

Shape constrained estimators in inverse regression models with convolution-type operator

Melanie Birke*

Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: melanie.birke@rub.de

Nicolai Bissantz

Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: nicolai.bissantz@rub.de

November 3, 2007

Abstract

In this paper we are concerned with shape restricted estimation in inverse regression problems with convolution-type operator. We use increasing rearrangements to compute increasing and convex estimates from an (in principle arbitrary) unconstrained estimate of the unknown regression function. An advantage of our approach is that it is not necessary that prior shape information is known to be valid on the complete domain of the regression function. Instead, it is sufficient if it holds on some compact interval. A simulation study shows that the shape restricted estimate on the respective interval is significantly less sensitive to moderate undersmoothing than the unconstrained estimate, which substantially improves applicability of estimates based on data-driven bandwidth estimators. Finally, we demonstrate the application of the increasing estimator by the estimation of the luminosity profile of an elliptical galaxy. Here, a major interest is in reconstructing the central peak of the profile, which, due to its small size, requires to select the bandwidth as small as possible.

Key words: Convexity, increasing rearrangements, image reconstruction, inverse problems, monotonicity, order restricted inference, regression estimation, shape restrictions

1 Introduction

We consider the indirect regression model

$$Y_k = (K\theta)(z_k) + \varepsilon_k, \quad k = -n, \dots, n \quad (1)$$

where $z_k = k/(na_n)$, $a_n \rightarrow 0$ for $n \rightarrow \infty$, are fixed design points and ε_k are i.i.d. with $E[\varepsilon_k] = 0$, $E[\varepsilon_k^2] = \sigma^2 < \infty$. The design is selected such that we asymptotically estimate θ on the full real

line \mathbb{R} . We assume that K is a known linear, injective operator and θ is an at least p times differentiable regression function which is unknown. Moreover, the function of interest θ cannot be observed directly but only $K\theta$. Statistical inverse problems of this form have received a lot of attention in recent years (cf. e.g. Mair & Ruymgaart, 1996, Cavalier & Tsybakov, 2002, or Bissantz et al., 2007b) and are of considerable interest for many practical applications. Examples include deconvolution problems (cf. Fan, 1991, Johnstone et al., 2004), tomographic problems in medical imaging (Johnstone & Silverman, 1990, Cavalier, 2000) or problems related to satellite gradiometry (Bissantz et al., 2007b). Similar problems with convolution-type operator K occur in nonparametric errors-in-variables regression (cf. e.g. Fan & Truong, 1993, who determine convergence rates in a random design errors-in-variables model). Another example is the recovery of the luminosity profile of an elliptical galaxy from a telescopic image (e.g. Lauer et al. 1995, 2007) which establishes a connection with shape restrictions in such models. The profile is in general monotonically increasing with decreasing radius, but also has a sharply peaked central structure. Whereas this would strongly suggest using a rather small bandwidth to estimate the profile, it is well known in inverse estimation, that even moderate undersmoothing yields unreasonable estimators which in particular will not be monotone with decreasing radius if unconstrained (spectral-type) estimators are used (e.g. Bissantz et al., 2007a). The knowledge about an increasing part of the luminosity profile motivates the usage of a shape restricted estimate. In inverse problems, shape restricted estimates have already been considered for density deconvolution (van Es et al., 1998) and for estimating a concave distribution function from data corrupted with noise (Jongbloed and van der Meulen, 2007). But, so far, to the best of our knowledge, there are no results for estimating a (partly) shape restricted regression function θ in inverse problems.

Hence, in this article we propose estimators for the regression function θ which meet shape restrictions on monotonicity or convexity on certain compact intervals based on two steps. The first step consists of an unconstrained estimation of the regression function which is, in a second step, modified with respect to the shape constraints by using increasing rearrangements. Properties of such methods in a direct regression setting on compact intervals have already been discussed by Dette, Neumeyer and Pilz (2006), Birke and Dette (2005), Birke and Dette (2007) and Chernozhukov, Fernández-Val and Galichon (2007).

As it will turn out in our subsequent analysis, the shape restricted estimator significantly reduces sensitivity of the estimate on (moderate) undersmoothing, and hence improves substantially estimates of such sharply structured signals. Moreover, from the methodological point-of-view, it makes data-driven bandwidth estimation more feasible, as the devastating consequences of undersmoothing are reduced. Another advantage of this method is, that it is still applicable if the shape constraints are met only on a compact interval. It would not be possible to include such restrictions in standard methods like spectral regularisation, which are basically constructed from an expansion of the solution by eigenfunctions of the operator K .

We restrict the convergence analysis of our method to the cases of isotone and convex regression functions in the case of a convolution-type operator K , i.e. where the operator K in model (1) is given by the convolution of the true function θ with another function Ψ . Such problems arise straightforwardly for example in non-parametric errors-in-variables models, and in the reconstruction of signals from devices with limited resolution due to the convolution with a point-spread-function.

Our method for shape restricted estimation requires an unconstrained (inverse) estimate of θ in

the first step. Moreover, our proofs of convergence properties of the shape restricted estimators are based on uniform and pointwise convergence rates of the unconstrained estimator. Here, we use the results of Bissantz and Birke (2007) and Birke, Bissantz and Holzmann (2007), who analyse spectral regularisation estimators (e.g. Mair & Ruymgaart, 1996, and Bissantz et al., 2007b) for the error-in-variables model, under the assumption that the Fourier transform $\Phi_\Psi(\omega)$ of Ψ is non-zero for all $\omega \in \mathbb{R}$, and that there exist $\beta \geq 0$ and $C \in C \setminus 0$ such that

$$\Phi_\Psi(\omega)|\omega|^\beta \rightarrow C, \quad |\omega| \rightarrow \infty.$$

The last condition implies that the inverse problem to be solved here is *ordinary smooth*. If we further call Φ_k the Fourier transform of the kernel \mathcal{K} and assume that Φ_k has compact support, the unconstrained estimator on which our shape restricted estimators will be based is

$$\hat{\theta}^{(j)}(x) = \sum_{k=-n}^n Y_k w_{j,k,n}(x) \tag{2}$$

with

$$w_{j,k,n}(x) = \frac{1}{2\pi n h^{j+1} a_n} \int_{\mathbb{R}} (-i\omega)^j e^{-i\omega(x-z_k)/h} \frac{\Phi_k(\omega)}{\Phi_\Psi(\omega/h)} d\omega,$$

where $h \rightarrow 0$ is a smoothing parameter for $n \rightarrow \infty$, and a_n has been defined after eq. (1). In the following we will write $\hat{\theta}$ for $\hat{\theta}^{(0)}$ if it is unmistakable that we mean the original regression estimate and only use the superscript (j) to point out which derivative of $\hat{\theta}$ we mean. Bissantz and Birke (2007) provide pointwise convergence rates and Birke, Bissantz and Holzmann (2007) uniform convergence rates for the estimator $\hat{\theta}^{(j)}$ under the condition that $j \leq p$. It will turn out that the shape restricted estimates achieve the same pointwise convergence rates as the unconstrained estimators, similar to the direct regression case (see Dette, Neumeyer and Pilz, 2006, Birke and Dette, 2005 and Birke and Dette, 2007).

Our assertions remain true for other estimators of the type (2) and can easily be transferred to the problem of estimating decreasing or concave regression functions for convolution-type operators in model (1). In many applications, such as the reconstruction of the luminosity profile of an elliptical galaxy from an astronomical image discussed in section 6, other estimators such as the iterative Richardson-Lucy method are in common use also (cf. Lucy, 1974 in the astronomical context and Bewersdorf et al., 2006, for the application in fluorescence microscopy). Although we do not provide convergence results for this case here, our method, i.e. plugging in the respective unconstrained estimator into the rearrangement, is again applicable. Moreover, the plug-in approach which we use here for the determination of a shape restricted estimate can be used in any inverse regression model (1), if an unconstrained estimate of θ is available.

This article is organized as follows. Since our method is based on increasing and convex rearrangements we review these methods briefly in section 2. Then, in section 3 we discuss increasing estimators for the problem (1) with convolution-type operators, followed by convex estimators in section 4. The results of a small simulation study of the increasing estimator are given in section 5. An application of the method to astronomical data is discussed in section 6. Finally, to improve readability of the article, we defer proofs to an appendix.

2 Increasing and Convex rearrangements

In this section we briefly review the methods of increasing and convex rearrangements which will be used as part of the shape-restricted inverse estimators introduced in sections 3 and 4. We begin with a statistically motivated definition of increasing rearrangements and then relate it to nondecreasing rearrangements first introduced by Hardy, Littlewood and Polya (1978). Assume that the function $g : \mathbb{R} \rightarrow \mathbb{R}$ is already strictly increasing and differentiable on the interval $[0, 1]$ and U is a uniformly distributed random variable on $[0, 1]$. Then the random variable $g(U)$ has density

$$(g^{-1})'(u)I_{[g(0),g(1)]}(u) \tag{3}$$

and its distribution function is given by

$$\int_{-\infty}^t (g^{-1})'(\tilde{u})I_{[g(0),g(1)]}(\tilde{u})d\tilde{u} = \int_0^1 I\{g(u) \leq t\}du = g^{-1}(t)$$

for $t \in [g(0), g(1)]$. Therefore the quantile function of $g(U)$ is g . Both, the distribution function and the quantile function, are increasing. Now, if g is not strictly increasing, we again define the distribution function by

$$\phi(g)(t) = \int_0^1 I\{g(u) \leq t\}du$$

and call the quantile function

$$\phi(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \phi(g)(t) \geq x\}$$

increasing rearrangement of g . Following the principle first introduced by Hardy, Littlewood and Polya (1978) the last two steps are exactly the way how increasing rearrangements are defined there. For the sake of completeness we include the definition here.

Figure 1 shows the increasing rearrangements (full line) of the functions

$$\begin{aligned} g_1(x) &= 3(2x - 1)^2 \\ g_2(x) &= x + \frac{1}{4} \sin(4\pi x) \end{aligned}$$

which are displayed as dashed line. The principle of increasing rearrangements to simply bring the function values in an ascending order is quite obvious in these pictures. However, also the disadvantage of increasing rearrangements, that they are not necessarily smooth although the true function g is smooth becomes clear (see middle picture of Figure 1). Therefore we define a smoothed increasing rearrangement. For a positive kernel K_d of order 2 and a bandwidth h_d

$$\frac{1}{h_d} \int_0^1 K_d\left(\frac{g(v) - u}{h_d}\right)dv$$

is a smoothed version of the density in (3), and the corresponding distribution function is given by

$$\phi_{h_d}(g)(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{g(v) - u}{h_d}\right)dudv.$$

Figure 1: Left and middle picture: Increasing rearrangements (full line) of two different functions g_1 and g_2 (dashed line) together with the corresponding distribution functions (dotted line). Right picture: Convex rearrangements for $a = 0, 0.25, 0.5$ and 0.75 (dotted lines) and the L_2 -optimal convex rearrangement (dashed line) together with the true function (full line).

Following the structure of increasing rearrangements we call the inverse of the distribution function $\phi_{h_d}(g)^{-1}(x)$ smoothed increasing rearrangement.

We now turn to the problem of making arbitrary functions convex. To this end note that a differentiable function is convex if and only if its first derivative is increasing. Therefore, if we want to make $\theta : [0, 1] \rightarrow \mathbb{R}$ convex, we use the (smoothed) increasing rearrangement for θ' . For $a \in [0, 1]$ a (smoothed) convex rearrangement is then given by

$$\rho(\theta)(x, a) = \int_a^x \phi(\theta')^{-1}(u) du + \theta(a)$$

and

$$\rho_{h_d}(\theta)(x, a) = \int_a^x \phi_{h_d}(\theta')^{-1}(u) du + \theta(a),$$

respectively (see Birke, 2007 or Birke and Dette, 2007). Note that the convex rearrangement strongly depends on the choice of the parameter a . To avoid the problem, that a convex rearrangement is not well-defined, occuring by different choices of a , we suggest to use L^2 -optimal convex rearrangements in the following which minimize the L^2 -distance to the true function θ . A closed form of them is given by

$$\rho^*(\theta)(x) = \int_0^1 \rho(x, a) da$$

and

$$\rho_{h_d}^*(\theta)(x) = \int_0^1 \rho_{h_d}(x, a) da,$$

respectively (see e.g. Birke and Dette, 2007 or Birke, 2007). The right picture in Figure 1 shows the true function

$$\theta(x) = \frac{1}{2}(2x - 1)^3 + \frac{1}{2}$$

– which is a primitive of g_1 – as full line together with several convex rearrangements for different lower bounds a as dotted lines and the L_2 -optimal convex rearrangement as dashed line.

3 Increasing Estimators

The definition of an increasing estimator is now straightforward. To this end assume that θ is increasing on the interval $[0, 1]$. We apply the increasing or smoothed increasing rearrangement to

$g = \hat{\theta}_{|[0,1]}$. Then increasing estimators of the function $\theta_{|[0,1]}$ are given by $\phi(\hat{\theta})^{-1}$ with

$$\phi(\hat{\theta})(t) = \int_0^1 I\{\hat{\theta}(v) \leq t\} dv$$

or $\phi_{h_d}(\hat{\theta})^{-1}$ with

$$\phi_{h_d}(\hat{\theta})(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{\theta}(v) - u}{h_d}\right) dudv.$$

One characteristic of the increasing rearrangement is, that it preserves the L_p -norm of the function to which it is applied. Therefore we conclude that the increasing estimator on the interval $(0, 1)$ has the same L_p -norm as the unconstrained estimator on that interval. But the real interest lies in the L_p estimation error, that is in the distance of the estimator to the true but unknown regression function θ . Chernozhukov, Fernández-Val und Galichon (2007) show for a measurable and weakly increasing function g and any estimator \hat{g} of g , that the increasing rearrangement $\phi(\hat{g})$ reduces the L_p estimation error for any $p \in [1, \infty]$. This also holds for the deconvolution estimator defined in (2), that is

$$\left(\int_0^1 |\phi(\hat{\theta})^{-1}(x) - \theta(x)|^p dx \right)^{1/p} \leq \left(\int_0^1 |\hat{\theta}(x) - \theta(x)|^p dx \right)^{1/p}$$

for all $p \in [1, \infty]$. It can be shown with similar methods as in Bissantz and Holzmann (2007), who consider the pointwise asymptotic behavior of spectral regularization estimators on a compact domain, that for the deconvolution problem the unconstrained estimator $\hat{\theta}$ is pointwise asymptotically normal by using a central limit theorem for weighted sums of independent random variables (for details in this setting see Bissantz and Birke, 2007). Under certain assumptions this asymptotic distribution is preserved while applying the smoothed increasing rearrangement to $\hat{\theta}$. First we state a central limit theorem for the distribution function of $\hat{\theta}$. To this end assume that the regression function θ , the regression estimate $\hat{\theta}$ and the kernel function K_d are twice continuously differentiable and that the bandwidths h and h_d and the sequence a_n fulfill the conditions $h \rightarrow 0$, $a_n \rightarrow 0$, $h_d \rightarrow 0$, $nh \rightarrow \infty$, $nh_d \rightarrow \infty$, $na_n \rightarrow \infty$, and furthermore

$$nh^{2\beta+5}a_n = O(1) \tag{4}$$

$$\frac{\log h^{-1}}{nh^{5/2}a_n^{5/2}} = o(1) \tag{5}$$

$$\frac{(\log h^{-1})^3}{nh^3a_n} = o(1) \tag{6}$$

$$\frac{na_n^{2r+1}}{h \log h^{-1}} = O(1) \tag{7}$$

$$\frac{(\log n)^2}{nh^{2\beta+1}a_n h_d^2} = o(1) \tag{8}$$

$$\frac{h_d}{h} = o(1). \tag{9}$$

Conditions (4) - (7) assure, that uniform rates of convergence for the unconstrained estimator and its first derivative exist (see Birke, Bissantz and Holzmann, 2007). Note that (4) - (9) include the

conditions

$$\begin{aligned}\frac{\log n}{nh^{2\beta+2j+1}a_n} &= o(1), \quad j = 0, 1 \\ \frac{\log n}{nh^{2\beta+1}a_nh_d^2} &= o(1) \\ \frac{h^2}{h_d} &= o(1)\end{aligned}$$

which we will need in the proofs of the following theorems. To see that the conditions (4) - (9) can be fulfilled simultaneously, assume that $a_n = c_a n^{-a}$, $h = c_h n^{-(1-a)/(2\beta+5)}$ and $h_d = c_d n^{-b}$ for some constants c_a, c_h and $c_d > 0$. This yields the bounds

$$\begin{aligned}\frac{\beta + 3}{(2\beta + 5)r + \beta + 3} &< a < \frac{4\beta + 5}{5(2\beta + 4)} \\ \frac{3(\beta + 3)}{4\beta + 5} &< r \\ \frac{1 - a}{2\beta + 5} &< b < \frac{2(1 - a)}{2\beta + 5}\end{aligned}$$

for a, b and r .

In the case of the inverse regression model (1) for convolution type operators we have

$$\sum_{k=-n}^n w_{0,k,n}^2(x) \sim \frac{1}{2\pi nh^{2\beta+1}a_n} \int_{-\infty}^{\infty} |\omega|^{2\beta} |\Phi_k(\omega)|^2 d\omega$$

and the following theorem holds.

Theorem 1 *If θ is two times continuously differentiable and strictly increasing on $(0, 1)$ and the bandwidths fulfill (4) - (9) then for any $t \in (\theta(0), \theta(1))$ with $(\theta^{-1})'(t) > 0$ and*

$$\frac{\max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|}{\left(\sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t))\right)^{1/2}} \rightarrow 0$$

we have

$$\sqrt{nh^{2\beta+1}a_n} \left(\phi_{h_d}(\hat{\theta})(t) - \theta^{-1}(t) - b_n(\theta^{-1}(t))(\theta^{-1})'(t) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2)$$

where

$$\begin{aligned}b_n(x) &= \frac{1}{2\pi h} \int_{-\infty}^{\infty} e^{\frac{-ix\omega}{h}} (1 - \Phi_k(\omega)) \Phi_f\left(\frac{\omega}{h}\right) d\omega \\ &= + \frac{1}{2\pi h} \int_{-\infty}^{\infty} e^{\frac{-ix\omega}{h}} \frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)} \cdot \left\{ \int_{-\infty}^{-1/a_n} e^{\frac{i\omega y}{h}} (K\theta)(y) dy + \int_{1/a_n}^{\infty} e^{\frac{i\omega y}{h}} (K\theta)(y) dy \right\} d\omega\end{aligned}$$

and

$$\gamma^2 = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} \omega^{2\beta} |\Phi_k(\omega)|^2 d\omega.$$

are the bias and asymptotic variance of the unconstrained estimator $\hat{\theta}(x)$.

It now follows from Theorem 1 that also the inverse of the distribution function, which is the increasing estimator, is asymptotically normal. This is the content of our next theorem.

Theorem 2 *Let the assumptions of Theorem 1 be fulfilled, then we have for any $x \in (0, 1)$ with $\theta'(x) > 0$*

$$\sqrt{nh^{2\beta+1}a_n} \left(\phi_{h_d}(\hat{\theta})^{-1}(x) - \theta(x) - b_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2),$$

where b_n and γ^2 are defined in Theorem 1.

In addition we can conclude from Theorem 2 that the increasing estimator is consistent for θ .

Remark 1 The above results can be generalized to other estimators of the type

$$\hat{g}(x) = \sum_{k=-n}^n Y_k w_{k,n}(x) \tag{10}$$

where the random variables Y_k , $k = -n, \dots, n$ are independently distributed, the weights $w_{k,n}$ are deterministic and the uniform rate of convergence s_n of $\hat{\theta}$ is available. The assumptions we have to check for an analogon of Theorem 2 are that the function g is two times continuously differentiable and strictly increasing on $(0, 1)$, and that

$$\begin{aligned} \frac{\max_{-n \leq k \leq n} w_{k,n}(x)}{\left(\sum_{k=-n}^n w_{k,n}^2(x) \right)^{1/2}} &= o(1) \\ \frac{h_d \max_{-n \leq k \leq n} w_{k,n}^{(1)}(x)}{\max_{-n \leq k \leq n} w_{k,n}(x)} &= o(1) \\ h_d^4 &= o\left(\sum_{k=-n}^n w_{k,n}^2(x) \right) \\ \frac{s_n}{h_d} &= o(1) \\ h_d^2 \left(\sum_{k=-n}^n (w_{k,n}^{(2)}(x))^2 \right)^{1/2} \left(\sum_{k=-n}^n (w_{k,n}(x))^2 \right)^{1/2} &= o\left(\sum_{k=-n}^n w_{k,n}^2(x) \right) \end{aligned}$$

with $\sup |\hat{g}(x) - g(x)| = O_P(s_n)$.

Then we have for any $x \in (0, 1)$ with $g'(x) > 0$

$$\left(\sum_{k=-n}^n w_{k,n}^2(x) \right)^{-1/2} \left(\phi_{h_d}(\hat{g})^{-1}(x) - g(x) - b_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

where b_n is the bias of the unconstrained estimator \hat{g} . For example, we can check those conditions for the j^{th} derivative of a spectral regularisation estimator, under the assumption that θ is at least $(j+2)$ -times continuously differentiable. As long as conditions (4) - (7) are fulfilled, we have from Birke, Bissantz and Holzmann (2007) the uniform convergence rates of the j^{th} derivative

$$\sup |\hat{\theta}^{(j)}(x) - \theta^{(j)}(x)| = O_P\left(\frac{\log n}{nh^{2\beta+2j+1}a_n}\right)^{1/2}$$

and we have for the sum occuring in the variance of the estimator

$$\sum_{k=-n}^n w_{j,k,n}^2(x) = O\left(\frac{1}{nh^{2\beta+2j+1}a_n}\right).$$

Note, that in this case $w_{j,k,n}^{(l)}(t) = w_{j+l,k,n}(t)$. If

$$\begin{aligned} nh^{2\beta+2j+5}a_n &= O(1), \\ \frac{h_d}{h} &= o(1), \text{ and} \\ \frac{\log n}{nh^{2\beta+2j+1}a_nh_d^2} &= o(1), \end{aligned}$$

it follows that

$$nh^{2\beta+2j+1}a_nh_d^4 = o(1)$$

$$h_d^2 \left(\sum_{k=-n}^n (w_{k,n}''(x))^2 \right)^{1/2} \left(\sum_{k=-n}^n (w_{k,n}(x))^2 \right)^{1/2} = \frac{h_d^2}{nh^{2\beta+2j+3}a_n} = o\left(\frac{1}{nh^{2\beta+2j+1}a_n}\right)$$

and

$$\frac{h_d \max_{-n \leq k \leq n} w_{j+1,k,n}(x)}{\max_{-n \leq k \leq n} w_{j,k,n}(x)} = o(1).$$

From Bissantz and Birke (2007) we have

$$\frac{\max_{-n \leq k \leq n} w_{j,k,n}(x)}{\left(\sum_{k=-n}^n w_{j,k,n}^2(x) \right)^{1/2}} = o(1)$$

and therefore for every $x \in (0, 1)$ with $\theta^{(j+1)}(x) > 0$ the asymptotic normality

$$\sqrt{nh^{2\beta+2j+1}a_n} \left(\phi_{h_d}(\hat{\theta}^{(j)})^{-1}(x) - \theta^{(j)}(x) - b_{n,j}(x) \right) \xrightarrow{D} \mathcal{N}(0, \gamma_j)$$

where

$$b_{n,j}(x) = \frac{1}{2\pi h^{j+1}} \int_{-\infty}^{\infty} (-i\omega)^j e^{-\frac{ix\omega}{h}} (1 - \Phi_k(\omega)) \Phi_f\left(\frac{\omega}{h}\right) d\omega \quad (11)$$

$$+ \frac{1}{2\pi h^{j+1}} \int_{-\infty}^{\infty} (-i\omega)^j e^{-\frac{ix\omega}{h}} \frac{\Phi_k(\omega)}{\Phi_\Psi\left(\frac{\omega}{h}\right)} \cdot \left\{ \int_{-\infty}^{-1/a_n} e^{\frac{i\omega y}{h}} (K\theta)(y) dy + \int_{1/a_n}^{\infty} e^{\frac{i\omega y}{h}} (K\theta)(y) dy \right\} d\omega$$

$$\gamma_j = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} |\omega|^{2\beta+2j} |\Phi_k(\omega)|^2 d\omega \quad (12)$$

are the leading term of the bias and the asymptotic variance of $\hat{\theta}^{(j)}(x)$.

4 Convex Estimators

In this section we assume that the regression function θ is at least two times continuously differentiable. If we define

$$\hat{\theta}'(u) = \frac{\partial}{\partial u} \hat{\theta}(u),$$

we obtain a convex estimator on the interval $[0, 1]$ by applying the (L^2 -optimal) smoothed convex rearrangement to the unconstrained estimator $\hat{\theta}$ and obtain

$$\hat{\rho}(x, a) = \rho_{h_d}(\hat{\theta})(x, a) = \int_a^x \phi_{h_d}(\hat{\theta}')^{-1}(u) du + \hat{\theta}(a) \quad (13)$$

and

$$\hat{\rho}^*(x) = \rho_{h_d}^*(\hat{\theta})(x),$$

respectively. Again we can show that the convex estimator shares the asymptotic properties of the unconstrained estimator which we state in the following theorem. To this end assume that the bandwidths h , h_d and a_n fulfill the conditions h , h_d and $a_n \rightarrow 0$, nh , nh_d and $na_n \rightarrow \infty$ and furthermore

$$nh^{2\beta+5+\delta}a_n = O(1) \quad (14)$$

$$\frac{\log h^{-1}}{nh^{5/2}a_n^{5/2}} = o(1) \quad (15)$$

$$\frac{(\log h^{-1})^3}{nh^3a_n} = o(1) \quad (16)$$

$$\frac{na_n^{2r+1}}{h \log h^{-1}} = O(1) \quad (17)$$

$$\frac{\log n}{nh^{2\beta+3}a_nh_d^2} = o(1) \quad (18)$$

$$\frac{h_d}{h^{1+\delta/4}} = o(1). \quad (19)$$

for any $\delta > 0$. Again, conditions (14) - (17) guarantee, that uniform convergence rates for $\hat{\theta}$ exist. Note, that it follows from (14) - (19) that

$$\frac{h_d}{h}(\log n)^{1/2} = o(1)$$

Conditions (14) - (19) can be simultaneously fulfilled if we choose $a_n = c_a n^{-a}$, $h = c_h n^{-(1-a)/(2\beta+5+\delta)}$ and $h_d = c_d n^{-b}$ with

$$\frac{2\beta + 6 + \delta}{2(2\beta + 5 + \delta)r + 2\beta + 6 + \delta} < a < \frac{4\beta + 5 + 2\delta}{5(2\beta + 4 + \delta)}$$

$$\begin{aligned} \frac{3(2\beta + 6 + \delta)}{2(4\beta + 5 + 2\delta)} &< r \\ \frac{(4 + \delta)(1 - a)}{4(2\beta + 5 + \delta)} &< b < \frac{(2 + \delta)(1 - a)}{2(2\beta + 5 + \delta)} \end{aligned}$$

for some constants c_a , c_h and $c_d > 0$.

Theorem 3 *Assume that the assumptions (14)-(19) are satisfied. If the regression function θ is strictly convex, we have for any $x \in (0, 1)$ with $\theta''(x) > 0$ and any $a \in (0, 1)$*

$$\hat{\rho}(x, a) - \theta(x) = \hat{\theta}(x) - \theta(x) + o_p\left(\frac{1}{\sqrt{nh^{2\beta+1}a_n}}\right)$$

Corollary 1 *If the assumptions of Theorem 3 are satisfied, then*

$$\sqrt{nh^{2\beta+1}a_n}(\hat{\rho}(x, a) - \theta(x) - b_n(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2)$$

where $b_n(x)$ and γ^2 are the bias and asymptotic variance of the unconstrained estimator $\hat{\theta}(x)$ already defined in Theorem 1

Remark 2 In this section we assume that θ is at least two times continuously differentiable. We can prove Theorem 3 for a two times continuously differentiable regression function but then we cannot choose the L_2 -optimal bandwidth $h = O(n^{-(1-a)/(2\beta+5)})$ for the unconstrained estimate and therefore do not obtain the optimal convergence rate of the convex estimator. The choice of $h = O(n^{-(1-a)/(2\beta+5)})$ contradicts condition (14). We have to choose a slightly undersmoothing bandwidth but come arbitrarily close to the optimal one ($\delta > 0$ small). In the case of a three times continuously differentiable regression function the optimal bandwidth is proportional to $n^{-(1-a)/(2\beta+7)}$ for which (14) - (19) hold ($\delta = 2$). Therefore we obtain the optimal rates of convergence in Theorem 3 if we can assume that the regression function is three times continuously differentiable. The structure of the proof given here can also be applied to the direct regression setting considered in Birke and Dette (2007) and therefore improves the results given there where the regression function has to be three times continuously differentiable.

Remark 3 Similar to the results for increasing estimators (see Remark 1) we can formulate the results of Theorem 3 and Corollary 1 for general estimators of the type (10). If the regression function is at least two times continuously differentiable and the uniform convergence rates of $\hat{g}^{(l)}$, $l = 0, 1$, are available, say

$$\sup_{x \in [0,1]} |\hat{g}^{(l)}(x) - g^{(l)}(x)| = O_P(s_{n,l}) = o_P(1)$$

and we additionally have

$$\begin{aligned} s_{n,k} &= o(s_{n,l}) \text{ for } k < l \\ s_{n,1}^2 &= o\left(\sum_{k=-n}^n w_{k,n}^2(x)\right)^{-1/2} \\ h_d^2 &= o\left(\sum_{k=-n}^n w_{k,n}^2(x)\right)^{-1/2} \end{aligned}$$

then it follows that for any $x \in (0, 1)$ with $\theta''(x) > 0$ and any $a \in (0, 1)$

$$\rho_{h_d}(\hat{g})(x, a) - g(x) = \hat{g}(x) - g(x) + o_P\left(\sum_{k=-n}^n w_{k,n}^2(x)\right)^{-1/2}.$$

Hence, if also

$$\frac{\max_{-n \leq k \leq n} |w_{k,n}(x)|}{\left(\sum_{k=-n}^n w_{k,n}^2(x)\right)^{1/2}} = o(1)$$

is fulfilled, then for any $x \in (0, 1)$ with $\theta''(x) > 0$ and any $a \in (0, 1)$ we have

$$\left(\sum_{k=-n}^n w_{k,n}^2(x)\right)^{1/2} \left(\rho_{h_d}(\hat{g})(x, a) - g(x) - b_{n,g}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

where $b_{n,g}$ is the bias of the unconstrained estimator \hat{g} .

For example, if we are interested in convex estimators of the j^{th} derivative of θ , $j \leq p - 2$ and assume that θ is at least $j + 2$ times continuously differentiable, we have from Birke, Bissantz and Holzmann (2007) the uniform convergence rates for $l = 0, 1$

$$\sup_{x \in [0,1]} |\hat{\theta}^{(j+l)}(x) - \theta^{(j+l)}(x)| = O_P\left(\frac{\log n}{nh^{2\beta+2(j+l)+1}a_n}\right)^{1/2}.$$

The above conditions are then fulfilled if

$$\begin{aligned} nh^{2\beta+2j+5+\delta}a_n &= O(1), \\ \frac{h_d}{h^{1+\delta/4}} &= o(1), \\ \text{and } \frac{\log n}{nh^{2\beta+2j+3}a_nh_d^2} &= o(1). \end{aligned}$$

In consequence,

$$\rho_{h_d}(\hat{\theta}^{(j)})(x, a) - \theta^{(j)}(x) = \hat{\theta}^{(j)}(x) - \theta^{(j)}(x) + o_P\left(\frac{1}{nh^{2\beta+2j+1}a_n}\right)^{1/2},$$

and since we have in this case

$$\max_{-n \leq k \leq n} |w_{j,k,n}(x)| = o\left(\frac{1}{nh^{2\beta+2j+1}a_n}\right)^{1/2}$$

(see Bissantz and Birke, 2007), we also find

$$\sqrt{nh^{2\beta+2j+1}a_n} \left(\rho_{h_d}(\hat{\theta}^{(j)})(x, a) - \theta^{(j)}(x) - b_{n,j}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma_j)$$

with $b_{n,j}$ and γ_j defined in (11) and (12), respectively.

Figure 2: Typical examples for increasing estimators of θ_1 (upper row, dashed curves) and convex estimators of θ_2 and θ_3 (middle and lower row, dashed curves) compared to the corresponding unconstrained estimators (dotted curves) and the true regression function θ (full curves). The regularization parameters are (from left to right) 0.1, 0.35 and 0.59, representing the undersmoothing, approximately L_2 -optimal and oversmoothing case, respectively. All estimates in a row have been computed from the same random dataset with sample size $2n + 1 = 201$ and noise std $\sigma = 0.1$.

5 Simulations

In this section we investigate the performance of the increasing and convex inverse estimators, respectively. In the first part, we discuss in a graphical way the impact of the shape constraints on the estimate, compared with the unconstrained estimator. Then, in the second part, we present the results of a simulation study for the increasing inverse estimator. To this end we compare the simulated integrated mean square errors of the unconstrained estimates with those for the increasing estimator. The simulations confirm the expectations from theory that the increasing estimator improves the L_p estimation error of the estimate (see Chernozhukov, Fernández-Val and Galichon, 2007).

5.1 A qualitative comparison

We begin our numerical investigation of the shape constrained estimators with a graphical comparison of the estimates for some example functions and regularization parameters. To this end we assume that at our disposal are observations

$$Y_k = (K\theta)(z_k) + \varepsilon_k, \quad k = -n, \dots, n$$

where K is the convolution operator on \mathbb{R} which causes the convolution of the function of interest θ with the function

$$\Psi(t) = \frac{3}{2} \cdot e^{-|3t|}.$$

Moreover, we assume that the noise terms ε_i are i.i.d. centered normal with variance σ^2 , and consider sample sizes $2n + 1 = 201$ (i.e. $n = 100$) and $2n + 1 = 2001$ (i.e. $n = 1000$). Finally, the design points are $z_k = \frac{k}{na_n}$ with $a_n = 0.25$.

As function of interest in our simulations we consider

$$\begin{aligned} \theta_1(t) &= \exp\left(-\frac{(x-1.1)^2}{2 \cdot 0.64}\right), \\ \theta_2(t) &= -\exp\left(-\frac{(x-1)^2}{2 \cdot 0.64}\right), \text{ and} \\ \theta_3(t) &= \left(x - \frac{1}{2}\right)^2 \cdot \exp\left(-\frac{(x-0.5)^2}{1.2^2}\right), \end{aligned}$$

Figure 3: Bandwidth dependent simulated L_2 -performance of the unconstrained (dotted) and the increasing (dashed) inverse estimator of θ_1 . Both distances have been scaled by a factor of 10 in the upper left and right panel, and a factor of 2 in the other cases. Full curves: ratio between the mean squared error of the increasing and the unconstrained estimator. Values below 1 indicate that the increasing estimator is better.

where θ_1 is strictly increasing on the interval $[0, 1]$, and θ_2 and θ_3 are convex on the interval $[0, 1]$. We will also use θ_1 in the subsequent simulation study of the increasing estimator (cf. section 5.2). Figure 2 compares the true regression functions θ_1 , θ_2 , and θ_3 with both the unconstrained inverse and the associated shape restricted estimates. The bandwidths for the deconvolution estimates have been selected by visual inspection such that the left panel in each row represents the undersmoothing case, the middle row the approximately optimal case, and the right panel the oversmoothing case. In the rearrangements we use $h_d = 0.03$. For the increasing regression function θ_1 , the shape restricted estimate improves the unconstrained estimate significantly for bandwidths below the approximately L_2 -optimal bandwidth. Here, the estimator apparently smoothes in a feasible way artificial oscillations in the unconstrained estimator due to the undersized bandwidth. In contrast to this, the oversmoothing estimate is already monotone. So, in this case it is obvious that the increasing estimator is nearly identical to the unconstrained input estimator, up to some additional smoothing dependent on the choice of h_d .

This is in general not the case for the convex estimator; here it is not obvious that the shape restricted estimate is better than the unconstrained version. However, if the unconstrained estimator is already convex, the convex estimator nearly coincides with the unconstrained one because of the same reasons as in the increasing case (see middle and right panel of the middle and lower row of Figure 2). In addition, the convex estimator still shares the important advantage with the increasing estimator that it provides estimates for a subsequent use which strictly obey the respective shape property. This is e.g. important in numerical simulations, which rely on estimated input quantities (such as simulations of stellar pulsations, cf. e.g. Grott et al., 2005), where monotonicity or convexity of the input function is required. Hence, the real advantage of the convex estimate is that it better represents θ w.r.t. its geometric properties.

5.2 A simulation study

We now present the results from a simulation study of the increasing estimator. To this end we have performed 100 simulations each with the true regression function θ_1 for a number of combinations of the sample size $2n + 1$, the noise variance σ^2 and the smoothing parameter h_d . Fig. 3 shows the simulated mean L_2 -error of the increasing and the unconstrained estimators, in dependence of the bandwidth h , and their ratio, where the value is below 1 if the increasing estimator is better than its unconstrained counterpart w.r.t. the L_2 -error.

From Figure 3, we draw the following conclusions. Independent of the choice of the parameters n, σ and h_d , the unconstrained and shape restricted estimators perform very similarly for large (oversmoothing) bandwidth. In more detail, for oversmoothing bandwidths, i.e. bandwidths larger than the L_2 -optimal bandwidth (which can be estimated visually from the figure) the mean L_2 -

Figure 4: Hubble Space Telescope/Wide Field Planetary Camera 2 [HST/WFPC2] R-band image of the elliptical galaxy NGC5017 at a distance of 40.3 Mpc (Lauer et al., 2007). Units along x and y axis are pixels. Image source: NASA/STScI, dataset U3CM1A01R.

Figure 5: Central luminosity profile of the elliptical galaxy NGC5017. Full curve: observations, dotted curve: unconstrained estimate, dashed curve: increasing estimate computed by the increasing rearrangements algorithm from the unconstrained estimate. See text for details.

error of the increasing estimator resembles closely the value of the increasing estimator. This is due to the fact, that in this case the unconstrained estimate is very smooth and in general already monotone on $[0, 1]$ (see also right column of Fig. 2). For undersmoothing bandwidths, the situation is different, that is the ratio is substantially smaller than 1. Here, the unconstrained estimator in general possesses artificial oscillations (also cf. left column of Fig. 2). Hence, in this case, the increasing estimator improves the resulting mean L_2 error substantially, which is in line with the expectation from theory (see Chernozhukov, Fernández-Val and Galichon, 2007).

A comparison of the relative performance of the unconstrained and increasing estimator in the various panels suggests an approximate independence of the performance on sample size $2n + 1$, or the variance of the noise σ^2 . In conclusion, the increasing estimate improves the inverse estimate for undersmoothing unconstrained input estimators, if it is known that the true function is monotone on the interval of interest. This is of substantial interest, since in practical applications with inverse estimators it is desired to be able to select the bandwidth as small as possible in order to reconstruct regions of high curvature in the true function (cf. e.g. the application in section 6). However, in the case of inverse estimation the estimates very rapidly decrease in quality for undersized bandwidth. This leads to substantial problems with data-driven bandwidth selection algorithms, since already selecting the bandwidth too small by some 10 – 20% yields unreasonable unconstrained estimates (cf. Bissantz et al., 2007a). Hence, we suggest to improve the reliability of data-driven inverse (regression) estimates by replacing the unconstrained estimate by the corresponding shape-restricted estimate. Whereas for undersized bandwidths the shape restricted estimator often significantly improves the estimate, it does not modify the result substantially for oversized bandwidths (in most cases). Finally, we remark that the observed improvement of the L_2 -error by the increasing estimator is consistent with the theoretical expectations, in contrast to the convex case, where this would not be expected. Indeed, some preliminary simulations showed no significant improvement in the mean L_2 -error of the estimator, independent of the bandwidth used for the unconstrained pilot estimator.

6 Luminosity profiles of elliptical galaxies

In this section we apply the increasing estimator for model (1) with convolution-type operators to the estimation of the luminosity profile of an elliptical galaxy for illustrative purposes. Fig. 4 shows

a HST/WFPC2 R-band image of the elliptical galaxy NGC5017, which will be considered in this example. An important problem for elliptical galaxies is the determination of the exact shape of the inner(most) luminosity profile of the galaxy, in particular the steepness of the galaxy’s brightness profile with decreasing radius: If the galaxy shows a steep inner increase, it belongs to the class of ”power-law” galaxies, but if the increase is significantly less steep in the inner part, it is a ”core galaxy” (e.g. Trujillo et al., 2004). The analysis of the profile is substantially complicated by the fact that an image from a telescope is not showing the true image of the galaxy; rather it shows its convolution with the point-spread-function [PSF] of the imaging device. If the image Θ is recorded on an equidistant grid of points, e.g. the pixels of a CCD (charge-coupled device, i.e. a digital imaging device), the data can be described by the model

$$Y_{j,k} = (\Theta * \Psi)(z_{j,k}) + \varepsilon_{j,k},$$

where Ψ is the (two-dimensional) PSF of the telescope, and the noise term $\varepsilon_{j,k}$ is assumed to be centered and with finite variance σ^2 . We note that the image data $Y_{j,k}$ are actually (approximately) Poisson distributed, which would e.g. make it impossible to compute the precise variance of the pointwise estimate of the luminosity profile with our methods. A possible approach to cope with this problem would be to apply the Anscombe transform to the data, which would yield data with nearly constant variance surface (Anscombe, 1948). However, the transformed data is not the convolution of the (transformed) true image Ψ with the PSF anymore. On the other hand, the unconstrained estimates of the spectral transform of $\Theta * \Psi$ include information from a large number of observations $Y_{j,k}$. For example, a spectral (Fourier) estimator is a weighted sum of the estimated Fourier coefficients, which by themselves are weighted sums of all observations. Hence, we expect the unconstrained estimator (i.e. the input to the rearrangement algorithm) to be approximately normal, as long as a sufficient number of observations is available. In our case, the image consists of 169×169 (i.e. close to 30000) pixel, which appears to be clearly sufficient for our purpose. The full curve in Figure 5 shows the observed luminosity profile of the galaxy along the x -coordinate of the image and through and to the left of the estimated center of the galaxy.

Our aim is now to recover the central part of the profile, where we use the fact, that the luminosity profile of an elliptical galaxy is at least approximately monotonically increasing towards the center. To this end, we proceed in two steps. First, we estimate the deconvolved image with an iterative inversion algorithm, and second, we use the increasing rearrangement algorithm to monotonize the profile.

For the reconstruction of the unconvolved image we use the Richardson-Lucy algorithm (e.g. Lucy, 1974), which is an iterative EM-algorithm. The PSF is estimated by fitting a bivariate Gaussian to the (very probable stellar) point-source on the left-middle of the image. We note that a more detailed analysis of the galaxy would require using a more sophisticated model for the PSF; moreover we use the HST data based on preview calibrations. However, such an analysis is beyond the scope of the exemplification of our method presented here. The dotted curve in Figure 5 shows the estimated luminosity profile, extracted from the reconstructed image. Since we are mainly interested in reconstructing the inner part of this profile, we have selected the stopping index of the Richardson-Lucy algorithm (i.e. the regularization parameter) by visual inspection of the image somewhat smaller than optimal, which results in some non-monotonic oscillations. The selection of a moderately undersmoothing bandwidth is supported by the simulation results presented in the

preceding section, which showed that the increasing estimator performs well even for moderately undersmoothing regularization parameters. This is important, because the larger the regularization parameter is, the more difficult it is to recover the central part of a galaxy's profile, where a peak in the profile would be smoothed out. Finally, the dashed line in Figure 5 shows the increasing estimator of the luminosity profile. A more detailed analysis of the shape of the luminosity profile, e.g. the range of radii where a Sérsic profile fits its shape well, could now be performed from the increasing estimate. It is important to note that this deconvolution analysis provides a smooth estimate of the shape, which is monotone as desired, without a priori assuming any parametric shape of the profile as it is commonly done when a power-law or Sérsic profile is fitted to the data.

Acknowledgements

The authors would like to thank Holger Dette for some helpful comments. Financial support of SFB 475 is gratefully acknowledged.

References

- F. J. Anscombe (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* 35, 246–254.
- J. Bewersdorf, A. Egner and S. W. Hell (2006). 4Pi Microscopy (in: Handbook of Biological Confocal Microscopy, ed. J. B. Pawley), New York: Springer.
- M. Birke (2007). Schätz- und Testverfahren in der nichtparametrischen Regression unter qualitativen Annahmen. PhD thesis, Ruhr-Universität Bochum (in German).
- M. Birke and H. Dette (2005). A note on estimating a monotone regression function by combining kernel and density estimates. *Technical report*.
- M. Birke and H. Dette (2007). Estimating a convex function in nonparametric regression. *Scand. J. Statist.* **34**, 384 - 404.
- N. Bissantz and H. Holzmann (2007). Asymptotic normality of spectral regularization estimators in statistical inverse problems. *Technical Report*.
- M. Birke, N. Bissantz and H. Holzmann (2007). Confidence bands and testing in error-in-variables regression models. *Preprint*
- N. Bissantz and M. Birke (2007). Asymptotic normality of spectral regularization estimators in error-in-variable models. *In preparation*
- N. Bissantz, L. Dümbgen, H. Holzmann and A. Munk (2007a). Nonparametric confidence bands in deconvolution density estimation. *J. Royal Statist. Society Ser. B.* 69, 483-506.

- N. Bissantz, T. Hohage, A. Munk and F. Ruymgaart (2007b). Convergence rates of general regularization methods for statistical inverse problems. *SIAM Journ. Numer. Anal.* (to appear)
- L. Cavalier (2000). Efficient estimation of a density in a problem of tomography. *Ann. Statist.* 28, 630-647.
- L. Cavalier and A. Tsybakov (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123, 323–354.
- V. Chernozhukov, I. Fernández-Val, A. Galichon (2007). Improving estimates of monotone functions by rearrangement. *Preprint*, arXiv:0704.368v1
- H. Dette, N. Neumeyer and K. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12, 469-490.
- B. van Es, G. Jongbloed and M. van Zuijlen (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.* 26, 2395-2406.
- R.L. Eubank (1999). *Nonparametric regression and spline smoothing*. Marcel Dekker, Inc., New York.
- J. Fan (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* 19, 1257-1272.
- J. Fan and Y. Truong (1993). Nonparametric regression with errors-in-variables. *Ann. Statist.* 21, 1900-1925.
- M. Grott, S. Chernigovski and W. Glatzel (2005). The simulation of non-linear stellar pulsations. *Monthly Not. Roy. Astron. Soc.* 360, 1532-1544.
- G.H. Hardy, J.E. Littlewood, G. Polya (1978). *Inequalities*. Cambridge Univ. Press, Cambridge.
- I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B* 66, 547-573.
- I.M. Johnstone and B. W. Silverman (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* 18, 251-280.
- G. Jongbloed and F.H. van der Meulen (2007). Estimating a concave distribution function from data corrupted with additive noise. Preprint
- T. Lauer, E. A. Ajhar, Y. Byun, A. Dressler, S.M. Faber, C. Grillmair, J. Kormendy, D. Richstone and S. Tremaine (1995). The Centers of Early-Type Galaxies with HST. I. An Observational Survey. *Astron. J.* 110, 2622-2654.

T. R. Lauer, K. Gebhardt, S.M. Faber, D. Richstone, S. Tremaine, J. Kormendy, M.C. Aller, R. Bender, A. Dressler, A.V. Filippenko, R. Green and L.C. Ho (2007). The Centers of Early-Type Galaxies with HST. VI. Bimodal Central Surface Brightness Profiles. *Astroph. J.* 664, 226-256.

L.B. Lucy (1974). An iterative technique for the rectification of observed distributions. *Astron. Journ.* 79, 745-754.

B.A. Mair and F. H. Ruymgaart (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56, 1424-1444.

I. Trujillo, P. Erwin, A.A. Ramos and A.W. Graham (2004). Evidence for a new elliptical-galaxy paradigm: Sérsic and core galaxies. *Astron. J.* 127, 1917-1942.

7 Appendix

Proof of Theorem 1. Similar to the approach in Dette, Neumeyer and Pilz (2006) we use the decomposition

$$\begin{aligned} \phi_{h_d}(\hat{\theta})(t) - \theta^{-1}(t) &= \phi_{h_d}(\theta)(t) - \theta^{-1}(t) - \frac{1}{h_d} \int_0^1 K_d\left(\frac{\theta(v) - t}{h_d}\right) (\hat{\theta}(v) - \theta(v)) dv \\ &\quad - \frac{1}{2h_d} \int_0^1 K_d'\left(\frac{\xi_n(v) - t}{h_d}\right) (\hat{\theta}(v) - \theta(v))^2 dv \\ &= \Delta_n^{(0)}(t) + \Delta_n^{(1)}(t) + \frac{1}{2} \Delta_n^{(2)}(t) \end{aligned}$$

where the last equation explains the notations $\Delta_n^{(i)}(t)$, $i = 0, 1, 2$ and $|\xi_n(v) - \theta(v)| \leq |\hat{\theta}(v) - \theta(v)|$ for all $v \in (0, 1)$. In the following we will show that $\Delta_n^{(1)}$ determines the asymptotic distribution and that $\Delta_n^{(0)}$ and $\Delta_n^{(2)}$ are asymptotically negligible. We start with $\Delta_n^{(1)}$. Note that

$$\begin{aligned} \Delta_n^{(1)}(t) &= - \sum_{k=-n}^n \tilde{w}_{k,n}(t) Y_k + \frac{1}{h_d} \int_0^1 K_d\left(\frac{\theta(v) - t}{h_d}\right) \theta(v) dv \\ &= \Delta_n^{(1.1)}(t) + \Delta_n^{(1.2)}(t) \end{aligned}$$

with

$$\tilde{w}_{k,n}(t) = \frac{1}{h_d} \int_0^1 K_d\left(\frac{\theta(v) - t}{h_d}\right) w_{0,k,n}(v) dv.$$

Therefore $\Delta_n^{(1.1)}$ is a weighted sum of the random variables Y_1, \dots, Y_n and we can apply a central limit theorem for weighted random variables (see e.g. Eubank, 1999). So if we can show that

$$\frac{\max_{-n \leq k \leq n} |\tilde{w}_{k,n}(t)|}{\left(\sum_{k=-n}^n \tilde{w}_{k,n}^2(t)\right)^{1/2}} \rightarrow 0,$$

then

$$\left(\sigma^2 \sum_{k=-n}^n \tilde{w}_{k,n}^2(t)\right)^{-1/2} \left(\Delta_n^{(1.1)}(t) - \mathbb{E}[\Delta_n^{(1.1)}(t)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Since the weights $w_{0,k,n}(t)$ are continuously differentiable in t we have for some ζ, ξ

$$\begin{aligned} \max_{-n \leq k \leq n} |\tilde{w}_{k,n}(t)| &\leq |(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))| \\ &\quad + h_d \int_{-1}^1 |v| K_d(v) |(\theta^{-1})''(\zeta)| dv \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))| \\ &\quad + h_d \int_{-1}^1 |v| K_d(v) |(\theta^{-1})'(t) (\theta^{-1})''(\xi)| \max_{-n \leq k \leq n} |w_{1,k,n}(\theta^{-1}(\xi))| dv \\ &\quad + h_d^2 \int_{-1}^1 |v| K_d(v) |(\theta^{-1})'(\xi) (\theta^{-1})''(\zeta)| \max_{-n \leq k \leq n} |w_{1,k,n}(\theta^{-1}(\xi))| dv \\ &= |(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))| \\ &\quad \times \left(1 + h_d \int_{-1}^1 |v| K_d(v) |(\theta^{-1})''(\zeta)| dv \frac{\max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|}{|(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|}\right) \\ &\quad + h_d \int_{-1}^1 |v| K_d(v) \frac{|(\theta^{-1})'(t) (\theta^{-1})''(\xi)| \max_{-n \leq k \leq n} |w_{1,k,n}(\theta^{-1}(\xi))|}{|(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|} dv \\ &\quad + h_d^2 \int_{-1}^1 |v| K_d(v) \frac{|(\theta^{-1})'(\xi) (\theta^{-1})''(\zeta)| \max_{-n \leq k \leq n} |w_{1,k,n}(\theta^{-1}(\xi))|}{|(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|} dv \end{aligned}$$

We know from Bissantz and Birke (2007) that

$$\max_{-n \leq k \leq n} |w_{j,k,n}(z)| = O\left(\frac{1}{nh^{\beta+j+1}a_n}\right),$$

$j = 0, 1$ and with condition (9) we get

$$\max_{-n \leq k \leq n} |\tilde{w}_{k,n}(t)| = |(\theta^{-1})'(t)| \max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))| (1 + o(1)).$$

A similar calculation yields for the denominator

$$\sum_{k=-n}^n \tilde{w}_{k,n}^2(t) = (\theta^{-1})'^2(t) \sum_{k=-n}^n w_{0,k,n}^2(t) (1 + o(1))$$

since

$$\sum_{k=-n}^n w_{j,k,n}^2(x) = O\left(\frac{1}{nh^{2\beta+2j+1}a_n}\right)$$

and (9) hold. This results in

$$\frac{\max_{-n \leq k \leq n} |\tilde{w}_{k,n}(t)|}{\left(\sum_{k=-n}^n \tilde{w}_{k,n}^2(t)\right)^{1/2}} = \frac{\max_{-n \leq k \leq n} |w_{0,k,n}(\theta^{-1}(t))|}{\left(\sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t))\right)^{1/2}} (1 + o(1)) \rightarrow 0$$

because of the condition on the weights of the unconstrained estimator. To show that it suffices to standardize with

$$\left(\sigma^2(\theta^{-1})'^2(t) \sum_{j=-n}^n w_{0,j,n}^2(\theta^{-1}(t))\right)^{-1/2}$$

instead of

$$\left(\sigma^2 \sum_{j=1}^n \tilde{w}_{j,n}^2(t)\right)^{-1/2}$$

we estimate

$$\frac{\Delta_n^{(1.1)}(t) - \mathbb{E}[\Delta_n^{(1.1)}(t)]}{\left(\sigma^2 \sum_{k=-n}^n \tilde{w}_{k,n}^2(t)\right)^{1/2}} = \frac{\Delta_n^{(1.1)}(t) - \mathbb{E}[\Delta_n^{(1.1)}(t)]}{\left(\sigma^2(\theta^{-1})'(t) \sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t))\right)^{1/2}} + A_n \frac{\Delta_n^{(1.1)}(t) - \mathbb{E}[\Delta_n^{(1.1)}(t)]}{\left(\sigma^2 \sum_{k=-n}^n \tilde{w}_{k,n}^2(t)\right)^{1/2}}$$

with

$$A_n = \frac{\left(\sigma^2(\theta^{-1})'^2(t) \sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t))\right)^{1/2} - \left(\sigma^2 \sum_{j=-n}^n \tilde{w}_{j,n}^2(t)\right)^{1/2}}{\left(\sigma^2(\theta^{-1})'(t) \sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t))\right)^{1/2}}.$$

For A_n we get

$$\begin{aligned} & |A_n| \\ &= \frac{|\left(\theta^{-1}\right)'(t) \sum_{j=-n}^n w_{0,j,n}^2(\theta^{-1}(t)) - \sum_{j=-n}^n \tilde{w}_{j,n}^2(t)|}{\left(\sigma^2(\theta^{-1})'(t) \sum_{j=-n}^n w_{0,j,n}^2(\theta^{-1}(t))\right)^{1/2} \left(\left(\theta^{-1}\right)'(t) \sum_{j=-n}^n w_{0,j,n}^2(\theta^{-1}(t)) + \sum_{j=-n}^n \tilde{w}_{j,n}^2(t)\right)^{1/2}} \\ &= \frac{N_n}{D_n}. \end{aligned}$$

The denominator D_n largely is the variance of $\hat{\theta}(\theta^{-1}(t))$ and therefore of order $O(1/nh^{2\beta+1}a_n)$. Hence, it suffices to show, that the numerator is of order $o(1/(nh^{2\beta+1}a_n))$. A straightforward calculation yields for the numerator

$$\begin{aligned} N_n &\leq (\theta^{-1})'^2(t) \sum_{k=-n}^n \left| \left(\int_{-1}^1 K_d(v) \{w_{0,k,n}(\theta^{-1}(t)) - w_{0,k,n}(\theta^{-1}(t + h_d v))\} dv \right) \right. \\ &\quad \left. \times \left(\int_{-1}^1 K_d(v) \{w_{0,k,n}(\theta^{-1}(t)) + w_{0,k,n}(\theta^{-1}(t + h_d v))\} dv \right) \right| (1 + o(1)) \end{aligned}$$

Since the weights are differentiable in an open interval around $\theta^{-1}(t)$ we get from a Taylor expansion of order 2

$$\begin{aligned} & \left| \int_{-1}^1 K_d(v) \{w_{0,k,n}(\theta^{-1}(t)) - w_{0,k,n}(\theta^{-1}(t + h_d v))\} dv \right| \\ &= h_d^2 \left((\theta^{-1})''(t) w_{1,k,n}(\theta^{-1}(t)) w_{0,k,n}(\theta^{-1}(t)) + (\theta^{-1})'^2(t) w_{2,k,n}(\theta^{-1}(t)) w_{0,k,n}(\theta^{-1}(t)) + o(1) \right) \end{aligned}$$

and

$$\int_{-1}^1 K_d(v) \{w_{0,k,n}(\theta^{-1}(t)) + w_{0,k,n}(\theta^{-1}(t + h_d v))\} dv = 2w_{0,k,n}(\theta^{-1}(t)) + o(1)$$

which eventually yields

$$N_n \leq O(h_d^2) \sum_{k=-n}^n \left(|w_{1,k,n}(\theta^{-1}(t))w_{0,k,n}(\theta^{-1}(t))| + |w_{2,k,n}(\theta^{-1}(t))w_{0,k,n}(\theta^{-1}(t))| \right).$$

An application of the Cauchy-Schwarz inequality now gives

$$N_n \leq O\left(h_d^2 \left(\sum_{k=-n}^n w_{2,k,n}^2(\theta^{-1}(t)) \right)^{1/2} \left(\sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t)) \right)^{1/2}\right).$$

The sums above are essentially the standard deviations of $\theta^{(j)}$, $j = 0, 2$, respectively which are of order $O(1/(nh^{2\beta+2j+1}a_n))^{1/2}$ (see Bissantz and Birke, 2007). Therefore we conclude that

$$N_n = O\left(\frac{h_d^2}{nh^{2\beta+3}a_n}\right) = o\left(\frac{1}{nh^{2\beta+1}a_n}\right)$$

and $|A_n| = o(1)$.

The second part $\Delta_n^{(1,2)}(t)$ together with the expectation of $\Delta_n^{(1,1)}(t)$ is

$$\begin{aligned} \mathbb{E}[\Delta_n^{(1,1)}(t)] + \Delta_n^{(1,2)}(t) &= -\frac{1}{h_d} \int_0^1 K_d\left(\frac{\theta(v) - t}{h_d}\right) \left[\sum_{k=-n}^n w_{0,k,n}(v)(K\theta)(z_k) - \theta(v) \right] dv \\ &= (\theta^{-1})'(t)b_n(\theta^{-1}(t)) + o\left(\frac{1}{nh^{2\beta+1}a_n}\right)^{1/2} \end{aligned}$$

Considering the deterministic part $\Delta_n^{(0)}$ we obtain as in Dette, Neumeyer and Pilz (2006)

$$\begin{aligned} \Delta_n^{(0)}(t) &= \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\theta(v) - u}{h_d}\right) dudv - \theta^{-1}(t) \\ &= \frac{1}{h_d} \int_0^{\theta^{-1}(t+h_d)} \int_{\theta(v)-h_d}^t K_d\left(\frac{\theta(v) - u}{h_d}\right) dudv - \theta^{-1}(t) \\ &= \theta^{-1}(t - h_d) \\ &\quad + \frac{1}{h_d} \int_0^1 I\{\theta^{-1}(t - h_d) \leq v \leq \theta^{-1}(t + h_d)\} \int_{\theta(v)-h_d}^t K_d\left(\frac{\theta(v) - u}{h_d}\right) dudv - \theta^{-1}(t) \\ &= \theta^{-1}(t - h_d) + h_d \int_{-1}^1 (\theta^{-1})'(t + zh_d) \int_z^1 K_d(w) dw dz - \theta^{-1}(t). \end{aligned}$$

Integration by parts now yields

$$\begin{aligned} \Delta_n^{(0)}(t) &= \int_{-1}^1 K_d(z)\theta^{-1}(t + h_d z) dz - \theta^{-1}(t) \\ &= \frac{h_d^2}{2} \int_{-1}^1 z^2 K_d(z) dz (\theta^{-1})''(t) + o(h_d^2). \end{aligned}$$

Therefore $\Delta_n^{(0)}$ is negligible since

$$\left(\sum_{k=-n}^n w_{0,k,n}^2(\theta^{-1}(t)) \right)^{1/2} = O(\sqrt{nh^{2\beta+1}a_n})$$

and with (9)

$$h_d^2 = o\left(\frac{1}{\sqrt{nh^{2\beta+1}a_n}}\right)$$

Because of the uniform convergence rates

$$\sup_{x \in [0,1]} |\hat{\theta}^{(j)}(x) - \theta^{(j)}(x)| = O_P\left(\frac{\log n}{nh^{2\beta+2j+1}a_n}\right)^{1/2}, \quad j = 0, 1, 2, 3 \quad (20)$$

- see Birke, Bissantz and Holzmann (2007) - we have

$$\left| \frac{\xi_n(v) - t}{h_d} - \frac{\theta(v) - t}{h_d} \right| \leq \frac{\sup_{v \in [0,1]} |\hat{\theta}(v) - \theta(v)|}{h_d} = O_P\left(\frac{\log n}{nh^{2\beta+1}a_nh_d^2}\right)^{1/2} \rightarrow 0$$

and hence, the second remainder $\Delta_n^{(2)}$ gives

$$\begin{aligned} \Delta_n^{(2)}(t) &= -\frac{1}{h_d^2} \int_0^1 K'_d\left(\frac{\xi(v) - t}{h_d}\right) (\hat{\theta}(v) - \theta(v))^2 dv \\ &= -\frac{1}{h_d^2} \int_0^1 K'_d\left(\frac{\theta(v) - t}{h_d}\right) (\hat{\theta}(v) - \theta(v))^2 dv (1 + o_P(1)) \\ &= O_P\left(\frac{\log n}{nh^{2\beta+1}a_nh_d}\right) = o_P\left(\frac{1}{nh^{2\beta+1}a_n}\right)^{1/2}. \end{aligned}$$

□

Proof of Theorem 2. In a first step we need a representation of $\phi_{h_d}(\hat{\theta})^{-1}(x) - \theta(x)$ in terms of $\phi_{h_d}(\hat{\theta})(\theta(x)) - \theta^{-1}(\theta(x))$. Then we can conclude from Theorem 1 that the assertion of Theorem 2 holds. As in Birke and Dette (2005) we obtain by the mean value theorem

$$\phi_{h_d}(\hat{\theta})^{-1}(x) - \theta(x) = -\frac{\phi_{h_d}(\hat{\theta})(\theta(x)) - \theta^{-1}(\theta(x))}{(\phi_{h_d}(\hat{\theta}))'(\xi_n(x))}$$

where $|\xi_n(x) - \theta(x)| \leq |\phi_{h_d}(\hat{\theta})^{-1}(x) - \theta(x)|$. By Slutsky's theorem it now suffices to show that

$$|\phi_{h_d}(\hat{\theta})'(\xi_n(x)) - (\theta^{-1})'(\theta(x))| \rightarrow 0$$

in probability for $n \rightarrow \infty$. As in Theorem 3.5 of Birke and Dette (2005) we obtain the uniform convergence

$$\sup_{t \in [0,1]} |\phi_{h_d}(\hat{\theta})(t) - \theta^{-1}(t)| = O_P\left(\frac{\log n}{nh^{2\beta+1}a_n}\right)^{1/2} + O_P\left(\frac{\log n}{nh^{2\beta+1}a_nh_d^2}\right) + O(h_d^2) = o_P(1) \quad (21)$$

$$\sup_{t \in [0,1]} |\phi_{h_d}(\hat{\theta})'(t) - (\theta^{-1})'(t)| = O_P\left(\frac{\log n}{nh^{2\beta+1}a_nh_d^2}\right)^{1/2} + O_P\left(\frac{\log n}{nh^{2\beta+1}a_nh_d^2}\right) + O(h_d) = o_P(1) \quad (22)$$

by again using the uniform convergence rates in (20) (for a detailed proof of equivalent results for kernel regression estimates see Birke, 2007).

From (21) it follows, that $\xi_n(x)$ converges in probability to $\theta(x)$. Observe that

$$|\phi_{h_d}(\hat{\theta})'(\xi_n(x)) - (\theta^{-1})'(\theta(x))| \leq |\phi_{h_d}(\hat{\theta})'(\xi_n(x)) - (\theta^{-1})'(\xi_n(x))| + |(\theta^{-1})'(\xi_n(x)) - (\theta^{-1})'(\theta(x))|.$$

For big enough n the quantity $\xi_n(x)$ lies in $[0, 1]$ and therefore, with (22), the first part on the right side of the above equation converges to 0. The second part converges to 0 because of the continuity of $(\theta^{-1})'$.

This assures the convergence in probability of $\phi_{h_d}(\hat{\theta})'(\xi_n(x))$ to $(\theta^{-1})'(\theta(x))$ and we complete the proof of Theorem 2 by applying Slutsky's theorem. \square

Proof of Theorem 3. The first part of the proof is similar to that in Birke and Dette (2007). We have to estimate

$$\hat{\rho}(x, a) - \theta(x) = \int_a^x (\phi_{h_d}(\hat{\theta}')^{-1}(z) - \theta'(z)) dz + \hat{\theta}(a) - \theta(a)$$

In a first step we need a representation of the increasing rearrangement of $\hat{\theta}'$. As in the proof of Theorem 2 this is

$$\phi_{h_d}(\hat{\theta}')^{-1}(u) - \theta'(u) = -\frac{\phi_{h_d}(\hat{\theta}')(\theta'(u)) - \theta'^{-1}(\theta'(u))}{[\phi_{h_d}(\hat{\theta}')]'\xi_n(u)}$$

with $|\xi_n(u) - \theta'(u)| \leq |\phi_{h_d}(\hat{\theta}')^{-1}(u)|$. This yields for the estimate $\hat{\rho}$ in (13) the representation

$$\hat{\rho}(x, a) - \theta(x) = A_n(x) + \left(\hat{\theta}(a) - \theta(a) \right)$$

with

$$A_n(x) = - \int_a^x \frac{\phi_{h_d}(\hat{\theta}')(\theta'(u)) - \theta'^{-1}(\theta'(u))}{(\theta^{-1})'(u)} du (1 + o_p(1))$$

where we used the fact, that $[\phi_{h_d}(\hat{\theta}')]'\xi_n(u)$ converges uniformly to $(\theta'^{-1})'(u)$. This can be seen from a similar argumentation as in the proof of Theorem 2,

$$\begin{aligned} \sup_{t \in J} |\phi_{h_d}(\hat{\theta}')(t) - \theta'^{-1}(t)| &= O_P\left(\frac{\log n}{nh^{2\beta+3}a_n}\right)^{1/2} + O(h_d) = o_P(1) \\ \sup_{t \in J} |\phi_{h_d}(\hat{\theta}')'(t) - (\theta'^{-1})'(t)| &= O_P\left(\frac{\log n}{nh^{2\beta+3}a_n h_d^2}\right)^{1/2} + o(1) = o_P(1) \end{aligned}$$

and

$$|(\theta'^{-1})'(\xi_n(u)) - (\theta'^{-1})'(\theta'(u))| = o(1)$$

because ξ_n uniformly converges to θ' and $(\theta'^{-1})'$ is uniformly continuous. For $A_n(x)$ we obtain the decomposition

$$A_n(x) = \int_a^x A_n(u) du = A_{n,1}(x) + A_{n,2}(x) + A_{n,3}(x)$$

with

$$\begin{aligned}
A_{n,1}(x) &= -\frac{1}{h_d} \int_0^1 \int_{\theta'(a)}^{\theta'(x)} K_d\left(\frac{\theta'(v) - u}{h_d}\right) du (\hat{\theta}(v) - \theta(v))' dv \\
A_{n,2}(x) &= -\frac{1}{h_d^2} \int_0^1 \int_{\theta'(a)}^{\theta'(x)} K_d'\left(\frac{\xi(v) - u}{h_d}\right) (\hat{\theta}'(v) - \theta'(v))^2 dv du \\
A_{n,3}(x) &= -\int_{\theta'(a)}^{\theta'(x)} (\phi_{h_d}(\theta')(u) - \theta'^{-1}(u)) du
\end{aligned}$$

Except for the last step, the estimation of the terms $A_{n,i}$, $i = 1, 2, 3$ does not depend on the structure of the unconstrained estimator $\hat{\theta}$. Therefore we obtain similarly as in Birke and Dette (2007) or Birke (2007)

$$\begin{aligned}
A_{n,1}(x) &= (\hat{\theta} - \theta)(x) - (\hat{\theta} - \theta)(a) + h_d \int_{-1}^1 u K_d(u) (\theta'^{-1})'(\zeta_{x,u}) (\hat{\theta} - \theta)'(\theta'^{-1}(\zeta_{x,u})) du \\
&\quad - h_d \int_{-1}^1 u K_d(u) (\theta'^{-1})'(\zeta_{a,u}) (\hat{\theta} - \theta)'(\theta'^{-1}(\zeta_{a,u})) du \\
A_{n,2}(x) &= -\int_{-1}^1 K_d(v) (\theta'^{-1})'(\theta'(x) + h_d v) (\hat{\theta}' - \theta')^2 (\theta'^{-1}(\theta'(x) + h_d v)) dv \\
&\quad + \int_{-1}^1 K_d(v) (\theta'^{-1})'(\theta'(a) + h_d v) (\hat{\theta}' - \theta')^2 (\theta'^{-1}(\theta'(a) + h_d v)) dv \\
A_{n,3}(x) &= -h_d^2 \left(\kappa_2(K_d) \left(\frac{1}{\theta''(x)} - \frac{1}{\theta''(a)} \right) + o(1) \right).
\end{aligned}$$

with $|\zeta_{x,t} - \theta'(x)| \leq h_d |t| \leq h_d$.

Now observe, that from Birke, Bissantz and Holzmann (2007) we have the uniform convergence rates

$$\sup_{z \in [0,1]} |\hat{\theta}^{(j)}(z) - \theta^{(j)}(z)| = O_P\left(\frac{\log n}{nh^{2\beta+2j+1}a_n}\right)^{1/2} \text{ for } j = 1, 2$$

and hence

$$\begin{aligned}
A_{n,1} &= \hat{\theta}(x) - \theta(x) - (\hat{\theta}(a) - \theta(a)) + O_P\left(\frac{h_d \log n}{nh^{2\beta+5}a_n}\right)^{1/2} \\
&= \hat{\theta}(x) - \theta(x) - (\hat{\theta}(a) - \theta(a)) + o_P\left(\frac{1}{nh^{2\beta+1}a_n}\right)^{1/2} \\
A_{n,2} &= O_P\left(\frac{\log n}{nh^{2\beta+3}a_n}\right) = o_P\left(\frac{1}{nh^{2\beta+1}a_n}\right)^{1/2} \\
A_{n,3} &= O(h_d^2) = o\left(\frac{1}{nh^{2\beta+1}a_n}\right)^{1/2}
\end{aligned}$$

This finishes the proof of Theorem 3.