Statistica Sinica **30** (2020), 1255-1283
doi:https://doi.org/10.5705/ss.202018.0160

REGULARIZATION PARAMETER SELECTION IN INDIRECT REGRESSION BY RESIDUAL BASED BOOTSTRAP

Nicolai Bissantz, Justin Chown and Holger Dette

Ruhr-Universität Bochum

Abstract: Residual-based analysis is generally considered a cornerstone of statistical methodology. For a special case of indirect regression, we investigate a residual-based empirical distribution function and provide a uniform expansion of this estimator, which is also shown to be asymptotically most precise. This investigation naturally leads to a completely data-driven technique for selecting the regularization parameter used in our indirect regression function estimator. The resulting methodology is based on a smooth bootstrap of the model residuals. A simulation study demonstrates the effectiveness of our approach.

Key words and phrases: Bandwidth selection, indirect regression estimator, inverse problems, regularization, residual-based empirical distribution function, smooth bootstrap.

1. Introduction

In many experiments, we can make only indirect observations of the physical process being observed. However, although important quantities of interest are not directly available for statistical inferences in these so-called *inverse problems*, images of these quantities under some transformation, such as a convolution, can be used instead. Here, we consider an inverse regression model. We observe a signal of interest from indirect observations

$$Y_j = \big[K\theta\big](x_j) + \varepsilon_j, \qquad j = -n, \ldots, n, \tag{1.1}$$

where K is an operator specifying a convolution of the true underlying regression θ with a distortion function ψ; that is,

$$\big[K\theta\big](x_j) = \int_{-1/2}^{1/2} \theta(u)\psi(x_j - u)\, du.$$

The resulting function $K\theta$ can be viewed as a distorted regression function. We assume that θ is a smooth periodic function, a common assumption in many

inverse problems. According to Tsybakov (2009), this means that the Fourier coefficients of θ are assumed to satisfy a crucial technical summability requirement (see Section 2 for further details).

We assume that ψ is known and behaves like a probability density function on the interval $[-1/2, 1/2]$, that is, ψ is positive-valued on the interval $[-1/2, 1/2]$ and integrates to one. Later, we specify further technical requirements for ψ. However, to ensure that the convolution operation is well defined, ψ must be extended periodically to the intervals $[x-1/2, x+1/2]$, for each $x \in [-1/2, 1/2]$. The covariates x_j in model (1.1) are uniformly distributed design points in the interval $[-1/2, 1/2]$; that is, $x_j = j/2n$, for $j = -n, \ldots, n$. In addition, the errors ε_j are assumed to be independent, have mean zero, and have a common distribution function F. Note that the assumptions given above only guarantee that model (1.1) is a well defined indirect regression model, where θ is identifiable (see, e.g.,Cavalier and Golubev (2006); Mair and Ruymgaart (1996)).

Statistical inverse problems have received much attention related to constructing estimators for various densities and indirect regression models. In particular, early works consider the properties of estimators for a range of important statistical inverse problems. Examples of such works include Masry (1991), who investigates estimators of a multivariate density function in an errors-in-variables model, using a deconvolution technique; Fan (1991), who derives the optimal rates of convergence for the density estimators in these models; and Masry (1993), who investigates estimators of a smooth multivariate regression function using deconvolution techniques when the estimation includes contaminated covariates.

Later studies on statistical inverse problems such as those considered in (1.1) yield a better understanding of the asymptotic properties of the estimators from a theoretical perspective. Here, important results include those of Mair and Ruymgaart (1996), who estimate an indirect regression function in a flexible model based on Hilbert scales. This includes the popular case of Sobolev classes, where the authors describe general regularization approaches for operator inversion, and show that the considered estimators are, in fact, minimax optimal. Cavalier and Tsybakov (2002) investigate an indirect heteroscedastic regression model, and prove the minimax optimality of their estimators. Moreover, Cavalier (2008) surveys the available literature for deconvolution estimators, and provides minimax rates of reconstructions for several models, including that in (1.1).

More recently, statistical testing and model selection properties have been considered in statistical inverse problems of the type in (1.1). Bissantz and Holzmann (2008) describe how to construct confidence intervals and confidence bands

in univariate statistical inverse problems. Later, Proksch, Bissantz and Dette (2015) generalize the univariate case in the previous study, and construct confidence bands for an indirect regression function of multiple covariates. Marteau and Mathé (2014) test for distorted signals using general regularization schemes.

The aforementioned deconvolution estimators are all based on projections of the data and result in kernel-type estimators that depend on some kind of regularization parameter. This quantity is analogous to the bandwidth found in the usual nonparametric function estimators. The data-driven selection of this parameter is an important problem that we examine closely in this article. In general, the techniques used to choose a sequence of regularization parameters focus on choosing a suitable estimator of the integrated mean squared error of an indirect regression estimator, or on choosing some other related quantity (see Section 3). Cavalier and Golubev (2006) make a particularly important contribution to this problem by investigating the integrated mean squared error of indirect regression estimators. Furthermore, they propose a suitably penalized quantity based on a threshold of this important estimation performance metric. The authors call this a *risk hull* approach because of the resulting bowl-shaped objective function used to choose the parameter sequence. From another perspective, we can consider potential bootstrap approaches to this problem, where we instead calculate the integrated mean squared error of a bootstrap version of the indirect regression estimator. The bootstrap method for choosing the regularization parameter sequence appears to be particularly promising compared with the risk hull approach.

We provide a statistical methodology for selecting a best fitting (most feasible) regression estimator from a sequence of function estimators, based on observations from model (1.1), and using the resulting model residuals constructed from the estimator $\hat{\theta}$; see (2.2):

$$\hat{\varepsilon}_j = Y_j - \big[K\hat{\theta}\big](x_j), \qquad j = -n, \ldots, n.$$

Many statistical procedures are residual-based, including the bootstrap methodology for selecting the regularization that we investigate. This requires that we first study the distribution function F of the model errors, which is usually unknown and must be estimated.

To the best of our knowledge, this topic has not been studied before with respect to statistical deconvolutions in a completely nonparametric setting. We form an estimator of F using the empirical distribution function of the model

residuals:

$$\hat{\mathbb{F}}(t) = \frac{1}{2n+1} \sum_{j=-n}^{n} \mathbf{1}\big[\hat{\varepsilon}_j \le t\big] = \frac{1}{2n+1} \sum_{j=-n}^{n} \mathbf{1}\big[Y_j - \big[K\hat{\theta}\big](x_j) \le t\big], \qquad t \in \mathbb{R}.$$

The estimator $\hat{\theta}$ is shown to be a suitable estimator of θ, such that we can study the limiting behavior of $\hat{\mathbb{F}}$, which is new. In addition, this work reveals that new and stronger conditions are required on the smoothness of θ in order for $\hat{\mathbb{F}}$ to be a consistent estimator of F. Hence, any residual-based inference procedure relying on $\hat{\mathbb{F}}$ also requires this stronger smoothness condition, for example, Kolmogorov–Smirnov-type and Cramér–von-Mises-type statistics.

Studying these problems requires new results related to the estimator $\hat{\theta}$ and its bootstrap analog. The literature on statistical deconvolution problems is extensive, and, hence, some results will be familiar. In particular, we show that the estimator $\hat{\theta}$ has a strong uniform rate of consistency for the function θ that is analogous to the already known minimax optimal rate of convergence (see Theorem 1 in Section 2.1 and Remark 3 in Section 3). There are also many results in the literature on residual-based empirical distribution functions for direct regression models; for example, uniform consistency and asymptotic optimality. We show that the estimator $\hat{\mathbb{F}}$ satisfies both of these properties (see Theorem 2 and Remark 2 in Section 2.1). The residual-based empirical distribution functions resulting from a wide class of semiparametric direct regression models are studied by Müller, Schick and Wefelmeyer (2007), and we derive comparable results for the indirect regression model (1.1).

The remainder of the paper proceeds as follows. Further notation and the estimation method are introduced in Section 2, along with the asymptotic results for the estimators $\hat{\theta}$ and $\hat{\mathbb{F}}$. In Section 3, we consider the problem of finding an optimal regularization parameter for the estimator $\hat{\theta}$. Here, we provide a rule-of-thumb approach in the spirit of Silverman (1986) and, in Section 3.1, we develop a data-driven approach for selecting this parameter using a smooth bootstrap of the model residuals, following Neumeyer (2009). We conclude the article with a numerical study in Section 4, which indicates good finite-sample performance of the proposed data-driven regularization against that of the theoretically optimal regularization. In addition, we consider a comparative technique for choosing the regularization for spectral cutoff estimators (a special case of our approach) proposed by Cavalier and Golubev (2006). All proofs are available in the online Supplementary Material.

2. Estimation Using The Indirect Regression Model

We begin with the space of square integrable functions $\mathcal{L}_2([-1/2,\, 1/2])$ with domain $[-1/2,\, 1/2]$. This function space has the well known and countable orthonormal basis

$$\left\{ e^{i2\pi kx} \,:\, x \in \left[-\frac{1}{2}, \frac{1}{2}\right] \right\}_{k\in\mathbb{Z}}.$$

In order to construct an estimator for the function θ, we need to restrict θ to a smooth class of functions from $\mathcal{L}_2([-1/2,\, 1/2])$. Thus, we only consider functions θ that are weakly differentiable in $\mathcal{L}_2([-1/2,\, 1/2])$.

For clarity, we now introduce some notation. Let $d \in \mathbb{N}$. We call $q^{(i)}$, for $1 \leq i \leq d$, a weak derivative of q in $\mathcal{L}_2([-1/2,\, 1/2])$ of order i, if $q^{(i)} \in \mathcal{L}_2([-1/2,\, 1/2])$ and $q^{(i)}$ satisfies

$$\int_{-1/2}^{1/2} q(x) \frac{d^i}{dx^i}\phi(x)\, dx = (-1)^i \int_{-1/2}^{1/2} q^{(i)}(x)\phi(x)\, dx,$$

for every infinitely differentiable function ϕ with support $[-1/2,\, 1/2]$ that has evaluations of ϕ, $(d^{i'}\phi)/(dx^{i'})$, for $i' = 1, \ldots, i$, at $1/2$ and $-1/2$ equal to zero. The corresponding space of smooth periodic functions is the Sobolev space $\mathcal{W}^{2,d}([-1/2,\, 1/2])$, where

$$\begin{aligned}\mathcal{W}^{2,d}\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) &= \left\{ q \in \mathcal{L}_2\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) \,:\, q^{(1)}, \ldots, q^{(d)} \in \mathcal{L}_2\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) \right\} \\ &= \left\{ q \in \mathcal{L}_2\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) \,:\, \sum_{k=-\infty}^{\infty} \left(1 + k^2\right)^d |\rho(k)|^2 < \infty \right\}.\end{aligned}$$

Here, $\{\rho(k)\}_{k\in\mathbb{Z}}$ are the Fourier coefficients of q:

$$\rho(k) = \int_{-1/2}^{1/2} q(u) e^{-i2\pi ku} du, \qquad k \in \mathbb{Z}.$$

Replacing d with a positive real number motivates the consideration of smoothness orders $s > 0$; that is, $\mathcal{W}^{2,s}([-1/2,\, 1/2])$ is defined in the same way as $\mathcal{W}^{2,d}([-1/2,\, 1/2])$; but with s in place of d. We require that θ satisfies a stronger series condition than that stated for $\mathcal{W}^{2,d}([-1/2,\, 1/2])$ above. Following Condition C_1 in Politis and Romano (1999), which is similar to a condition imposed in

Watson and Leadbetter (1963), we restrict $\mathscr{R}_s$ to a subspace of $\mathcal{W}^{2,s}([-1/2,\, 1/2])$:

$$\mathscr{R}_s = \left\{ q \in \mathcal{W}^{2,s}\left(\left[-\frac{1}{2}, \frac{1}{2}\right]\right) \,:\, \sum_{k=-\infty}^{\infty} |k|^s |\rho(k)| < \infty \right\}.$$

Note that $\theta \in \mathscr{R}_s$ implies a restriction on the Fourier coefficients $\{\Theta(k)\}_{k\in\mathbb{Z}}$ of θ, which are defined similarly to the Fourier coefficients $\{\rho(k)\}_{k\in\mathbb{Z}}$ above.

Another important note related to the Fourier basis $\{\exp(i2\pi kx) : x \in [-1/2,\, 1/2]\}_{k\in\mathbb{Z}}$ is that it decomposes the operator K into a singular value decomposition along each of the orthonormal basis functions. Here, we need only consider the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ of the distortion function ψ, which are defined similarly to the Fourier coefficients $\{\rho(k)\}_{k\in\mathbb{Z}}$ above. Much of the research in the area of deconvolution problems has focused on two important cases. The first case is that of the so-called ordinarily smooth distortion functions. Here, we assume that the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ decay at a polynomic rate: there is some $b > 0$, such that $|\Psi(k)| \sim |k|^{-b}$. Here, "$\sim$" denotes asymptotic similarity. Under this assumption, we can construct an estimator $\hat{\theta}$ for θ with a strong uniform consistency rate that is comparable, albeit worse, with the rates expected in the usual nonparametric regression case. In addition, we can show that the estimator $\hat{\mathbb{F}}$ is both root-n consistent for F, uniformly in $t \in \mathbb{R}$, and asymptotically most precise. The second case is that of the so-called super smooth distortion functions. Here, we assume that the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ decay at an exponential rate, for example, $|\Psi(k)| \sim \exp(-|k|^b)$. Under this assumption, the resulting indirect regression estimator has a strong uniform consistency rate that is polynomic in the logarithm of n only, which we expect is too slow for us to maintain the root-n consistency of $\hat{\mathbb{F}}$. Therefore, in this paper, we focus on the first case of ordinarily smooth distortion functions ψ, employing a similar assumption to (1.4) of Fan (1991).

Assumption 1. There are finite constants $b > 0$, $\Gamma > 0$, and $0 < C_\Psi < C_\Psi^*$, such that for every $|k| > \Gamma$, the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ of ψ satisfy $C_\Psi < |k|^b |\Psi(k)| < C_\Psi^*$.

Example 1. Suppose ψ is known to be the standard Laplace density function restricted to the interval $[-1/2,\, 1/2]$; that is,

$$\psi(x) = \frac{(1/2)e^{-|x|}}{\int_{-1/2}^{1/2} (1/2)e^{-|x|}\, dx} = \frac{(1/2)e^{-|x|}}{1 - e^{-1/2}}, \qquad x \in \left[-\frac{1}{2}, \frac{1}{2}\right].$$

Then, the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ are given by

$$\Psi(k) = \frac{1}{1+4\pi^2k^2}\frac{1-e^{i\pi|k|}e^{-1/2}}{1-e^{-1/2}}, \qquad k\in\mathbb{Z}.$$

Thus, Assumption 1 is satisfied for the choices $b=2$, $\Gamma=1$, $C_\Psi=(1+4\pi^2)^{-1}$, and $C_\Psi^*=5(4\pi^2)^{-1}$.

Recall that we use a uniform fixed design on the interval $[-1/2, 1/2]$. Writing Q for the conditional distribution of a response Y, given a fixed design point x, results in the equivalence $Q(y\,|\,x)=P_x(Y\le y)$, where P_x denotes the distribution of Y depending on x, which is not random. It follows that we can write the Fourier coefficients $\{R(k)\}_{k\in\mathbb{Z}}$ of $K\theta$ as

$$R(k) = \int_{-1/2}^{1/2}\int_{-\infty}^{\infty} ye^{-i2\pi kx}\,Q(dy\,|\,x)\,dx, \qquad k\in\mathbb{Z}. \tag{2.1}$$

The double integral on the right-hand side of (2.1) is an average. We can construct an estimator for this average from an empirical average using data (x_j, Y_j), for $j=-n,\dots,n$, obtaining

$$\hat{R}(k) = \frac{1}{2n+1}\sum_{j=-n}^{n} Y_j e^{-i2\pi kx_j}, \qquad k\in\mathbb{Z}.$$

To recover θ from the convolution $K\theta$, we use the convolution theorem for the Fourier transformation: $R(k)=\Theta(k)\Psi(k)$, $k\in\mathbb{Z}$. Because ψ is a probability density function that is bounded away from zero on $[-1/2, 1/2]$, it follows that $\{\Psi(k)\}_{k\in\mathbb{Z}}$ is bounded away from zero in absolute value on any bounded region $\mathcal{Z}\subset\mathbb{Z}$. Hence, Ψ^{-1} is well defined (see, e.g., the discussion on preconditioning on page 1425 of Mair and Ruymgaart (1996)). Observing that the Fourier transformation reduces the convolution to a multiplication, we exploit the Fourier inversion formula by writing

$$\theta(x) = \sum_{k=-\infty}^{\infty}\frac{R(k)}{\Psi(k)}e^{i2\pi kx}, \qquad x\in\left[-\frac{1}{2},\,\frac{1}{2}\right].$$

To substitute our estimated Fourier coefficients $\{\hat{R}(k)\}_{k\in\mathbb{Z}}$ for the Fourier coefficients $\{R(k)\}_{k\in\mathbb{Z}}$, we need to control the random fluctuations that occur at high frequency spectra. That is, the inversion of the operator K in (1.1) requires regularization; see Cavalier and Golubev (2006) for a clear discussion on

regularization and ill-posedness.

Politis and Romano (1999) introduce spectral smoothing to control these fluctuations at higher frequencies, which amounts to regularizing the inversion operator, following Mair and Ruymgaart (1996). Consider the ratio $|\hat{R}(k)|/|\Psi(k)|$, which becomes large as $|k|$ increases. We wish to utilize lower frequencies and dampen the contributions from higher frequencies by introducing a sequence of weights. The most striking difference between the approaches taken by Politis and Romano (1999) and that of Mair and Ruymgaart (1996) is that the latter require the regularization in order to preserve the fundamental Fourier frequency. That is, the regularization must be equal to one around some neighborhood of the zero-th Fourier frequency. This approach to regularization is easy to specify for applications and leads to the desired optimality properties.

Let us now introduce some notation. Write $\{h_n\}_{n\geq 1}$ for a regularizing sequence that satisfies $h_n \to 0$, as $n \to \infty$. Consider smoothing kernel functions similar to those used in typical nonparametric function estimators, that is, maps $x \mapsto h_n^{-1}\delta(x/h_n)$, where δ is a suitably constrained probability density function. Politis and Romano (1999) observe that the Fourier transform of a smoothing kernel $h_n^{-1}\delta(x/h_n)$ takes the form $\Lambda(h_n k)$, where Λ is the Fourier transform of the desired kernel function K_Λ. Thus, the Fourier transform of a smoothing kernel depends on n only through the regularizing sequence $\{h_n\}_{n\geq 1}$ by shrinking the Fourier frequency from k to $h_n k$. We require that our smoothing kernel has a Fourier transform Λ that satisfies the following general assumption.

Assumption 2. The region $I = \{k \in \mathbb{Z} : |k| \leq M\}$ exists for some integer $M \geq 1$, such that $\Lambda(k) = 1$ when $k \in I$, and $|\Lambda(k)| \leq 1$ otherwise. Let Λ satisfy $\int_{-\infty}^{\infty} |u|^b |\Lambda(u)|\, du < \infty$, where $b > 0$ is the degree of ill-posedness introduced in Assumption 1.

To facilitate discussion, we introduce the order notation $O(a_n)$ when there are sequences $\{a_n\}_{n\geq 1}$ (of positive real numbers) and $\{b_n\}_{n\geq 1}$ satisfying $a_n^{-1} b_n \to L$, for some finite constant L, and we write $o(a_n)$ when $L = 0$. Similarly, we write O_P and o_P when the analogous statements hold on an event with probability tending to one as the sample size (depending on n) increases. Assumption 2 ensures that only the estimation bias has a desirable rate of convergence: order $O(h_n^s)$, when $\theta \in \mathscr{R}_s$. This is comparable with the direct estimation setting of sufficiently high-order kernels, or the so-called "superkernels" (see, e.g., the discussion on page 3 of Politis and Romano (1999)). The idea of restricting the choice of the smoothing kernel function based on obtaining a suitable rate of

convergence in the estimation bias dates back to Parzen (1962).

An estimator of θ is given by a kernel smoother:

$$\hat{\theta}(x) = \sum_{k=-\infty}^{\infty} \Lambda(h_n k) \frac{\hat{R}(k)}{\Psi(k)} e^{i2\pi kx} = \frac{1}{2n+1} \sum_{j=-n}^{n} Y_j W_{h_n}\big(x - x_j\big), \quad x \in \left[-\frac{1}{2}, \frac{1}{2}\right], \tag{2.2}$$

where the smoothing kernel W_{h_n} is given by

$$W_{h_n}\big(x - x_j\big) = \sum_{k=-\infty}^{\infty} \frac{\Lambda(h_n k)}{\Psi(k)} \exp\Big(i2\pi k\big(x - x_j\big)\Big).$$

The smoothing kernel W_{h_n} is sometimes called a deconvolution kernel (see, e.g., Birke, Bissantz and Holzmann (2010)).

2.1. Asymptotic results for the deconvolution estimator and the empirical distribution function of the residuals

Our first result specifies the asymptotic order of the bias of $\hat{\theta}$.

Lemma 1. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$, and let Assumptions 1 and 2 hold. Then, for any regularizing sequence $\{h_n\}_{n\geq 1}$ satisfying $h_n \to 0$ and $nh_n^{b+1} \to \infty$, as $n \to \infty$, we have*

$$\sup_{x\in[-1/2,\,1/2]} \Big|E\big[\hat{\theta}(x)\big] - \theta(x)\Big| = O\big(h_n^s + (nh_n^{b+1})^{-1}\big).$$

The asymptotic order of the bias of $\hat{\theta}$ is affected by the degree of ill-posedness of the inverse problem, which we expect can be made negligible by the choice of regularization parameters $\{h_n\}_{n\geq 1}$. In the following result, we observe this detrimental effect in the asymptotic order of consistency as well.

Lemma 2. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$, and let Assumptions 1 and 2 hold. Assume that Λ satisfies $\int_{-\infty}^{\infty} |u|^{b+1}|\Lambda(u)|\,du < \infty$, and that the random variables $Y_{-n}, \ldots, Y_n$ have a finite absolute moment of order $\kappa > 2 + 1/b$. Finally, let the regularizing sequence $\{h_n\}_{n\geq 1}$ satisfy $h_n \to 0$, such that $(nh_n^{2b+1})^{-1}\log(n) \to 0$, as $n \to \infty$. Then,*

$$\sup_{x\in[-1/2,\,1/2]} \Big|\hat{\theta}(x) - E\big[\hat{\theta}(x)\big]\Big| = O\Big(\big(nh_n^{2b+1}\big)^{-1/2}\log^{1/2}(n)\Big), \qquad a.s..$$

The two lemmas above imply that we can obtain a strong uniform rate of convergence of the estimator $\hat{\theta}$ for θ by choosing a regularizing sequence $\{h_n\}_{n\geq 1}$

that balances the asymptotic orders of both the bias and consistency; that is,

$$h_n = O\big(n^{-1/(2s+2b+1)} \log^{1/(2s+2b+1)}(n)\big). \tag{2.3}$$

For this choice of regularizing parameters, we have $(nh_n^{b+1})^{-1} = o(h_n^s)$, which implies that the bias of $\hat{\theta}$ is of order $O(h_n^s)$. Note that Lemma 2 requires that the responses have a finite moment of order larger than $2 + 1/b$, which is only a sufficient condition. One can easily show that $\kappa > 2+1/(s+b)$ is necessary when $\{h_n\}_{n\geq 1}$ satisfies (2.3), which is more reasonable for situations when $b \to 0$. We now state the uniform rate of convergence of $\hat{\theta}$ for θ when the parameter sequence $\{h_n\}_{n\geq 1}$ satisfies (2.3), as well as two additional properties of the estimator $\hat{\theta}$.

Theorem 1. *Let the assumptions of Lemma* 2 *hold, but now require only* $\kappa > 2 + 1/(s + b)$. *Choose the regularizing sequence* $\{h_n\}_{n\geq 1}$ *to satisfy* (2.3). *Then,*

$$\sup_{x\in[-1/2,\,1/2]} \Big|\hat{\theta}(x) - \theta(x)\Big| = O\big(n^{-s/(2s+2b+1)} \log^{s/(2s+2b+1)}(n)\big), \qquad a.s..$$

In addition, if $s > (2b + 1)/(2\gamma)$, *for some* $0 < \gamma \leq 1$, *then*

$$\left[\sup_{x\in[-1/2,\,1/2]} \Big|\hat{\theta}(x) - \theta(x)\Big|\right]^{1+\gamma} = o(n^{-1/2}), \qquad a.s..$$

If Λ *satisfies* $\int_{-\infty}^{\infty} |u|^{s+b-1/2}|\Lambda(u)|\, du < \infty$, *then, for sufficiently large* n,

$$\hat{\theta} - \theta \in \mathscr{R}_{s-1/2,1}, \qquad a.s.,$$

where $\mathscr{R}_{s-1/2,1} = \{q \in \mathscr{R}_{s-1/2} : \|q\|_\infty \leq 1\}$ *is the unit ball of the metric space* $(\mathscr{R}_{s-1/2}, \|\cdot\|_\infty)$.

Remark 1. The second statement of Theorem 1 requires that the smoothness index s of the function space $\mathscr{R}_s$ be larger than the degree of ill-posedness b of the inverse problem, which is a stronger requirement than that specified in the literature. The additional smoothness is simply explained by the entanglement of the smoothness index s and the degree of ill-posedness b in the strong uniform consistency rate given in the first statement of Theorem 1: $O(n^{-s/(2s+2b+1)} \log^{s/(2s+2b+1)}(n))$. This entanglement also occurs for indirect regression estimators that satisfy minimax optimality, where the integrated mean squared error is now of order $O(n^{-(2s)/(2s+2b+1)})$.

We are now ready to state our main results for the estimator $\hat{\mathbb{F}}$.

Theorem 2. *Assume the error distribution function F admits a bounded Lebesgue density function f that is Hölder continuous with exponent $0 < \gamma \leq 1$, and let $\varepsilon_{-n}, \ldots, \varepsilon_n$ have a finite absolute moment of order $\kappa > 2 + 1/(s+b)$. Let the remaining assumptions of Theorem 1 be satisfied, but now with $s > \max\{(2b+1)/(2\gamma), 3/2\}$. Then,*

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \Big\{ \mathbf{1}\big[\hat{\varepsilon}_j \leq t\big] - \mathbf{1}\big[\varepsilon_j \leq t\big] - \varepsilon_j f(t) \Big\} \right| = o_P(n^{-1/2}).$$

Corollary 1. *Under the conditions of Theorem 2, the process*

$$\begin{aligned} \mathbb{G}_n(t) &= (2n+1)^{1/2}\{\hat{\mathbb{F}}(t) - F(t)\} \\ &= (2n+1)^{-1/2} \sum_{j=-n}^{n} \Big\{ \mathbf{1}\big[\varepsilon_j \leq t\big] - F(t) + \varepsilon_j f(t) \Big\} + o_P(1), \end{aligned}$$

for $t \in \mathbb{R}$, weakly converges to a mean zero Gaussian process $\{Z(t) : t \in \mathbb{R}\}$, with the following covariance function, for $u, v \in \mathbb{R}$:

$$\begin{aligned} \Sigma(u,v) =& F\big(\min\{u, v\}\big) - F(u)F(v) \\ &+ f(u)E\big[\varepsilon \mathbf{1}[\varepsilon \leq v]\big] + f(v)E\big[\varepsilon \mathbf{1}[\varepsilon \leq u]\big] + \sigma^2 f(u)f(v). \end{aligned}$$

Here, we write $\sigma^2 = E[\varepsilon^2]$ and ε for a generic random variable with distribution function F.

Remark 2. Model (1.1) is a nonparametric regression. The estimator $\hat{\mathbb{F}}$ has influence function $\mathbf{1}[\varepsilon \leq t] - F(t) + \varepsilon f(t)$, where ε is a generic random variable with distribution function F. If we additionally assume that F has finite Fisher information for location, it follows that $\hat{\mathbb{F}}$ is efficient for estimating F, in the sense of Hájek and Le Cam, from the results of Müller, Schick and Wefelmeyer (2004).

3. Regularization Parameter Selection and The Smooth Bootstrap of residuals

We now consider the problem of choosing an appropriate sequence of regularization parameters $\{h_n\}_{n\geq 1}$ required by the estimator $\hat{\theta}$. Popular approaches in the literature suggest that a practical choice of regularization is a scheme that minimizes the integrated mean squared error (IMSE) of $\hat{\theta}$. However, the selection of such a parameter can also be viewed as a model selection problem, where

we select the *most feasible* regression model from a sequence of regression function estimators generated from a sequence of regularization parameters. In the case of iterative estimation procedures, a suitable stopping iteration is sought. Multiscale and related methods based on partial sums of normalized residuals have been thoroughly investigated in the literature (see, e.g., González-Manteiga, Martinez-Miranda and Pérez-González (2004); Bissantz, Mair and Munk (2006, 2008); Davies and Meise (2008); Hotz et al. (2012)). The Lepski methodology has recently become a popular approach in this context, where the IMSE of the indirect regression estimator is replaced by a suitable nonrandom objective function using oracle inequalities (see, e.g., Goldenshluger (1999); Cavalier and Tsybakov (2002); Mathé and Pereverzev (2006); Blanchard and Mathé (2012); Blanchard, Hoffmann and Reiß (2016)). An important approach for spectral cutoff estimators, based on assessing a risk hull, is investigated by Cavalier and Golubev (2006), which we have already discussed. In contrast to previous works, we propose a methodology based on a smooth bootstrap of the model residuals to form a consistent estimator of the IMSE of $\hat{\theta}$. Furthermore, from the perspective of conducting model selection, we propose choosing the regularization parameter sequence that minimizes this quantity.

In the following result, we give the asymptotic order of the integrated variance and the integrated squared bias of the estimator $\hat{\theta}$ that will lead to a rule-of-thumb approach for selecting regularization parameters that approximately minimize the IMSE of $\hat{\theta}$.

Proposition 1. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$, and let Assumptions 1 and 2 hold. Assume that $\varepsilon_{-n}, \ldots, \varepsilon_n$ have finite variance σ^2. Then, for any regularizing sequence $\{h_n\}_{n\geq 1}$ satisfying $h_n \to 0$, such that both $nh_n^{2b+1} \to \infty$, as $n \to \infty$, and $(nh_n^{b+1})^{-1} = o(h_n^s)$ hold, there are constants $C_\Lambda > 0$ and $C_R > 0$, such that*

$$\int_{-1/2}^{1/2} E\Big[\big\{\hat{\theta}(x) - E\big[\hat{\theta}(x)\big]\big\}^2\Big]\, dx = C_\Lambda \sigma^2 (nh_n^{2b+1})^{-1} + o\big((nh_n^{2b+1})^{-1}\big)$$

and

$$\int_{-1/2}^{1/2} \Big\{E\big[\hat{\theta}(x)\big] - \theta(x)\Big\}^2\, dx = C_R h_n^{2s} + o\big(h_n^{2s}\big).$$

Remark 3. From the results of Proposition 1, we can obtain an approximately optimal regularizing sequence, in the sense of minimizing the IMSE of $\hat{\theta}$:

$$h_{n,opt} \approx \left(\frac{2b+1}{2s}\frac{C_\Lambda}{C_R}\sigma^2\right)^{1/(2s+2b+1)} n^{-1/(2s+2b+1)}.$$

Consequently, the integrated mean squared error of $\hat{\theta}$ is of order $O(n^{-(2s)/(2s+2b+1)})$. Setting $\epsilon = \epsilon_n = O(n^{-1/2})$ in Table 1 on page 9 of Cavalier (2008) yields that $\hat{\theta}$ is indeed minimax optimal for estimating θ.

The conclusion that $\hat{\theta}$, formed from a regularizing sequence of order $O(n^{-1/(2s+2b+1)})$, is minimax optimal only guarantees that the estimation strategy is optimal, in the sense that it both minimizes the rate of convergence for the integrated mean squared error, a measure of estimation performance, and that no other estimator will achieve a faster rate of convergence for this performance metric. However, as we can see from Remark 3, the choice of regularizing parameters $\{h_{n,opt}\}_{n\geq 1}$ requires further investigation using numerical methods, because some unknown constants are not directly estimable. For example, working with the approximately optimal bandwidth choice in Remark 3, the constant C_Λ is proportional to the limit of $h_n^{2b+1}\sum_{k=-\infty}^{\infty}\{\Lambda(h_n k)/\Psi(k)\}^2$, which can be approximated by a finite series and a pilot regularizing sequence. On the other hand, C_R is essentially an asymptotically stabilized bias. Usually, this is not observable and, hence, a numerical method such as bootstrap or cross-validation is required to estimate it. In addition, and more generally, the optimal bandwidth depends on the unknown smoothness index s of the function space $\mathscr{R}_s$. Estimating this quantity is very difficult and likely not even possible without harsh and confining assumptions. However, an educated guess yields the optimal bandwidth choice corresponding to the fastest possible decay of the IMSE of $\hat{\theta}$. This means choosing s as large as possible. Unfortunately, the resulting methodology is still arbitrary.

3.1. Smooth bootstrap of residuals

Computational approaches for automated and data-driven bandwidth selection methods have been well studied in the literature for many nonparametric function estimators. In general, the approaches focus on estimating the IMSE of the estimator using either a cross-validation or a bootstrap approach, which can then be minimized with respect to the choice of bandwidth in an exact or approximate way. Cao (1993) studies two methods for selecting a bandwidth in a kernel density estimator using a smooth bootstrap of their univariate data. More recently, Neumeyer (2009) has proven the general validity of a smooth bootstrap process of the model residuals from a nonparametric regression. Owing to its simplicity, we introduce a similar smooth bootstrap process that admits a consistent estimator of the IMSE of $\hat{\theta}$, which requires mirroring the restrictions

given by Theorem 2 on model (1.1) in the bootstrap scheme. Throughout this section, we describe the stochastic properties of our random quantities using a smooth bootstrap measure P^*, which, for a single bootstrap response Y^*, is the conditional probability function

$$P_x^*(Y^* \le t) = P_x(Y^* \le t \,|\, \mathbb{D}) = P_x(\varepsilon^* \le t - [K\hat{\theta}](x) \,|\, \mathbb{D}),$$

given the original sample of data $\mathbb{D} = \{(x_{-n}, Y_{-n}), \ldots, (x_n, Y_n)\}$. Here, ε^* is a smooth bootstrap model residual, which we construct as follows.

We begin by examining the requirements imposed by Theorem 2 on model (1.1). We need to ensure our smooth bootstrap model residual ε^* satisfies having a mean equal to zero, independence, a finite moment of order $\kappa > 2 + 1/(s+b)$, and a common distribution function F_n^* that admits a bounded Lebesgue density function f_n^* that is Hölder continuous. The first requirement is satisfied merely by centering our original model residuals:

$$\tilde{\varepsilon}_j = \hat{\varepsilon}_j - \frac{1}{2n+1} \sum_{l=-n}^{n} \hat{\varepsilon}_l, \qquad j = -n, \ldots, n.$$

Turning our attention to the next constraint, we can see that conditioning on the original sample $\mathbb{D}$ and selecting from $\tilde{\varepsilon}_{-n}, \ldots, \tilde{\varepsilon}_n$, completely at random and with replacement, satisfies independence under P^* (and, therefore, conditionally on the observed data $\mathbb{D}$). However, the remaining assumptions are not satisfied because resampling in this way results in the bootstrap model residuals $\tilde{\varepsilon}_j^*$ having a discrete distribution.

To fulfill the last requirements imposed on model (1.1), we contaminate the randomly selected centered model residual $\tilde{\varepsilon}_j^*$ using an independent, centered random variable U_j that has a finite moment of order $\kappa > 2 + 1/(s+b)$ and a common distribution function characterized by a bounded Lebesgue density function w. Hence, we construct our smooth bootstrap model residuals $\varepsilon_{-n}^* = \tilde{\varepsilon}_{-n}^* + c_n U_{-n}, \ldots, \varepsilon_n^* = \tilde{\varepsilon}_n^* + c_n U_n$. Here, the sequence $\{c_n\}_{n \ge 1}$ is a scaling sequence similar to a bandwidth for a kernel density estimation. Consequently, ε_j^* has the common distribution function

$$F_n^*(t) = P^*(\varepsilon_j^* \le t) = \frac{1}{(2n+1)c_n} \sum_{j=-n}^{n} \int_{-\infty}^{t} w\left(\frac{u - \tilde{\varepsilon}_j}{c_n}\right) du, \qquad t \in \mathbb{R}, \quad (3.1)$$

and density function

$$f_n^*(t) = \frac{1}{(2n+1)c_n} \sum_{j=-n}^{n} w\left(\frac{t-\tilde{\varepsilon}_j}{c_n}\right), \qquad t \in \mathbb{R}.$$

We can see that F_n^* is a smooth estimator of F based on a kernel density estimator f_n^* of the original error density f. Hence, the remaining requirement imposed by Theorem 2 on F can be mirrored in the bootstrap process by our choice of w; that is, we can choose w to be Hölder continuous with the desired exponent. Using model (1.1), we obtain our bootstrap sample $(x_{-n}, Y_{-n}^*), \ldots, (x_n, Y_n^*)$, where

$$Y_j^* = \big[K\hat{\theta}\big](x_j) + \varepsilon_j^*, \qquad j = -n, \ldots, n.$$

Define $\hat{\theta}^*$ as in (2.2), but substitute Y_j^* for Y_j and the regularizing sequence $\{g_n\}_{n\geq 1}$ for the regularizing sequence $\{h_n\}_{n\geq 1}$, which is also chosen to satisfy (2.3). Choosing the scaling sequence $\{c_n\}_{n\geq 1}$ such that $c_n = O(n^{-\alpha})$, for some $0 < \alpha < 1/2 + 1/\kappa < 1$, results in the bootstrap indirect regression estimator $\hat{\theta}^*$ satisfying similar properties to those of $\hat{\theta}$ given in Theorem 1. We summarize these results in Proposition 4 in the Supplementary Material.

In practice, an important use of bootstrapping is to find suitable quantiles for test statistics. In the case of a residual-based analysis, one is typically interested in continuous functionals $T(Z_{F_0})$ of a Gaussian process Z_{F_0}, parameterized under a null hypothesis $H_0 : F = F_0$ against an alternative hypothesis $H_a : F = F_a \neq F_0$. A test statistic $T_n = T(\hat{\mathbb{G}}_n)$ is convenient for assessing the adequacy of the null hypothesis H_0, but this quantity depends on the unknown error distribution F. Neumeyer (2009) uses a smooth bootstrapping of residuals obtained from nonparametric smoothing in a direct regression model to approximate quantiles of the limiting distribution of T_n, using a bootstrap version $T_n^* = T(\mathbb{G}_n^*)$ of this quantity, where $\mathbb{G}_n^* = (2n+1)^{1/2}\{\hat{\mathbb{F}}^* - F_n^*\}$ is the smooth bootstrap analog of $\mathbb{G}_n$. We therefore expect our results to be analogous to those of Neumeyer (2009).

In the following, we work with residuals constructed from the following bootstrap data:

$$\hat{\varepsilon}_j^* = Y_j^* - \big[K\hat{\theta}^*\big](x_j), \qquad j = -n, \ldots, n.$$

The following result is the analog of Theorem 2 for the empirical distribution function of these residuals. The proof of this result follows along the same lines as the proof of Theorem 2 and its supporting results (see the online Supplementary Material). These have been omitted, for brevity.

Theorem 3. *Assume the density function w is Hölder continuous, with exponent $0 < \gamma \leq 1$. Let the assumptions of Proposition 4 from the Supplementary Material be satisfied, with $s > \max\{(2b+1)/(2\gamma), 3/2\}$. Then,*

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \left\{ \mathbf{1}\big[\hat{\varepsilon}_j^* \leq t\big] - \mathbf{1}\big[\varepsilon_j^* \leq t\big] - \varepsilon_j^* f_n^*(t) \right\} \right| = o_{P^*}(n^{-1/2}).$$

Note that this result always includes the optimal bandwidth choice $c_n = O(n^{-1/5})$ for density estimation. This fact, together with the results of Proposition 3 from the Supplementary Material yield the following analog of Corollary 1.

Corollary 2. *Let the assumptions of Theorem 3 be satisfied. If, additionally, both densities f and w are Hölder continuous with exponent $2/3 < \gamma \leq 1$, the scaling sequence $\{c_n\}_{n\geq 1}$ satisfies $c_n = O(n^{-1/5})$, and $s > (1+\gamma)(2b+1)/(3\gamma-2)$, then the process*

$$\begin{aligned}(2n+1)^{-1/2} &\sum_{j=-n}^{n} \left\{ \mathbf{1}\big[\hat{\varepsilon}_j^* \leq t\big] - F_n^*(t) \right\} \\ &= (2n+1)^{-1/2} \sum_{j=-n}^{n} \left\{ \mathbf{1}\big[\varepsilon_j^* \leq t\big] - F_n^*(t) + \varepsilon_j^* f_n^*(t) \right\} + o_{P^*}(1),\end{aligned}$$

for $t \in \mathbb{R}$, weakly converges, conditionally on the sample $(x_{-n}, Y_{-n}), \ldots, (x_n, Y_n)$, to a mean zero Gaussian process $\{Z^*(t) : t \in \mathbb{R}\}$, with covariance function, for $u, v \in \mathbb{R}$,

$$\begin{aligned}\Sigma^*(u,v) = F_n^*\big(\min\{u, v\}\big) - F_n^*(u)F_n^*(v) &+ f_n^*(u)E^*\big[\varepsilon^*\mathbf{1}[\varepsilon^* \leq v]\big] \\ &+ f_n^*(v)E^*\big[\varepsilon^*\mathbf{1}[\varepsilon^* \leq u]\big] + \sigma^{2,*} f_n^*(u) f_n^*(v),\end{aligned}$$

where ε^* is a generic random variable with distribution function F_n^* and $\sigma^{2,*} = E^*[(\varepsilon^*)^2]$. Additionally, we have

$$\sup_{u,v\in\mathbb{R}} \Big| \Sigma^*(u,v) - \Sigma(u,v) \Big| = o_P(1),$$

where Σ is given in Corollary 1.

Following the observations on pages 207–209 in Neumeyer (2009), we immediately obtain valid smooth bootstrap approximations of the quantiles of the test statistics T_n. We conclude this section with the following remark.

Remark 4. The residual-based empirical process $\mathbb{G}_n$ and its smooth bootstrap analog $\mathbb{G}_n^*$ have the same limiting distribution when the conditions of Corollary 2 are satisfied. This limiting distribution is given by the Gaussian process described in Corollary 1, which has continuous sample paths. It then follows that statistics $T_n = T(\mathbb{G}_n)$ and their smooth bootstrap version $T_n^* = T(\mathbb{G}_n^*)$, obtained from continuous functionals, satisfy the following consistency property. Define $q_{n,\alpha}^*$ by $P^*(T_n^* \leq q_{n,\alpha}^*) = \alpha$. Combining the continuity of the functional used to construct T_n and T_n^* and the continuous sample paths of Gaussian processes using the continuous mapping theorem, we obtain

$$P\big(T_n \leq q_{n,\alpha}^*\big) = \alpha + o(1),$$

which characterizes the validity of the proposed smooth bootstrap of the model residuals. Hence, the bootstrap described here can be used to approximate the unknown quantiles of test statistics obtained from continuous functionals of Z_{F_0}.

3.2. Regularization parameter selection by bootstrap

Now, we turn our attention to a different choice of regularization parameters that also approximately minimizes the IMSE of the indirect regression estimator $\hat{\theta}$. For clarity, throughout this section, we subscript the estimators $\hat{\theta}$ and $\hat{\theta}^*$ using the regularization parameters used to form them; that is, we write $\hat{\theta}_{h_n}$ to indicate that the regularizing sequence $\{h_n\}_{n\geq 1}$ is used to form the estimator $\hat{\theta}$. The IMSE of $\hat{\theta}_{h_n}$, which we want to minimize with respect to the parameter sequence $\{h_n\}_{n\geq 1}$, is given by

$$IMSE\big(\hat{\theta}_{h_n}\big) = \int_{-1/2}^{1/2} E\Big[\big\{\hat{\theta}_{h_n}(x) - \theta(x)\big\}^2\Big]\, dx, \tag{3.2}$$

which can be viewed as an objective function with respect to the mapping $h_n \mapsto IMSE(\hat{\theta}_{h_n})$.

Following Cao (1993), we arbitrarily choose the original regularizing sequence $\{h_n\}_{n\geq 1}$ according to Theorem 1 as a pilot sequence to form an initial and consistent estimator $\hat{\theta}_{h_n}$. A practical choice for $\{h_n\}_{n\geq 1}$ is the rule-of-thumb parameter sequence given in Remark 3, where the unknown constants are estimated and the smoothness index s is chosen to be small. It is crucial that, in order for our approach to admit an asymptotically nearly optimal choice of regularizing parameters, the pilot sequence $\{h_n\}_{n\geq 1}$ is chosen such that $s_0 - 1/2 \leq s < s_0$, where s_0 is the largest possible (finite) smoothness index such that $\theta \in \mathscr{R}_{s_0}$.

Consider the IMSE objective, but for the bootstrap data, where we instead have $\hat{\theta}_{h_n}$ for the unknown function θ in (3.2). Hence, we have an analogous form of (3.2) in the smooth bootstrap measure P^* that can be approximated via Monte Carlo simulation:

$$IMSE^*\big(\hat{\theta}^*_{g_n}\big) = \int_{-1/2}^{1/2} E^*\Big[\big\{\hat{\theta}^*_{g_n}(x) - \hat{\theta}_{h_n}(x)\big\}^2\Big]\, dx. \tag{3.3}$$

Because both $\hat{\theta}_{h_n}$ and $\hat{\theta}^*_{g_n}$ satisfy the projective representation (2.2), it follows that the expected values on the far right-hand sides of (3.2) and (3.3) are averages, taken with respect to the distribution functions F and F^*_n, respectively. We can then use standard arguments to show

$$\begin{aligned} &E^*\bigg[\int_{-1/2}^{1/2}\big\{\hat{\theta}^*_{g_n}(x) - \hat{\theta}_{h_n}(x)\big\}^2\, dx\bigg] \\ &\qquad = E\bigg[\int_{-1/2}^{1/2}\big\{\hat{\theta}_{g_n}(x) - \theta(x)\big\}^2\, dx\bigg]\Big\{C + o_P(1)\Big\} + o_P(1), \end{aligned}$$

for a constant $C > 0$. Hence, we obtain $IMSE^*(\hat{\theta}^*_{g_n}) = CIMSE(\hat{\theta}_{g_n}) + o_P(1)$. This implies (3.3) is a close predictor of (3.2), and, thus, we can use the mapping $g_n \mapsto IMSE^*(\hat{\theta}^*_{g_n})$ as an objective criterion for finding a nearly optimal regularizing sequence. It follows that we can choose $\{g_{n,opt}\}_{n\geq 1}$, such that

$$g_{n,opt} = \operatorname*{arg\,min}_{g\in(0,\,\hbar]} E^*\bigg[\int_{-1/2}^{1/2}\big\{\hat{\theta}^*_{g}(x) - \hat{\theta}_{h_n}(x)\big\}^2\, dx\bigg], \tag{3.4}$$

where $\hbar > 0$ is a constant chosen larger than the optimal regularization parameter. Consequently, the resulting regularization parameters $\{g_{n,opt}\}_{n\geq 1}$ can be viewed as objective corrections to the subjective pilot regularization parameters $\{h_n\}_{n\geq 1}$.

Recall the Fourier frequency smoothing kernel Λ used in the deconvolution estimators $\hat{\theta}_{h_n}$ and $\hat{\theta}^*_{g_n}$. It is easy to see that restricting the choice of Λ, and, hence, restricting the choice of the resulting deconvolution smoothing kernel from (2.2), leads to unique minimizers for each of (3.2) and (3.3). For example, choosing Λ as an indicator function (e.g., working with spectral cutoff estimators) leads to the deconvolution smoothing kernel in (2.2) being a smooth function with infinitely many derivatives.

Remark 5. The bootstrap methodology for finding an asymptotically nearly

optimal regularizing sequence critically requires that the pilot sequence $\{h_n\}_{n\geq 1}$ be chosen to undersmooth the estimator $\hat{\theta}$. Interestingly, this contrasts with the bootstrap methodology outlined in Cao (1993), who recommends using an oversmoothing pilot bandwidth to identify an optimal bandwidth to use with a kernel density estimator. Apparently, pilot bandwidths g_n satisfying $ng_n^3 \to \infty$, $g_n = O(n^{-1/7})$, or $g_n = O(n^{-2/13})$ are satisfactory in that case.

4. Finite-sample Properties

We conclude this article with a small numerical study of the previous results, and investigate the effectiveness of our smooth bootstrap methodology for selecting a regularization parameter. In the following simulations, we choose two regression functions θ_1 and θ_2, where

$$\theta_1(x) = \frac{6}{5} + \sqrt{2}\cos\big(2\pi x\big) + \frac{\sqrt{2}}{4}\cos\big(4\pi x\big)$$

and

$$\theta_2(x) = \frac{3}{2} + \frac{\sqrt{2}}{2}\cos\big(2\pi x\big) - \frac{\sqrt{2}}{8}\cos\big(4\pi x\big) - \frac{4\sqrt{2}}{3}\cos\big(6\pi x\big),$$

with $x \in [-1/2, 1/2]$. On this interval, the function θ_1 is similar in shape to a unimodal probability density function, whereas the function θ_2 is more complicated. For greater transparency, one can easily check that θ_1 and θ_2 belong to the restricted space $\mathscr{R}_s$ for $s > 0$ using their respective Fourier coefficients $\{\Theta_1(k)\}_{k\in\mathbb{Z}}$ and $\{\Theta_2(k)\}_{k\in\mathbb{Z}}$, given by

$$\Theta_1(k) = \frac{6}{5}\mathbf{1}[k=0] + \frac{\sqrt{2}}{2}\mathbf{1}[|k|=1] - \frac{\sqrt{2}}{8}\mathbf{1}[|k|=2], \quad k \in \mathbb{Z},$$

and

$$\Theta_2(k) = \frac{3}{2}\mathbf{1}[k=0] + \frac{\sqrt{2}}{4}\mathbf{1}[|k|=1] - \frac{\sqrt{2}}{16}\mathbf{1}[|k|=2] - \frac{2\sqrt{2}}{3}\mathbf{1}[|k|=3], \quad k \in \mathbb{Z}.$$

The distortion function ψ is taken as the Laplace density with mean zero and a scale of $1/10$ that has been restricted to the interval $[-1/2, 1/2]$, as in Example 1, which also satisfies Assumption 1 for the choice $b = 2$. The fixed covariates are taken as $x_j = j/(2n+1)$, which is asymptotically equivalent to $j/(2n)$. This choice allows us to use the fast Fourier transform algorithm to estimate the functions θ_1 and θ_2. Finally, we consider two cases for the model errors: normally distributed errors, with mean zero and scale $2/3$, and t-distributed errors, with

four degrees of freedom and scale 2/3. Our simulations consider samples of sizes 51, 101, 201, and 301; that is, n is taken as 25, 50, 100, and 150.

We work with the smoothing kernel that has Fourier coefficients satisfying

$$\Lambda(k) = \begin{cases} 1 & \text{if } |k| \le 7, \\ \left(\frac{|k|}{7}\right)^{-6} & \text{if } 7 < |k| \le n, \\ 0 & \text{otherwise,} \end{cases}$$

which leads to considering function spaces $\mathscr{R}_s$ with $5/2 < s < 7/2$. In order to select an appropriate regularization parameter for the indirect regression function estimators, we use the pilot sequences $h_{n,1} = 5(2n+1)^{-1/11}\log^{1/11}(2n+1)$, which correspond to a choice $s = 3$ in (2.3), to estimate θ_1, and use $h_{n,2} = 2.5(2n+1)^{-1/11}\log^{1/11}(2n+1)$ to estimate θ_2.

To create the smooth bootstrap of the residuals, we use standard normally distributed contaminates U_j and Silverman's rule to select a bandwidth in kernel density estimation; that is, we take the scaling sequence $c_n = 1.06\hat{\sigma}(2n+1)^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation of the model residuals obtained using the pilot regularizing sequence. Using 200 smooth bootstrap replications to construct suitable approximations of the IMSE of the estimators of θ_1 and θ_2, we take 100 equally spaced candidate regularization parameters in an interval $[l_n, u_n]$, where $l_n = (2n+1)^{-1/10}$, which results in undersmoothed estimators, and $u_n = 10(2n+1)^{-1/12}\log^{1/12}(2n+1)$, which results in oversmoothed estimators. Following the discussion in Section 3.2, we choose the optimal regularization parameter $g_{n,opt}$ as the grid point that minimizes this approximate IMSE, which we then use to construct the resulting function estimators of θ_1 and θ_2.

The assumptions of Theorem 2 are satisfied for the choices made above. Figure 1 displays the results of our indirect regression estimator for a typical data set obtained from the indirect regression θ_2 and t-distributed errors based on a sample size of 201. The scatter plot of the data shows the function estimators $\hat{\theta}$ and $K\hat{\theta}$ work well in terms of estimating θ_2 and $K\theta_2$, respectively. Clearly, the indirect regression estimator, constructed using the proposed data-driven regularization methodology, is explaining the data very well, which is supported by the appearance of the completely random scatter in the plot of the residuals. The plot of the distribution functions shows that the empirical distribution function of the residuals $\hat{\mathbb{F}}$ matches very closely the true error distribution function F, as expected.

Turning our attention to the numerical summaries of the estimator $\hat{\mathbb{F}}$, we can

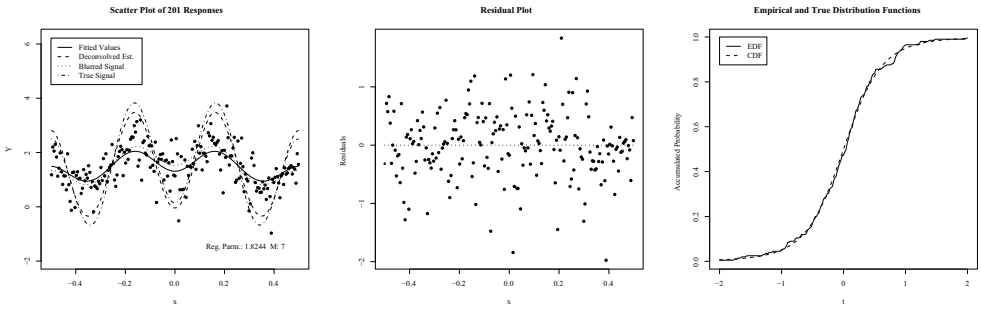

Figure 1. From left to right: A scatter plot of the data overlaid with the fitted blurred regression (solid), estimated regression (dashed), true blurred regression function (dotted), and the true regression function (dot-dashed); A scatter plot of the model residuals overlaid with a line at zero; A plot of the residual-based empirical distribution function (solid), overlaid with the true error distribution function (dashed).

Table 1. Simulated asymptotic bias and variance (in parentheses) of $(2n+1)^{1/2}\{\hat{\mathbb{F}}(t) - F(t)\}$ at the points -2, -1, 0, 1, and 2 for the case of normally distributed errors. The results from each regression θ_1 and θ_2 are given as rows within each sample size, with the first row corresponding to θ_1, and the second to θ_2.

n	t				
	-2	-1	0	1	2
51	−0.001 (0.001)	−0.020 (0.047)	0.000 (0.092)	0.022 (0.046)	−0.001 (0.002)
	−0.003 (0.001)	−0.083 (0.042)	0.005 (0.095)	0.078 (0.045)	0.005 (0.001)
101	−0.001 (0.001)	−0.036 (0.044)	0.002 (0.091)	0.031 (0.047)	0.000 (0.001)
	−0.005 (0.001)	−0.063 (0.043)	0.015 (0.086)	0.066 (0.047)	0.002 (0.001)
201	−0.001 (0.001)	−0.041 (0.046)	0.003 (0.090)	0.044 (0.047)	0.001 (0.001)
	−0.004 (0.001)	−0.067 (0.045)	−0.001 (0.087)	0.056 (0.044)	0.005 (0.001)
301	−0.001 (0.001)	−0.022 (0.044)	0.001 (0.092)	0.019 (0.045)	0.000 (0.001)
	−0.003 (0.001)	−0.035 (0.048)	−0.015 (0.091)	0.047 (0.046)	0.002 (0.001)

plainly see this estimator is performing well. Beginning with the case of normally distributed errors, Table 1 shows the figures for the simulated asymptotic biases and variances of $\hat{\mathbb{F}}$ at the points -2, -1, 0, 1, and 2. The simulated asymptotic biases are calculated by computing the simulated biases of $\hat{\mathbb{F}}$, and multiplying these by the square root of the corresponding sample size. The simulated asymptotic variance is similarly calculated, but now we multiply by the corresponding sample size. Inspecting Table 1, we find the squared asymptotic bias of $\hat{\mathbb{F}}$ becomes negligible to the asymptotic variance of $\hat{\mathbb{F}}$ at larger sample sizes, which is expected. In Table 2, we give the asymptotic mean squared error (AMSE)

Table 2. Asymptotic mean squared error of $(2n+1)^{1/2}\{\hat{\mathbb{F}}(t)-F(t)\}$ at the points -2, -1, 0, 1, and 2 for the case of normally distributed errors. The results from each regression θ_1 and θ_2 are given as rows within each sample size, with the first row corresponding to θ_1, and the second to θ_2.

n	t				
	-2	-1	0	1	2
51	0.001	0.048	0.092	0.047	0.002
	0.001	0.049	0.095	0.051	0.001
101	0.001	0.046	0.091	0.048	0.001
	0.001	0.047	0.086	0.052	0.001
201	0.001	0.047	0.090	0.049	0.001
	0.001	0.049	0.087	0.047	0.001
301	0.001	0.045	0.092	0.045	0.001
	0.001	0.049	0.091	0.048	0.001
∞	0.001	0.046	0.091	0.046	0.001

Table 3. Asymptotic integrated mean squared error of $(2n+1)^{1/2}\{\hat{\mathbb{F}}-F\}$, by sample size, for the case of normally distributed errors. The results from each regression θ_1 and θ_2 are given as rows, with the first row corresponding to θ_1, and the second to θ_2.

51	101	201	301	∞
0.193	0.191	0.195	0.188	0.188
0.208	0.196	0.196	0.193	

Table 4. Simulated asymptotic bias and variance (in parentheses) of $(2n+1)^{1/2}\{\hat{\mathbb{F}}(t)-F(t)\}$ at the points -2, -1, 0, 1, and 2 for the case of t-distributed errors. The results from each regression θ_1 and θ_2 are given as rows within each sample size, with the first row corresponding to θ_1, and the second to θ_2.

n	t				
	-2	-1	0	1	2
51	-0.005 (0.006)	-0.011 (0.042)	0.025 (0.124)	-0.008 (0.042)	0.006 (0.005)
	-0.016 (0.004)	-0.044 (0.037)	0.011 (0.122)	0.042 (0.037)	0.011 (0.005)
101	-0.004 (0.006)	-0.015 (0.037)	0.003 (0.135)	0.008 (0.038)	0.006 (0.005)
	-0.014 (0.005)	-0.036 (0.037)	-0.005 (0.138)	0.034 (0.039)	0.013 (0.005)
201	-0.005 (0.006)	-0.016 (0.039)	0.004 (0.158)	0.018 (0.039)	0.008 (0.006)
	-0.014 (0.005)	-0.029 (0.034)	-0.004 (0.133)	0.023 (0.035)	0.008 (0.006)
301	-0.004 (0.006)	-0.013 (0.037)	-0.012 (0.143)	0.015 (0.035)	0.004 (0.006)
	-0.010 (0.005)	-0.023 (0.036)	0.001 (0.129)	0.023 (0.036)	0.006 (0.006)

of $\hat{\mathbb{F}}$, which is calculated by multiplying the simulated mean squared error of $\hat{\mathbb{F}}$ by the corresponding sample size. The figures corresponding to the sample

Table 5. Asymptotic mean squared error of $(2n+1)^{1/2}\{\hat{\mathbb{F}}(t) - F(t)\}$ at the points -2, -1, 0, 1, and 2 for the case of t-distributed errors. The results from each regression θ_1 and θ_2 are given as rows within each sample size, with the first row corresponding to θ_1, and the second to θ_2.

n	t				
	-2	-1	0	1	2
51	0.006	0.042	0.124	0.042	0.005
	0.004	0.039	0.122	0.039	0.005
101	0.006	0.037	0.135	0.038	0.005
	0.005	0.039	0.138	0.040	0.006
201	0.006	0.039	0.158	0.039	0.006
	0.005	0.035	0.133	0.035	0.006
301	0.006	0.037	0.143	0.036	0.006
	0.006	0.036	0.129	0.036	0.006
∞	0.006	0.036	0.156	0.036	0.006

Table 6. Asymptotic integrated mean squared error of $(2n+1)^{1/2}\{\hat{\mathbb{F}} - F\}$ by sample size for the case of t-distributed errors. The results from each regression θ_1 and θ_2 are given as rows, with the first row corresponding to θ_1, and the second to θ_2.

51	101	201	301	∞
0.233	0.227	0.232	0.222	0.228
0.223	0.228	0.216	0.223	

size ∞ are calculated using the results of Theorem 2. Comparing the results in Table 2, we find the theoretical prediction made in Theorem 2 concerning the asymptotic pointwise precision of $\hat{\mathbb{F}}$ corresponds well with the simulated results. Finally, turning our attention to Table 3, we give the asymptotic integrated mean squared error (AIMSE) of $\hat{\mathbb{F}}$, which is calculated similarly to the AMSE of $\hat{\mathbb{F}}$, but where we integrate with respect to t. These results also confirm that $\hat{\mathbb{F}}$ performs well as an estimator of F, even at the smaller sample sizes 51 and 101. A possible explanation for this observation is the use of the smooth bootstrap methodology for choosing the regularization parameter in the estimate $\hat{\theta}$. Table 4, Table 5, and Table 6 show the figures related to Table 1, Table 2, and Table 3, respectively, when the model errors follow a t-distribution. The results are analogous to the case of normally distributed errors.

The results concerning our indirect regression estimator are interesting. In addition to finding an asymptotically nearly optimal regularization parameter using the proposed bootstrap methodology, we also conducted a similar grid

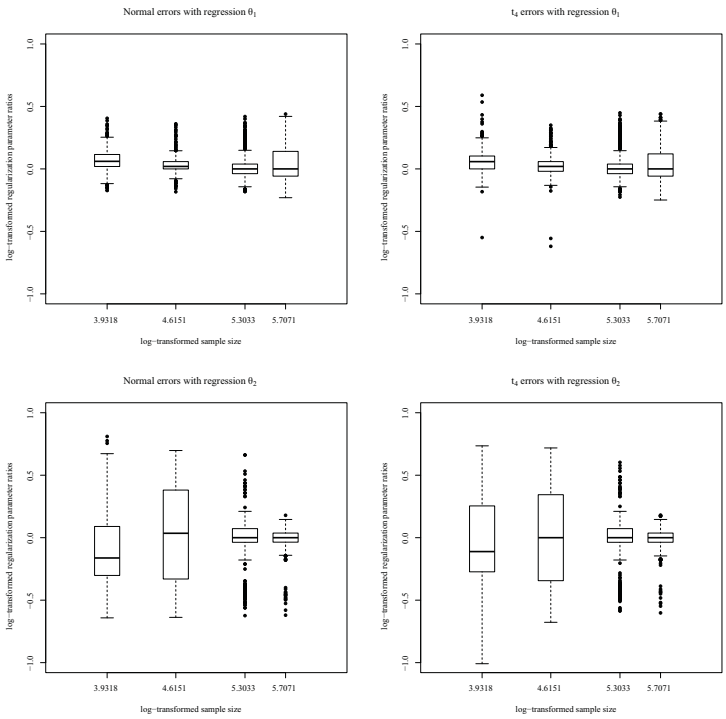

Figure 2. Box plots of log-transformed ratios of regularization parameters (bootstrap-based selection to ISE-based selection) by log-transformed sample size. Plots on the left correspond to normally distributed errors, and plots on the right correspond to t-distributed errors. The plots on the top correspond to the regression θ_1, and plots on the bottom correspond to the regression θ_2.

Table 7. Integrated mean squared error of the indirect regression estimator by sample size for each regression θ_1 and θ_2 in the case of normally distributed errors. Figures corresponding to "Bootstrap" are IMSE estimates based on the proposed smooth bootstrap methodology for selecting the regularization parameter, and the figures corresponding to "Best" are the IMSE estimates corresponding to selecting the regularization parameter by minimizing the ISE. The results for each regression θ_1 and θ_2 are given as rows within each regularization selection method, with the first row corresponding to θ_1, and the second to θ_2.

Regularization	51	101	201	301
Bootstrap	0.168	0.095	0.056	0.041
	0.741	0.596	0.343	0.245
Best	0.131	0.079	0.049	0.036
	0.573	0.438	0.277	0.201

Table 8. Integrated mean squared error of the indirect regression estimator by sample size for each regression θ_1 and θ_2 in the case of t-distributed errors. Figures corresponding to "Bootstrap" are the IMSE estimates based on the proposed smooth bootstrap methodology for selecting the regularization parameter, and the figures corresponding to "Best" are the IMSE estimates corresponding to selecting the regularization parameter by minimizing the ISE. The results for each regression θ_1 and θ_2 are given as rows within each regularization selection method, with the first row corresponding to θ_1, and the second to θ_2.

Regularization	51	101	201	301
Bootstrap	0.159	0.097	0.056	0.041
	0.733	0.587	0.340	0.249
Best	0.125	0.081	0.049	0.037
	0.569	0.434	0.273	0.205

search procedure choosing an optimal regularization parameter that minimizes the integrated squared error (ISE) between the indirect regression estimator and the regression function for each of θ_1 and θ_2. In general, this methodology is not available in applications, but we expect it to produce the best indirect regression estimate with respect to the IMSE of these estimators.

Figure 2 shows box plots of the log-transformed ratios of the nearly optimal regularization parameter selected from the proposed bootstrap methodology to the regularization parameter chosen from the ISE methodology, at each log-transformed sample size. At the larger sample sizes, we can plainly see the boxes are beginning to include zero. It appears that with increasing sample size, both the bootstrap selection methodology and the ISE selection methodology choose similar regularizations for each of θ_1 and θ_2, for both normally distributed and t-distributed errors.

We also numerically measured the performance of the indirect regression estimator by simulating the IMSE using both regularization techniques for each regression θ_1 and θ_2 for both cases of normally distributed errors and t-distributed errors. The results are given in Table 7 for the case of normally distributed errors, and in Table 8 for the case of t-distributed errors. We can plainly see that the IMSE of the estimators using each regularization method are decreasing to zero as the sample size increases, and the IMSE values between the bootstrap-based method and the ISE-based method appear to be very similar, even at the smaller sample sizes 51 and 101. In summary, we find that the residual-based empirical distribution function is performing well as an estimator of the error distribution function, and that the proposed smooth bootstrap methodology for selecting the

regularization parameter required for the indirect regression estimator provides a useful and convenient tool for precise indirect regression function estimation.

4.1. Example: comparison between regularization methods for spectral cutoff estimators

Consider the special case of indirect regression estimators formed using the so-called spectral cutoff method. This means we consider the simpler spectral smoothing kernel

$$\Lambda(k) = \mathbf{1}\big[-1 \le k \le 1\big], \qquad k \in \mathbb{Z}.$$

Here, one seeks a regularization that essentially decides how many Fourier frequencies to include in the indirect regression estimator, which follows from observing that (2.2) evaluates Λ at the product $h_n k$, where the regularizing parameter h_n is small. Cavalier and Golubev (2006) investigate a penalized estimator of the integrated mean squared error of indirect regression estimators obtained from the spectral cutoff method called a *risk hull*; see (1.9)–(1.11) on pages 1656–1657. The authors propose selecting a regularization that minimizes this quantity, and call this approach to regularize the risk hull method. Note that the risk hull method requires choosing a tuning parameter α that influences the strength of the penalty. The authors suggest using $\alpha = 1.1$, which we use as well.

In this example, we simulated a comparison between the risk hull method and the proposed bootstrap regularization selection method from Section 3.2, for both regressions θ_1 and θ_2. The distortion function ψ is specified in Section 4, and the errors are again normally distributed with mean zero and scale $2/3$. As before, we considered sample sizes 51, 101, 201, and 301. We used the same pilot sequences as in the previous example for the bootstrap selection method.

The results of our numerical study are summarized in the box plots displayed in Figure 3. At smaller sample sizes 51 and 101, we can see that both approaches choose similar regularizations; that is, both procedures suggest similar spectral cuts. The larger sample sizes 201 and 301, however, show that the risk hull method begins to favor regularizations that include fewer Fourier frequencies than those of the bootstrap method. Consequently, for θ_1, the simulated IMSE values are 0.311, 0.286, 0.276, and 0.274 using the risk hull method, and 0.312, 0.279, 0.056, and 0.042 using the proposed bootstrap procedure, for each sample size 51, 101, 201, and 301, respectively. Similarly, for θ_2, the simulated IMSE values are 2.354, 2.229, 2.100, and 2.049 using the risk hull method, and 0.737, 0.613, 0.539, and 0.209, respectively, using the proposed bootstrap procedure. We

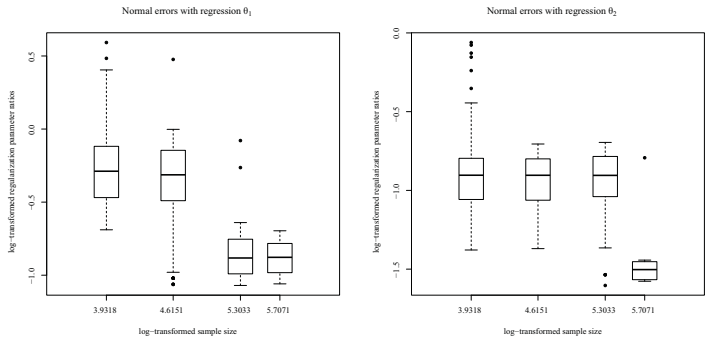

Figure 3. Box plots of log-transformed ratios of regularization parameters (bootstrap-based selection to risk-hull-based selection) by log-transformed sample size. The box plot on the left corresponds to the regression θ_1, and the box plot on the right to the regression θ_2.

can plainly see the proposed bootstrap selection procedure compares favorably with the risk hull method.

Unfortunately, using the risk hull method together with the spectral cutoff estimator produces unsatisfactory results for larger sample sizes. This phenomenon can also be observed in Rochet (2013) (see Case 1 on page 491, and compare the lines in the tables corresponding to $\hat{x}^*_{sco}$ for the sample sizes 50 and 200; note the lack of substantial decay in their performance metric, despite the increase in sample size). Because the proposed bootstrap procedure is widely applicable and has good finite-sample performance, we recommend it to practitioners considering data-driven regularization selection procedures.

Supplementary Material

The online Supplementary Material contains the proofs of the technical results.

Acknowledgements

We would like to thank the Referees and an Associate Editor for their careful reading and helpful comments, which resulted in several improvements to our approach, including incorporating several important works in the literature that were previously unknown to us. This work was supported, in part, by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Projects C1 and C4) of the German Research Foundation (DFG)

and, in part, by the Bundesministerium für Bildung und Forschung through the project "MED4D: Dynamic medical imaging: Modeling and analysis of medical data for improved diagnosis, supervision and drug development."

References

Birke, M., Bissantz, N. and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems* **26**, 115020.

Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24**, 034009.

Bissantz, N., Mair, B. and Munk, A. (2006). A multi-scale stopping criterion for MLEM reconstructions in PET. In *IEEE Nuclear Science Symposium Conference Record*, 3376-3379.

Bissantz, N., Mair, B. and Munk, A. (2008). A statistical stopping rule for MLEM reconstructions in PET. In *IEEE Nuclear Science Symposium Conference Record*, 4198-4200.

Blanchard, G., Hoffmann, M. and Reiß, M. (2016). Optimal adaptation for early stopping in statistical inverse problems. `arXiv:1606.07702v1`.

Blanchard, G. and Mathé, P. (2012). Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems* **28**, 115011.

Cao, R. (1993). Bootstrapping the mean integrated squared error. *J. Multivariate Anal.* **45**, 137-160.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems* **24**, 034004.

Cavalier, L. and Golubev, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34**, 1653-1677.

Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123**, 323-354.

Davies, P.L. and Meise, M. (2008). Approximating data with weighted smoothing splines. *J. Nonparametr. Stat.* **20**, 207-228.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5**, 907-925.

González-Manteiga, W., Martinez-Miranda, M.D. and Pérez-González, A. (2004). The choice of smoothing parameter in nonparametric regression through wild bootstrap. *Comput. Statist. Data Anal.* **47**, 487-515.

Hotz, T., Marnitz, P., Stichtenoth, R., Davies, L., Kabluchko, Z. and Munk, A. (2012). Locally adaptive image denoising by a statistical multiresolution criterion. *Comput. Statist. Data Anal.* **56**, 543-558.

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

Marteau, C. and Mathé, P. (2014). General regularization schemes for signal detection in inverse problems. *Math. Methods Statist.* **23**, 176-200.

Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37**, 1105-1115.

Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary pro-

cesses. *J. Nonparametr. Stat.* **3**, 13-36.

Mathé, P. and Pereverzev, S.V. (2006). Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comp.* **75**, 1913-1929.

Müller, U.U., Schick, A. and Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plann. Inference* **119**, 75-93.

Müller, U.U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions* **25**, 1-18.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scand. J. Stat.* **36**, 204-228.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.

Politis, D.N. and Romano, J.P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal.* **68**, 1-25.

Proksch, K., Bissantz, N. and Dette, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli* **21**, 144-175.

Rochet, P. (2013). Adaptive hard-thresholding for linear inverse problems. *ESAIM Probab. Stat.* **17**, 485-499.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* CRC press.

Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation.* Springer series in statistics. Springer-Verlag, New York.

Watson, G.S. and Leadbetter, M.R. (1963). On the estimation of the probability density, 1. *Ann. Math. Statist.* **34**, 480-491.

Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE.

E-mail: nicolai.bissantz@web.de

Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE.

E-mail: justin.chown@rub.de

Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE.

E-mail: holger.dette@rub.de

(Received October 2016; accepted August 2018)