Bernoulli **17**(2), 2011, 714–735
DOI: 10.3150/10-BEJ289

Adaptivity and optimality of the monotone least-squares estimator

ERIC CATOR

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands. E-mail: e.a.cator@tudelft.nl

In this paper, we will consider the estimation of a monotone regression (or density) function in a fixed point by the least-squares (Grenander) estimator. We will show that this estimator is locally asymptotic minimax, in the sense that, for each f_0, the attained rate of the probabilistic error is uniform over a shrinking L^2-neighborhood of f_0 and there is no estimator that attains a significantly better uniform rate over these shrinking neighborhoods. Therefore, it adapts to the individual underlying function, not to a smoothness class of functions. We also give general conditions for which we can calculate a (non-standard) limiting distribution for the estimator.

Keywords: adaptivity; least squares; monotonicity; optimality

1. Introduction

There exists an extensive literature on the problem of estimating a monotone increasing regression function or monotone decreasing density. Our results will concern the NPMLE or Grenander estimator for a monotone density, see [7], and the least-squares estimator for a monotone regression function. Prakasa Rao obtained the rate and the limiting distribution for the Grenander estimator in a fixed point in [13], and in [1] a similar result was obtained for the least-squares estimator, both under differentiability conditions on the underlying function. Results for global measures of convergence were obtained in [8] and [11] for the density case and in [4] for the regression case. A unified approach that incorporates some other well-known monotone estimators is given in [5]. A common problem with the global results is that they can only be proved under quite strong conditions, in which case there exist other non-isotonic estimators with faster rates.

We will focus on the pointwise convergence of the least-squares (or Grenander) estimator for general underlying monotone functions (or densities) and show a specific type of adaptivity and optimality of this estimator. We will start by giving our result, and then compare it to other results available in the literature. In this paper, we will consider the white noise model:

$$Y(t) = \int_0^t f_0(s)\,\mathrm{d}s + \frac{1}{\sqrt{n}} W(t),$$

for $t \in [-1, 1]$ and $W(t)$ standard two-sided Brownian motion. Here, f_0 will be a monotone non-decreasing function. However, our results can be generalized to three other models, namely monotone regression with measurements on a grid (not necessarily with normal errors), measurements on random design points and a sample from a decreasing density. Since the general arguments are very similar for all these models, the exact statements and proofs for the other three models can be found in the technical report [3]. The proofs in the white noise model show the important ideas most clearly.

Denote the least-squares estimator at 0 by $\hat{f}(0)$. Let $\mathcal{F}_m$ be the space of monotone increasing functions on $[-1, 1]$, and define $B(f_0, \varepsilon)$ as the L^2-ball around f_0 with radius ε.

Theorem 1.1. *Let $f_0 \in \mathcal{F}_m$ be continuous at 0. Choose two significance levels, $\alpha \in (0, 1)$ and $\beta \in (0, 1/2)$. There exist $\delta > 0$ and $\eta > 0$, such that for all n large enough, we can find a rate γ_n with*

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}_m\cap B(f_0;\delta n^{-1/2})} \mathbb{P}_f\big(|\hat{f}(0) - f(0)| \geq \gamma_n\big) \leq \alpha$$

and

$$\liminf_{n\to\infty} \inf_{\hat{\theta}} \sup_{f\in\mathcal{F}_m\cap B(f_0;\delta n^{-1/2})} \mathbb{P}_f\big(|\hat{\theta}(Y) - f(0)| \geq \eta\cdot\gamma_n\big) > \beta,$$

where $\hat{\theta}(Y)$ is any estimator of $f(0)$ based on the data Y.

There are several aspects of this result that deserve our attention. In words, our result could be formulated as: The rate of the estimator at f_0 is uniform for L^2-neighborhoods of size $\mathrm{O}(n^{-1/2})$, and on these shrinking neighborhoods, no estimator can attain a significantly better uniform rate (the rate γ_n depends explicitly on f_0 (and α), and will be specified in Section 3). The idea of using the minimax concept for shrinking neighborhoods was used by Hájek in [9] when he introduced local asymptotic minimax risk bounds. His result, when restricted to estimating a one-dimensional parameter θ_0, was formulated as follows: Under suitable regularity conditions (LAN) and a symmetric loss function l, there exists a Fisher information $I(\theta_0)$, such that for any sequence of estimators T_n,

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{|\theta-\theta_0|<\delta} \mathbb{E}_\theta\big(l\big(\sqrt{n}(T_n - \theta_0)\big)\big) \geq \mathbb{E}(l(Z)), \tag{1.1}$$

where $Z \sim N(0, I(\theta_0)^{-1})$. An important difference with our approach is the fact that the rate, $\sqrt{n}$, is fixed here (it does not depend on θ_0). This approach has also been extended to semi-parametric models with non-smooth rates (see [6]), but there the rate is fixed through a restriction of the parameter space over which the supremum is taken. The strength of the local asymptotic mimimax result of Hájek is not the rate; it's the fact that the *constant* is optimal. This might be the reason for the limited number of papers we found applying this concept in semi-parametric setups with non-smooth parameters. In those cases, it is usually not feasible to control the constants. We prove optimality with respect to the *rate*, in which case the concept might be more widely applicable.

Another difference between (1.1) and Theorem 1.1 is that our neighborhoods shrink with n, which leads to a stronger result. However, this has also been done for the classical smooth parametric setup in [14], page 118. He extends Hájek's result, under similar conditions, such that if

$\mathcal{I}$ denotes all finite subsets of $\mathbb{R}$, then for any bowl-shaped loss function l,

$$\sup_{I\in\mathcal{I}}\liminf_{n\to\infty}\sup_{h\in I}\mathbb{E}_{\theta_0+h/\sqrt{n}}l\left(\sqrt{n}\left(T_n-\theta_0-\frac{h}{\sqrt{n}}\right)\right)\geq \mathbb{E}l(Z). \tag{1.2}$$

Here we see the shrinking neighborhoods of order $1/\sqrt{n}$. However, the supremum inside the lim inf is not over a full neighborhood, but merely over a finite subset. In Section 3 we will prove a slightly stronger version of Theorem 1.1, where the full neighborhood is replaced by only two functions (depending on n). Also note that in our result, the infimum over all estimators is taken *inside* the liminf over n, which is stronger than the result in (1.2). We expect that we are able to get this stronger result because we concentrate only on the rate and not on optimal constants.

An important difference from previous applications is also that Theorem 1.1 gives a lower bound for all estimators of the probabilistic error, instead of the expected risk of some loss function. Note that a lower bound on the probabilistic error implies a lower bound for (the rate of) the risk of all increasing loss functions; for example, the L^q-loss

$$\mathbb{P}_f\big(|\hat{\theta}(Y)-f(0)|\geq \eta\cdot\gamma_n\big)>\beta \quad\Rightarrow\quad \mathbb{E}_f|\hat{\theta}(Y)-f(0)|^q>\eta^q\gamma_n^q\beta.$$

The drawback is that Theorem 1.1 does not show that the least-squares estimator $\hat{f}(0)$ attains the rate γ_n for the L^q-loss. However, we can show, under mild conditions, that $\hat{f}(0)$ does attain the same rate in L^q-loss; see the technical report [3].

Considering the rate in terms of the probabilistic error is essential if we wish to get the full generality of our results, as was also observed in [2]. In that paper, they consider the white noise model where f_0 (now not necessarily monotone) is an element of a union of k convex parameter classes $\mathcal{F}_i$. Then they consider the classes $\mathcal{G}_j=\bigcup_{i\leq j}\mathcal{F}_i$, which are now not necessarily convex. They construct an estimator for a linear functional Tf_0 (e.g., $f_0(0)$) that attains an optimal rate (which they describe in terms of the modulus of continuity of T) for the probabilistic error for each of the classes $\mathcal{G}_j$ simultaneously. It is known that this type of adaptivity is not possible in general for the L^q error rate. Although their result is very general, it considers only a finite number of classes, and they consider the supremum of the probabilistic error over the entire class $\mathcal{G}_j$. Our results give more detailed information about the rate at any specific f_0, and about the local optimality of that rate.

Another approach that addresses the adaptive estimation of a monotone function at a fixed point can be found in [12]. Here the authors define an estimation procedure $\hat{f}_n$ for $f_0(0)$, where f_0 is a monotone regression function in the white noise model. This estimation procedure is rate-adaptive in a minimax sense with respect to a Lipschitz parameter α. This means that if we define

$$\mathcal{F}_m(\alpha,M)=\{f\in\mathcal{F}_m\colon \forall x,y\in[-1,1]\colon |f(x)-f(y)|\leq M|x-y|^{\alpha}\},$$

then they show that there exist constants $C>0$ and $0<\alpha_0<1$ such that

$$\sup_{f\in\mathcal{F}_m(\alpha,M)}\mathbb{E}_f|\hat{f}_n-f(0)|^2\leq CM^{2/(2\alpha+1)}n^{-2\alpha/(2\alpha+1)},$$

for all $\alpha_0 \le \alpha \le 1$ and $M \ge 1$. It was already known that there exists a constant $D > 0$ such that for any estimator δ_n, we have

$$\sup_{f\in\mathcal{F}_m(\alpha,M)} \mathbb{E}_f|\delta_n - f(0)|^2 \ge DM^{2/(2\alpha+1)}n^{-2\alpha/(2\alpha+1)}.$$

A serious drawback of this procedure compared to the estimator we consider, is that it does not, in general, give a monotone function as an estimate when the procedure is applied to an interval of fixed points. Furthermore, the rate is described in terms of a *global* Lipschitz parameter, and the adaptivity is only shown for $\alpha \in [\alpha_0, 1]$, whereas we will show in Section 3 that the rate of the LSE $\hat{f}(0)$ depends on the *local* behavior of f_0 around 0. This will result in, for example, fast rates when the function f_0 has derivative 0 in 0. Also, we will allow (locally optimal) rates that cannot be described by the Lipschitz parameter alone (think of logarithmic corrections), or even functions that are not Lipschitz at all. Finally, we show adaptivity in the sense of local asymptotic minimax without any restriction on f_0 other than monotonicity and continuity at 0.

An attractive feature of our results is that the rate is described explicitly for each f_0, so that we do not need smoothness classes to specify the rate. This is natural for our setup, since no matter how smooth f_0 is, in the L^2 sense (which is equivalent in this case to the Hellinger distance between the different models) there are a lot of very non-smooth increasing functions close to f_0. In fact, we believe that when we replace the parameter space $\mathcal{F}_m$ by a different convex and compact subset $\mathcal{K}$ of $L^2[-1, 1]$, and we consider a linear functional that is continuous on $\mathcal{K}$, it might be possible to prove the analogue of Theorem 1.1 for the least-squares plug-in estimator under some regularity conditions on $\mathcal{K}$. However, this is future work.

We feel that Theorem 1.1 is the most important result of the paper. However, in proving this theorem, we will also give an explicit expression for the rate γ_n of the least-squares estimator in terms of the local behavior of the underlying f_0. This result can be found in Theorem 3.1 of Section 3, just before the proof of Theorem 1.1. Furthermore, in Section 4, we give weak regularity conditions on f_0 near 0, under which we can determine the non-standard limiting distribution of the least-squares estimator. These conditions imply that f_0 behaves similarly on each scale in a neighborhood of 0, and they are weaker than differentiability conditions. The limiting distributions found for the least-squares estimator in [1] and [15] are special cases of our result. Section 2 is preliminary, and describes properties of certain left- and right-scale functions of f_0 that play an important role in our proofs.

2. The scale functions ψ_r and ψ_l

Without loss of generality, we will assume in the rest of the paper that $f_0(0) = 0$. Define

$$F_0(t) = \int_0^t f_0(s)\,\mathrm{d}s.$$

Note that this function is convex, and due to the continuity of f_0 in 0 (without which, we cannot estimate $f_0(0)$ consistently), we have that $F_0'(0) = 0$. We are interested in how F_0 behaves close to 0, and, therefore, we define the functions ψ_l and ψ_r by

$$\psi_r(s) = \limsup_{t\downarrow 0} \frac{F_0(st)}{F_0(t)} \quad \text{and} \quad \psi_l(s) = \limsup_{t\uparrow 0} \frac{F_0(st)}{F_0(t)} \qquad (s \in [0,1]). \tag{2.1}$$

If $F_0(t) = 0$ for some $t > 0$, we define $\psi_r(s) = 0$ for $s \in [0, 1)$ and $\psi_r(1) = 1$, and likewise for ψ_l. In this section, we will take a closer look at the functions ψ_r and ψ_l defined in (2.1). We will concentrate on ψ_r, since completely analogous statements will hold for ψ_l. Since the function $s \mapsto F_0(st)$ is convex and increasing for all $t > 0$, we get that $\psi_r(s)$ is also an increasing and convex function on $[0, 1]$ (this is true for the lim sup of convex functions, not necessarily for the lim inf). Furthermore, we clearly have that $\psi_r(0) = 0$ and $\psi_r(1) = 1$. Finally, since F_0 is convex, we know that for $s \in [0, 1]$,

$$F_0(st) \le sF_0(t) + (1-s)F_0(0) = sF_0(t).$$

This shows that for any F_0 we have that $\psi_r(s) \le s$.

Lemma 2.1. *For each $\tau \in [0, 1)$, there exists a positive continuous non-decreasing function η with*

$$F_0(t) = 0 \quad \Rightarrow \quad \eta(t) = 0 \qquad (\forall t \in [0,1]),$$

such that for all $t \in (0, 1]$ and for all $s \in [0, \tau]$

$$F_0(st) \le \bigl(\psi_r(s) + \eta(t)\bigr)F_0(t).$$

Proof. Suppose $F_0(t) > 0$ for all $t > 0$. Define the auxiliary functions

$$G_t(s) = \sup_{u\le t} \frac{F_0(su)}{F_0(u)}.$$

These functions are all convex and they decrease pointwise to ψ_r on $[0, 1]$. Since $G_t(0) = 0$ for all $t > 0$, we conclude that G_t converges uniformly to ψ_r on $[0, \tau]$. Define

$$\eta(0) = 0 \quad \text{and} \quad \eta(t) = \sup_{s\in[0,\tau]} |\psi_r(s) - G_t(s)| \qquad (t \in (0,1])$$

and note that

$$\frac{F_0(st)}{F_0(t)} \le G_t(s) \le \psi_r(s) + \eta(t),$$

to conclude the statement of the lemma (note that ψ_r is continuous on $[0, \tau]$, so η is indeed a continuous increasing function). Now suppose that $F_0(t) = 0$ for some $t > 0$. We defined $\psi_r(s) = 0$ for $s \in [0, 1)$ in this case. Define

$$r_0 = \sup\{t \in [0,1]\colon\ F_0(t) = 0\}.$$

If $r_0 = 1$, then the statement of the lemma holds with $\eta = 0$. Suppose $r_0 < 1$. Since $s \le \tau$, we have for each $t \le r_0/\tau$, $F_0(st) = 0$. Define $\eta(t) = 0$ for $t \in [0, r_0]$, $\eta(t) = 1$ for $t \in [r_0/\tau, 1]$, and continuous in between, and the statement of the lemma holds trivially. □

Let $\{W_s\colon\ s \in \mathbb{R}\}$ be a two-sided Brownian motion. We will encounter the probabilities in the next lemma throughout the rest of the paper.

Lemma 2.2. *For any F_0 we have that*

$$\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le 1} W_s - C\big(s - \psi_r(s)\big)\Big) \le \frac{1}{\sqrt{2\pi}C}.$$

If there exists $s \in (0, 1)$ such that $\psi_r(s) < s$, then there exist $\tau \in (0, 1)$ and $\rho \in (0, 1]$ such that

$$\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le 1} W_s - C\big(s - \psi_r(s)\big)\Big) \le \sqrt{\frac{2}{\pi\tau}} \frac{1}{C\rho(2-\rho)} \mathrm{e}^{-C^2\tau\rho^2/2}.$$

Proof. First note that the left-hand side and the right-hand side of two-sided Brownian motion are independent. It is, therefore, enough to consider the two sides within the probability separately. It is well known that

$$\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le -v\Big) = \mathbb{P}\Big(\sup_{s\ge 0} W_s - Cs \ge v\Big) = \mathrm{e}^{-2Cv}.$$

This follows from the hitting time of a linear boundary. Since, for all F_0, we have that $\psi_r(s) \le s$, we also need that

$$\mathbb{P}\Big(\inf_{0\le s\le 1} W_s \le -w\Big) = \mathbb{P}\Big(\sup_{0\le s\le 1} W_s \ge w\Big) = 2\big(1 - \Phi(w)\big),$$

where Φ is the distribution function of the standard normal distribution. We get

$$\begin{aligned}
\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le 1} W_s - C\big(s - \psi_r(s)\big)\Big) &\le \mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le 1} W_s\Big) \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{-2Cw}\mathrm{e}^{-w^2/2}\,\mathrm{d}w \\
&= 2\mathrm{e}^{2C^2}\big(1 - \Phi(2C)\big) \\
&\le \frac{1}{\sqrt{2\pi}C}.
\end{aligned}$$

Now suppose that for some $s \in (0, 1)$, $\psi_r(s) < s$. Since ψ_r is convex, $\psi_r(0) = 0$ and $\psi_r(1) = 1$, this implies that for any $\tau \in (0, 1)$ and any $s \in (0, \tau]$, $\psi_r(s) \le s\psi_r(\tau)/\tau < s$. Choose $\tau \in (0, 1)$ and define $\rho = 1 - \psi_r(\tau)/\tau > 0$. Then

$$\forall s \in [0, \tau]\colon\ s - \psi_r(s) \ge \rho s.$$

Now use that

$$\begin{aligned}\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le 1} W_s - C\big(s-\psi_r(s)\big)\Big) &\le \mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le \inf_{0\le s\le \tau} W_s - C\rho s\Big)\\ &\le \mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le W_\tau - C\rho\tau\Big).\end{aligned}$$

This last probability we can calculate exactly:

$$\begin{aligned}&\mathbb{P}\Big(\inf_{s\le 0} W_s - Cs \le W_\tau - C\rho\tau\Big)\\ &= 1-\Phi\big(C\sqrt{\tau}\rho\big) + \frac{1}{\sqrt{2\pi\tau}}\int_0^\infty \mathrm{e}^{-2Cv}\mathrm{e}^{-(v-C\tau\rho)^2/(2\tau)}\,\mathrm{d}v\\ &= 1-\Phi\big(C\sqrt{\tau}\rho\big) + \big(1-\Phi\big(C\sqrt{\tau}(2-\rho)\big)\big)\mathrm{e}^{C^2\tau(2-\rho)^2/2}\mathrm{e}^{-C^2\tau\rho^2/2}\\ &\le \sqrt{\frac{2}{\pi\tau}}\frac{1}{C\rho(2-\rho)}\mathrm{e}^{-C^2\tau\rho^2/2}.\end{aligned} \tag{2.2}$$

□

We will now consider the case where $F_0(st)/F_0(t)$ actually has a limit. This is comparable to saying that F_0 is a regularly varying function in 0, but we have the extra information that F_0 is convex.

Lemma 2.3. *Suppose for each $s \in (0,1]$ we have $F_0(s) > 0$ and*

$$\lim_{t\downarrow 0}\frac{F_0(st)}{F_0(t)} = \psi_r(s).$$

Then either $\psi_r(s) = 0$ on $[0,1)$, or $\psi_r(s) = s^\alpha$, for some $\alpha \ge 1$. In the latter case, we have that for each $\tau > 0$ (also for $\tau \ge 1$)

$$\sup_{s\in[0,\tau]}\left(\frac{F_0(st)}{F_0(t)} - s^\alpha\right) \xrightarrow{t\downarrow 0} 0.$$

Proof. Suppose $0 \le u \le s \le 1$. Then

$$\psi_r(us) = \lim_{t\downarrow 0}\frac{F_0(ust)}{F_0(t)} = \lim_{t\downarrow 0}\frac{F_0(ust)}{F_0(st)}\frac{F_0(st)}{F_0(t)} = \psi_r(u)\psi_r(s).$$

Since ψ_r is continuous and convex on $[0,1)$ and $\psi_r(0) = 0$, we conclude that either $\psi_r(s) = 0$ on $[0,1)$ or $\psi_r(s) = s^\alpha$ with $\alpha \ge 1$. In this last case, choose $s > 1$. Then

$$\lim_{t\downarrow 0}\frac{F_0(st)}{F_0(t)} = \left(\lim_{t\downarrow 0}\frac{F_0(t)}{F_0(st)}\right)^{-1} = \left(\lim_{t\downarrow 0}\frac{F_0(s^{-1}t)}{F_0(t)}\right)^{-1} = s^\alpha.$$

The family of convex functions $\{s \mapsto F_0(st)/F_0(t)\}$ converges pointwise to the convex function $s \mapsto s^\alpha$, and all functions are 0 in 0, so the convergence is actually uniform on compact subsets of $[0,\infty)$. □

3. Local asymptotic minimax optimality

Our data $Y(t)$ satisfies

$$\mathrm{d}Y(t) = f_0(t)\,\mathrm{d}t + n^{-1/2}\,\mathrm{d}W(t),$$

where f_0 is a monotone L^2-function on $[-1,1]$ and $W(t)$ is standard two-sided Brownian motion. We wish to study the least-squares estimator, but we will define for a realization of $W(t)$,

$$Y(t) = \int_0^t f_0(t)\,\mathrm{d}t + n^{-1/2}W(t),$$

and the convex function

$$\hat{F}(t) = \sup\{\phi(t)\colon \phi \text{ affine and } \forall s \in [-1,1]\colon \phi(s) \le Y(s)\}.$$

So $\hat{F}$ is the greatest convex minorant of Y. Now we define the estimator $\hat{f}$ as the left-derivative of the convex function $\hat{F}$, so for $t \in (-1,1)$

$$\hat{f}(t) = \lim_{h\downarrow 0} \frac{\hat{F}(t) - \hat{F}(t-h)}{h}.$$

This is a monotone function and can be seen as a limit of least-squares estimators over the class of monotone functions absolutely bounded by M, as $M \to \infty$.

As stated before, we will assume without loss of generality that $f_0(0) = 0$. Furthermore, to ensure that our estimator $\hat{f}(0)$ is consistent as $n \to \infty$, we assume that f_0 is continuous at 0. We are interested in the probability of the event $\{\hat{f}(0) \ge a\}$, for $a > 0$. Remember that

$$F_0(t) = \int_0^t f_0(s)\,\mathrm{d}s.$$

Fix $C > 0$ not depending on n, and choose $a, b > 0$ and $r_a, r_b > 0$ such that

$$F_0(r_a) = ar_a, \qquad F_0(-r_b) = br_b \quad \text{and} \quad r_a^{1/2}a = r_b^{1/2}b = Cn^{-1/2}. \tag{3.1}$$

Since F_0 is convex and continuous, and f_0 is continuous at 0, this can always be done if $F_0(1) > 0$ and $F_0(-1) > 0$, simply by choosing n large enough. We will consider the special (and simpler) case $f_0(t) = 0$ for all $t > 0$ (or for all $t < 0$) separately.

The sequence $\max(a,b) = \max(a(n), b(n))$ will constitute the rate of the least-squares estimator $\hat{f}(0)$; see Theorem 3.1 below. We would, therefore, like to give some feel for equations (3.1). Suppose f_0 is Lipschitz continuous at 0 with parameter $\alpha > 0$, so for x in a neighbourhood of 0, we have (remember that $f_0(0) = 0$)

$$|f_0(x)| \lesssim |x|^\alpha.$$

Here, $g(x) \lesssim h(x)$ denotes that there exists a constant $M > 0$ such that $g(x) \le Mh(x)$ for all relevant x. Then $F_0(x) \lesssim |x|^{\alpha+1}$, so (3.1) gives us

$$ar_a \lesssim r_a^{\alpha+1}.$$

This means that $r_a^{-1} \lesssim a^{-1/\alpha}$. Together with the second equality for a in (3.1),

$$a \lesssim r_a^{-1/2} n^{-1/2},$$

this leads to

$$a \lesssim n^{-\alpha/(2\alpha+1)}.$$

For b we can derive the same bound. This corresponds to the rate found in [12]. Another interesting case is when $\lim_{a\to 0} r_a = r_0 > 0$. This means that f_0 is flat to the right of 0 on the interval $[0, r_0)$. Then

$$a \lesssim r_0^{-1/2} n^{-1/2},$$

so this corresponds to a parametric rate.

Theorem 3.1. *With the notations as above, we have that*

$$\limsup_{n\to\infty} \mathbb{P}\big(\hat{f}(0) \geq a\big) \leq \mathbb{P}\Big(\inf_{s\leq 0} W_s - Cs \leq \inf_{0<s\leq 1} W_s - C\big(s - \psi_r(s)\big)\Big) \leq \frac{1}{C\sqrt{2\pi}}$$

and

$$\limsup_{n\to\infty} \mathbb{P}\big(\hat{f}(0) \leq -b\big) \leq \mathbb{P}\Big(\inf_{s\leq 0} W_s - Cs \leq \inf_{0<s\leq 1} W_s - C\big(s - \psi_l(s)\big)\Big) \leq \frac{1}{C\sqrt{2\pi}}.$$

Since both probabilities tend to zero when $C \to \infty$, it follows that equations (3.1) determine an upper bound for the rate of convergence of $\hat{f}(0)$.

Proof. We will only show the result for a; the proof for b is similar. Note that we have the following "switch relation" for the greatest convex minorant:

$$\{\hat{f}(0) \geq a\} = \Big\{\inf_{-1\leq t\leq 0} \big(n^{-1/2} W_t + F_0(t) - at\big) \leq \inf_{0<t\leq 1} \big(n^{-1/2} W_t + F_0(t) - at\big)\Big\}. \tag{3.2}$$

We can rewrite (3.2) as follows:

$$\begin{aligned}\{\hat{f}(0) \geq a\} = \Big\{&\inf_{-r_a^{-1}\leq s\leq 0} \big(r_a^{-1/2} W_{r_a s} + n^{1/2} r_a^{-1/2} F_0(r_a s) - n^{1/2} r_a^{1/2} as\big)\\ &\leq \inf_{0<s\leq r_a^{-1}} \big(r_a^{-1/2} W_{r_a s} + n^{1/2} r_a^{-1/2} F_0(r_a s) - n^{1/2} r_a^{1/2} as\big)\Big\}.\end{aligned} \tag{3.3}$$

Define

$$\tilde{W}_s = r_a^{-1/2} W_{r_a s}.$$

Clearly, $\tilde{W}_s$ is also a two-sided Brownian motion. Now we can use Lemma 2.1: For any $\tau \in (0, 1)$ there exists a positive continuous function η with $F_0(t) = 0 \Rightarrow \eta(t) = 0$, such that

$$\inf_{0<s\leq r_a^{-1}} \big(r_a^{-1/2} W_{r_a s} + n^{1/2} r_a^{-1/2} F_0(r_a s) - n^{1/2} r_a^{1/2} as\big) \leq \inf_{0<s\leq\tau} \tilde{W}_s - C\big(s - \psi_r(s)\big) + C\eta(r_a).$$

Now remark that for $s < 0$, $F_0(s) \geq 0$, so that

$$\inf_{-r_a^{-1} \leq s \leq 0} \bigl(r_a^{-1/2} W_{r_a s} + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs\bigr) \geq \inf_{s \leq 0} (\tilde{W}_s - Cs).$$

In view of (3.3), we have shown that

$$\mathbb{P}\bigl(\hat{f}(0) \geq a\bigr) \leq \mathbb{P}\Bigl(\inf_{s \leq 0} (\tilde{W}_s - Cs) \leq \inf_{0 < s \leq \tau} \tilde{W}_s - C\bigl(s - \psi_r(s)\bigr) + C\eta(r_a)\Bigr). \tag{3.4}$$

Define $r_0 = \lim_{a \downarrow 0} r_a$. Since we always have that $F_0(r_0) = 0$, we conclude that $\lim_{a \downarrow 0} \eta(r_a) = 0$, so

$$\limsup_{n \to \infty} \mathbb{P}\bigl(\hat{f}(0) \geq a\bigr) \leq \mathbb{P}\Bigl(\inf_{s \leq 0} (\tilde{W}_s - Cs) \leq \inf_{0 < s \leq \tau} \tilde{W}_s - C\bigl(s - \psi_r(s)\bigr)\Bigr).$$

Since this is true for all $\tau \in (0, 1)$, and since ψ_r is increasing on $[0, 1]$, we conclude that

$$\limsup_{n \to \infty} \mathbb{P}\bigl(\hat{f}(0) \geq a\bigr) \leq \mathbb{P}\Bigl(\inf_{s \leq 0} (\tilde{W}_s - Cs) \leq \inf_{0 < s \leq 1} \tilde{W}_s - C\bigl(s - \psi_r(s)\bigr)\Bigr).$$

The final inequality of the theorem is given in Lemma 2.2.

When $f_0(t) = 0$ for all $t > 0$, we choose $a = Cn^{-1/2}$, and (3.2) implies that

$$\mathbb{P}\bigl(\hat{f}(0) \geq a\bigr) \leq \mathbb{P}\Bigl(\inf_{-1 \leq t \leq 0} (W_t - Ct) \leq \inf_{0 < t \leq 1} (W_t - Ct)\Bigr).$$

This shows that in this case, the upper confidence limit for $\hat{f}(0)$ is of order $n^{-1/2}$ (parametric rate). This also happens when $r_0 > 0$, which is the case when f_0 is flat to the right of 0. $\square$

We wish to show that the rate for the least-squares estimator is local asymptotic minimax; see the Introduction for a discussion of this concept. Remember that $\mathcal{F}_m$ denotes the space of all monotone increasing L^2-functions on $[-1, 1]$, and that $B(f_0; \varepsilon)$ denotes the L^2-ball around f_0 of radius ε.

Proof of Theorem 1.1. We will show the following stronger version of Theorem 1.1: Let $f_0 \in \mathcal{F}_m$ be continuous at 0. Choose two significance levels $\alpha \in (0, 1)$ and $\beta \in (0, 1/2)$. There exist $\delta > 0$ and $\eta > 0$, such that for all n large enough, we can find a rate γ_n with

$$\limsup_{n \to \infty} \sup_{f \in \mathcal{F}_m \cap B(f_0; \delta n^{-1/2})} \mathbb{P}_f\bigl(|\hat{f}(0) - f(0)| \geq \gamma_n\bigr) \leq \alpha \tag{3.5}$$

and a sequence of functions $f_n \in \mathcal{F}_m \cap B(f_0; \delta n^{-1/2})$ such that

$$\liminf_{n \to \infty} \inf_{\hat{\theta}} \max_{i = 0, n} \mathbb{P}_{f_i}\bigl(|\hat{\theta}(Y) - f_i(0)| \geq \eta \cdot \gamma_n\bigr) > \beta, \tag{3.6}$$

where $\hat{\theta}(Y)$ is any estimator of $f(0)$ based on the data Y.

Remark 1. For a general concept of local optimality, it might be more natural to replace the last statement by

$$\liminf_{n\to\infty} \inf_{\hat\theta} \sup_{f\in\mathcal{F}_m\cap B(f_0;\delta n^{-1/2})} \mathbb{P}_{f_i}\big(|\hat\theta(Y) - f_i(0)| \geq \eta\cdot\gamma_n\big) > \beta,$$

as was done in Theorem 1.1. Clearly, (3.6) is stronger, since the sequence f_n is the same for all possible estimators $\hat\theta$.

Remark 2. Choose an event $A \subset C([-1,1])$ such that $\mathbb{P}_{f_0}(Y\in A)\geq 1/2$ and $\mathbb{P}_{f_1}(Y\notin A)\geq 1/2$. Define the estimator

$$\hat\theta(Y) = f_0(0)1_A(Y) + f_1(0)1_{A^c}(Y).$$

Then for any choice of η and γ, we would have

$$\max_{i=0,1}\mathbb{P}_{f_i}\big(|\hat\theta(Y) - f_i(0)|\geq \eta\cdot\gamma\big)\leq \frac{1}{2},$$

which is why we choose $\beta\in(0,1/2)$.

Without loss of generalization, we can assume $f_0(0)=0$. Choose n large enough such that the equations

$$F_0(r_a) = ar_a, \qquad F_0(-r_b) = br_b \quad \text{and} \quad r_a^{1/2}a = r_b^{1/2}b = Cn^{-1/2}$$

have solutions for some fixed $C>0$ with

$$\frac{1}{2\sqrt{2\pi}C}\leq\alpha.$$

Define

$$\gamma_n = 4\max(a,b).$$

Fix $\delta\in(0,C]$ and suppose $f_1\in\mathcal{F}_m\cap B(f_0;\delta n^{-1/2})$. Define

$$\tilde F_1(t) = \int_0^t f_1(s)\,\mathrm{d}s.$$

Note that

$$\begin{aligned}\left|\frac{1}{r_a}\int_0^{r_a}\big(f_1(t)-f_0(t)\big)\,\mathrm{d}t\right| &\leq \sqrt{\frac{1}{r_a}\int_0^{r_a}\big(f_1(t)-f_0(t)\big)^2\,\mathrm{d}t}\\ &\leq \delta n^{-1/2}r_a^{-1/2}\\ &= \delta a/C.\end{aligned}$$

This means that

$$\tilde{F}_1(r_a) \le 2ar_a. \tag{3.7}$$

Analogously, we conclude that

$$\tilde{F}_1(-r_b) \le 2br_b.$$

Since

$$\tilde{F}_1(r_a) \ge f_1(0)r_a \quad \text{and} \quad \tilde{F}_1(-r_b) \ge -r_b f_1(0),$$

we conclude that

$$f_1(0) \le 2a \quad \text{and} \quad f_1(0) \ge -2b.$$

Now define

$$F_1(t) = \int_0^t \big(f_1(s) - f_1(0)\big)\,\mathrm{d}s.$$

Since $F_1(t) = \tilde{F}_1(t) - f_1(0)t$, we get

$$F_1(r_a) \le \gamma_n r_a \quad \text{and} \quad F_1(-r_b) \le \gamma_n r_b.$$

Now use equation (3.3) for the situation where the underlying function is f_1:

$$\begin{aligned}\{\hat{f}(0) - f_1(0) \ge \gamma_n\} = \Big\{ &\inf_{-r_a^{-1} \le s \le 0} \big(r_a^{-1/2} W_{r_a s} + n^{-1/2} r_a^{-1/2} F_1(r_a s) - n^{-1/2} r_a^{1/2} \gamma_n s\big) \\ &\le \inf_{0 < s \le r_a^{-1}} \big(r_a^{-1/2} W_{r_a s} + n^{-1/2} r_a^{-1/2} F_1(r_a s) - n^{-1/2} r_a^{1/2} \gamma_n s\big)\Big\}.\end{aligned}$$

Again we have that $F_1(s) \ge 0$ for $s \le 0$. Also, since F_1 is convex, $F_1(r_a s) \le \gamma_n r_a s$ for all $s \in [0, 1]$. Now we can follow the exact same steps as in the proof of Theorem 3.1, starting at equation (3.3) and realizing that $n^{-1/2} r_a^{1/2} \gamma_n \ge 4C$, to conclude that

$$\mathbb{P}_{f_1}\big(\hat{f}(0) - f_1(0) \ge \gamma_n\big) \le \frac{1}{4\sqrt{2\pi}C}.$$

Analogously, we can prove that

$$\mathbb{P}_{f_1}\big(\hat{f}(0) - f_1(0) \le -\gamma_n\big) \le \frac{1}{4\sqrt{2\pi}C}.$$

This clearly shows that

$$\mathbb{P}_{f_1}\big(|\hat{f}(0) - f_1(0)| > \gamma_n\big) \le \alpha.$$

So we have shown that our rate γ_n satisfies (3.5).

Now we choose our sequence f_n. For fixed n, suppose $a \geq b$; the case $b > a$ can be handled analogously by perturbing f_0 to the left of 0. Define

$$f_n(t) = \begin{cases} \delta a/C, & \text{if } t \geq 0 \text{ and } f_0(t) \leq \delta a/C, \\ f_0(t), & \text{otherwise.} \end{cases}$$

Then f_n is a monotone L^2-function. Note that f_n will be discontinuous at 0 and that $\gamma_n = 4a$. To show that $f_n \in B(f_0, \delta n^{-1/2})$, define $s_a = \inf\{t \geq 0\colon\ f_0(t) \geq a\}$. Clearly,

$$F_0(s_{\delta a/C}) \leq \delta a C^{-1} s_{\delta a/C}.$$

Since $F_0(r_a) = a r_a$ and F_0 is convex, we would get for $s > r_a$:

$$F_0(s) > as \geq \delta a C^{-1} s.$$

This proves that $s_{\delta a/C} \leq r_a$. Therefore,

$$\int_{-1}^{1} \big(f_n(t) - f_0(t)\big)^2 \,\mathrm{d}t \leq \delta^2 C^{-2} a^2 s_{\delta a/C} \leq \delta^2 C^{-2} a^2 r_a \leq \delta^2 n^{-1}.$$

Define μ as the probability measure on $C([-1, 1])$ that corresponds to standard two-sided Brownian motion, and denote with P_0 and P_n the measures corresponding to the model with f_0 and f_n, respectively. It is well known that

$$\frac{\mathrm{d}P_i}{\mathrm{d}\mu}(W) = \exp\left(n^{1/2} \int f_i(t)\,\mathrm{d}W(t) - \frac{1}{2} n \int f_i(t)^2 \,\mathrm{d}t\right).$$

Therefore,

$$\frac{\mathrm{d}P_n}{\mathrm{d}P_0}(W) = \exp\left(n^{1/2} \int \big(f_n(t) - f_0(t)\big)\,\mathrm{d}W(t) - \frac{1}{2} n \int f_n(t)^2 \,\mathrm{d}t + \frac{1}{2} n \int f_0(t)^2 \,\mathrm{d}t\right).$$

This means that

$$\begin{aligned}
\|P_n - P_0\|_1^2 &\leq \mathbb{E}_{P_0}\left(\frac{\mathrm{d}P_n}{\mathrm{d}P_0}(W) - 1\right)^2 = \mathbb{E}_{P_0}\left(\frac{\mathrm{d}P_n}{\mathrm{d}P_0}(W)\right)^2 - 1 \\
&= \mathbb{E}_{\mu}\Bigg(\exp\Bigg(n^{1/2} \int \big(2 f_n(t) - f_0(t)\big)\,\mathrm{d}W(t) \\
&\qquad\qquad - n \int f_n(t)^2 \,\mathrm{d}t + \frac{1}{2} n \int f_0(t)^2 \,\mathrm{d}t\Bigg)\Bigg) - 1 \\
&= \exp\left(\frac{1}{2} n \int \big(2 f_n(t) - f_0(t)\big)^2 \,\mathrm{d}t - n \int f_n(t)^2 \,\mathrm{d}t + \frac{1}{2} n \int f_0(t)^2 \,\mathrm{d}t\right) - 1 \\
&= \exp\left(n \int \big(f_n(t) - f_0(t)\big)^2 \,\mathrm{d}t\right) - 1.
\end{aligned}$$

Since

$$\int \big(f_1(t) - f_0(t)\big)^2 \,\mathrm{d}t \le \delta^2 n^{-1},$$

we conclude that

$$\|P_1 - P_0\| \le \sqrt{\exp(\delta^2) - 1}.$$

Choose $\delta \in (0, C]$ small enough, such that $\|P_1 - P_0\| < 2 - 4\beta$. Choose $\eta = \delta/8C$. Denote with p_i the density of P_i with respect to μ $(i = 0, 1)$. We have that for any estimator $\hat\theta$

$$\begin{aligned}
\max_{i=0,n} \mathbb{P}_{f_i}\big(|\hat\theta(Y) - f_i(0)| \ge 4\eta a\big) &\ge \frac{1}{2}\sum_{i=0,n} \mathbb{P}_{f_i}\big(|\hat\theta(Y) - f_i(0)| \ge 4\eta a\big) \\
&= \frac{1}{2}\mathbb{E}_\mu\big(1_{\{|\hat\theta(Y)| \ge 4\eta a\}} p_0(W) + 1_{\{|\hat\theta(Y) - \delta a/C| \ge 4\eta a\}} p_1(W)\big) \\
&\ge \frac{1}{2}\mathbb{E}_\mu(\min(p_0(W), p_1(W))) \\
&= \frac{1}{2}\left(1 - \frac{1}{2}\|P_1 - P_0\|\right) \\
&> \beta.
\end{aligned}$$

This proves (3.6). □

4. Limiting distribution of the least-squares estimator

Our methods also allow us to derive non-standard limiting distributions for the least-squares estimator. These limiting distributions only exist when f_0 is somehow "regular" near 0. The precise conditions are described in the following theorem and will use Lemma 2.3. We start with the rate equations: For $n > 0$ and $C > 0$, we define a, r_a, b and r_b by

$$F_0(r_a) = a r_a, \qquad F_0(-r_b) = b r_b \quad \text{and} \quad r_a^{1/2} a = r_b^{1/2} b = C n^{-1/2}.$$

Theorem 4.1. *Suppose that*

$$\lim_{n\to\infty} \frac{r_a}{r_b} = \gamma, \tag{4.1}$$

with $\gamma \in [0, \infty)$. *Furthermore, suppose that for* $s \ge 0$,

$$\lim_{t\downarrow 0} \frac{F_0(st)}{F_0(t)} = s^\alpha \tag{4.2}$$

for $\alpha > 1$ (*see also Lemma* 2.3). *Then, if* W_s $(s \in \mathbb{R})$ *denotes two-sided standard Brownian motion*,

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{f}(0) > a\big) = \mathbb{P}\Big(\inf_{s\le 0}(W_s + C\gamma^{\alpha-1/2}|s|^\alpha - Cs) \le \inf_{s\ge 0}(W_s + C|s|^\alpha - Cs)\Big).$$

If $\lim_{n\to\infty} r_a/r_b = +\infty$, *then*

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{f}(0) > 0\big) = 0.$$

Remark. Condition (4.1) says that the rates to the left and the right of 0 are well behaved with respect to each other, which is a natural condition for a limiting distribution to exist. Furthermore, Condition (4.2) says that F_0 scales properly near 0, which is another natural condition. We do not want different behavior of F_0 for different scales.

Proof of Theorem 4.1. We start with assuming that $\gamma > 0$. Since $r_a/r_b \to \gamma$ and $ar_a^{1/2} = br_b^{1/2}$, we see that $a/b \to \gamma^{-1/2}$ and $F_0(r_a)/F_0(-r_b) \to \gamma^{1/2}$ (since $F_0(r_a) = ar_a$ and $F_0(-r_b) = br_b$). For each $\eta > 0$, we have that $(\gamma - \eta)r_b \le r_a \le (\gamma + \eta)r_b$, for n large enough. Therefore,

$$\limsup_{n\to\infty} \frac{F_0(r_a)}{F_0(\gamma r_b)} \le \lim_{n\to\infty} \frac{F_0((\gamma+\eta)r_b)}{F_0(\gamma r_b)} = \left(\frac{\gamma+\eta}{\gamma}\right)^\alpha.$$

We used that, since $\psi_r(s) > 0$ for $s \in (0,1]$, we have that $r_a \to 0$ and $r_b \to 0$. The inequality holds for all $\eta > 0$, and we can show a similar inequality for the lim inf, which means that

$$\lim_{n\to\infty} \frac{F_0(r_a)}{F_0(\gamma r_b)} = 1.$$

Since r_a and r_b are decreasing continuous functions of n, we have shown that

$$\lim_{t\downarrow 0} \frac{F_0(\gamma t)}{F_0(-t)} = \gamma^{1/2}.$$

This implies that

$$\lim_{t\downarrow 0} \frac{F_0(-st)}{F_0(-t)} = \lim_{t\downarrow 0} \frac{F_0(+\gamma st)}{F_0(+\gamma t)} = s^\alpha.$$

So the rescaled behavior of F_0 to the left of zero is equal to the behavior of F_0 to the right of zero.

The rest of the proof is based on equation (3.3):

$$\mathbb{P}\big(\hat{f}(0) \ge a\big) = \mathbb{P}\Big(\inf_{-r_a^{-1}\le s\le 0}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs\big)$$
$$\le \inf_{0<s\le r_a^{-1}}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs\big)\Big).$$

Here, W_s is two-sided Brownian motion. Note that we can rewrite this equation as

$$\mathbb{P}\big(\hat{f}(0) \geq a\big) = \mathbb{P}\Big(\operatorname*{arg\,min}_{s \in [-r_a^{-1}, r_a^{-1}]} \big(W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs\big) \leq 0\Big).$$

Using Lemma 2.3, we conclude that there exists a family of functions $\eta_t(s)$ on $[0, \infty)$, such that $\eta_t \to 0$ uniformly on compacta as $t \to 0$, with

$$F_0(st) = s^\alpha F_0(t) + \eta_t(s) F_0(t) (t \in \mathbb{R}).$$

This shows that for $s \in [0, \infty)$, we have

$$\begin{aligned} W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs &= W_s + Cs^\alpha + C\eta_t(s) - Cs \\ &\to W_s + Cs^\alpha - Cs, \end{aligned}$$

uniformly on compacta. For $s \in (-\infty, 0]$, we have to be a bit more careful:

$$\begin{aligned} n^{1/2} r_a^{-1/2} F_0(r_a s) &= n^{1/2} r_a^{-1/2} \left(\frac{r_a}{r_b}\right)^\alpha |s|^\alpha F_0(-r_b) + n^{1/2} r_a^{-1/2} \eta_t(|s| r_a / r_b) F_0(-r_b) \\ &= \left(\frac{r_a}{r_b}\right)^{\alpha - 1/2} |s|^\alpha n^{1/2} r_b^{1/2} b + \left(\frac{r_a}{r_b}\right)^{-1/2} \eta_t(|s| r_a / r_b) n^{1/2} r_b^{1/2} b \\ &\to C\gamma^{\alpha - 1/2} |s|^\alpha, \end{aligned}$$

uniformly on compacta. We have shown that uniformly on compacta

$$W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs \to \begin{cases} W_s + Cs^\alpha - Cs, & \text{for } s \geq 0, \\ W_s + C\gamma^{\alpha - 1/2} |s|^\alpha - Cs, & \text{for } s \leq 0. \end{cases}$$

Now we wish to use Theorem 2.7 from [10], page 198. This theorem implies that the location of the minimum of the process $W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs$ converges in distribution to the location of the minimum of its limiting process, provided that this location is $O_p(1)$. To show this last condition, we consider for $M > 1$

$$\begin{aligned} &\mathbb{P}\Big(\operatorname*{arg\,min}_{s \in [-r_a^{-1}, r_a^{-1}]} \big(W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs\big) > M\Big) \\ &\quad \leq \mathbb{P}\Big(\inf_{s \geq M} \big(W_s + n^{1/2} r_a^{-1/2} F_0(r_a s) - Cs\big) < 0\Big). \end{aligned}$$

Now we use that for n large enough, $F_0(M r_a) \geq M^\alpha F_0(r_a) - F_0(r_a)$. So for $s \geq M$, using convexity of F_0, we get

$$F_0(s r_a) \geq s F_0(M r_a)/M \geq M^{\alpha - 1} a r_a s - a r_a s / M.$$

Using this, we get

$$\mathbb{P}\Big(\inf_{s\geq M}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs\big) < 0\Big) \leq \mathbb{P}\Big(\inf_{s\geq M}(W_s + CM^{\alpha-1}s - Cs/M - Cs) < 0\Big).$$

Clearly, this last probability goes to zero exponentially fast as $M \to +\infty$, since $\alpha > 1$. Now we have to check the lower bound for the location of the minimum:

$$\begin{aligned}
&\mathbb{P}\Big(\operatorname*{arg\,min}_{s\in[-r_a^{-1},r_a^{-1}]}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs\big) < -M\Big)\\
&\leq \mathbb{P}\Big(\inf_{s\leq -M}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs\big) < 0\Big)\\
&\leq \mathbb{P}\Big(\inf_{s\leq -M}(W_s - Cs) < 0\Big).
\end{aligned}$$

This last probability again goes to zero exponentially fast as $M \to +\infty$. This proves the theorem for $\gamma > 0$. When $\gamma = 0$, so $r_a/r_b \to 0$, the above reasoning goes through, except for the convergence of the process $W_s + n^{1/2}r_a^{-1/2}F_0(r_a s) - Cs$ for $s \in (-\infty, 0]$. We need to show that

$$n^{1/2}r_a^{-1/2}F_0(r_a s) \to 0,$$

uniformly on compact subsets of $(-\infty, 0]$. Fix a compact set $[-M, 0]$ and choose n so large that $Mr_a \leq r_b$. Then, for all $s \in [-M, 0]$,

$$\begin{aligned}
|n^{1/2}r_a^{-1/2}F_0(r_a s)| &\leq n^{1/2}r_a^{-1/2}F_0(-r_b)|s|\frac{r_a}{r_b}\\
&\leq C\left(\frac{r_a}{r_b}\right)^{1/2}M\\
&\to 0.
\end{aligned}$$

Finally we need to prove the last statement. For this, we directly use equation (3.2):

$$\mathbb{P}\big(\hat{f}(0) \geq 0\big) = \mathbb{P}\Big(\inf_{-1\leq t\leq 0}\big(n^{-1/2}W_t + F_0(t)\big) \leq \inf_{0<t\leq 1}\big(n^{-1/2}W_t + F_0(t)\big)\Big).$$

Now we take the usual rescaling, replacing t with $r_a s$ and multiplying with $r_a^{-1/2}$:

$$\mathbb{P}\big(\hat{f}(0) \geq 0\big) = \mathbb{P}\Big(\inf_{-r_a^{-1}\leq s\leq 0}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s)\big) \leq \inf_{0<s\leq r_a^{-1}}\big(W_s + n^{1/2}r_a^{-1/2}F_0(r_a s)\big)\Big).$$

Choose n large such that $r_a \geq r_b$. Then if $s \leq -r_b/r_a$, we have $F_0(r_a s) \geq |s|F_0(-r_b)r_a/r_b$, whereas if $-r_b/r_a \leq s \leq 0$, we still have that $F_0(r_a s) \geq 0$, so

$$\begin{aligned}
\mathbb{P}\big(\hat{f}(0) \geq 0\big) \leq{}& \mathbb{P}\left(\inf_{-r_a^{-1}\leq s\leq -r_b/r_a}\left(W_s + \left(\frac{r_a}{r_b}\right)^{1/2}C|s|\right) \leq \inf_{0<s\leq 1} W_s + Cs\right)\\
&+ \mathbb{P}\Big(\inf_{-r_b/r_a\leq s\leq 0} W_s \leq \inf_{0<s\leq 1} W_s + Cs\Big).
\end{aligned}$$

Since $r_a/r_b \to +\infty$, these two probabilities clearly go to zero, since $\inf_{0\le s\le 1} W_s + Cs < 0$ with probability 1. Note that for this last result, we do not need any other assumptions on F_0. □

We introduce the auxiliary function G_0 and H_0 on a full neighborhood of 0: fix $\delta > 0$ and for $t \in (-\delta, \delta)$

$$G_0(t) = F_0(t)/t \quad \text{and} \quad H_0(t) = t\sqrt{|G_0^{-1}(t)|}.$$

We have that both G_0 and H_0 are strictly increasing functions on $(-\delta, \delta)$. We also know that the rate equations (3.1) imply that

$$H_0(a) = Cn^{-1/2} \quad \text{and} \quad H_0(-b) = -Cn^{-1/2}.$$

Corollary 4.2. *Suppose Conditions* (4.1) *and* (4.2). *If* W_s $(s \in \mathbb{R})$ *denotes two-sided standard Brownian motion*, *define the process*

$$X(s) = \begin{cases} W_s + s^\alpha, & \text{for } s \ge 0, \\ W_s + \gamma^{\alpha-1/2}|s|^\alpha, & \text{for } s \le 0, \end{cases}$$

and the process $\hat{X}(s)$ *as the greatest convex minorant of* X. *Then*

$$n^{1/2} H_0(\hat{f}(0)) \xrightarrow{\text{d}} \operatorname{sgn}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right)\left|\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right|^{(2\alpha-1)/(2\alpha-2)}.$$

Here, $\operatorname{sgn}(x)$ *denotes the sign of* $x \in \mathbb{R}$.

Proof. We start by considering $\mathbb{P}(n^{1/2}H_0(\hat{f}(0)) \ge C)$, for $C > 0$. We get

$$\begin{aligned} \mathbb{P}\big(n^{1/2}H_0(\hat{f}(0)) \ge C\big) &= \mathbb{P}\big(\hat{f}(0) \ge a\big) \\ &\to \mathbb{P}\Big(\inf_{s\le 0}(W_s + C\gamma^{\alpha-1/2}|s|^\alpha - Cs) \le \inf_{s\ge 0}(W_s + C|s|^\alpha - Cs)\Big), \end{aligned}$$

according to Theorem 4.1. Now replace s by $C^{2/(1-2\alpha)}s$, multiply left and right by $C^{-1/(1-2\alpha)}$ and use Brownian scaling to get

$$\begin{aligned} &\mathbb{P}\big(n^{1/2}H_0(\hat{f}(0)) \ge C\big) \\ &\quad \to \mathbb{P}\Big(\inf_{s\le 0}\big(W_s + \gamma^{\alpha-1/2}|s|^\alpha - C^{(2\alpha-2)/(2\alpha-1)}s\big) \le \inf_{s\ge 0}\big(W_s + |s|^\alpha - C^{(2\alpha-2)/(2\alpha-1)}s\big)\Big). \end{aligned}$$

Using the switch relation for the greatest convex minorant, we see that

$$\begin{aligned} \mathbb{P}\big(n^{1/2}H_0(\hat{f}(0)) \ge C\big) &\to \mathbb{P}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0) \ge C^{(2\alpha-2)/(2\alpha-1)}\right) \\ &= \mathbb{P}\left(\operatorname{sgn}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right)\left|\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right|^{(2\alpha-1)/(2\alpha-2)} \ge C\right). \end{aligned}$$

When $\gamma = 0$, the proof is finished, since in that case

$$\mathbb{P}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0) \geq 0\right) = 1.$$

Now suppose $\gamma > 0$. We have seen in the proof of Theorem 4.1 that the scaling of F_0 to the left of 0 is the same as the scaling to the right, so for all $s \geq 0$

$$\lim_{t\to 0} \frac{F_0(st)}{F_0(t)} = s^{\alpha}.$$

Consider for $C > 0$

$$\begin{aligned}
\mathbb{P}\big(n^{1/2} H_0(\hat{f}(0)) \leq -C\big) &= \mathbb{P}\big(\hat{f}(0) \leq -b\big) \\
&\to \mathbb{P}\Big(\inf_{s\leq 0}(W_s + C\gamma^{-\alpha+1/2}|s|^{\alpha} - Cs) \leq \inf_{s\geq 0}(W_s + C|s|^{\alpha} - Cs)\Big),
\end{aligned}$$

using Theorem 4.1 for the left-hand side of the origin (that is, interchange a and b and replace γ with $1/\gamma$). Now replace s with $-\gamma C^{2/(1-2\alpha)}s$, multiply left and right by $\gamma^{-1/2}C^{-1/(1-2\alpha)}$ and use Brownian scaling to get

$$\begin{aligned}
&\mathbb{P}\big(n^{1/2} H_0(\hat{f}(0)) \leq -C\big) \\
&\quad \to \mathbb{P}\Big(\inf_{s\geq 0}\big(W_s + |s|^{\alpha} + C^{(2\alpha-2)/(2\alpha-1)}s\big) \leq \inf_{s\leq 0}\big(W_s + \gamma^{\alpha-1/2}|s|^{\alpha} + C^{(2\alpha-2)/(2\alpha-1)}s\big)\Big).
\end{aligned}$$

Note that the two infima have switched sides because of the scaling with a negative constant. Again, using the switch relation, we get

$$\begin{aligned}
\mathbb{P}\big(n^{1/2} H_0(\hat{f}(0)) \leq -C\big) &\to \mathbb{P}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0) \leq -C^{(2\alpha-2)/(2\alpha-1)}\right) \\
&= \mathbb{P}\left(\operatorname{sgn}\left(\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right)\left|\frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)\right|^{(2\alpha-1)/(2\alpha-2)} \leq -C\right).
\end{aligned}$$

This proves the corollary. □

The condition that F_0 is regularly varying around 0 with parameter $\alpha > 1$ implies that the function H_0 is regularly varying around 0 with parameter $\beta = (2\alpha - 1)/(2\alpha - 2)$; so for all $s \geq 0$

$$\lim_{t\to 0} \frac{H_0(st)}{H_0(t)} = s^{\beta}.$$

It is well known from the theory of regularly varying functions that this limit is uniform for $s \in [1/M, M]$, for any $M > 1$. This will help us prove the next corollary.

Corollary 4.3. *With the conditions and notations from Corollary 4.2, we can show that*

$$\frac{\hat{f}(0)_+}{H_0^{-1}(n^{-1/2})} \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)_+ \quad \textit{and} \quad \frac{\hat{f}(0)_-}{-H_0^{-1}(-n^{-1/2})} \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)_-.$$

Proof. We wish to show that

$$\frac{H_0(\hat{f}(0))}{n^{-1/2}} \cdot \left| \frac{H_0^{-1}(\mathrm{sgn}(\hat{f}(0))n^{-1/2})}{\hat{f}(0)} \right|^{\beta} \mathrm{sgn}(\hat{f}(0)) \to 1 \qquad \text{in probability.} \tag{4.3}$$

Suppose $\eta > 0$. Using Corollary 4.2, there exists $M > 1$ such that for all n large enough

$$\mathbb{P}\big(H_0(\hat{f}(0)) \in [-n^{-1/2}M, -n^{-1/2}/M] \cup [n^{-1/2}/M, n^{-1/2}M]\big) \geq 1 - \eta.$$

If $H_0(\hat{f}(0)) \in [n^{-1/2}/M, n^{-1/2}M]$, we know that

$$\frac{H_0^{-1}(n^{-1/2}/M)}{H_0^{-1}(n^{-1/2})} \leq \frac{\hat{f}(0)}{H_0^{-1}(n^{-1/2})} \leq \frac{H_0^{-1}(n^{-1/2}M)}{H_0^{-1}(n^{-1/2})}.$$

Since H_0^{-1} is regularly varying around 0 with parameter $1/\beta$, we then know that for n large enough,

$$\frac{1}{2}M^{-1/\beta} \leq \frac{\hat{f}(0)}{H_0^{-1}(n^{-1/2})} \leq 2M^{1/\beta}.$$

A similar reasoning shows that if $H_0(\hat{f}(0)) \in [-n^{-1/2}/M, -n^{-1/2}M]$, then for n large enough,

$$\frac{1}{2}M^{-1/\beta} \leq \frac{\hat{f}(0)}{H_0^{-1}(-n^{-1/2})} \leq 2M^{1/\beta}.$$

Now consider

$$\begin{aligned} n^{1/2}H_0(\hat{f}(0)) = \mathrm{sgn}(\hat{f}(0))H_0\left(H_0^{-1}(\mathrm{sgn}(\hat{f}(0))n^{-1/2})\frac{\hat{f}(0)}{H_0^{-1}(\mathrm{sgn}(\hat{f}(0))n^{-1/2})}\right) \\ \Big/ H_0(H_0^{-1}(\mathrm{sgn}(\hat{f}(0))n^{-1/2})). \end{aligned}$$

Since $H_0(st)/H_0(t) \to s^{\beta}$ uniform for s in compact subsets of $(0, \infty)$, we can conclude with probability higher than $1 - \eta$, that for n large enough,

$$\left| n^{1/2}H_0(\hat{f}(0)) - \mathrm{sgn}(\hat{f}(0))\left(\frac{\hat{f}(0)}{H_0^{-1}(\mathrm{sgn}(\hat{f}(0))n^{-1/2})}\right)^{\beta} \right| < \eta/M$$

and

$$|n^{1/2}H_0(\hat{f}(0))| \geq 1/M.$$

This proves (4.3). Corollary 4.2 then immediately shows that

$$\frac{\operatorname{sgn}(\hat{f}(0))\hat{f}(0)}{H_0^{-1}(\operatorname{sgn}(\hat{f}(0))n^{-1/2})} \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0).$$

This can be written in a nicer way when we look at $\hat{f}(0)_+$ and $\hat{f}(0)_-$:

$$\frac{\hat{f}(0)_+}{H_0^{-1}(n^{-1/2})} \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)_+$$

and

$$\frac{\hat{f}(0)_-}{-H_0^{-1}(-n^{-1/2})} \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0)_-. \qquad \square$$

Suppose f_0 is differentiable in 0 with $f_0'(0) > 0$. Then

$$\frac{F_0(st)}{F_0(t)} = \frac{s^2t^2 f_0'(0)/2 + \mathrm{o}(t^2)}{t^2 f_0'(0)/2 + \mathrm{o}(t^2)} \to s^2 \qquad (t \to 0).$$

Furthermore, $G_0(t) = F_0(t)/t = \frac{1}{2} f_0'(0)t + \mathrm{o}(t)$, which implies that $G_0^{-1}(t) = 2 f_0'(0)^{-1}t + \mathrm{o}(t)$, so

$$H_0(t) = \sqrt{2} f_0'(0)^{-1/2} t^{3/2} + \mathrm{o}(t^{3/2}).$$

This means that

$$H_0^{-1}(n^{-1/2}) = \left(\frac{1}{2} f_0'(0)\right)^{1/3} n^{-1/3} + \mathrm{o}(n^{-1/3}).$$

Define $X(s) = W_s + s^2$, with W_s two-sided Brownian motion, and define $\hat{X}$ as the greatest convex minorant of X. Then Corollary 4.3 tells us that

$$\left(\frac{1}{2} f_0'(0)\right)^{-1/3} n^{1/3} \hat{f}(0) \xrightarrow{\mathrm{d}} \frac{\mathrm{d}\hat{X}}{\mathrm{d}s}(0),$$

in accordance with the classical result by Brunk in [1], when translated to the white noise model, except that we do not need a continuous derivative of f_0 in a neighborhood of 0, we just need the existence of the derivative in 0. Also the limit distributions derived in [15], where $f_0(x) = A|x|^\alpha(1 + \mathrm{o}(x))$, follow from our general approach. Note that even when F_0 is regularly varying near 0, it is still possible to have rates that contain slowly varying functions, like logarithmic corrections; apparently, these corrections do not change the limiting distribution itself, only the rate.

Acknowledgement

We would like to thank an anonymous referee for making a lot of helpful suggestions, greatly improving the readability of the paper, and clarifying the connections with other research.

References

[1] Brunk, H.D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (M.L. Puri, ed.) 177–197. London: Cambridge Univ. Press. MR0277070

[2] Cai, T.T. and Low, M.G. (2006). Adaptation under probabilistic error for estimating linear functionals. *J. Multivariate Anal.* **97** 231–245. MR2208851

[3] Cator, E.A. (2009). Technical report: Adaptivity and optimality of the monotone least squares estimator for four different models. Arxiv:0805.1855.

[4] Durot, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122** 222–240. MR1894068

[5] Durot, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699

[6] Golubev, G.K. (1991). Local asymptotic normality in problems of nonparametric estimation of functions, and lower bounds for quadratic risks. *Teor. Veroyatnost. i Primenen.* **36** 143–149. MR1109023

[7] Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415

[8] Groeneboom, P., Hooghiemstra, G. and Lopuhaä, H.P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. MR1740109

[9] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. I*: *Theory of Statistics* 175–194. Berkeley, CA: Univ. California Press. MR0400513

[10] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391

[11] Kulikov, V.N. and Lopuhaä, H.P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. MR2211085

[12] Low, M.G. and Kang, Y.G. (2002). Estimating monotone functions. *Statist. Probab. Lett.* **56** 361–367. MR1898714

[13] Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

[14] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

[15] Wright, F.T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

Received May 2008 and revised March 2010