)|^\rho I\left(|U_{s'}(\theta)| > \xi/2\right)\right)\right].
\end{aligned}$$

For $s' \in \mathcal{S}_1$, using (55), denoting $r' = 4(s'-\kappa)/(4s'+1)$, $r = 4(s-\kappa)/(4s+1)$,

$$\begin{aligned}
\frac{\tilde{\delta}^{r'}(s')}{\delta^r(s)} & \leq C_3(s^\# - s_{m-1})^{-C_4} \exp\left\{C_5(s - s^-)\log\delta^{-1}\right\} \\
& = C_3(s^\# - s_{m-1})^{-C_4} \exp\left\{C_5 \log\log\log\delta^{-1}\right\} \\
& \leq \left((s^\# - s_{m-1})^{-1} \log\log\delta^{-1}\right)^{C_6}.
\end{aligned}$$

On the other hand, for $s' \in \mathcal{S}_1$, $d_\varepsilon(s, s^\#)/d_\varepsilon(s', s^\#) \leq C_7$ for sufficiently small ε. Thus, for $s' \in \mathcal{S}_1$, for sufficiently small ε, by (57),

$$\frac{\sigma_{s'}}{\sigma_s} = \frac{\eta(s')}{\eta(s)} \frac{d_\varepsilon(s, s^\#)}{d_\varepsilon(s', s^\#)} \leq C_8 \frac{\tilde{\delta}^{r'}(s')}{\delta^r(s)} \leq \left((s^\# - s_{m-1})^{-1} \log\log\delta^{-1}\right)^{C_9}. \quad (68)$$

Now, by (57) and (22), $\psi_\nu \geq C\eta(s) = C_{10}\sigma_s d_\varepsilon(s, s^\#)$. Thus, for $s' \in \mathcal{S}_1$, for sufficiently small ε,

$$\frac{\sigma_s}{\psi_\nu} \leq \frac{C_{11}}{d_\varepsilon(s, s^\#)} \leq \frac{C_{12}}{(s^\# - s_{m-1})\log\delta^{-1}}. \quad (69)$$

Thus, firstly, by (69) and (16),

$$\frac{\xi}{\psi_\nu} = \left(\frac{\sigma_s}{\psi_\nu}\right)^{1/2} \leq C_{13}$$

for sufficiently small ε. Secondly, by (68), (69), and (16),

$$\begin{aligned}\psi_\nu^{-\rho}\sum_{i=0}^{\lceil\rho\rceil}\xi^{\rho-i}\sigma_{s'}^i &= \psi_\nu^{-\rho}\sum_{i=0}^{\lceil\rho\rceil}(\sigma_s\psi_\nu)^{(\rho-i)/2}\sigma_{s'}^i = \sum_{i=0}^{\lceil\rho\rceil}\left(\frac{\sigma_s}{\psi_\nu}\right)^{(\rho+i)/2}\left(\frac{\sigma_{s'}}{\sigma_s}\right)^i \\ &\le \frac{((s^\# - s_{m-1})^{-1}\log\log\delta^{-1})^{C_{14}}}{(\log\delta^{-1})^{C_{15}}} \le C_{16}\end{aligned}$$

for sufficiently small ε. Thirdly, by (68) and (69), because by (14) $\mathrm{Card}(B_{grid}) \le C_{17}(\log\delta^{-1})^{C_{18}}$,

$$\begin{aligned}&\mathrm{Card}(B_{grid})\exp\left\{-\frac{1}{4}\frac{\xi}{\sigma_{s'}}\right\} \le C_{19}(\log\delta^{-1})^{C_{18}}\exp\left\{-\frac{1}{4}\frac{(\sigma_s\psi_\nu)^{1/2}}{\sigma_{s'}}\right\} \\ &\le C_{19}(\log\delta^{-1})^{C_{18}}\exp\left\{-\frac{\sqrt{\log\delta^{-1}}}{((s^\# - s_{m-1})^{-1}\log\log\delta^{-1})^{C_{18}}}\right\} \longrightarrow 0\end{aligned}$$

by condition (16), as $\varepsilon \to 0$. Fourthly, by (83), $\mathrm{Var}(U_{s'}(\theta)) \le C_{19}\sigma_{s'}^2$. Then, by Lemma 8,

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\sup_{\nu\in B}\sup_{\theta\in\Theta_\nu} E_\theta\left(\psi_\nu^{-\rho}\left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s}\in\mathcal{S}_1\right)\right) \\ &\le \limsup_{\varepsilon\to 0}\sup_{\nu\in B}\sup_{\theta\in\Theta_\nu}\Big[\left(1+\gamma_{\varepsilon 13}+\psi_\nu^{-1}\xi\right)^\rho P_\theta\left(\hat{s}\in\mathcal{S}_1\right) \\ &\quad + C_{20}\sum_{s'\in\mathcal{S}_1}\Bigg[\left(1+\gamma_{\varepsilon 13}+\psi_\nu^{-\rho}\xi^\rho+\psi_\nu^{-\rho}\sum_{i=0}^{\lceil\rho\rceil}\xi^{\rho-i}\sigma_{s'}^i\right)\exp\left\{-\frac{1}{4}\frac{\xi}{\sigma_{s'}}\right\} \\ &\quad + \left(1+\gamma_{\varepsilon 13}+\psi_\nu^{-\rho}\xi^\rho+\psi_\nu^{-\rho}(\mathrm{Var}(U_{s'}(\theta)))^{\rho/2}\right)\exp\left\{-\frac{1}{8}\frac{\xi^2}{\mathrm{Var}(U_{s'}(\theta))}\right\}\Bigg]\Big] \\ &= \limsup_{\varepsilon\to 0}\sup_{\nu\in B}\sup_{\theta\in\Theta_\nu} P_\theta\left(\hat{s}\in\mathcal{S}_1\right).\end{aligned}$$

Let us turn to $\mathcal{R}^2_{\varepsilon,\nu}$. Let $\hat{s} = s' \in \mathcal{S}_2$ and $\theta \in \Theta_\nu$. We have by (14) and Lemma 4 (iii) (see also Lemma 7),

$$R(s^+,1,\tilde{\delta}(s^+))+\mathcal{B}_{s,L}(s^+,1,\tilde{\delta}(s^+)) \le (1+\gamma_{\varepsilon 14})\left(R(\bar{s},1,\tilde{\delta}(\bar{s}))+\mathcal{B}_{s,L}(\bar{s},1,\tilde{\delta}(\bar{s}))\right).$$

Then, by the definition of $\hat{s}$, (66), (67),

$$\begin{aligned}|Q_\varepsilon^* - Q(\theta)| &\le \left|\hat{Q}_{s'} - \hat{Q}_{s^+}\right| + \left|\hat{Q}_{s^+} - Q(\theta)\right| \\ &\le \eta(s^+) + \left|E_\theta\hat{Q}_{s^+} - Q(\theta)\right| + |Z_{s^+}| + |U_{s^+}(\theta)| \\ &\le R(s^+,1,\tilde{\delta}(s^+)) + \mathcal{B}_{s,L}(s^+,1,\tilde{\delta}(s^+))(1+\gamma_{\varepsilon 15})\end{aligned}$$

$$
\begin{aligned}
& + |Z_{s^+}| + |U_{s^+}(\theta)| \\
\leq\ & (1+\gamma_{\varepsilon 16}) \left[R(\bar{s}, 1, \tilde{\delta}(\bar{s})) + \mathcal{B}_{s,L}(\bar{s}, 1, \tilde{\delta}(\bar{s})) \right] + |Z_{s^+}| + |U_{s^+}(\theta)| \\
\leq\ & (1+\gamma_{\varepsilon 17}) \left[R(s, L, \tilde{\delta}(s)) + B(s, L, \tilde{\delta}(s)) \right] + |Z_{s^+}| + |U_{s^+}(\theta)| \\
\leq\ & \psi_\nu (1+\gamma_{\varepsilon 18}) + |Z_{s^+}| + |U_{s^+}(\theta)| \,.
\end{aligned}
$$

Then, for $\theta \in \Theta_\nu$,

$$
\begin{aligned}
& E_\theta \left(\psi_\nu^{-\rho} \, |Q_\varepsilon^* - Q(\theta)|^\rho \, I\left(\hat{s} \in \mathcal{S}_2\right) \right) \\
& \quad \leq\ \left(1+\gamma_{\varepsilon 18} + \psi_\nu^{-1}\xi\right)^\rho P_\theta\left(\hat{s} \in \mathcal{S}_2\right) + C_{21}\mathrm{Card}(\mathcal{S}_2) \\
& \quad\quad \times E\left[\left(1+\gamma_{\varepsilon 5} + \psi_\nu^{-\rho}|Z_{s^+}|^\rho + \psi_\nu^{-\rho}|U_{s^+}(\theta)|^\rho\right) I\left(|Z_{s^+}| + |U_{s^+}(\theta)| > \xi\right) \right].
\end{aligned}
$$

And then, similarily as before, because $s^+ \in \mathcal{S}_1$,

$$
\begin{aligned}
& \limsup_{\varepsilon\to 0} \sup_{\nu\in B} \sup_{\theta\in\Theta_\nu} E_\theta \left(\psi_\nu^{-\rho} \, |Q_\varepsilon^* - Q(\theta)|^\rho \, I\left(\hat{s} \in \mathcal{S}_2\right) \right) \\
& \qquad\qquad \leq \limsup_{\varepsilon\to 0} \sup_{\nu\in B} \sup_{\theta\in\Theta_\nu} P_\theta\left(\hat{s} \in \mathcal{S}_2\right).
\end{aligned}
$$

This leads to

$$
\limsup_{\varepsilon\to 0} \sup_{\nu\in B} \left(R^{(1)}_{\varepsilon,\nu} + R^{(2)}_{\varepsilon,\nu} \right) \leq \limsup_{\varepsilon\to 0} \sup_{\nu\in B} \sup_{\theta\in\Theta_\nu} \left[P_\theta\left(\hat{s} \in \mathcal{S}_1\right) + P_\theta\left(\hat{s} \in \mathcal{S}_2\right) \right] \leq 1.
$$

□

4.4 Lower bound in Theorem 1

Let us sketch a proof of inequality (31). The complete proof of the lower bound is given in the Appendix K of the technical report. Let us denote

$$
\widetilde{\psi}_{s,L} = c_{s,p}\varphi_{s,L}
$$

where $c_{s,p}$ is defined in (27) and $\varphi_{s,L}$ is defined in (25).

Let us note that l_p bodies are orthosymmetric, that is, when $\theta \in \Theta_{s,L}$, then $\tau_i\theta \in \Theta_{s,L}$, where τ_i changes the sign of the i:th element. The proof will start noting a fact given by Efroimovich and Low (1996), which says that when the parameter space is orthosymmetric, we can restrict attention to estimators which are functions of (y_i^2) only, where $y_i = \theta_i + \varepsilon z_i$ are our observations. After that we can proceed as in the linear case. Let us thus note that

$$
\inf_{Q_\varepsilon\in F} \sup_{\nu\in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \geq \inf_{Q_\varepsilon\in \tilde{F}} \sup_{\nu\in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \tag{70}
$$

where F is the set of measurable functions from $\mathbf{R}^\infty \to \mathbf{R}$ where $\mathbf{R}^\infty = \{(x_1, x_2, \ldots) : x_i \in \mathbf{R}\}$ and $\tilde{F}$ is the set of those measurable functions from $\mathbf{R}^\infty \to \mathbf{R}$ which are functions of x_i^2 only,

$$\tilde{F} = \left\{ \tilde{Q} : \mathbf{R}^\infty \to \mathbf{R} \;\middle|\; \tilde{Q}((x_i)) = Q((x_i^2)), \text{ for some } Q : \mathbf{R}^\infty \to \mathbf{R} \right\}.$$

Let $s' = s_1$ and $s'' = s_{m-1}$. Let $L', L'' \in [L_*, L^*]$. Let us denote $\nu' = (s', L')$ and $\nu'' = (s'', L'')$. There is such $\tilde{\theta} \in \mathbf{R}^\infty$ that

$$\tilde{\theta} \in \Theta_{s',L'}, \qquad Q(\tilde{\theta}) \sim (L')^{1-r(s')} \tilde{\delta}^{r(s')} c_{s,p}, \qquad \sum_{i=1}^{\infty} \tilde{\theta}_i^4 = \tilde{\delta}^2 \tag{71}$$

where

$$\tilde{\delta} = \delta d_\varepsilon(s', s^\#), \;\; \delta = \sqrt{2}\varepsilon^2.$$

Consider the sequencies $\bar{\theta}_0, \bar{\theta}_1 \in \mathbf{R}^\infty$,

$$\bar{\theta}_0 = 0, \qquad \bar{\theta}_1 = (1-\xi)^{1/2} \tilde{\theta}$$

where $0 < \xi < 1/2$ is arbitrary. Now $\bar{\theta}_0 \in \Theta_{\nu''}$ and $\bar{\theta}_1 \in \Theta_{\nu'}$. Also,

$$Q(\bar{\theta}_0) = 0, \qquad Q(\bar{\theta}_1) \sim (1-\xi) \widetilde{\psi}_{\nu'}.$$

Thus for any estimator Q_ε,

$$\left| Q_\varepsilon - Q(\bar{\theta}_0) \right| = \widetilde{\psi}_{\nu'} \, D\left((1-\xi)^{-1} \widetilde{\psi}_{\nu'}^{-1} Q_\varepsilon, 0 \right)$$

and

$$\left| Q_\varepsilon - Q(\bar{\theta}_1) \right| \sim \psi_{\nu'} \, D\left((1-\xi)^{-1} \widetilde{\psi}_{\nu'}^{-1} Q_\varepsilon, 1 \right),$$

where $D(u,v) = (1-\xi)|u-v|$, $u, v \in \mathbf{R}$. Thus, denoting $q = \widetilde{\psi}_{\nu'} / \widetilde{\psi}_{\nu''}$, denoting by $\tilde{P}_\theta$ the distribution of (y_i^2), when $y_i = \theta_i + \varepsilon z_i$, denoting $\tilde{P}_i = \tilde{P}_{\bar{\theta}_i}$, and letting $\tilde{E}_i$ denote the corresponding expectations,

$$\begin{aligned}
&\inf_{Q_\varepsilon \in F} \sup_{\nu \in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \\
&\geq \inf_{Q_\varepsilon \in \tilde{F}} \sup_{\nu \in B} \sup_{\theta \in \Theta_\nu} \tilde{E}_\theta \left(\widetilde{\psi}_\nu^{-\rho} \left| Q_\varepsilon - Q(\theta) \right|^\rho \right) \\
&\geq \inf_{Q_\varepsilon \in \tilde{F}} \max \left\{ \tilde{E}_0 \left(\widetilde{\psi}_{\nu''}^{-\rho} \left| Q_\varepsilon - Q(\bar{\theta}_0) \right|^\rho \right), \tilde{E}_1 \left(\widetilde{\psi}_{\nu'}^{-\rho} \left| Q_\varepsilon - Q(\bar{\theta}_1) \right|^\rho \right) \right\} \\
&\sim \inf_{Q_\varepsilon \in \tilde{F}} \max \left\{ q^\rho \tilde{E}_0 D^\rho(Q_\varepsilon, 0), \tilde{E}_1 D^\rho(Q_\varepsilon, 1) \right\}.
\end{aligned}$$

Let

$$\tau = \exp\left\{-\frac{1-\xi}{2}\, d_\varepsilon(s', s^{\#})\right\}.$$

By Markov's inequality, for $0 < \alpha < 1$,

$$\tilde{P}_1\left(\frac{d\tilde{P}_0}{d\tilde{P}_1} \geq \tau\right) \geq 1 - \alpha \tag{72}$$

for sufficiently small ε. Applying Theorem 6 (i) in Tsybakov (1998), with the fact (72),

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \inf_{Q_\varepsilon \in F} \max\left\{q^\rho \tilde{E}_0 D^\rho(Q_\varepsilon, 0), \tilde{E}_1 D^\rho(Q_\varepsilon, 1)\right\} \\
&\geq \frac{(1-\alpha)(1-2\xi)^\rho \tau (q\xi)^\rho}{(1-2\xi)^\rho + \tau(q\xi)^\rho} = (1-\alpha)(1-2\xi)^\rho.
\end{aligned}$$

The lower bound is proved because $0 < \xi < 1/2$ and $0 < \alpha < 1$ were chosen arbitrarily. □

4.5 Finishing the proof of Theorem 1

We may write, denoting $\nu = (s, L)$, $s^- = s^-(s)$,

$$\begin{aligned}
&E_\theta\left(\psi_\nu^{-\rho} \left|Q_\varepsilon^* - Q(\theta)\right|^\rho\right) \\
&= E_\theta\left(\psi_\nu^{-\rho} \left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s} < s^-\right)\right) + E_\theta\left(\psi_\nu^{-\rho} \left|Q_\varepsilon^* - Q(\theta)\right|^\rho I\left(\hat{s} \geq s^-\right)\right).
\end{aligned}$$

Then the upper bound (30) follows from Lemma 10 with $B' = B = [s_*, s_{m-1}] \times [L_*, L^*]$ and from Lemma 11. Lower bound (31) is proved in Section 4.4. Thus Theorem 1 is proved. □

Acknowledgments

I wish to thank the referees for the helpful comments.

References

[1] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381-393.

[2] Butucea C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM: Probability and Statistics* **5** 1-31.

[3] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

[4] Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. The Annals of Statistics, In press.

[5] Donoho, D. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238-270.

[6] Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668-701.

[7] Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290-323.

[8] Efroimovich, S. Y. (1994). On adaptive estimation of nonlinear functionals. *Statist. Probab. Lett.* **19** 57-63.

[9] Efroimovich, S. Y. and Low, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106-1125.

[10] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273-1294.

[11] Gayraud, G. and Tribouley, K. (1999). Wavelet methods to estimate an integrated quadratic functional: Adaptivity and asymptotic law. *Statist. Probab. Lett.* **44** 109-122.

[12] Huang, L. and Fan, J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli* **5** 927-949.

[13] Ibragimov, I. A. and Khas'minskii, R. Z. (1980). Some estimation problems for stochastic differential equations. *Lecture Notes in Control and Inform. Sci. B* 1-12. Springer, Berlin.

[14] Ibragimov, I. A. and Khas'minskii, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18-32.

[15] 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.

[16] Johnstone, I. M. (2001). Chi-square oracle inequalities. In State of the Art in Probability and Statistics, Festchrift for Willem R. van Zwet, M. de Gunst, C. Klaassen and A. van der Waart, editors. IMS Lecture Notes - Monographs, 36, 399-418.

[17] Klemelä, J. (2002). Sharp adaptive estimation of quadratic functionals. Technical report. http://www.denstruct.net

[18] Klemelä, J. and Tsybakov, A. B. (2001). Sharp adaptive estimation of linear functionals. *Ann. Statist.* **29** 1567-1600.

[19] Klemelä, J. and Tsybakov, A. B. (2004). Exact constants for pointwise adaptive estimation under the Riesz transform. *Probab. Theory Relat. Fields* **129** (3) 441-467.

[20] Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli* **5** 1099-1118.

[21] Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** No. 5.

[22] Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459-470.

[23] Lepski, O. V. (1991). Asymptotically minimax adaptive estimation I: upper bounds. Optimal adaptive estimates. *Theory Probab. Appl.* **36** 682-697.

[24] Lepski, O. V. (1992a). Asymptotically minimax adaptive estimation II: statistical models without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37** 433-468.

[25] Lepski, O. V. (1992b). On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation. Advances in Soviet Math.* Vol. 12. Khasminskii R. Z., ed., Amer. Math. Soc., Providence, R. I., 87-106.

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

[27] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399-2430.

[28] Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420-2469.

A Proof of Theorem 2

We need the following lemma.

Lemma 12 *(i) for $s \geq 2\kappa + 1/4$, for integers $M \geq 1$,*

$$\sup_{\theta \in \Theta_{s,L}} \sum_{i=M}^{\infty} i^{2\kappa} \theta_i^2 \leq L M^{2(\kappa - s)},$$

(ii) for $s \geq 2\kappa + 1/4$, $\sup_{\theta \in \Theta_{s,L}} Q_{2\kappa}(\theta) < \infty$.

Lemma 12 is proved in the Appendix D of the technical report.

Proof of Theorem 2. Consider first the case where $s = s_m = 2\kappa + 1/4$. We may write

$$\begin{aligned} & E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 \right) \\ & \quad = E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 I \left(\hat{s} \leq s_{m-1} \right) \right) + E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 I \left(\hat{s} = s_m \right) \right). \end{aligned}$$

Because $s_{m-1} < s^-(s_m)$ for sufficiently small ε, where s^- is defined in (59), we get by Lemma 10 that

$$\lim_{\varepsilon \to 0} \left(\varepsilon^2 b_\varepsilon \right)^{-1} \sup_{L \in [L_*, L^*]} \sup_{\theta \in \Theta_{s_m, L}} E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 I \left(\hat{s} \leq s_{m-1} \right) \right) = 0.$$

We have also

$$\begin{aligned} & \lim_{\varepsilon \to 0} \left(\varepsilon^2 b_\varepsilon \right)^{-1} \sup_{L \in [L_*, L^*]} \sup_{\theta \in \Theta_{s_m, L}} E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 I \left(\hat{s} = s_m \right) \right) \\ & \leq \lim_{\varepsilon \to 0} \left(\varepsilon^2 b_\varepsilon \right)^{-1} \sup_{L \in [L_*, L^*]} \sup_{\theta \in \Theta_{s_m, L}} E_\theta \left(\left| \bar{Q}_\varepsilon - Q(\theta) \right|^2 \right) < \infty \end{aligned}$$

where the finitness of the right hand side was proved by Efroimovich (1994, Theorem 2.3) for L_2 and Hölder bodies. It is easy to see that the finitness of the right hand side holds under items (i) and (ii) of Lemma 12. Equation (32) is proved. We may also write

$$\begin{aligned} & \varepsilon^{-2} E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 \right) \\ & = \varepsilon^{-2} E_\theta \left(\left| \bar{Q}_\varepsilon - Q(\theta) \right|^2 \right) + \varepsilon^{-2} E_\theta \left(|Q_\varepsilon^* - Q(\theta)|^2 I \left(\hat{s} \leq s_{m-1} \right) \right) \\ & \quad + \varepsilon^{-2} E_\theta \left(\left| \bar{Q}_\varepsilon - Q(\theta) \right|^2 \left(I \left(\hat{s} = s_m \right) - 1 \right) \right). \end{aligned}$$

Now

$$E_\theta\left(\left|\bar{Q}_\varepsilon - Q(\theta)\right|^2 (I(\hat{s} = s_m) - 1)\right) \le \left[E_\theta \left|\bar{Q}_\varepsilon - Q(\theta)\right|^4 (1 - P_\theta(\hat{s} = s_m))\right]^{1/2}$$

and this vanishes asymptotically, uniformly over $L \in [L_*, L^*]$ and $\theta \in \Theta_{s,L}$, for $s > s_m$, by Lemma 9 and by the fact that $E_\theta \left|\bar{Q}_\varepsilon - Q(\theta)\right|^4 \le C$. Because $s_{m-1} < s^-(s_m)$ for sufficiently small ε, we get for $s > s_m$, by Lemma 10, that

$$\lim_{\varepsilon\to 0} \varepsilon^{-2} \sup_{L\in[L_*,L^*]} \sup_{\theta\in\Theta_{s,L}} E_\theta\left(|Q^*_\varepsilon - Q(\theta)|^2 I(\hat{s} \le s_{m-1})\right) = 0$$

where $s > s_m$. Again similarily as Efroimovich (1994, Theorem 2.3), for $s > s_m$,

$$\lim_{\varepsilon\to 0} \sup_{L\in[L_*,L^*]} \sup_{\theta\in\Theta_{s,L}} \left[\varepsilon^{-2} E_\theta\left(\left|\bar{Q}_\varepsilon - Q(\theta)\right|^2\right) - 4Q_{2\kappa}(\theta)\right] = 0.$$

Equation (33) is proved. □

B Properties of the filter

We state next properties of weights $g = (g_i)$, defined in (8).

Lemma 13 *(i) There exists a positive constant C such that for $0 < \gamma, \gamma' \le C$, for $s_* \le s' \le s \le s^{\#}$, $L, L' \in [L_*, L^*]$,*

$$\|g(s, L, \gamma) - g(s', L', \gamma')\|_2 \le \|g(s', L', \gamma')\|_2 ,$$

(ii) for $(s, L) \in [s_*, s^{\#}] \times [L_*, L^*]$, $\gamma \le 1$, $c = 1 - 4s^{\#}/(4s^{\#} + 1)$,

$$\gamma^{1-2r(s)} \sup_{\theta\in\Theta_{s,L}} \sum_{i=1}^{\infty} g_i^2(s, L', \gamma)\theta_i^2 \le C \max\left\{\gamma^c, \gamma^{1-2r(s)}\right\} ,$$

for a positive constant C and for $L' \in [L_*, L^*]$,

(iii) for $c(s) = 4\kappa/(4s + 1)$

$$\sup_{L\in[L_*,L^*]} |g_i(s, L, \gamma)| \le C\gamma^{-c(s)}, \;\; i = 1, 2, \ldots$$

when $s \in [s_*, s^{\#}]$ and C is a positive constant.

Note that because $s^{\#} = 2\kappa + 1/4$ then $r(s) \le 1/2$ in Lemma 13. Heteroscedastic term $U_{s,L,\gamma}(\theta)$ is defined in (37). We have that $\mathrm{Var}(U_{s,L,\gamma}(\theta)) = 4\varepsilon^2 \sum_{i=1}^{\infty} g_i^2\theta_i^2$ and the Lemma 13 (ii) implies that term $U_{s,L,\gamma}(\theta)$ is negligible in a certain sense. Lemma 13 (iii) is applied to prove an exponential bound for $Z_{s,L,\gamma}$, defined in (36).

Proof of item (i). Denote

$$h = h(s, L, \gamma) = b^{1/q} \tag{73}$$

where b is defined in (9). Let us prove item (i). Let $s' \le s$ and $\gamma, \gamma' \sim 0$. Let us denote $h_0 = h(s', L', \gamma')$ and $h_1 = h(s, L, \gamma)$. Now

$$\begin{aligned}
&\|g(s', L', \gamma') - g(s, L, \gamma)\|_2^2 \\
&= \sum_{i=1}^{\infty} \left(h_0^{-2\kappa}\phi_{s',p}(h_0 i) - h_1^{-2\kappa}\phi_{s,p}(h_1 i)\right)^2 \\
&\sim h_0^{-4\kappa-1} \int_0^{\infty} \left(\phi_{s',p}(t) - \left(\frac{h_0}{h_1}\right)^{2\kappa} \phi_{s,p}\left(\frac{h_1}{h_0} t\right)\right)^2 dt \\
&= h_0^{-4\kappa-1} \left[\int_0^{\infty} \phi_{s',p}^2 + \left(\frac{h_0}{h_1}\right)^{4\kappa+1} \int_0^{\infty} \phi_{s,p}^2 \right. \\
&\quad \left. - 2\left(\frac{h_0}{h_1}\right)^{2\kappa+1} \int_0^{\infty} \phi_{s',p}\left(\frac{h_0}{h_1} t\right) \phi_{s,p}(t) dt\right] \\
&\le h_0^{-4\kappa-1} \left[\int_0^{\infty} \phi_{s',p}^2 + \left(\frac{h_0}{h_1}\right)^{4\kappa+1} \left(\int_0^{\infty} \phi_{s,p}^2 - 2\int_0^{\infty} \phi_{s',p}\phi_{s,p}\right)\right] \\
&< h_0^{-4\kappa-1} \int_0^{\infty} \phi_{s',p}^2 \\
&\sim \|g(s', L', \gamma')\|_2^2
\end{aligned}$$

because

$$\left(\frac{h_0}{h_1}\right)^{-2\kappa} \int_0^{\infty} \phi_{s',p}\left(\frac{h_0}{h_1} t\right) \phi_{s,p}(t) dt \ge \int_0^{\infty} \phi_{s',p}\phi_{s,p}$$

and

$$\int_0^{\infty} \phi_{s,p}^2 \le 2\int_0^1 \phi_{s',p}\phi_{s,p} + \int_1^{\infty} \phi_{s',p}\phi_{s,p} < 2\int_0^{\infty} \phi_{s',p}\phi_{s,p}.$$

The item (i) has been proved. □

Proof of item (ii). Let first $2\kappa p < q = p(s - 1/p + 1/2)$. Then, for $\theta \in \Theta_{s,L}$, by the convexity of $x \mapsto x^{p/2}$,

$$\sum_{i=1}^{\infty} g_i^2 \theta_i^2 \le \left(\sum_{i=1}^{\infty} |g_i \theta_i|^p\right)^{2/p} \le \left(\sum_{i=1}^{\infty} i^{2\kappa p} |\theta_i|^p\right)^{2/p} \le L.$$

Let secondly $q \le 2\kappa p$. That is, $s_* < s < 2\kappa + 1/p - 1/2$. We have, for $\theta \in \Theta_{s,L}$, for $r = p/2$, $r' = p/(p-2)$,

$$\begin{aligned}
\sum_{i=1}^{\infty} g_i^2 \theta_i^2 &\le \left[\sum_{i=1}^{\infty} i^q |\theta_i|^p\right]^{2/p} \left[\sum_{i=1}^{\infty} [g_i^2 i^{-q/r}]^{r'}\right]^{1/r'} \\
&\le L h^{q/r - 4\kappa - 1/r'} \left\{ h \sum_{i=1}^{\infty} \left[\frac{(hi)^{4\kappa}}{(1+(hi)^q)^2} (hi)^{-q/r}\right]^{r'} \right\}^{1/r'} \\
&\le C_1 h^{2s-4\kappa} \left\{ \int_0^{\infty} \left[\phi_{s,p}^2(t) t^{-q/r}\right]^{r'} dt \right\}^{1/r'} \\
&\le C_2 \gamma^{(4s-8\kappa)/(4s+1)}
\end{aligned}$$

for some positive constants C_1, C_2, where $\phi_{s,p}(t) = t^{2\kappa}/(1+t^q)$. We used the fact that $b \asymp \gamma^{1/(\alpha+\beta)}$ where $\alpha + \beta = (4s+1)/(2q)$ and thus $h \asymp \gamma^{2/(4s+1)}$. Also,

$$\int_0^1 \left[\phi_{s,p}^2(t) t^{-q/r}\right]^{r'} dt \le \int_0^1 \left[t^{4\kappa} t^{-q/r}\right]^{r'} dt < \infty$$

because $r'[4\kappa - q/r] > -1$. This follows from $q \le 2\kappa p$. Also,

$$\int_1^{\infty} \left[\phi_{s,p}^2(t) t^{-q/r}\right]^{r'} dt \le \int_1^{\infty} \left[t^{2(2\kappa - q)} t^{-q/r}\right]^{r'} dt < \infty$$

because $r'[2(2\kappa - q) - q/r] < -1$. This follows from the assumption $s \ge s_* > (3 - p + 4\kappa)/(2p)$. In fact, it would be enough that $s > (2 - p + 4\kappa)/(2p)$. Thus for $s_* < s < 2\kappa + 1/p - 1/2$, for $\theta \in \Theta_{s,L}$, for $\gamma \le 1$,

$$\begin{aligned}
\gamma^{1-2r(s)} \sum_{i=1}^{\infty} g_i^2 \theta_i^2 &\le C_2 \gamma^{1-2r(s)} \gamma^{(4s-8\kappa)/(4s+1)} = C_2 \gamma^{1-4s/(4s+1)} \\
&\le C_2 \gamma^{1-4s^{\#}/(4s^{\#}+1)}.
\end{aligned}$$

Thus item (ii) is satisfied with $c = 1 - 4s^{\#}/(4s^{\#}+1)$. □

Proof of item (iii). Function $t \mapsto t^{2\kappa}/(1+bt^q)$ is maximized by $t = (2\kappa/[b(q-2\kappa)])^{1/q}$. Thus

$$g_i = i^{2\kappa}(1+bi^q)^{-1} \le C_1 b^{-2\kappa/q} \le C_2 \gamma^{-4\kappa/(4s+1)} \tag{74}$$

for some positive constants C_1, C_2, because $b \asymp \gamma^{2q/(4s+1)}$. We have proved item (iii). □

C Proof of Lemma 5

We have defined $\tilde{s} = \tilde{s}(s, s') = \min\{s, s' + p/4 - 1/2\}$. Then $\tilde{s} \le s$ and $\kappa < \tilde{s} < ps'/2 + \kappa + p/4 - 1/2$. We will denote $q' = p(s' - 1/p + 1/2)$ and $\tilde{q} = p(\tilde{s} - 1/p + 1/2)$. Define

$$\mathcal{B}_{s,L}(s', L', \gamma) = Lh^{2(\tilde{s}-\kappa)}(s', L', \gamma) \left[\int_0^\infty \left(\frac{t^{2\kappa + q' - \tilde{q}/r}}{1 + t^{q'}} \right)^{r'} dt \right]^{1/r'}$$

where h is defined in (73), $r = p/2$, and $r' = p/(p-2)$. Now

$$\int_0^\infty \left(\frac{t^{2\kappa + q' - \tilde{q}/r}}{1 + t^{q'}} \right)^{r'} dt \le \int_0^1 t^{r'(2\kappa + q' - \tilde{q}/r)} dt + \int_1^\infty t^{r'(2\kappa - \tilde{q}/r)} dt < \infty$$

because $r'(2\kappa + q' - \tilde{q}/r) > -1$ (here we used the fact that $\tilde{s} < ps'/2 + \kappa + p/4 - 1/2$) and $r'(2\kappa - \tilde{q}/r) < -1$ (here we used the fact that $\tilde{s} > \kappa$). Lemma follows, because for $\theta \in \Theta_{s,L}$, denoting $h' = h(s', L', \gamma)$,

$$\begin{aligned}
&\sum_{i=1}^\infty (i^{2\kappa} - g_i(s', L', \gamma))\theta_i^2 \\
&= \sum_{i=1}^\infty \frac{i^{2\kappa}(h'i)^{q'}}{1 + (h'i)^{q'}} \theta_i^2 \\
&\le \left\{ \sum_{i=1}^\infty i^{\tilde{q}} |\theta_i|^p \right\}^{2/p} \left\{ \sum_{i=1}^\infty \left[\frac{i^{2\kappa}(h'i)^{q'}}{1 + (h'i)^{q'}} i^{-\tilde{q}/r} \right]^{r'} \right\}^{1/r'} \\
&\le L(h')^{\tilde{q}/r - 1/r' - 2\kappa} \left[h' \sum_{i=1}^\infty \left(\frac{(h'i)^{2\kappa + q'}}{1 + (h'i)^{q'}} (h'i)^{-\tilde{q}/r} \right)^{r'} \right]^{1/r'} \\
&\sim \mathcal{B}_{s,L}(s', L', \gamma)
\end{aligned}$$

uniformly in $(s, L) \in A_0$. $\square$

D Proof of Lemma 12

First, by Hölder's inequality, for $\theta \in \Theta_{s,L}$,

$$\sum_{i=M}^\infty i^{2\kappa}\theta_i^2 \le \left[\sum_{i=M}^\infty i^{p(s-1/p+1/2)} |\theta_i|^p \right]^{2/p} \left[\sum_{i=M}^\infty \left(i^{2\kappa - 2(s-1/p+1/2)} \right)^{p/(p-2)} \right]^{(p-2)/p}$$

where

$$\sup_{\theta\in\Theta_{s,L}} \left[\sum_{i=M}^{\infty} i^{p(s-1/p+1/2)}|\theta_i|^p\right]^{2/p} \le L$$

and

$$\begin{aligned}\left[\sum_{i=M}^{\infty}\left(i^{2\kappa-2(s-1/p+1/2)}\right)^{p/(p-2)}\right]^{(p-2)/p} &\le \left[M^{2p(\kappa-s+1/p-1/2)/(p-2)+1}\right]^{(p-2)/p}\\ &= M^{2(\kappa-s)}.\end{aligned}$$

Thus Lemma 12 (i) has been proved. Similarily,

$$\sup_{\theta\in\Theta_{s,L}} Q_{2\kappa}(\theta) = \sup_{\theta\in\Theta_{s,L}} \sum_{i=1}^{\infty} i^{4\kappa}\theta_i^2 \le L\left[\sum_{i=1}^{\infty}\left(i^{4\kappa-2(s-1/p+1/2)}\right)^{p/(p-2)}\right]^{(p-2)/p} < \infty$$

because $[4\kappa-2(s-1/p+1/2)]p/(p-2) < -1$ for $s \ge 2\kappa+1/4$, $p > 2$. Thus Lemma 12 (ii) has been proved. □

E Proof of Lemma 7

Denoting for shortness $r = r(s)$, $r' = r(s')$,

$$\frac{\tilde{\delta}^r(s'')}{\tilde{\delta}^{r'}(s''')} = \exp\left\{-\frac{1}{2}(r-r')\log\delta^{-1}\right\}\frac{(d_\varepsilon(s'',s^{\#})\delta^{1/2})^r}{(d_\varepsilon(s''',s^{\#})\delta^{1/2})^{r'}} \tag{75}$$

and for $s''' \le s'' < s^{\#}$ and $s' < s < s^{\#}$,

$$\frac{(d_\varepsilon(s'',s^{\#})\delta^{1/2})^r}{(d_\varepsilon(s''',s^{\#})\delta^{1/2})^{r'}} \le \left(d_\varepsilon(s'',s^{\#})\delta^{1/2}\right)^{r-r'} \le 1$$

for sufficiently small ε. Equation (54) follows from this. Also,

$$\frac{\tilde{\delta}^{r'}(s')}{\tilde{\delta}^r(s)} = \exp\left\{(r-r')\log\delta^{-1}\right\}\frac{d_\varepsilon^{r'}(s',s^{\#})}{d_\varepsilon^r(s,s^{\#})}$$

and for $s' < s \le s_{m-1}$,

$$\frac{d_\varepsilon^{r'}(s',s^{\#})}{d_\varepsilon^r(s,s^{\#})} \le \frac{[2\rho(r(s^{\#})-r(s_*))\log\delta^{-1}]^{r(s)}}{[2\rho(r(s^{\#})-r(s_{m-1}))\log\delta^{-1}]^{r(s)}} \le C_5(r(s^{\#})-r(s_{m-1}))^{-r(s^{\#})}.$$

Equation (55) follows from this. Equation (56) follows similarily. The statement (57) follows from definitions of $\psi_{s,L}$, $\eta(s)$ and from Lemma 4. □

F Proof of Lemma 8

Let us prove Lemma 8 (i). The case (ii) is more easy than the case (i). Let first $\rho = 0$. Let us denote $a_i = \varepsilon^2 g_i$. Then $Z_s = \sum_{i=1}^{\infty} a_i(z_i^2 - 1)$. Let $0 < t < 1/(2a_i)$ for $i = 1, 2, \ldots$. Now

$$E \exp\left\{t a_i z_i^2\right\} = (1 - 2a_i t)^{-1/2}.$$

Thus,

$$\begin{aligned} P(Z_s \geq x_\varepsilon) &\leq \exp\{-tx_\varepsilon\} E \exp\{tZ_s\} \\ &= \exp\{-tx_\varepsilon - t\sum_{i=1}^{\infty} a_i\} \prod_{i=1}^{\infty} (1 - 2a_i t)^{-1/2} \\ &= \exp\{-tx_\varepsilon\} \exp\left\{\sum_{i=1}^{\infty} \left[-ta_i - \frac{1}{2}\log(1 - 2a_i t)\right]\right\} \\ &\leq \exp\{-tx_\varepsilon\} \exp\left\{\sum_{i=1}^{\infty} \frac{(ta_i)^2}{1 - 2a_i t}\right\} \end{aligned}$$

because, for $0 < u < 1/2$,

$$-u - \frac{1}{2}\log(1 - 2u) = 2u^2 \sum_{i=0}^{\infty} \frac{(2u)^i}{i + 2} \leq u^2 \sum_{i=0}^{\infty} (2u)^i = \frac{u^2}{1 - 2u}.$$

Let $0 < \chi < 1/2$. Let us denote

$$D_{s,\varepsilon} = \frac{\chi \sigma_s^2}{2|a|_\infty}$$

where $|a|_\infty = \sup_{i=1,2,\ldots} |a_i|$. By assumption $x_{s,\varepsilon}/D_{s,\varepsilon} \to 0$. Let ε be so small that x_ε is smaller than $D_{s,\varepsilon}$. Suppose first that also

$$x_\varepsilon \geq \frac{1 - \chi}{1 - 2\chi}\, \sigma_s \stackrel{def}{=} D_{0,s,\varepsilon}. \tag{76}$$

By remark (58), $D_{0,s,\varepsilon} < D_{s,\varepsilon}$ for sufficiently small ε. Then, because $\sigma_s^2 = 2\sum_{i=1}^{\infty} a_i^2$,

$$\frac{\chi}{|a|_\infty} \sum_{i=1}^{\infty} a_i^2 = D_{s,\varepsilon} \geq x_\varepsilon.$$

Thus, choosing $t = x_\varepsilon/\sigma_s^2$,

$$1 - 2a_i t = 1 - \frac{2x_\varepsilon a_i}{2\sum_{k=1}^{\infty} a_k^2} \geq 1 - \frac{2x_\varepsilon |a|_\infty}{2\sum_{k=1}^{\infty} a_k^2} \geq 1 - \chi$$

(and also $t \leq \chi/(2a_i) \leq 1/(2a_i)$) for $i = 1, 2, \ldots$. Thus,

$$\begin{aligned} P(Z_s \geq x_\varepsilon) &\leq \exp\{-tx_\varepsilon\} \exp\left\{\frac{t^2}{1-\chi} \sum_{i=1}^{\infty} a_i^2\right\} \\ &= \exp\left\{-\frac{x_\varepsilon^2}{2\sigma_s^2}\left(2 - \frac{1}{1-\chi}\right)\right\} \\ &\leq \exp\left\{-\frac{x_\varepsilon}{2\sigma_s}\right\} \end{aligned}$$

because, by (76),

$$\frac{x_\varepsilon}{\sigma_s}\left(2 - \frac{1}{1-\chi}\right) = \frac{x_\varepsilon}{\sigma_s}\frac{1-2\chi}{1-\chi} \geq 1.$$

Similarily, for $t > 0$ and $2a_i t \leq 1$, $i = 1, 2, \ldots$,

$$\begin{aligned} P(Z_s \leq -x_\varepsilon) &\leq \exp\{-tx_\varepsilon\} E \exp\{-tZ_s\} \\ &= \exp\{-tx_\varepsilon\} \exp\left\{\sum_{i=1}^{\infty}\left[ta_i - \frac{1}{2}\log(1 + 2a_i t)\right]\right\} \\ &\leq \exp\{-tx_\varepsilon\} \exp\left\{t^2 \sum_{i=1}^{\infty} a_i^2\right\} \end{aligned}$$

because, for $0 < u < 1/2$, $u - 2^{-1}\log(1 + 2u) \leq u^2$. Let us choose $t = x_\varepsilon/\sigma_s^2$. Then,

$$P(Z_s \leq -x_\varepsilon) \leq \exp\left\{-\frac{1}{2}\frac{x_\varepsilon^2}{\sigma_s^2}\right\}.$$

For $0 < x_\varepsilon < D_{0,s,\varepsilon}$ and for C_2 large enough,

$$C_2 \exp\left\{-\frac{1}{2}\frac{x_\varepsilon}{\sigma_s}\right\} \geq C_2 \exp\left\{-\frac{1}{2}\frac{1-\chi}{1-2\chi}\right\} \geq 1 \geq P\left(|Z_s| \geq x_\varepsilon\right).$$

The case $\rho = 0$ has been proved. Let then $\rho \geq 1$. Let $T > \limsup_{\varepsilon\to 0} x_\varepsilon$ and let ε be so small that $x_\varepsilon < T < D_{\varepsilon,s}$. For $x_\varepsilon = x_0 < x_1 < \cdots < x_N = T$,

$$\begin{aligned} E\left(|Z_s|^\rho I(x_\varepsilon \leq Z_s \leq T)\right) &\leq \sum_{i=1}^{N} x_i^\rho P(x_{i-1} < Z_s < x_i) \\ &= \sum_{i=1}^{N} x_i^\rho \left(P(Z_s > x_{i-1}) - P(Z_{s,\varepsilon} > x_i)\right) \end{aligned}$$

$$
\begin{aligned}
&= x_1^\rho P(Z_s > x_0) + \sum_{i=1}^{N-1} (x_{i+1}^\rho - x_i^\rho) P(Z_s > x_i) \\
&\leq x_1^\rho P(Z_s > x_0) + \rho \sum_{i=1}^{N-1} x_{i+1}^{\rho-1} (x_{i+1} - x_i) P(Z_s > x_i).
\end{aligned}
$$

Let us choose such a sequence $x_\varepsilon = x_0 < x_{1,N} < \ldots < x_{N-1,N} < x_N = T$ that $\sup_{i=0,\ldots,N-1} |x_{i+1,N} - x_{i,N}| \to 0$ when $N \to \infty$. Now, for sufficiently large N,

$$
\begin{aligned}
& E\left(|Z_s|^\rho I(x_\varepsilon \leq Z_s \leq T)\right) \\
&\leq x_1^\rho P(Z_s > x_0) + \rho \sum_{i=1}^{N-1} x_{i+1}^{\rho-1} (x_{i+1} - x_i) P(Z_s > x_i) \\
&\leq C_3 \left[x_\varepsilon^\rho \exp\left\{ -\frac{1}{2} \frac{x_\varepsilon}{\sigma_s} \right\} + \rho \int_{x_\varepsilon}^{T} t^{\rho-1} \exp\left\{ -\frac{1}{2} \frac{t}{\sigma_s} \right\} dt \right] \\
&\leq C_3 \left[x_\varepsilon^\rho \exp\left\{ -\frac{1}{2} \frac{x_\varepsilon}{\sigma_s} \right\} + \rho \int_{x_\varepsilon}^{\infty} t^{\rho-1} \exp\left\{ -\frac{1}{2} \frac{t}{\sigma_s} \right\} dt \right].
\end{aligned}
$$

Thus, because T is arbitrary,

$$
E\left(|Z_s|^\rho I(Z_s \geq x_\varepsilon)\right) \leq C_3 \left[x_\varepsilon^\rho \exp\left\{ -\frac{1}{2} \frac{x_\varepsilon}{\sigma_s} \right\} + \rho \int_{x_\varepsilon}^{\infty} t^{\rho-1} \exp\left\{ -\frac{1}{2} \frac{t}{\sigma_s} \right\} dt \right].
$$

Also,

$$
\begin{aligned}
& \int_{x_\varepsilon}^{\infty} t^{\rho-1} \exp\left\{ -\frac{1}{2} \frac{t}{\sigma_s} \right\} dt \\
&= 2\sigma_s x_\varepsilon^{\rho-1} \exp\left\{ -\frac{1}{2} \frac{x_\varepsilon}{\sigma_s} \right\} + 2\sigma_s(\rho - 1) \int_{x_\varepsilon}^{\infty} t^{\rho-2} \exp\left\{ -\frac{1}{2} \frac{t}{\sigma_s} \right\} dt \\
&= \exp\left\{ -\frac{1}{2} \frac{x_\varepsilon}{\sigma_s} \right\} \sum_{i=0}^{\lceil \rho \rceil} (2\sigma_s)^i \rho(\rho - 1) \cdots (\rho - i + 1) x_\varepsilon^{\rho - i}.
\end{aligned}
$$

The case

$$
E\left(|Z_s|^\rho I(Z_s \leq -x_\varepsilon)\right)
$$

is calculated similarily. The case $\rho \geq 1$ has been proved. □

G Proof of Lemma 9

By Lemma 6 it is enough to prove lemma for $s < s^{\#}$. Let $\bar{s'} = \bar{s'}(s')$ be the smallest element of B_{grid} greater than s'. We have that

$$\begin{aligned}
\sup_{\theta\in\Theta_\nu} P_\theta\left(\hat{s}=s'\right) &\leq \sum_{s''\in B_{grid}, s''\leq s'} \sup_{\theta\in\Theta_\nu} P_\theta\left(\left|\hat{Q}_{\bar{s'}}-\hat{Q}_{s''}\right| > \eta(s'')\right) \\
&\leq \operatorname{Card}(B_{grid}) \max_{s''\in B_{grid}, s''\leq s'} \sup_{\theta\in\Theta_\nu} P_\theta\left(\left|\hat{Q}_{\bar{s'}}-\hat{Q}_{s''}\right| > \eta(s'')\right).
\end{aligned}$$

We have for $s'' \leq s'$, by Lemma 5, that $\mathcal{B}_{s,L}(s'', 1, \tilde{\delta}(s'')) \leq C_1 \tilde{\delta}^{\tilde{r}}(s'')$, where $\tilde{r} = 4(\tilde{s}-\kappa)/(4s''+1)$ and $\tilde{s} = \tilde{s}(s, s'')$ is as in Lemma 5. By (57), $\eta(s'') \geq C_2 \tilde{\delta}^{r''}(s'')$ where $r'' = 4(s''-\kappa)/(4s''+1)$. Then, by (54),

$$\begin{aligned}
\frac{\mathcal{B}_{s,L}(s'', 1, \tilde{\delta}(s''))}{\eta(s'')} &\leq C_3 \exp\left\{-C_4(\tilde{s}-s'')\log\delta^{-1}\right\} \\
&\leq C_3 \exp\left\{-C_4 \min\{c, s-s^-\}\log\delta^{-1}\right\} \\
&= C_3 \exp\left\{-C_4 \log\log\log\delta^{-1}\right\} \\
&\leq (\log\log\delta^{-1})^{-C_5} \\
&\stackrel{def}{=} \gamma''_\varepsilon
\end{aligned}$$

where c is as in Lemma 5. Using similar inference for $\mathcal{B}_{s,L}(\bar{s'}, 1, \tilde{\delta}(\bar{s'}))$ we have for $\theta \in \Theta_\nu$, for $s'' \leq s'$, for sufficiently small ε,

$$\begin{aligned}
\left|E_\theta\hat{Q}_{s''}-Q(\theta)\right| + \left|E_\theta\hat{Q}_{\bar{s'}}-Q(\theta)\right| &\leq C_6\left[\mathcal{B}_{s,L}(s'', 1, \tilde{\delta}(s'')) + \mathcal{B}_{s,L}(\bar{s'}, 1, \tilde{\delta}(\bar{s'}))\right] \\
&\leq C_7\gamma''_\varepsilon\eta(s'').
\end{aligned}$$

Denote $\gamma'_\varepsilon = C_7\gamma''_\varepsilon$. Now, for $\theta \in \Theta_\nu$, for sufficiently small ε,

$$\begin{aligned}
\left|\hat{Q}_{\bar{s'}}-\hat{Q}_{s''}\right| &\leq \left|E_\theta\hat{Q}_{\bar{s'}}-Q(\theta)\right| + \left|E_\theta\hat{Q}_{s''}-Q(\theta)\right| \\
&\quad + |Z_{\bar{s'}}-Z_{s''}| + |U_{\bar{s'}}(\theta)| + |U_{s''}(\theta)| \\
&\leq \gamma'_\varepsilon\eta(s'') + |Z_{\bar{s'}}-Z_{s''}| + |U_{\bar{s'}}(\theta)| + |U_{s''}(\theta)|.
\end{aligned}$$

Thus, for $\theta \in \Theta_\nu$, for sufficiently small ε,

$$\begin{aligned}
&P_\theta\left(\left|\hat{Q}_{\bar{s'}}-\hat{Q}_{s''}\right| > \eta(s'')\right) \\
&\leq P_\theta\left(|Z_{\bar{s'}}-Z_{s''}| + |U_{\bar{s'}}(\theta)| + |U_{s''}(\theta)| > \eta(s'')(1-\gamma'_\varepsilon)\right) \\
&\leq P_\theta\left(|Z_{\bar{s'}}-Z_{s''}| > \eta(s'')(1-2\gamma'_\varepsilon)\right) + P_\theta\left(|U_{s''}(\theta)| > \eta(s'')\gamma'_\varepsilon/2\right) \\
&\quad + P_\theta\left(|U_{\bar{s'}}(\theta)| > \eta(s'')\gamma'_\varepsilon/2\right).
\end{aligned}$$

We have by Lemma 13 (i), for sufficiently small ε,

$$\begin{aligned}\mathrm{std}\left(Z_{\bar{s}'} - Z_{s''}\right) &= \delta \left\| g\left(s'', 1, \tilde{\delta}(s'')\right) - g\left(\bar{s}', 1, \tilde{\delta}(\bar{s}')\right) \right\|_2 \\ &\leq \delta \left\| g\left(s'', 1, \tilde{\delta}(s'')\right) \right\|_2 \\ &= \sigma_{s''}.\end{aligned}$$

Now from (22), $\eta(s'') = \sigma_{s''} d_\varepsilon(s'', s^\#)$. Thus, for sufficiently small ε,

$$\frac{\eta(s'')}{\mathrm{std}\left(Z_{\bar{s}'} - Z_{s''}\right)} \geq d_\varepsilon(s'', s^\#) \geq d_\varepsilon(s', s^\#).$$

Thus, applying Lemma 8 (i) and remark (58),

$$\begin{aligned}P_\theta\left(|Z_{\bar{s}'} - Z_{s''}| > \eta(s'')(1 - 2\gamma'_\varepsilon)\right) &\leq C_8 \exp\left\{ -\frac{1}{2} \frac{\eta(s'')(1 - 2\gamma'_\varepsilon)}{\mathrm{std}\left(Z_{\bar{s}'} - Z_{s''}\right)} \right\} \\ &\leq C_8 \exp\left\{ -\frac{1}{2} d_\varepsilon(s', s^\#)(1 - 2\gamma'_\varepsilon) \right\}\end{aligned}$$

for sufficiently small ε. By Lemma 6, $\Theta_{s,L} \subset \Theta_{s'',1}$. Thus, denoting again $r'' = 4(s'' - \kappa)/(4s'' + 1)$, for $\theta \in \Theta_{s,L}$, using Lemma 13 (ii),

$$\begin{aligned}\mathrm{Var}(U_{s''}(\theta)) &= 4\varepsilon^2 \sum_{i=1}^{\infty} g_i^2(s'', 1, \tilde{\delta}(s''))\theta_i^2 \\ &= \frac{4}{\sqrt{2}} \frac{\tilde{\delta}^{2r''}(s'')}{d_\varepsilon(s'', s^\#)} \tilde{\delta}^{1-2r''}(s'') \sum_{i=1}^{\infty} g_i^2(s'', 1, \tilde{\delta}(s''))\theta_i^2 \\ &\leq \frac{4}{\sqrt{2}} \frac{\tilde{\delta}^{2r''}(s'')}{d_\varepsilon(s'', s^\#)} C'_8 \max\left\{ \tilde{\delta}^c(s''), \tilde{\delta}^{1-2r''}(s'') \right\} \\ &\leq \frac{4}{\sqrt{2}} \frac{\tilde{\delta}^{2r''}(s'')}{d_\varepsilon(s'', s^\#)} C'_8 \tilde{\delta}^{1-2r(s_{m-1})}(s_{m-1})\end{aligned}$$

because $s'' \leq s' \leq s_{m-1}$. Furthermore,

$$\begin{aligned}\tilde{\delta}^{1-2r(s_{m-1})}(s_{m-1}) &\leq C_9 \exp\{-C_{10}(1 - r(s_{m-1})) \log \delta^{-1}\} \\ &\leq C_9 \exp\{-C_{11}(2\kappa + 1/4 - s_{m-1}) \log \delta^{-1}\} \\ &\leq C_9 \exp\{-C_{12} \log\log\log \delta^{-1} \cdot \log \delta^{-1}\} \\ &\stackrel{def}{=} \alpha(\delta).\end{aligned}$$

Combining two previous displays, for $\theta \in \Theta_{s,L}$, for sufficiently small ε,

$$\mathrm{Var}(U_{s''}(\theta)) \leq C''_8 \frac{\tilde{\delta}^{2r''}(s'')}{d_\varepsilon(s'', s^\#)} \alpha(\delta). \tag{77}$$

We have that $(\gamma'_\varepsilon)^2/\alpha(\delta) \to \infty$ as $\varepsilon \to 0$. Thus, because by (57) $\eta(s'') \geq C_9 \tilde{\delta}^{r''}(s'')$,

$$\frac{\eta^2(s'')(\gamma'_\varepsilon)^2}{4\mathrm{Var}(U_{s''}(\theta))} \geq C_{13}(\gamma'_\varepsilon)^2 d_\varepsilon(s'', s^\#)/\alpha(\delta) \geq d_\varepsilon(s'', s^\#) \geq d_\varepsilon(s', s^\#)$$

for sufficiently small ε. Then, applying Lemma 8 (ii) with the fact that $U_s(\theta) \sim N(0, \mathrm{Var}(U_s(\theta)))$,

$$\begin{aligned} P_\theta\left(|U_{s''}(\theta)| > \eta(s'')\gamma'_\varepsilon/2\right) &\leq C_{11} \exp\left\{-\frac{1}{2}\frac{(\eta(s'')\gamma'_\varepsilon/2)^2}{\mathrm{Var}(U_{s''}(\theta))}\right\} \\ &\leq C_{11} \exp\left\{-\frac{1}{2}\, d_\varepsilon(s', s^\#)\right\} \end{aligned}$$

for sufficiently small ε. Similarily,

$$P_\theta\left(|U_{\bar{s'}}(\theta)| > \eta(s'')\gamma'_\varepsilon/2\right) \leq C_{12} \exp\left\{-\frac{1}{2}\, d_\varepsilon(s', s^\#)\right\}$$

for sufficiently small ε. □

H Proof of (63).

Let $s' \in B_{grid}$, $s' < s^-$. Because $s' \leq s_{m-1}$, then by the definition of η in (22), condition (16) on the grid, and (57),

$$\begin{aligned} \tau(s') &\leq 2\eta(s')\left(1 + \frac{(\log\delta^{-1})^{1/2}}{d_\varepsilon(s', s^\#)}\right) \\ &\leq 2\eta(s')\left(1 + \left[2\rho(r(s^\#) - r(s_{m-1}))(\log\delta^{-1})^{1/2}\right]^{-1}\right) \\ &\leq 2\eta(s')\left(1 + \left[C_1(s^\# - s_{m-1})(\log\delta^{-1})^{1/2}\right]^{-1}\right) \\ &\leq C_2\eta(s') \\ &\leq C_3\psi_{s',L}. \end{aligned} \tag{78}$$

Thus,

$$\rho_1(\nu) \leq C_5 \sum_{s' \in B_{grid}, s' < s^-} \sup_{\theta\in\Theta_\nu} P_\theta\left(\hat{s} = s'\right)\left(\frac{\psi_{s',L}}{\psi_\nu}\right)^\rho.$$

First, when $s \leq s_{m-1}$, then from (55) and (57) it follows that

$$\frac{\psi_{s',L}}{\psi_\nu} \leq C_4(s^\# - s_{m-1})^{-C_5} \exp\left\{\frac{1}{2\rho}\, d_\varepsilon(s', s)\right\}. \tag{79}$$

Thus, using Lemma 9 and the fact that by (14)

$$\mathrm{Card}(B_{grid}) \le C_6(\log\delta^{-1})^{C_7},$$

$$\begin{aligned}\rho_1(\nu) &\le C_8\mathrm{Card}(B_{grid})(s^{\#}-s_{m-1})^{-\rho C_5}\\ &\quad \sum_{s'\in B_{grid}, s'<s^-}\left[\exp\left\{\frac{1}{2}\,d_\varepsilon(s',s)\right\}\exp\left\{-\frac{1}{2}\,d_\varepsilon(s',s^{\#})(1-\gamma_\varepsilon)\right\}\right]\\ &\le \left(\log\delta^{-1}(s^{\#}-s_{m-1})^{-1}\right)^{C_9}\exp\left\{-c_\varepsilon\log\delta^{-1}\right\}\end{aligned}$$

where we denoted

$$c_\varepsilon = 2\rho(4\kappa+1)\left[\frac{s^{\#}-s_{m-1}}{(4s^{\#}+1)^2}-\gamma_\varepsilon\,\frac{s^{\#}-\kappa}{(4s^{\#}+1)^2}\right].$$

Note that by (16) and by the definition of γ_ε given in Lemma 9, $c_\varepsilon \ge C_{10}(s^{\#}-s_{m-1})$ for sufficiently small ε. The lemma follows for the case $s \le s_{m-1}$, because

$$\lim_{\varepsilon\to 0}\frac{(s^{\#}-s_{m-1})\log\delta^{-1}}{\log[(s^{\#}-s_{m-1})^{-1}\log\delta^{-1}]}=\infty$$

by (16). For the case $s = s_m = 2\kappa + 1/4$, we apply (52), (56), (57) to get

$$\frac{\psi_{s',L}}{\psi_\nu} \le C_3 b_\varepsilon^{-1/2}(\log\delta^{-1})^{C_4}\exp\left\{\frac{1}{2\rho}\,d_\varepsilon(s',s)\right\} \tag{80}$$

and lemma follows for this case. Finally for the case $s > s_m = 2\kappa + 1/4$, we apply again (52), (56), (57) to get

$$\frac{\psi_{s',L}}{\psi_\nu} \le C_3(\log\delta^{-1})^{C_4}\exp\left\{\frac{1}{2\rho}\,d_\varepsilon(s',2\kappa+1/4)\right\}. \tag{81}$$

and lemma follows also for this case. We have proved (63). □

I Proof of (64).

Let again $s' \in B_{grid}$, $s' < s^-$. Now,

$$\begin{aligned}&E\left[(\psi_{s',L}+|Z_{s'}|+|U_{s'}(\theta)|)^\rho\, I\left(|Z_{s'}|+|U_{s'}(\theta)|>\tau(s')\right)\right]\\ &\le C_1 E\left[\left(\psi^\rho_{s',L}+|Z_{s'}|^\rho+|U_{s'}(\theta)|^\rho\right) I\left(|Z_{s'}|+|U_{s'}(\theta)|>\tau(s')\right)\right]\\ &\le C_2\left(\psi^\rho_{s',L}P\left(|Z_{s'}|>\tau(s')/2\right)+E\left[|Z_{s'}|^\rho I\left(|Z_{s'}|>\tau(s')/2\right)\right]\right.\\ &\qquad \left.+\,\psi^\rho_{s',L}P\left(|U_{s'}(\theta)|>\tau(s')/2\right)+E\left[|U_{s'}(\theta)|^\rho I\left(|U_{s'}(\theta)|>\tau(s')/2\right)\right]\right)\end{aligned}$$

by using (78) and

$$
\begin{aligned}
& I\left(|Z_{s'}| + |U_{s'}(\theta)| > \tau(s')\right) \\
& \quad \leq \; I\left(|Z_{s'}| > \tau(s')/2,\, |U_{s'}(\theta)| \leq \tau(s')/2\right) \\
& \qquad + I\left(|Z_{s'}| \leq \tau(s')/2,\, |U_{s'}(\theta)| > \tau(s')/2\right) \\
& \qquad + I\left(|Z_{s'}| > \tau(s')/2,\, |U_{s'}(\theta)| > \tau(s')/2\right). \qquad (82)
\end{aligned}
$$

Now, again by (78),

$$
\begin{aligned}
\sum_{i=0}^{\lceil\rho\rceil} \tau^{\rho-i}(s')\sigma_{s'}^{i} &= \sigma_{s'}^{\rho} \sum_{i=0}^{\lceil\rho\rceil} 2^{\rho-i}\left[d_\varepsilon(s', \min\{s, s^{\#}\}) + (\log\delta^{-1})^{1/2}\right]^{\rho-i} \\
&\leq (\lceil\rho\rceil + 1)(2\sigma_{s'})^{\rho}\left[d_\varepsilon(s', \min\{s, s^{\#}\}) + (\log\delta^{-1})^{1/2}\right]^{\rho} \\
&= (\lceil\rho\rceil + 1)\tau^{\rho}(s') \\
&\leq C_3\psi_{s',L}^{\rho}.
\end{aligned}
$$

Also, by (77), because $\lim_{\gamma\to 0}\alpha(\gamma) = 0$, and using also (57), denoting $r' = 4(s' - \kappa)/(4s' + 1)$,

$$
\mathrm{Var}(U_{s'}(\theta)) \leq C_4 \frac{\tilde{\delta}^{2r'}(s')}{d_\varepsilon^2(s', s^{\#})} \leq C_5 \frac{\eta^2(s')}{d_\varepsilon^2(s', s^{\#})} = C_5\sigma_{s'}^2. \qquad (83)
$$

Also, $\sigma_{s'}^2 \leq C_6\psi_{s',L}^2$. By definition, $\tau(s')/\sigma_{s'} = 2[d_\varepsilon(s', \min\{s, s^{\#}\}) + (\log\delta^{-1})^{\alpha}]$. Also,

$$
\frac{\tau^2(s')}{\mathrm{Var}(U_{s'}(\theta))} \geq \frac{\tau^2(s')}{\sigma_{s'}^2} \geq 4\left[d_\varepsilon(s', \min\{s, s^{\#}\}) + (\log\delta^{-1})^{\alpha}\right].
$$

Thus, by Lemma 8 and comment (58),

$$
\begin{aligned}
\limsup_{\varepsilon\to 0}\sup_{\nu\in B'} \rho_2(\nu) \leq{}& C_7 \limsup_{\varepsilon\to 0}\sup_{\nu\in B'} \sum_{s'\in B_{grid}, s'<s^-} \psi_\nu^{-\rho} \sup_{\theta\in\Theta_\nu} \\
& \left[\left(\psi_{s',L}^{\rho} + \sum_{i=0}^{\lceil\rho\rceil}\tau^{\rho-i}(s')\sigma_{s'}^{i}\right)\exp\left\{-\frac{1}{4}\,\frac{\tau(s')}{\sigma_{s'}}\right\}\right. \\
& \left. + \left(\psi_{s',L}^{\rho} + \tau^{\rho}(s') + (\mathrm{Var}(U_{s'}(\theta)))^{\rho/2}\right)\exp\left\{-\frac{1}{8}\,\frac{\tau^2(s')}{\mathrm{Var}(U_{s'}(\theta))}\right\}\right] \\
\leq{}& C_8 \limsup_{\varepsilon\to 0}\sup_{\nu\in B'} \sum_{s'\in B_{grid}, s'<s^-} \left(\frac{\psi_{s',L}}{\psi_\nu}\right)^{\rho} \\
& \exp\left\{-\frac{1}{2}\left[d_\varepsilon(s', \min\{s, s^{\#}\}) + (\log\delta^{-1})^{1/2}\right]\right\}.
\end{aligned}
$$

Then, for the case $s \le s_{m-1}$, using (79)

$$\begin{aligned} &\lim_{\varepsilon\to 0}\sup_{\nu\in B'}\rho_2(\nu) \\ &\le\; C_9\lim_{\varepsilon\to 0}\mathrm{Card}(B_{grid})\left(s^{\#}-s_{m-1}\right)^{-\rho C_{10}}\exp\left\{-\frac{1}{2}\left(\log\delta^{-1}\right)^{1/2}\right\}=0 \end{aligned}$$

where we used (16). For two other cases, $s = s_m = s^{\#} = 2\kappa + 1/4$, $s > s_m = s^{\#} = 2\kappa + 1/4$, we use (80) and (81) to prove similarly that $\lim_{\varepsilon\to 0}\sup_{\nu\in B'}\rho_2(\nu) = 0$. We have proved (64). □

J A lemma needed to prove the lower bound of Theorem 1

To prove the lower bound we need the following lemma.

Lemma 14 *There exists such $\theta(s,L,\gamma) \in \Theta_{s,L}$ that*

$$\lim_{\gamma\to 0}\frac{Q(\theta(s,L,\gamma))}{\gamma^{r(s)}L^{1-r(s)}} = c_{s,p}, \tag{84}$$

$$\sum_{i=1}^{\infty}\theta_i^4(s,L,\gamma) = \gamma^2, \tag{85}$$

and

$$\sup_{(s,L)\in A_0}\theta_i^2(s,L,\gamma) \le C\gamma \tag{86}$$

for a positive constant C, where $A_0 = [s_*, s^{\#}] \times [L_*, L^*]$.

Proof. We will choose $\theta(s,L,\gamma)$ as in (44). Equation (84) follows from (51) and equation (85) follows from (45). Let us prove (86). Function $t \mapsto t^{2\kappa}/(1+b_1t^q)$ is maximized by $t = (2\kappa/[b_1(q-2\kappa)])^{1/q}$. Thus

$$|\theta_i(s,L,\gamma)| \le C'\left(a_1b_1^{-2\kappa/q}\right)^{1/(p-2)} \le C\gamma$$

where a_1 and b_1 are from (45). Here we used the facts $a_1b_1^{-2\kappa/q} \sim b_1^{(p-2)(1+1/q)/p}$ and $b_1 \sim \gamma$, which follow from (48) and (50). □

K The complete proof of the lower bound of Theorem 1

Let us denote

$$\widetilde{\psi}_{s,L} = c_{s,p}\varphi_{s,L}$$

where $c_{s,p}$ is defined in (27) and $\varphi_{s,L}$ is defined in (25).

Let us note that l_p bodies are orthosymmetric, that is, when $\theta \in \Theta_{s,L}$, then $\tau_i\theta \in \Theta_{s,L}$, where τ_i changes the sign of the i:th element. The proof will start with a reasoning given by Efroimovich and Low (1996), which shows that when the parameter space is orthosymmetric, we can restrict attention to estimators which are functions of (y_i^2) only, where $y_i = \theta_i + \varepsilon z_i$ are our observations. After that we can proceed as in the linear case.

Let us thus show that

$$\inf_{Q_\varepsilon \in F} \sup_{\nu \in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \geq \inf_{Q_\varepsilon \in \tilde{F}} \sup_{\nu \in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \tag{87}$$

where F is the set of measurable functions from $\mathbf{R}^\infty \to \mathbf{R}$ where $\mathbf{R}^\infty = \{(x_1, x_2, \ldots) : x_i \in \mathbf{R}\}$ and $\tilde{F}$ is the set of those measurable functions from $\mathbf{R}^\infty \to \mathbf{R}$ which are functions of x_i^2 only,

$$\tilde{F} = \left\{ \tilde{Q} : \mathbf{R}^\infty \to \mathbf{R} \;\middle|\; \tilde{Q}((x_i)) = Q((x_i^2)), \text{ for some } Q : \mathbf{R}^\infty \to \mathbf{R} \right\}.$$

Denote by τ_i the function $\mathbf{R}^\infty \to \mathbf{R}^\infty$ which changes the sign of the ith element. Let $\nu \in B$, $\theta \in \Theta_\nu$. By orthosymmetricity of Θ_ν, also $\tau_i\theta \in \Theta_\nu$. Assume that $Q_\varepsilon \in F$ is such that $Q_\varepsilon(x) \neq Q_\varepsilon(\tau_i x)$. Then, by convexity of the loss function, denoting $Y = (y_i)$,

$$\begin{aligned}
&\max\{E_\theta |Q_\varepsilon(Y) - Q(\theta)|^\rho, E_{\tau_i\theta} |Q_\varepsilon(Y) - Q(\tau_i\theta)|^\rho\} \\
&\geq \frac{1}{2} E_\theta |Q_\varepsilon(Y) - Q(\theta)|^\rho + \frac{1}{2} E_{\tau_i\theta} |Q_\varepsilon(Y) - Q(\tau_i\theta)|^\rho \\
&= \frac{1}{2} E_\theta |Q_\varepsilon(Y) - Q(\theta)|^\rho + \frac{1}{2} E_\theta |Q_\varepsilon(\tau_i Y) - Q(\theta)|^\rho \\
&\geq E_\theta \left| \frac{1}{2} (Q_\varepsilon(Y) - Q(\theta)) + \frac{1}{2} (Q_\varepsilon(\tau_i Y) - Q(\theta)) \right|^\rho \\
&= E_\theta \left| \frac{1}{2} (Q_\varepsilon(Y) + Q_\varepsilon(\tau_i Y)) - Q(\theta) \right|^\rho.
\end{aligned}$$

Now $(Q_\varepsilon + Q_\varepsilon(\tau_i\cdot))/2$ does not depend on the sign of the ith argument. The inequality (87) has been proved.

Let $s' = s_1$ and $s'' = s_{m-1}$. Let $L', L'' \in [L_*, L^*]$. Let us denote $\nu' = (s', L')$ and $\nu'' = (s'', L'')$. By (84) and (85) there is such $\tilde{\theta} \in \mathbf{R}^\infty$ that

$$\tilde{\theta} \in \Theta_{s',L'}, \qquad Q(\tilde{\theta}) \sim (L')^{1-r(s')} \tilde{\delta}^{r(s')} c_{s,p}, \qquad \sum_{i=1}^{\infty} \tilde{\theta}_i^4 = \tilde{\delta}^2 \tag{88}$$

where

$$\tilde{\delta} = \delta d_\varepsilon(s', s^{\#}), \ \ \delta = \sqrt{2}\varepsilon^2.$$

Consider the sequencies $\bar{\theta}_0, \bar{\theta}_1 \in \mathbf{R}^\infty$,

$$\bar{\theta}_0 = 0, \qquad \bar{\theta}_1 = (1-\xi)^{1/2} \tilde{\theta}$$

where $0 < \xi < 1/2$ is arbitrary. Now $\bar{\theta}_0 \in \Theta_{\nu''}$ and $\bar{\theta}_1 \in \Theta_{\nu'}$. Also, by (88),

$$Q(\bar{\theta}_0) = 0, \qquad Q(\bar{\theta}_1) \sim (1-\xi)\widetilde{\psi}_{\nu'}.$$

Thus for any estimator Q_ε,

$$\left|Q_\varepsilon - Q(\bar{\theta}_0)\right| = \widetilde{\psi}_{\nu'}\, D\left((1-\xi)^{-1}\widetilde{\psi}_{\nu'}^{-1} Q_\varepsilon, 0\right)$$

and

$$\left|Q_\varepsilon - Q(\bar{\theta}_1)\right| \sim \psi_{\nu'}\, D\left((1-\xi)^{-1}\widetilde{\psi}_{\nu'}^{-1} Q_\varepsilon, 1\right),$$

where $D(u,v) = (1-\xi)|u-v|$, $u, v \in \mathbf{R}$. Thus, denoting $q = \widetilde{\psi}_{\nu'}/\widetilde{\psi}_{\nu''}$, denoting by $\tilde{P}_\theta$ the distribution of (y_i^2), when $y_i = \theta_i + \varepsilon z_i$, denoting $\tilde{P}_i = \tilde{P}_{\bar{\theta}_i}$, and letting $\tilde{E}_i$ be the corresponding expectations,

$$\begin{aligned}
&\inf_{Q_\varepsilon \in F} \sup_{\nu \in B} \mathcal{R}_{\varepsilon,\nu}(Q_\varepsilon, \widetilde{\psi}_\nu) \\
&\quad \geq \inf_{Q_\varepsilon \in \tilde{F}} \sup_{\nu \in B} \sup_{\theta \in \Theta_\nu} \tilde{E}_\theta \left(\widetilde{\psi}_\nu^{-\rho} \left|Q_\varepsilon - Q(\theta)\right|^\rho\right) \\
&\quad \geq \inf_{Q_\varepsilon \in \tilde{F}} \max\left\{\tilde{E}_0 \left(\widetilde{\psi}_{\nu''}^{-\rho} \left|Q_\varepsilon - Q(\bar{\theta}_0)\right|^\rho\right), \tilde{E}_1 \left(\widetilde{\psi}_{\nu'}^{-\rho} \left|Q_\varepsilon - Q(\bar{\theta}_1)\right|^\rho\right)\right\} \\
&\quad \sim \inf_{Q_\varepsilon \in \tilde{F}} \max\left\{q^\rho \tilde{E}_0 D^\rho(Q_\varepsilon, 0), \tilde{E}_1 D^\rho(Q_\varepsilon, 1)\right\}.
\end{aligned}$$

Now the density of y_1^2 with respect to Lebesgue measure is

$$\frac{1}{2\sqrt{2\pi x_1}\varepsilon} \exp\left\{-\frac{x_1 + \theta_1^2}{2\varepsilon^2}\right\} \left(\exp\left\{\frac{\theta_1\sqrt{x_1}}{\varepsilon^2}\right\} + \exp\left\{-\frac{\theta_1\sqrt{x_1}}{\varepsilon^2}\right\}\right) I_{[0,\infty)}(x_1).$$

We have that

$$\exp\left\{\frac{\theta_1\sqrt{x_1}}{\varepsilon^2}\right\} + \exp\left\{-\frac{\theta_1\sqrt{x_1}}{\varepsilon^2}\right\} = 2\sum_{j=0}^{\infty} \frac{\left(\varepsilon^{-2}\theta_1\sqrt{x_1}\right)^{2j}}{(2j)!} \leq 2\exp\left\{\frac{\theta_1^2 x_1}{\varepsilon^4}\right\}.$$

Thus,

$$\begin{aligned}\frac{d\tilde{P}_0}{d\tilde{P}_1}(x) &= \prod_{i=1}^{\infty}\left[2\exp\left\{\frac{\bar{\theta}_{1i}^2}{2\varepsilon^2}\right\}\left(\exp\left\{\frac{\bar{\theta}_{1i}\sqrt{x_i}}{\varepsilon^2}\right\}+\exp\left\{-\frac{\bar{\theta}_{1i}\sqrt{x_i}}{\varepsilon^2}\right\}\right)^{-1} I_{[0,\infty)}(x_i)\right] \\ &\geq \prod_{i=1}^{\infty}\left[\exp\left\{\frac{\bar{\theta}_{1i}^2}{2\varepsilon^2}-\frac{\bar{\theta}_{1i}^2 x_i}{\varepsilon^4}\right\} I_{[0,\infty)}(x_i)\right].\end{aligned}$$

Denote

$$S = -\sum_{i=1}^{\infty}\bar{\theta}_{1i}^2 y_i^2.$$

Now $E_\theta y_i^2 = \theta_i^2 + \varepsilon^2$ and $\mathrm{Var}_\theta y_i^2 = 2\varepsilon^4 + 4\theta_i^2\varepsilon^2$ and thus

$$E_{\bar{\theta}_1}(S) = -\left(\sum_{i=1}^{\infty}\bar{\theta}_{1i}^4 + \varepsilon^2\sum_{i=1}^{\infty}\bar{\theta}_{1i}^2\right)$$

and

$$\mathrm{Var}_{\bar{\theta}_1}(S) = 2\varepsilon^4\sum_{i=1}^{\infty}\bar{\theta}_{1i}^4 + 4\varepsilon^2\sum_{i=1}^{\infty}\bar{\theta}_{1i}^6.$$

Let

$$\tau = \exp\left\{-\frac{1-\xi}{2}\, d_\varepsilon(s', s^{\#})\right\}.$$

Then, by Markov's inequality, for $0 < \alpha < 1$,

$$\begin{aligned}\tilde{P}_1\left(\frac{d\tilde{P}_0}{d\tilde{P}_1} \geq \tau\right) &\geq P_{\bar{\theta}_1}\left(S \geq \varepsilon^4\log\tau - \frac{\varepsilon^2}{2}\,\|\bar{\theta}_1\|_2^2\right) \\ &\geq 1 - \left(\varepsilon^4\log\tau + \frac{\varepsilon^2}{2}\,\|\bar{\theta}_1\|_2^2 + \sum_{i=1}^{\infty}\bar{\theta}_{1i}^4\right)^{-2}\mathrm{Var}_{\bar{\theta}_1}(S) \\ &\geq 1-\alpha \qquad (89)\end{aligned}$$

for sufficiently small ε, because

$$\left(\varepsilon^4\log\tau + \frac{\varepsilon^2}{2}\,\|\bar{\theta}_1\|_2^2 + \sum_{i=1}^{\infty}\bar{\theta}_{1i}^4\right)^{-2}\mathrm{Var}_{\bar{\theta}_1}(S) \leq \frac{C_1}{d_\varepsilon(s', s^{\#})} \longrightarrow 0 \qquad (90)$$

when $\varepsilon \to 0$. Indeed, by (88),

$$2\varepsilon^4\sum_{i=1}^{\infty}\bar{\theta}_{1i}^4 = 2\varepsilon^4\tilde{\delta}^2 = 4\varepsilon^8 d_\varepsilon^2(s', s^{\#})$$

and $\bar{\theta}_{1i}^2 \le C_2\tilde{\delta}$ for $i = 1, \ldots$, by (86). Thus

$$4\varepsilon^2 \sum_{i=1}^{\infty} \bar{\theta}_{1i}^6 \le C_3 d_\varepsilon(s', s^\#)\varepsilon^4 \sum_{i=1}^{\infty} \bar{\theta}_{1i}^4 \le C_4\varepsilon^8 d_\varepsilon^3(s', s^\#),$$

and thus

$$\mathrm{Var}_{\bar{\theta}_1}(S) \le C_5\varepsilon^8 d_\varepsilon^3(s', s^\#),$$

and also

$$\begin{aligned}
&\left(\varepsilon^4 \log\tau + \frac{\varepsilon^2}{2}\,\|\bar{\theta}_1\|^2 + \sum_{i=1}^{\infty} \bar{\theta}_{1i}^4\right)^2 \\
&\ge \left(-\frac{1-\xi}{2}\,\varepsilon^4 d_\varepsilon(s', s^\#) + 2\varepsilon^4 d_\varepsilon^2(s', s^\#)\right)^2 \ge C_6\varepsilon^8 d_\varepsilon^4(s', s^\#)
\end{aligned}$$

for sufficiently small ε. Thus (90) follows. Then, applying Theorem 6 (i) in Tsybakov (1998), with the fact (89),

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \inf_{Q_\varepsilon \in F} \max\left\{q^\rho \tilde{E}_0 D^\rho(Q_\varepsilon, 0), \tilde{E}_1 D^\rho(Q_\varepsilon, 1)\right\} \\
&\ge \frac{(1-\alpha)(1-2\xi)^\rho \tau(q\xi)^\rho}{(1-2\xi)^\rho + \tau(q\xi)^\rho} = (1-\alpha)(1-2\xi)^\rho
\end{aligned}$$

because, denoting $r' = 4(s' - \kappa)/(4s' + 1)$ and r'', $r^\#$ correspondingly,

$$\begin{aligned}
\tau q^\rho &= \exp\left\{-\frac{1-\xi}{2}\, d_\varepsilon(s', s^\#)\right\}\left(\frac{\widetilde{\psi}_{\nu'}}{\widetilde{\psi}_{\nu''}}\right)^\rho \\
&\sim \exp\left\{\rho[(r'' - r^\#) + \xi(r^\# - r')]\log\delta^{-1}\right\} \\
&\quad \exp\left\{\rho(r' - r'')\log\log\delta^{-1} + \rho r'\log[2\rho(r^\# - r')] - \rho r''\log[2\rho(r^\# - r'')]\right\} \\
&\quad \left(\frac{(L')^{1-r'}(v(s') + b(s'))}{(L'')^{1-r''}(v(s'') + b(s''))}\right)^\rho \longrightarrow \infty
\end{aligned}$$

as $\varepsilon \to 0$ because $(r'' - r^\#) + \xi(r^\# - r') > 0$ for sufficiently small ε. The lower bound is proved because $0 < \xi < 1/2$ and $0 < \alpha < 1$ were chosen arbitrarily.

□

L Reproducing the Figure 1

This material can also found at the address http://www.denstruct.net To reproduce and modify Figure 1 of the article, use the following R function.

```
kvadra<-function(svec,pvec){

z<-matrix(0,length(svec),length(pvec))

for (i in 1:length(svec)){
   s<-svec[i]
 for (j in 1:length(pvec)){
     p<-pvec[j]
     q<-p*(s-1/p+1/2)

#upper bound

I1<-beta(p/(p-2)-1/q-2*k*p/(q*(p-2))-1,1+1/q+2*k*p/(q*(p-2)))/q
I2<-beta(2-(4*k+1)/q,(4*k+1)/q)/q
pote1<-(p-2)*(4*k+1)/(p*(4*s+1))
pote2<-2*(s-k)/(4*s+1)
vakio<-((4*k+1)/(4*(s-k)))^(4*(s-k)/(4*s+1))*(4*s+1)/(4*k+1)
yla<-I1^pote1*I2^pote2*vakio

#lower bound

I3<-beta(4/(p-2)-1/q-8*k/(q*(p-2)),1/q+8*k/(q*(p-2)))/q
I4<-beta(2/(p-2)-1/q-4*k/(q*(p-2))-2*k/q,1/q+4*k/(q*(p-2))+2*k/q)/q
pote3<--2*(4*k+1)/(p*(4*s+1))
pote4<--2*(s-k)/(4*s+1)
ala<-I1^pote3*I3^pote4*I4

    z[i,j]<-yla/ala
  }
}
return(x=svec,y=pvec,z=z)
}
```

Then use for example the following commands (assuming that the above function is in the file kvadra.R).

```
source("kvadra.R")

k<-0

sala<-k+0.01
```

```
syla<-2*k+0.25
sstep<-0.01
svec<-seq(sala,syla,sstep)

pala<-3
pyla<-4
pstep<-0.25
pvec<-seq(pala,pyla,pstep)

kv<-kvadra(svec,pvec)
contour(kv$x,kv$y,kv$z,nlevels=20,xlab="s",ylab="p")
title(sub=expression(paste("a) ", 3<=p<=4)))

k<-0

sala<-k+0.01
syla<-2*k+0.25
sstep<-0.01
svec<-seq(sala,syla,sstep)

pala<-4
pyla<-100
pstep<-1
pvec<-seq(pala,pyla,pstep)

kv<-kvadra(svec,pvec)
contour(kv$x,kv$y,kv$z,nlevels=12,xlab="s",ylab="p")
title(sub=expression(paste("b) ", p>=4)))
```

M l_2 body

The l_2 body is defined as $\{\theta : \sum_{i=1}^{\infty} i^{2s}\theta_i^2 \leq L\}$. It was shown in Donoho and Nussbaum (1990) that the weigths

$$g_i(s, L, \gamma) = \left(i^{2\kappa} - h^{2(s-\kappa)} i^{2s}\right)_+$$

where $(x)_+ = \max\{x, 0\}$,

$$h = h(s, L, \gamma) = \left(\frac{\gamma}{L}\frac{\mu_s}{\lambda_s}\right)^{2/(4s+1)},$$

$$\mu_s = \frac{2(s-\kappa)}{[2(s+\kappa)+1](4s+1)}, \quad \lambda_s = \frac{8(s-\kappa)^2}{[2(s+\kappa)+1](4s+1)(4\kappa+1)}$$

solve the optimal recovery problem.

Note that for the weigths $g(s, L, \gamma)$ we do not have that for fixed s', $\sup_{\Theta_{s,L}} |\sum_{i=1}^{\infty} g_i(s', L', \gamma)\theta_i^2 - Q(\theta)|$ would be smaller than $\gamma^{r(s')}$, simultaneously for several $s > s'$. Thus Lemma 5 does not hold for the case $p = 2$.