

Estimation in Nonparametric Regression with Nonregular Errors

Ursula U. Müller
Texas A&M University

Wolfgang Wefelmeyer
University of Cologne

Abstract

For sufficiently nonregular distributions with bounded support, the extreme observations converge to the boundary points at a faster rate than the square root of the sample size. In a nonparametric regression model with such a nonregular error distribution, this fact can be used to construct an estimator for the regression function that converges at a faster rate than the Nadaraya–Watson estimator. We explain this in the simplest case, review corresponding results from boundary estimation that are applicable here, and discuss possible improvements in parametric and semiparametric models.

1 Introduction

To begin let X_1, \dots, X_n be i.i.d. random variables with finite variance. Then the sample mean estimates the mean $E[X]$ with rate $n^{-1/2}$. If the distribution is symmetric with finite support, then the mean coincides with the mid-range, and an alternative estimator is the *empirical mid-range* $\hat{\varrho} = (\min X_i + \max X_i)/2$. It can have a better rate than the sample mean if the distribution function is not smooth at the boundary of the support. Specifically, let $[-a, a]$ denote the support, and assume that the distribution function F of X fulfills $F(a - z) \sim z^\alpha$ for $z \downarrow 0$ with extreme value index $\alpha > 0$. Then $\min X_i + a = O_p(n^{-1/\alpha})$ and $\max X_i - a = O_p(n^{-1/\alpha})$. For these classical results on extrema of observations see e.g. Embrechts, Klüppelberg and Mikosch (1997) or de Haan and Ferreira (2006). It follows that $\hat{\varrho} - E[X] = O_p(n^{-1/\alpha})$. For $\alpha < 2$ this is faster than the rate $n^{-1/2}$ of the sample mean. If F has a density f , then the corresponding behavior at the boundary is $f(a - z) \sim z^{\alpha-1}$ for $z \downarrow 0$, and a power $\alpha - 1 < 1$ leads to a faster rate for $\hat{\varrho}$.

In Section 2 we apply this result to the nonparametric regression model $Y = r(X) + \varepsilon$ with independent X and ε . For a bandwidth $b = b_n$ with $b \rightarrow 0$ and $nb \rightarrow \infty$, we estimate $r(x)$ by a *local empirical mid-range* of the responses,

$$\hat{r}(x) = \frac{1}{2} \left(\min_{|X_i - x| \leq b} Y_i + \max_{|X_i - x| \leq b} Y_i \right).$$

We assume that the error ε has a non-regular distribution function as above, and that the regression function r and the covariate density g are Hölder at x with exponent β . We show in Theorem 2 that then

$$\hat{r}(x) - r(x) = O_p((nb)^{-1/\alpha}) + O(b^\beta).$$

The choice $b \sim n^{-1/(\alpha\beta+1)}$ gives the best rate $n^{-\beta/(\alpha\beta+1)}$. — For the Nadaraya–Watson estimator we have the best rate $n^{-\beta/(2\beta+1)}$; see Theorem 1. Hence the local empirical mid-range $\hat{r}(x)$ is faster than the classical estimator if $\alpha < 2$. This corresponds to the above result for $\hat{\nu}$. — If the covariate X is d -dimensional, the best convergence rate of the Nadaraya–Watson estimator is $n^{-\beta/(2\beta+d)}$, while the best convergence rate of the local empirical mid-range is $n^{-\beta/(\alpha\beta+d)}$. Again, the local empirical mid-range $\hat{r}(x)$ is faster than the classical estimator if $\alpha < 2$.

In Section 3 we describe extensions of this result to a more general setting. Let X denote a d -dimensional covariate and Y a one-dimensional response. Let (X, Y) have a density. We write it $g(x)h(x, y)$, where g is the density of the covariate X and $h(x, \cdot)$ is the conditional density of the response Y given $X = x$. Suppose that $h(x, \cdot)$ has support $[\ell(x), u(x)]$. Then the support of (X, Y) is bounded above and below by the surfaces ℓ and u , respectively. We assume that $h(x, \cdot)$ is nonregular,

$$\begin{aligned} h(x, \ell(x) + z) &\sim z^{\alpha(x)-1}, & z \downarrow 0, \\ h(x, u(x) - z) &\sim z^{\alpha(x)-1}, & z \downarrow 0. \end{aligned}$$

Suppose we observe i.i.d. copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) . The local empirical mid-range at $\hat{r}(x)$ now estimates the mid-range of the density $h(x, \cdot)$. If the density is symmetric, the mid-range coincides with the regression function $r(x) = E(Y|X = x)$. We estimate $\ell(x)$ and $u(x)$ separately and will restrict attention to estimating $u(x)$.

So far we have assumed that the density $g(x)h(x, y)$ of (X, Y) is unknown. If we know something about g or h , we can get better estimators, and sometimes even better rates. We describe some such results in Section 4.

2 The simplest case

Consider the nonparametric regression model

$$Y = r(X) + \varepsilon,$$

where X and ε are independent random variables and r is the (unknown) regression function. Let ε have mean zero and finite variance. Suppose we observe i.i.d. copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) and want to estimate the regression function r at a fixed point x . Write F for the distribution function of the error ε . Assume that the covariate X has a density g . To begin we recall briefly how one derives a convergence rate for the Nadaraya–Watson estimator. This makes it easier to see to which extent the arguments carry over to the local empirical mid-range, and where they differ.

The joint distribution of (X, Y) is $dF(y - r(z))g(z)dz$. Set

$$\mu(x) = \int y dF(y - r(x))g(x).$$

Since $E[\varepsilon] = 0$, the regression function has the representation

$$r(x) = E(Y|X = x) = \frac{\mu(x)}{g(x)}.$$

Kernel estimators for $g(x)$ and $\mu(x)$ are

$$\hat{g}(x) = \frac{1}{n} \sum_{i=1}^n k_b(x - X_i) \quad \text{and} \quad \hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n Y_i k_b(x - X_i),$$

where $k_b(z) = k(z/b)/b$ for a kernel k and a bandwidth b . The usual estimator for $r(x)$ is the *Nadaraya–Watson estimator*

$$(2.1) \quad \hat{r}_{NW}(x) = \frac{\hat{\mu}(x)}{\hat{g}(x)} = \frac{\sum_{i=1}^n Y_i k_b(x - X_i)}{\sum_{i=1}^n k_b(x - X_i)}.$$

In order to determine its convergence rate, we write

$$(2.2) \quad \hat{r}_{NW} - r = \frac{\hat{\mu} - \mu}{g} - \frac{\hat{\mu}}{\hat{g}g}(\hat{g} - g).$$

It remains to study the rates of $\hat{g}(x)$ and $\hat{\mu}(x)$ separately. We use the following assumptions.

Assumption K *The kernel k is bounded and has bounded support.*

Assumption B *The bandwidth $b = b_n$ fulfills $b \rightarrow 0$ and $nb \rightarrow \infty$.*

Assumption G_β The covariate density g is Hölder with exponent β at x ,

$$|g(z) - g(x)| \leq c|z - x|^\beta$$

for z in a neighborhood of x and for a positive constant c and $\beta \in (0, 1]$.

Assumption R_β The regression function r is Hölder with exponent β at x ,

$$|r(z) - r(x)| \leq c|z - x|^\beta$$

for z in a neighborhood of x and for a positive constant c and $\beta \in (0, 1]$.

We decompose $\hat{g}(x)$ into variance and bias terms,

$$\hat{g}(x) - g(x) = \hat{g}(x) - E[\hat{g}(x)] + E[\hat{g}(x)] - g(x).$$

Since g is continuous at x by Assumption G_β , we have

$$(2.3) \quad \text{Var } \hat{g}(x) \leq \frac{1}{n} E[k_b^2(x - X)] = \frac{1}{nb} \int g(x - bu) k^2(u) du = O\left(\frac{1}{nb}\right).$$

Similarly, since r is continuous at x by Assumption R_β , we have

$$(2.4) \quad \text{Var } \hat{\mu}(x) = O(1/(nb)).$$

It remains to treat the bias terms $E[\hat{g}(x)] - g(x)$ and $E[\hat{\mu}(x)] - \mu(x)$. From the Hölder condition G_β for g we obtain

$$(2.5) \quad E[\hat{g}(x)] - g(x) = E[k_b(x - X)] - g(x) = \int (g(x - bu) - g(x)) k(u) du = O(b^\beta).$$

Use $\mu(x) = r(x)g(x)$ and $E[\varepsilon] = 0$ to write

$$(2.6) \quad E[\hat{\mu}(x)] - \mu(x) = r(x)(E[k_b(x - X)] - g(x)) + E[(r(X) - r(x))k_b(x - X)].$$

The first right-hand term is of order b^β by (2.5). The second right-hand term is treated similarly, now using the Hölder condition R_β for r ,

$$E[(r(X) - r(x))k_b(x - X)] = \int (r(x - bu) - r(x))k(u) du = O(b^\beta).$$

We arrive at

$$(2.7) \quad E[\hat{\mu}(x)] - \mu(x) = O(b^\beta).$$

Applying (2.3)–(2.7) to relation (2.2), we obtain the following convergence rate for the Nadaraya–Watson estimator.

Theorem 1 *Under Assumptions K, B, G_β and R_β , the Nadaraya–Watson estimator fulfills*

$$\hat{r}_{NW}(x) - r(x) = O_p((nb)^{-1/2}) + O(b^\beta).$$

The choice $b \sim n^{-1/(2\beta+1)}$ gives the best convergence rate $n^{-\beta/(2\beta+1)}$. — In order to compare the Nadaraya–Watson estimator with the local empirical mid-range, we set $J(x) = \{i \in \{1, \dots, n\} : X_i \in [x - b/2, x + b/2]\}$ and write $|J(x)|$ for the cardinality of $J(x)$. For the box kernel $k(z) = \mathbf{1}(|z| \leq 1/2)$ we have

$$\hat{g}(x) = \frac{1}{nb}|J(x)| \quad \text{and} \quad \hat{\mu}(x) = \frac{1}{nb} \sum_{i \in J(x)} Y_i.$$

The Nadaraya–Watson estimator (2.1) simplifies to

$$\hat{r}_{NW}(x) = \frac{1}{|J(x)|} \sum_{i \in J(x)} Y_i.$$

Written this way, the Nadaraya–Watson estimator for the box kernel is seen to be the empirical mean of the responses Y_i for covariates close to x .

Suppose now that the distribution of ε is symmetric about zero and has bounded support (but is otherwise unknown). Then ε has again mean zero. Set

$$\hat{\ell}(x) = \min_{i \in J(x)} Y_i \quad \text{and} \quad \hat{u}(x) = \max_{i \in J(x)} Y_i.$$

We can estimate $r(x)$ by

$$\hat{r}(x) = \frac{1}{2}(\hat{\ell}(x) + \hat{u}(x)).$$

This is the empirical mid-range of the responses Y_i for covariates close to x . Let $[-a, a]$ denote the support of ε . The convergence rate of $\hat{r}(x)$ to $r(x)$ is given by the convergence rate of $\hat{\ell}(x)$ to $r(x) - a$ and of $\hat{u}(x)$ to $r(x) + a$. By symmetry, it suffices to look at $\hat{u}(x)$. We decompose $\hat{u}(x)$ into variance and bias terms,

$$(2.8) \quad \hat{u}(x) - r(x) - a \leq \max_{i \in J(x)} (Y_i - r(X_i)) - a + \max_{i \in J(x)} |r(X_i) - r(x)|.$$

From the Hölder condition R_β for r we obtain the same rate for the bias term as in the case of the Nadaraya–Watson estimator,

$$(2.9) \quad \max_{i \in J(x)} |r(X_i) - r(x)| = O(b^\beta).$$

If B and G_β hold, then

$$(2.10) \quad \frac{1}{nb}|J(x)| = g(x) + o_p(1).$$

This can also be seen from the treatment of $\hat{g}(x)$ above. — Now write

$$\hat{L}(x) = \min_{i \in J(x)} \varepsilon_i \quad \text{and} \quad \hat{U}(x) = \max_{i \in J(x)} \varepsilon_i.$$

The variance term of $\hat{u}(x)$ is then $\hat{U}(x) - a$.

Assumption F_α *The distribution function F of the error fulfills*

$$F(a - z) \sim z^\alpha, \quad z \downarrow 0,$$

for a positive constant α .

If G_β and F_α hold, then (2.10) together with the classical result on the convergence rate of a maximum mentioned in Section 1 imply

$$(2.11) \quad \hat{U}(x) - a = O_p((nb)^{-1/\alpha}).$$

By the same argument,

$$(2.12) \quad \hat{L}(x) + a = O_p((nb)^{-1/\alpha}).$$

Write

$$\hat{r}(x) - r(x) = \frac{1}{2}(\hat{\ell}(x) + \hat{u}(x)) - r(x) = \frac{1}{2}(\hat{\ell}(x) - r(x) + a + \hat{u}(x) - r(x) - a).$$

Applying (2.9)–(2.12) to relation (2.8), we obtain our main result.

Theorem 2 *Under Assumptions K , B , G_β and R_β , the empirical mid-range fulfills*

$$\hat{r}(x) - r(x) = O_p((nb)^{-1/\alpha}) + O(b^\beta).$$

The choice $b \sim n^{-1/(\alpha\beta+1)}$ gives the best convergence rate $n^{-\beta/(\alpha\beta+1)}$. For $\alpha < 2$ it is faster than the best rate $n^{-\beta/(2\beta+1)}$ of the Nadaraya–Watson estimator.

Higher-dimensional covariates. Suppose the covariate X is d -dimensional. Replace $k_b(z) = k(z/b)/b$ in the definitions of $\hat{g}(x)$ and $\hat{\mu}(x)$ by a multivariate version $k_b(z) = k(z/b)/b^d$. Let B and F_α hold, and assume d -dimensional versions of K , G_β and R_β . Then

$$\text{Var } \hat{g}(x) = O\left(\frac{1}{nb^d}\right) \quad \text{and} \quad \text{Var } \hat{\mu}(x) = O\left(\frac{1}{nb^d}\right).$$

The rates of the bias terms of the Nadaraya–Watson estimator remain the same, and we obtain the rate

$$\hat{r}_{NW}(x) - r(x) = O_p((nb^d)^{-1/2}) + O(b^\beta).$$

Hence the choice $b \sim n^{-1/(2\beta+d)}$ gives the best convergence rate $n^{-\beta/(2\beta+d)}$.

The definition of the local empirical mid-range also remains essentially unchanged, except that now $x = (x_1, \dots, x_d)^\top$ and

$$J(x) = \{i \in \{1, \dots, n\} : X_i \in [x_1 - b/2, x_1 + b/2] \times \dots \times [x_d - b/2, x_d + b/2]\}.$$

Similarly as in the one-dimensional case one proves that

$$\frac{1}{nb^d} |J(x)|(x) = g(x) + o_p(1).$$

We obtain a d -dimensional version of Theorem 2,

$$\hat{r}(x) - r(x) = O_p((nb^d)^{-1/\alpha}) + O(b^\beta).$$

The choice $b \sim n^{-1/(\alpha\beta+d)}$ gives the best convergence rate $n^{-\beta/(\alpha\beta+d)}$. For $\alpha < 2$ it is faster than the best rate $n^{-\beta/(2\beta+d)}$ of the Nadaraya–Watson estimator.

Simulations. We expect that the improvement of the empirical mid-range (MR) over the Nadaraya–Watson estimator (NW) is particularly drastic for small α . In the following table we consider the regression function $r(x) = |x|^{1/2}$, with “bowl shaped” error density $f(z) = (1 - |z|)^{-1/2}$ on $(-1, 1)$. The covariate density is the uniform distribution on $[-1, 1]$. The entries in the table are the simulated average mean squared errors for several bandwidths b and sample sizes $n = 50, 100$ and 200 . The minimal mean squared errors are in bold print.

Table 1: Simulated average mean squared errors

	n	b=0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
NW	50	0.1109	0.0762	0.0599	0.0500	0.0444	0.0405	0.0392	0.0386
	100	0.0560	0.0387	0.0309	0.0264	0.0241	0.0232	0.0235	0.0243
	200	0.0282	0.0198	0.0163	0.0145	0.0141	0.0145	0.0155	0.0174
MR	50	0.0610	0.0277	0.0171	0.0133	0.0126	0.0133	0.0148	0.0169
	100	0.0123	0.0060	0.0053	0.0059	0.0071	0.0087	0.0107	0.0134
	200	0.0024	0.0023	0.0030	0.0041	0.0054	0.0072	0.0093	0.0120

General boundary behavior of the error distribution. For simplicity we have assumed $F(a - z) \sim z^\alpha$ for $z \downarrow 0$. A version of Theorem 2 continues to hold if the density f fulfills $f(a - z) \sim z^{\alpha-1}L(z)$ for $z \downarrow 0$, where $L(z)$ varies slowly as $z \downarrow 0$,

or, more generally, if the observations are in the domain of attraction of a Weibull distribution. Such versions can be based on convergence results for extreme order statistics that go back to Gnedenko (1943) and Smirnov (1952).

Adapting to the extreme value index. In applications we may not know the extreme value index α of F . Then we do not know whether the local empirical mid-range \hat{r} or the Nadaraya–Watson estimator \hat{r}_{NW} is better. In this situation one should first estimate α by some estimator $\hat{\alpha}$ and choose the estimator for $\hat{r}(x)$ according to whether $\hat{\alpha} > 2$ or $\hat{\alpha} \leq 2$. For estimators of α we refer to the monographs of Embrechts, Klüppelberg and Mikosch (1997) and Beirlant, Goegebeur, Segers and Teugels (2004).

Known error range. If the support $[-a, a]$ of ε is known, we can use $\hat{u}(x) - a$ and $\hat{\ell}(x) + a$ separately as estimators for $r(x)$. The best combination of these two estimators is again the local empirical mid-range $\hat{r}(x) = (\hat{\ell}(x) + \hat{u}(x))/2$.

3 Extensions

Let (X, Y) have a density $g(x)h(x, y)$, where g is the density of the covariate X and $h(x, \cdot)$ is the conditional density of the response Y given $X = x$. Suppose that $h(x, \cdot)$ has support $[\ell(x), u(x)]$, and that $h(x, \cdot)$ is nonregular,

$$\begin{aligned} h(x, \ell(x) + z) &\sim z^{\alpha(x)-1}, & z \downarrow 0, \\ h(x, u(x) - z) &\sim z^{\alpha(x)-1}, & z \downarrow 0. \end{aligned}$$

Suppose we observe i.i.d. copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) and want to estimate $u(x)$. We can still use the local constant estimator $\hat{u}(x) = \max_{i \in J(x)} Y_i$. For $\alpha = 2$ its asymptotic limit distribution is obtained in Hall and Park (2004).

Instead of using just one extreme observation, we can use more. Set $N = |J(x)|$ and let $Z_1 \leq \dots \leq Z_N$ denote the ordered observations Y_i for $i \in J(x)$. Choose $m = m_n$ with $m/n \rightarrow 0$. Dekkers and de Haan (1989) and Gijbels and Peng (2000) study the estimator

$$\hat{u}(x) = \frac{Z_{N-m} - Z_{N-2m}}{2^{-\hat{\alpha}(x)} - 1} + Z_{N-m}$$

based on the Pickands estimator

$$\hat{\alpha}(x) = \log \frac{Z_{N-m} - Z_{N-2m}}{Z_{N-2m} - Z_{N-4m}}.$$

Another possibility is to use a local polynomial estimator rather than a local constant estimator. In particular, the local linear estimator is obtained from the

lowest polynomial above the observations (X_i, Y_i) with $|X_i - x| \leq b/2$,

$$u(x) = \min\{z : Y_i \leq q^\top(X_i - x) + z \text{ for some } q \in \mathbb{R}^d \text{ and for all } i \in J(x)\}.$$

For local polynomial estimators we refer to Korostelev and Tsybakov (1993), Härdle, Park and Tsybakov (1995), Hall, Park and Stern (1998) and Hall and Park (2004).

For $d = 1$, Bouchard, Girard, Iouditski and Nazin (2004) suggest an estimator of the form $\hat{u}(x) = \sum_{i=1}^n \lambda_i k_b(x - X_i)$ with $\lambda_1, \dots, \lambda_n$ chosen such that \hat{u} lies above the observations and minimizes the area below \hat{u} .

Girard and Jacob (2008) estimate u by a Nadaraya–Watson estimator based on power-transformed data,

$$\hat{u}(x) = \left(\frac{(p+1) \sum_{i=1}^n Y_i^p k_b(x - X_i)}{\sum_{i=1}^n k_b(x - X_i)} \right)^{1/p},$$

where $k_b(z) = k(z/b)/b^d$ for a d -dimensional kernel k and a bandwidth b .

4 Semiparametric and parametric models

If we have a parametric or semiparametric model for the density $g(x)h(x, y)$ of (X, Y) , we can get better estimators for the boundaries and the regression function, and sometimes even better rates.

Consider first the case of i.i.d. observations X_1, \dots, X_n , now from a *location model* with density $f(x - \vartheta)$, where f is known and the location parameter ϑ is unknown. Then we can estimate ϑ by the maximum likelihood estimator. Let f have support $[0, \infty)$ with $f(z) \sim z^{\alpha-1}$ for $z \downarrow 0$. For $\alpha > 2$ the maximum likelihood estimator has rate $n^{-1/2}$. For $\alpha = 2$ and $1 < \alpha < 2$ its rate is $(n \log n)^{-1/2}$ and $n^{-1/\alpha}$, respectively; see Woodroffe (1972) and (1974) and Akahira (1975a); for location models with additional parameters see Smith (1985). These results carry over to densities with bounded support. In particular, for $\alpha = 2$ the rate of the maximum likelihood estimator is slightly better than that of the sample mean and the sample extrema, and hence also better than that of the empirical mid-range. Weiss and Wolfowitz (1973) show that the maximum likelihood estimator is efficient. For $\alpha \geq 2$ the rate of the maximum likelihood estimator is sharp, and for $0 < \alpha < 2$ the rate of the the empirical mid-range is sharp; see Akahira (1975b) and also Akahira and Takeuchi (1981) and (1991), who also obtain lower bounds on the asymptotic distributions. For more general parametric models, lower bounds for variances are given in Polfeldt (1970a)–(1970c), Akahira, Puri and Takeuchi (1986) and Akahira and Takeuchi (1995); and

lower bounds for distributions are given in Akahira (1982) and Akahira and Takeuchi (1987).

Consider now a regression model $Y = r_{\vartheta}(X) + \varepsilon$, where ε and X are one-dimensional and independent with densities f and g . Then the conditional density of Y at y given $X = x$ is $h(x, y) = f(y - r_{\vartheta}(x))$. If f is known and has bounded support $[-a, a]$, say, we can use local versions of the above estimators to estimate a . If f depends on unknown parameters, local versions of Smith (1985) can be used.

If we observe i.i.d. copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of a two-dimensional random variable (X, Y) and the upper boundary u of the support is known to be convex or monotone or both, then it suggests itself to use an estimator for u that has the same properties. This is often the case in frontier estimation. We refer to Korostelev, Simar and Tsybakov (1995) and Gijbels, Mammen, Park and Simar (1999).

If there is a parametric model $u = u_{\vartheta}$ for the upper boundary, we can estimate ϑ by the minimizer $\vartheta = \hat{\vartheta}$ of

$$\sum_{i=1}^n w(u_{\vartheta}(X_i) - Y_i) \quad \text{subject to} \quad u_{\vartheta}(X_i) \geq Y_i, \quad i = 1, \dots, n,$$

a linear program. Examples for w are $w(z) = z$ and $w(z) = z^2$. This approach goes back to Aigner and Chu (1968) and Koenker and Bassett (1978). For results of this type see Smith (1994), Portnoy and Jurečková (1999) and Knight (2001, 2006).

Related problems are treated in Donald and Paarsch (2002), Hirano and Porter (2003), Chernozhukov and Hong (2004), Chernozhukov (2005), Florens and Simar (2005) and Kumbhakar, Park, Simar and Tsionas (2007).

References

- [1] Aigner, D. J. and Chu, S. F. (1968). On estimating the industry production function. *Amer. Economic Review* 58:826–893.
- [2] Akahira, M. (1975a). Asymptotic theory for estimation of location in non-regular cases I. Order of convergence of consistent estimators. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* 22(1):8–26.
- [3] Akahira, M. (1975b). Asymptotic theory for estimation of location in non-regular cases II. Bounds of asymptotic distributions of consistent estimators. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* 22(3):3–19.

- [4] Akahira, M. (1982). Asymptotic optimality of estimators in nonregular cases. *Ann. Inst. Statist. Math.* 34:69–82.
- [5] Akahira, M., Puri, M. L. and Takeuchi, K. (1986). Bhattacharyya bound of variances of unbiased estimators in nonregular cases. *Ann. Inst. Statist. Math.* 38:35–44.
- [6] Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*. Lecture Notes in Statistics 7. New York: Springer-Verlag.
- [7] Akahira, M. and Takeuchi, K. (1987). The lower bound for the variance of unbiased estimators for one-directional family of distributions. *Ann. Inst. Statist. Math.* 39:593–610.
- [8] Akahira, M. and Takeuchi, K. (1991). Asymptotic efficiency of estimators for a location parameter family of densities with the bounded support. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* 38(1):1–9.
- [9] Akahira, M. and Takeuchi, K. (1995). *Non-Regular Statistical Estimation*. Lecture Notes in Statistics 107. New York: Springer-Verlag.
- [10] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes. Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: Wiley.
- [11] Bouchard, G., Girard, S., Iouditski, A. B. and Nazin, A. V. (2004). Nonparametric frontier estimation by linear programming. *Autom. Remote Control* 65:58–64.
- [12] Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Statist.* 33:806–839.
- [13] Chernozhukov, V. and Hong, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* 72:1445–1480.
- [14] Dekkers, A. L. M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* 17:1795–1832.
- [15] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. An Introduction*. Springer Series in Operations Research and Financial Engineering. New York: Springer.

- [16] Donald, S. G. and Paarsch, H. J. (2002). Superconsistent estimation and inference in structural econometric models using extreme order statistics. *J. Econometrics* 109:305–340.
- [17] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events. For Insurance and Finance*. Applications of Mathematics 33. Berlin: Springer.
- [18] Florens, J.-P. and Simar, L. (2005). Parametric approximations of nonparametric frontiers. *J. Econometrics* 124:91–116.
- [19] Gijbels, I., Mammen, E., Park, B. U. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* 94:220–228.
- [20] Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics. *Extremes* 3:251–277.
- [21] Girard, S. and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *J. Multivariate Anal.* 99:403–420.
- [22] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. of Math. (2)* 44:423–453.
- [23] Hall, P. and Park, B. U. (2004). Bandwidth choice for local polynomial estimation of smooth boundaries. *J. Multivariate Anal.* 91:240–261.
- [24] Hall, P., Park, B. U. and Stern, S. E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* 66:71–98.
- [25] Härdle, W., Park, B. U. and Tsybakov, A. B. (1995). Estimation of non-sharp support boundaries. *J. Multivariate Anal.* 55:205–218.
- [26] Hirano, K. and Porter, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica* 71:1307–1338.
- [27] Knight, K. (2001). Limiting distributions of linear programming estimators. *Extremes* 4:87–103.
- [28] Knight, K. (2006). Asymptotic theory for M -estimators of boundaries. In: Sperlich, S., Härdle, W., Aydinli, G., eds., *The Art of Semiparametrics*, 1–21, Heidelberg: Physica-Verlag/Springer.

- [29] Koenker, R. and Bassett, G., Jr. (1978). Regression quantiles. *Econometrica* 46:33–50.
- [30] Korostelev, A. P., Simar, L. and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* 23:476–489.
- [31] Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction*. Lecture Notes in Statistics 82. New York: Springer-Verlag.
- [32] Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. *J. Econometrics* 137:1–27.
- [33] Polfeldt, T. (1970a). The order of the minimum variance in a non-regular case. *Ann. Math. Statist.* 41:667–672.
- [34] Polfeldt, T. (1970b). Minimum variance order when estimating the location of an irregularity in the density. *Ann. Math. Statist.* 41:673–679.
- [35] Polfeldt, T. (1970c). Asymptotic results in non-regular estimation. *Skand. Aktuarietidskr.* 1970, Supplement 1–2.
- [36] Portnoy, S. and Jurečková, J. (1999). On extreme regression quantiles. *Extremes* 2:227–243.
- [37] Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Translation* 67.
- [38] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72:67–90.
- [39] Smith, R. L. (1994). Nonregular regression. *Biometrika* 81:173–183.
- [40] Weiss, L. and Wolfowitz, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Statist.* 1:944–947.
- [41] Woodroffe, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Math. Statist.* 43:113–122.
- [42] Woodroffe, M. (1974). Maximum likelihood estimation of translation parameter of truncated distribution II. *Ann. Statist.* 2:474–488.