Statistica Sinica Preprint No: SS-2018-0160	
Title	Regularization parameter selection in indirect regression by residual based bootstrap
Manuscript ID	SS-2018-0160
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0160
Complete List of Authors	Nicolai Bissantz Justin Chown and Holger Dette
Corresponding Author	Justin Chown
E-mail	justin.chown@ruhr-uni-bochum.de
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Regularization parameter selection in indirect regression by residual based bootstrap

Nicolai Bissantz, Justin Chown* and Holger Dette

Abstract. Residual-based analysis is generally considered a cornerstone of statistical methodology. For a special case of indirect regression, we investigate the 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 a 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.

Keywords: bandwidth selection, indirect regression estimator, inverse problems, regularization, residual-based empirical distribution function, smooth bootstrap

2010 AMS Subject Classifications: Primary: 62G08, 62G09; Secondary: 62G20, 62G30.

1. Introduction

In many experiments one is only able to make indirect observations of the physical process being observed. Important quantities that are of interest to the study are not directly available for statistical inference in such so-called *inverse problems*, but images of these quantities under some transformation such as a convolution can be used instead. Here, we consider an inverse regression model, i.e. observing a signal of interest from indirect observations

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

where K is an operator specifying convolution of the true underlying regression θ with a distortion function ψ, i.e.

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

The resulting function $K\theta$ can be viewed as a distorted regression function. We assume that θ is a smooth periodic function, which is a common assumption taken in many inverse problems and further discussed below. As in pages 49-51 of Tsybakov (2009), this means the Fourier coefficients of θ are assumed to satisfy a crucial technical summability requirement (see Section 2 for further details).

* Correspondences may be addressed to Justin Chown (justin.chown@ruhr-uni-bochum.de).
Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

We will assume that ψ is known and behaves like a probability density function on the interval $[-1/2,\, 1/2]$, i.e. ψ is positive-valued on the interval $[-1/2,\, 1/2]$ and integrates to one. Later, we will specify further technical requirements regarding ψ. However, to ensure the convolution operation is well-defined, it is clear that ψ must be periodically extended 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 $[-1/2,\, 1/2]$, i.e. $x_j = j/2n$, $j = -n, \ldots, n$, and the errors ε_j are assumed to be independent, have mean equal to zero and have the common distribution function F. Note, the assumptions given above only guarantee that model (1.1) is a well-defined indirect regression model, where θ is identifiable (see, for example, Cavalier and Golubev, 2006, or Mair and Ruymgaart, 1996).

Statistical inverse problems have received a great amount of attention for the construction of estimators for various densities and indirect regression models. In particular, early work considers properties of estimators in a range of important statistical inverse problems. Important examples are Masry (1991), who investigates estimators of a multivariate density function in an errors-in-variables model using a deconvolution technique; Fan (1991), who derives optimal rates of convergence for density estimators in these models, and Masry (1993), who investigates estimators of a smooth multivariate regression function using deconvolution techniques for the case of regression function estimation with contaminated covariates.

A second sequence of publications about statistical inverse problems such as those considered in (1.1) yields a better understanding of the asymptotic properties of the estimators from a theoretical perspective. Important results are due to Mair and Ruymgaart (1996), who consider estimators of the indirect regression function in a flexible model based on Hilbert scales. This includes the popular case of Sobolev classes, where these 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 minimax optimality of their estimators. Moreover, Cavalier (2008) surveys the available literature for deconvolution estimators and provides minimax rates of reconstructions in several models including (1.1).

More recently, statistical testing and model selection properties have been considered in statistical inverse problems of the type (1.1). Bissantz and Holzmann (2008) provide an overview for constructing confidence intervals and confidence bands in univariate statistical inverse problems. Later, Proksch, Bissantz and Dette (2015) generalize the univariate case studied by the previous authors and construct confidence bands for an indirect regression function of multiple covariates. Marteau and Mathé (2014) consider the problem of testing for distorted signals using general regularization schemes.

All of the deconvolution estimators investigated in the previously mentioned articles are 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. Data-driven selection of this parameter is an important problem that we want to more closely examine in this article. Techniques for choosing the sequence of regularization parameters generally focus on choosing a suitable estimator of the integrated mean squared error of an indirect regression estimator or choosing some other related quantity (see Section 3). Cavalier and Golubev (2006) make a particularly important contribution to this problem, where these authors investigate the integrated mean squared error of indirect regression estimators and propose a suitably penalized quantity based on a threshold

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

of this important estimation performance metric. The authors call this a *risk hull* approach because of the resulting bowl-shaped objective function used for choosing their 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 investigated here appears to be particularly promising when compared with the risk hull approach previously mentioned.

In this article, 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) 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 generally unknown and must be estimated.

To the best of our knowledge, this object has not been studied before with respect to statistical deconvolution 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 here 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 stronger conditions regarding the smoothness of θ are required for $\hat{\mathbb{F}}$ to be a consistent estimator of F that have not appeared before in the literature. Hence, any residual-based inference procedure relying on $\hat{\mathbb{F}}$ would also require this stronger smoothness condition, e.g. Kolmogorov-Smirnov-type and Cramér-von-Mises-type statistics.

Studying these problems requires new results concerning the estimator $\hat{\theta}$ and its bootstrap analog. The literature on statistical deconvolution problems is vast, and, hence, some results will be familiar. In particular, we show 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 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).

We have organized the remainder of this article as follows. Further notation and the estimation method are introduced in Section 2, and the asymptotic results for the estimators $\hat{\theta}$ and $\hat{\mathbb{F}}$ may be found in Section 2.1. 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 that is 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 that is in the

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

spirit of 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 the theoretically optimal regularization, and we consider a comparative technique for choosing the regularization for spectral cut-off estimators (a special case of our approach) proposed by Cavalier and Golubev (2006) as an example in Section 4.1. All of the proofs of our results may be found in Section 5.

2. Estimation in the indirect regression model

Let us 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

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

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

For clarity, we will now introduce some notation. Let $d \in \mathbb{N}$. We will call $q^{(i)}$, $1 \leq i \leq d$, 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 have evaluations of ϕ, $(d^{i'}\phi)/(dx^{i'})$, $i' = 1, \ldots, i$, 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

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

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 considering smoothness orders $s > 0$, i.e. $\mathcal{W}^{2,s}([-1/2,\, 1/2])$ is defined exactly as $\mathcal{W}^{2,d}([-1/2,\, 1/2])$ is defined above but now with s in place of d. We will 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 will restrict $\mathscr{R}_s$ to a subspace of $\mathcal{W}^{2,s}([-1/2,\, 1/2])$:

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

Note, $\theta \in \mathscr{R}_s$ 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.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Another important note regarding the Fourier basis $\{\exp(i2\pi kx) : x \in [-1/2, 1/2]\}_{k\in\mathbb{Z}}$ is that it also decomposes the operator K into a singular value decomposition along each of the orthonormal basis functions, where now we only need to 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, which means assuming 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 whose strong uniform consistency rate is comparable, albeit worse, to the rates expected in the usual nonparametric regression case, and we can show the estimator $\hat{\mathbb{F}}$ is both root-n consistent for F, uniformly in $t \in \mathbb{R}$, and $\hat{\mathbb{F}}$ is asymptotically most precise. The second case is that of the so-called super smooth distortion functions, which means assuming the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ decay at an exponential rate, e.g. $|\Psi(k)| \sim \exp(-|k|^b)$. Under this assumption, the resulting indirect regression estimator has only a strong uniform consistency rate that is polynomic in the logarithm of n, which we expect is too slow for us to maintain the root-n consistency of $\hat{\mathbb{F}}$. Throughout this article, we will therefore focus on the first case of ordinarily smooth distortion functions ψ, and we work with a similar assumption as (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 $[-1/2, 1/2]$, i.e.

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

The Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ are then given by

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

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

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 in the right-hand side of (2.1) is an average, and we can form an estimator for it from an empirical average using data (x_j, Y_j), $j = -n, \ldots, n$, and obtain

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

To recover θ from the convolution $K\theta$, we will make use of the convolution theorem for Fourier transformation: $R(k) = \Theta(k)\Psi(k)$, $k \in \mathbb{Z}$. Since ψ is a probability density function

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 $\mathcal{Z} \subset \mathbb{Z}$, and, hence, Ψ^{-1} is well-defined (see, for example, the discussion on preconditioning on page 1425 of Mair and Ruymgaart, 1996). Observing that Fourier transformation reduces convolution to multiplication, we can exploit the Fourier inversion formula by writing

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

To plug-in 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, i.e. the inversion of the operator K in (1.1) requires regularization; see Cavalier and Golubev (2006) for a very 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 in same spirit as Mair and Ruymgaart (1996). To explain the idea consider the ratio $|\hat{R}(k)|/|\Psi(k)|$, which becomes large as k increases. The idea is 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 Mair and Ruymgaart (1996) is the previous authors require the regularization to preserve the fundamental Fourier frequency, i.e. the regularization must be equal to one around some neighborhood of the zeroth Fourier frequency. This approach to regularization is simple to specify for applications and leads to 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, i.e. maps $x \mapsto h_n^{-1}\delta(x/h_n)$, where δ is a suitably constrained probability density function. Politis and Romano (1999) observe 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. This means 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 will require our smoothing kernel to have 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 a more clear discussion we will now 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 only the estimation bias has a desirable rate of convergence: the order $O(h_n^s)$, when $\theta \in \mathscr{R}_s$. This is comparable to the direct estimation setting by sufficiently high-order kernels or the so-called "superkernels" (see, for example, 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 all the way back to Parzen (1962).

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

An estimator of θ is given by a kernel smoother:

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

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, for example, 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 impacted by the degree of ill-posedness of the inverse problem, which we expect can be made negligible by 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 additionally satisfies $\int_{-\infty}^{\infty} |u|^{b+1} |\Lambda(u)|\, du < \infty$ and 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, i.e.

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

For this choice of regularizing parameters, we have $(nh_n^{b+1})^{-1} = o(h_n^s)$, which implies the bias of $\hat{\theta}$ has the order $O(h_n^s)$. Note, Lemma 2 requires the responses to 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), and two additional properties of the estimator $\hat{\theta}$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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

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

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

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

If Λ *satisfies* $\int_{-\infty}^{\infty} |u|^{s+b-1/2}|\Lambda(u)|\,du < \infty$, *then, for large enough* 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 the smoothness index s of the function space $\mathscr{R}_s$ to be larger than the degree of ill-posedness b of the inverse problem, which is a stronger requirement than what has appeared in the literature before. 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 satisfying minimax optimality, where now the integrated mean squared error has the order $O(n^{-(2s)/(2s+2b+1)})$.

We are now ready to state our main results concerning 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*

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

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

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

writing $\sigma^2 = E[\varepsilon^2]$ *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).

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 a practical choice of regularization would be a scheme that minimizes the integrated mean squared error (IMSE) of $\hat{\theta}$. However, 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, for example, González-Manteiga, Martinez-Miranda and Pérez-González, 2004; Bissantz, Mair and Munk, 2006; Bissantz, Mair and Munk, 2008; Davies and Meise, 2008, and Hotz et al., 2012). Lepski methodology among others has recently become a popular approach in this context, where the IMSE of the indirect regression estimator is replaced by a suitable non-random objective function using oracle inequalities (see, for example, Goldenshluger, 1999; Cavalier and Tsybakov, 2002; Mathé and Pereverzev, 2006; Blanchard and Mathé, 2012, and Blanchard, Hoffmann and Reiß, 2016). An important approach for spectral cut-off 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 will propose a methodology based on a smooth bootstrap of the model residuals to form a consistent estimator of the IMSE of $\hat{\theta}$, and, using 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 the 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.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

The conclusion that $\hat{\theta}$, formed from a regularizing sequence of order $O(n^{-1/(2s+2b+1)})$, is minimax optimal only guarantees 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 by further numerical methods due to unknown constants that 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 could 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 like 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 $\mathscr{R}_s$. Estimation of this quantity is very difficult and likely not even possible without harsh and confining assumptions. However, an educated guess would lead to the optimal bandwidth choice corresponding with the fastest possible decay of the IMSE of $\hat{\theta}$. This means choosing s to be as large 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. The approaches generally focus on estimating the IMSE of the estimator using either a cross-validation or 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. Due to its simplicity, we will 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 will 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

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

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.

Let us begin with 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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 will contaminate the randomly selected centered model residual $\tilde{\varepsilon}_j^*$ by an independent, centered random variable U_j that has a finite moment of order $\kappa > 2 + 1/(s+b)$ and 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 $\{c_n\}_{n\geq 1}$ is a scaling sequence similar to a bandwidth for kernel density estimation. Consequently, ε_j^* has the common distribution function

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

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 choice of w, i.e. 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

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

Define $\hat{\theta}^*$ as in (2.2) but now with Y_j^* replacing Y_j and a regularizing sequence $\{g_n\}_{n\geq 1}$ replacing 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 as $\hat{\theta}$ given in Theorem 1. We summarize these results in Proposition 5 in Section 5.

A particularly important use of bootstrapping in practice is to find suitable quantiles for test statistics. In the case of residual-based analysis, one is typically interested in continuous functionals $T(Z_{F_0})$ of a Gaussian process of 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 analogue of $\mathbb{G}_n$. We therefore expect analogous results from Neumeyer (2009) to hold in the present context.

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

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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 Section 5). 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 5 be satisfied, with $s > \max\{(2b+1)/(2\gamma),\, 3/2\}$. Then*

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

Note, this result always includes the optimal bandwidth choice $c_n = O(n^{-1/5})$ for density estimation. This fact in combination with the results of Proposition 4 yield the following analog of Corollary 1:

COROLLARY 2. *Let the assumptions of Theorem 3 be satisfied. If, additionally, both of the 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*

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

$t \in \mathbb{R}$, 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 can immediately obtain valid smooth bootstrap approximations of quantiles for test statistics T_n previously discussed. We conclude this section with the following remark:

REMARK 4. Both the residual-based empirical process $\mathbb{G}_n$ and its smooth bootstrap analogue $\mathbb{G}_n^*$ 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 with 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 immediately used to approximate unknown quantiles of test statistics obtained from continuous functionals of F.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 will subscript the estimators $\hat{\theta}$ and $\hat{\theta}^*$ by the regularization parameters used to form them, i.e. we write $\hat{\theta}_{h_n}$ to indicate 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 will 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 for our approach to admit an asymptotically nearly optimal choice of regularizing parameters that 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 now 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:

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

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

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

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), which implies that we can use the mapping $g_n \mapsto IMSE^*(\hat{\theta}^*_{g_n})$ as an objective criteria 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), respectively. For example, choosing Λ as an indicator function, e.g. working

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

with spectral cut-off estimators, leads to the deconvolution smoothing kernel in (2.2) to be a smooth function with infinitely many derivatives.

REMARK 5. The bootstrap methodology for finding an asymptotically nearly optimal regularizing sequence previously discussed critically requires that the pilot sequence $\{h_n\}_{n\geq 1}$ is chosen to undersmooth the estimator $\hat{\theta}$. Interestingly, this is in contrast to the bootstrap methodology outlined in Cao (1993), where it is recommended to use an oversmoothing pilot bandwidth for finding 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 we investigate the effectiveness of our smooth bootstrap methodology for selecting a regularization parameter. In the following simulations, we chose 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 while 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 $\{\Theta_2(k)\}_{k\in\mathbb{Z}}$, which are 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 a mean of 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 for estimating 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 freedom and scaled to $2/3$ as well. Our simulations consider samples of sizes 51, 101, 201 and 301, i.e. n is taken as 25, 50, 100 and 150.

We work with the smoothing kernel that has Fourier coefficients satisfying

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

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 work

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

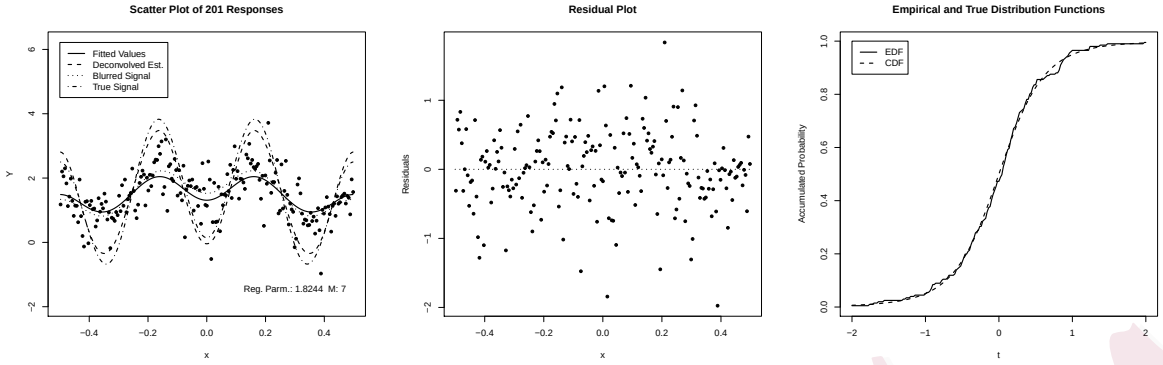

FIGURE 1. *From left to right: A scatter plot of the data overlaid with the fitted blurred regression (solid), estimated regression (dashed), true blurred regression function (dotted) and the true regression function (dot-dashed); A scatter plot of the model residuals overlaid with a line at zero; A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

with the pilot sequences $h_{n,1} = 5(2n+1)^{-1/11}\log^{1/11}(2n+1)$, which corresponds with the choice $s = 3$ in (2.3), for estimating θ_1 and $h_{n,2} = 2.5(2n+1)^{-1/11}\log^{1/11}(2n+1)$ for estimating θ_2.

To create the smooth bootstrap of the residuals we have used standard normally distributed contaminates U_j and Silverman's rule for selecting a bandwidth in kernel density estimation, i.e. 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 respectively estimating each of θ_2 and $K\theta_2$. We can plainly see that the indirect regression estimator, constructed with the proposed data-driven regularization methodology, is explaining the data very well, which follows from the appearance of completely random scatter in the plot of the residuals. The plot of the distribution functions shows the empirical distribution function of the residuals $\hat{\mathbb{F}}$ matches very closely to the true error distribution function F as expected.

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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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)

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	0.048	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 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.

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

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.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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)

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	0.006	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 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*.*

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

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*.*

biases of $\hat{\mathbb{F}}$ and multiplying these by the square-root of the corresponding sample size, and 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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

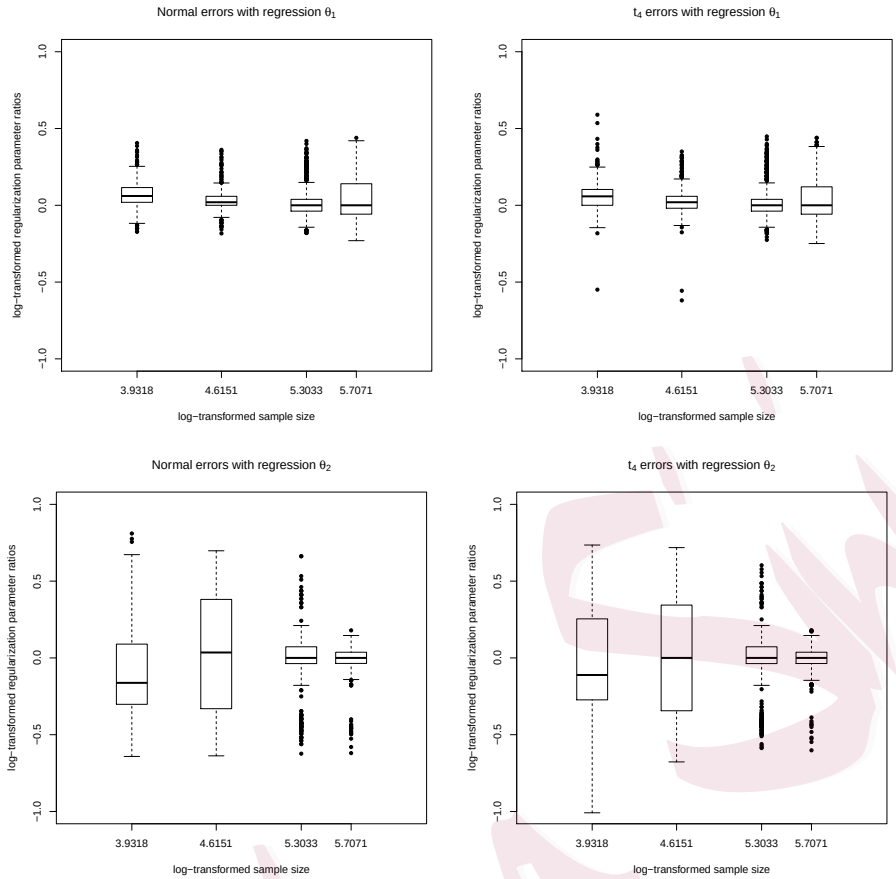

FIGURE 2. *Boxplots of log-transformed ratios of regularization parameters (bootstrap-based selection to ISE-based selection) by log-transformed sample size. Plots on the left correspond to normally distributed errors and plots on the right correspond to t-distributed errors, while plots on the top correspond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

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) 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 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 now integrating 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 related figures to Table 1, Table 2 and Table 3, respectively, when the model errors are t-distributed, and the results are analogous to the case of normally distributed errors.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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 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 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

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.*

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 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 case of θ_1 and θ_2. In general, this methodology is not available in applications, but we expect it to produce the best resulting indirect regression estimate with respect to the IMSE of these estimators.

In Figure 2 we give boxplots 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 in both cases of normally distributed and t-distributed errors.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

We have 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 in 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 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 the residual-based empirical distribution function is performing well as an estimator of the error distribution function, and 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 cut-off estimators. Consider the special case of indirect regression estimators formed from the so-called spectral cut-off method. This means we consider the simpler spectral smoothing kernel

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

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 cut-off method called a *risk hull*; see equations (1.9) – (1.11) on pages 1656-1657. These authors propose selecting a regularization that minimizes this quantity and call this approach to regularization 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 $\alpha = 1.1$ on page 1664, 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 that were used in the previous example for the bootstrap selection method.

The results of our numerical study are summarized in the boxplots displayed in Figure 3. At smaller sample sizes 51 and 101, we can see that both approaches generally choose similar regularizations, i.e. both procedures suggest similar spectral cuts. The larger sample sizes 201 and 301, however, show the risk hull method begins to favor regularizations that include fewer Fourier frequencies than the bootstrap method. Consequently, for the case 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, respectively for each sample size 51, 101, 201 and 301. Similarly, for the case 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 using the proposed bootstrap procedure. We can plainly see the proposed bootstrap selection procedure compares favorably to the risk hull method.

Unfortunately, using the risk hull method in combination with the spectral cut-off estimator produced 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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

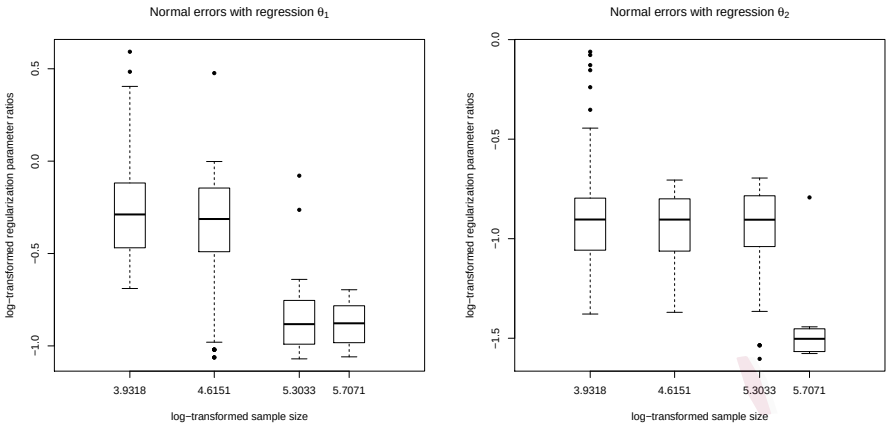

FIGURE 3. *Boxplots of log-transformed ratios of regularization parameters (bootstrap-based selection to risk-hull-based selection) by log-transformed sample size. The boxplot on the left corresponds to the regression θ_1 and the boxplot on the right to the regression θ_2.*

$\hat{x}^*_{sco}$ for the sample sizes 50 and 200 and note the lack of substantial decay in their performance metric despite the increase in sample size). Since the proposed bootstrap procedure is highly applicable and has good finite sample performance, we recommend practitioners to use it when considering data-driven regularization selection procedures.

Acknowledgements We would like to thank the Referees and an Associate Editor for their careful reading and helpful comments that improved the article, which resulted in several improvements to our approach as well as pointing out several important works in the literature that had been previously unknown to us. This work has been 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".

5. Technical details

The estimator $\hat{R}$ is biased only in the design points, which asymptotically exhaust the interval $[-1/2, 1/2]$ at the rate n^{-1}. We arrive at the following result concerning the bias of $\hat{R}$:

LEMMA 3. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$. Then*

$$\max_{k\in\mathbb{Z}} \left| E\big[\hat{R}(k)\big] - R(k) \right| = O\big(n^{-1}\big).$$

PROOF. For any $s_1 \leq s_2$, we have the inclusion $\mathscr{R}_{s_2} \subset \mathscr{R}_{s_1}$, and, therefore, we only need to prove the result for $s = 1$. For clarity, write $r = K\theta$. Without any loss of generality, we can

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

assume that $|r(0)| < \infty$. We can write

$$E\big[\hat{R}(k)\big] = \frac{1}{2n+1}\sum_{j=-n}^{n}\left\{\int_{-\infty}^{\infty} y\, Q(dy\,|\,x_j)\right\}e^{-i2\pi k x_j} = \frac{1}{2n+1}\sum_{j=-n}^{n} r(x_j)e^{-i2\pi k x_j} \tag{5.1}$$
$$= \frac{1}{2n+1}\sum_{j=1}^{n} r(x_j)e^{-i2\pi k x_j} + \frac{r(0)}{2n+1} + \frac{1}{2n+1}\sum_{j=-n}^{-1} r(x_j)e^{-i2\pi k x_j}.$$

The second equality in (5.1) shows that $\hat{R}$ is on the average estimating the discrete Fourier transform of r calculated on the design points, which is expected.

We can relate the discrete Fourier transform of r to its Fourier coefficients $\{R(k)\}_{k\in\mathbb{Z}}$ as follows. Partition the interval $[-1/2,\, 1/2]$ into

$$\left(\bigcup_{j=-n}^{-1}\left[\frac{2j-1}{4n+2}, \frac{2j+1}{4n+2}\right)\right)\bigcup\left[-\frac{1}{4n+2}, \frac{1}{4n+2}\right]\bigcup\left(\bigcup_{j=1}^{n}\left(\frac{2j-1}{4n+2}, \frac{2j+1}{4n+2}\right]\right).$$

Since $x_j = j/(2n)$, we have $v/(2n+1) + j/(2n+1) = x_j + (v-x_j)/(2n+1)$. It follows that $E[\hat{R}(k)] - R(k)$ is equal to

$$\frac{1}{2n+1}\sum_{j=1}^{n}\left\{\int_{-1/2}^{1/2}\left\{r(x_j) - r\left(x_j + \frac{v-x_j}{2n+1}\right)\right\}dv\right\}\exp(-i2\pi k x_j)$$
$$+\frac{1}{2n+1}\sum_{j=1}^{n}\int_{-1/2}^{1/2} r\left(x_j + \frac{v-x_j}{2n+1}\right)\left\{\exp(-i2\pi k x_j) - \exp\left(-i2\pi k\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right\}dv$$
$$+\frac{1}{2n+1}\int_{-1/2}^{1/2}\left\{r(0) - r\left(\frac{v}{2n+1}\right)\right\}\exp\left(-i2\pi k\frac{v}{2n+1}\right)dv$$
$$+\frac{r(0)}{2n+1}\int_{-1/2}^{1/2}\left\{1 - \exp\left(-i2\pi k\frac{v}{2n+1}\right)\right\}dv$$
$$+\frac{1}{2n+1}\sum_{j=-n}^{-1}\left\{\int_{-1/2}^{1/2}\left\{r(x_j) - r\left(x_j + \frac{v-x_j}{2n+1}\right)\right\}dv\right\}\exp(-i2\pi k x_j)$$
$$+\frac{1}{2n+1}\sum_{j=-n}^{-1}\int_{-1/2}^{1/2} r\left(x_j + \frac{v-x_j}{2n+1}\right)\left\{\exp(-i2\pi k x_j) - \exp\left(-i2\pi k\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right\}dv.$$

We can see that $|E[\hat{R}(k)] - R(k)|$ is bounded by

$$R_1(k) + R_2(k) + R_3(k) + R_4(k) + R_5(k) + O(n^{-1}),$$

where the error term $O(n^{-1})$ does not depend on k and

$$R_1(k) = \frac{1}{2n+1}\sum_{j=1}^{n}\int_{-1/2}^{1/2}\left|r(x_j) - r\left(x_j + \frac{v-x_j}{2n+1}\right)\right|dv,$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$R_2(k)$ is equal to

$$\frac{1}{2n+1}\sum_{j=1}^{n}\left|\int_{-1/2}^{1/2} r\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big\{\exp\big(-i2\pi kx_j\big)-\exp\Big(-i2\pi k\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big)\Big\}\,dv\right|,$$

$$R_3(k)=\frac{1}{2n+1}\int_{-1/2}^{1/2}\left|r(0)-r\Big(\frac{v}{2n+1}\Big)\right|dv,$$

$$R_4(k)=\frac{1}{2n+1}\sum_{j=-n}^{-1}\int_{-1/2}^{1/2}\left|r(x_j)-r\Big(x_j+\frac{v-x_j}{2n+1}\Big)\right|dv,$$

and $R_5(k)$ is equal to

$$\frac{1}{2n+1}\sum_{j=-n}^{-1}\left|\int_{-1/2}^{1/2} r\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big\{\exp\big(-i2\pi kx_j\big)-\exp\Big(-i2\pi k\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big)\Big\}\,dv\right|.$$

Hence, the result follows, if we can show $\max_{k\in\mathbb{Z}} R_i(k)=O(n^{-1})$, for each $i=1,\ldots,5$.

To continue, use Euler's formula to write

$$\exp\big(-i2\pi kx_j\big)=\cos\big(2\pi kx_j\big)-i\sin\big(2\pi kx_j\big).$$

Since sine and cosine are each Lipschitz functions with constant equal to 1, it follows that

$$\left|\exp\big(-i2\pi kx_j\big)-\exp\Big(-i2\pi k\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big)\right|^2\le 2^3\pi^2k^2\Big(\frac{v-x_j}{2n+1}\Big)^2.$$

Therefore we have the bound

$$\left|\exp\big(-i2\pi kx_j\big)-\exp\Big(-i2\pi k\Big(x_j+\frac{v-x_j}{2n+1}\Big)\Big)\right|\le 2^{3/2}\pi|k|\frac{|v-x_j|}{2n+1},$$

which will be used throughout the proof.

Beginning with $R_1(k)$, it follows from both $\theta\in\mathscr{R}_1$ and the equivalence $r=K\theta$ that $r\in\mathscr{R}_1$ as well. Using the Fourier inversion formula, write

$$\left|r(x_j)-r\Big(x_j+\frac{v-x_j}{2n+1}\Big)\right|\le\frac{2^{3/2}\pi}{2n+1}|v-x_j|\sum_{k=-\infty}^{\infty}|k|\big|R(k)\big|.$$

Hence, we can find an appropriate constant $C>0$ such that $R_1(k)$ is bounded by

$$Cn^{-2}\sum_{j=1}^{n}\int_{-1/2}^{1/2}\big|v-x_j\big|\,dv,$$

which both does not depend on k and is easily seen to be $O(n^{-1})$. This implies $\max_{k\in\mathbb{Z}} R_1(k)=O(n^{-1})$.

Turning our attention to $R_2(k)$, we can assume without loss of generality that $|k|>0$ as this term is equal to zero whenever $k=0$. The integral in $R_2(k)$ is equal to the sum of

$$\int_{-1/2}^{1/2}\Big\{r\Big(x_j+\frac{v-x_j}{2n+1}\Big)-r(x_j)\Big\}\,dv\exp\big(-i2\pi kx_j\big)$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

and

$$\int_{-1/2}^{1/2} \left\{ r(x_j) \exp\big(-i2\pi k x_j\big) - r\left(x_j + \frac{v-x_j}{2n+1}\right) \exp\left(-i2\pi k\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right\} dv.$$

Therefore, we can see that $R_2(k)$ is bounded by the sum of $\max_{k\in\mathbb{Z}} R_1(k)$, which we have already shown $\max_{k\in\mathbb{Z}} R_1 = O(n^{-1})$, and the quantity

$$\frac{1}{2n+1}\sum_{j=1}^{n} \left| \int_{-1/2}^{1/2} \left\{ r(x_j) \exp\big(-i2\pi k x_j\big) - r\left(x_j + \frac{v-x_j}{2n+1}\right) \exp\left(-i2\pi k\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right\} dv \right|. \tag{5.2}$$

We can use the Fourier inversion formula to write

$$\begin{aligned} & r(x_j) \exp\big(-i2\pi k x_j\big) - r\left(x_j + \frac{v-x_j}{2n+1}\right) \exp\left(-i2\pi k\left(x_j + \frac{v-x_j}{2n+1}\right)\right) \\ &= \sum_{\xi=-\infty}^{\infty} R(\xi) \left\{ \exp\big(i2\pi(\xi-k)x_j\big) - \exp\left(i2\pi(\xi-k)\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right\}. \end{aligned} \tag{5.3}$$

From (5.3) we can see that (5.2) is further bounded by

$$\begin{aligned} & \frac{1}{2n+1}\sum_{j=1}^{n} \int_{-1/2}^{1/2} \sum_{\xi=-\infty}^{\infty} |R(\xi)| \left| \exp\big(i2\pi(\xi-k)x_j\big) - \exp\left(i2\pi(\xi-k)\left(x_j + \frac{v-x_j}{2n+1}\right)\right)\right| dv \\ & \leq \frac{2^{3/2}\pi}{2n+1} \left\{ \frac{1}{2n+1}\sum_{j=1}^{n} \int_{-1/2}^{1/2} |v-x_j|\, dv \right\} \left\{ \sum_{|\xi-k|>0} |\xi-k||R(\xi)| \right\}. \end{aligned}$$

Since we have already shown $r \in \mathscr{R}_1$, we have, for $\zeta = \xi - k$, $\max_{|k|>0} \sum_{|\zeta|>0} |\zeta||R(k+\zeta)| < \infty$. Hence, we can find an appropriate constant $C > 0$ for (5.2) to be further bounded by

$$C n^{-2} \sum_{j=1}^{n} \int_{-1/2}^{1/2} |v - x_j|\, dv,$$

which both does not depend on k and is easily seen to be of order $O(n^{-1})$. Combining this fact with the result that $\max_{k\in\mathbb{Z}} R_1(k) = O(n^{-1})$ implies $\max_{k\in\mathbb{Z}} R_2(k) = O(n^{-1})$.

Using the previous arguments we can also show $\max_{k\in\mathbb{Z}} R_3(k) = O(n^{-1})$ and $\max_{k\in\mathbb{Z}} R_4(k) = O(n^{-1})$. Finally, a similar argument for showing $\max_{k\in\mathbb{Z}} R_2(k) = O(n^{-1})$ can be used to show $\max_{k\in\mathbb{Z}} R_5(k) = O(n^{-1})$. This concludes the proof of Lemma 3. □

With the result of Lemma 3, we can give the proof of Lemma 1 from Section 2.1:

Proof of Lemma 1. We begin with the decomposition

$$E\big[\hat{\theta}(x)\big] = \sum_{k=-\infty}^{\infty} \Lambda(h_n k)\Theta(k) \exp(i2\pi k x) + \sum_{k=-\infty}^{\infty} \frac{\Lambda(h_n k)}{\Psi(k)} \Big\{ E\big[\hat{R}(k)\big] - R(k) \Big\} \exp(i2\pi k x)$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

so that $E[\hat{\theta}(x)] - \theta(x)$ is equal to

$$\sum_{k=-\infty}^{\infty} \{\Lambda(h_n k) - 1\}\Theta(k)\exp(i2\pi kx) + \sum_{k=-\infty}^{\infty} \frac{\Lambda(h_n k)}{\Psi(k)}\Big\{E\big[\hat{R}(k)\big] - R(k)\Big\}\exp(i2\pi kx).$$

We can see that $\sup_{x\in[-1/2,\,1/2]} |E[\hat{\theta}(x)] - \theta(x)|$ is bounded by

$$\sum_{k=-\infty}^{\infty} |\Lambda(h_n k) - 1||\Theta(k)| + \max_{k\in\mathbb{Z}} \Big|E\big[\hat{R}(k)\big] - R(k)\Big| \sum_{k=-\infty}^{\infty} \frac{|\Lambda(h_n k)|}{|\Psi(k)|}. \tag{5.4}$$

Partition $\mathbb{Z}$ into $I(h_n) \cup I^c(h_n)$, where $I(h_n) = \{z \in \mathbb{Z} : h_n|z| \le M\} = \{z \in \mathbb{Z} : |z| \le Mh_n^{-1}\}$. Hence, for every $k \in I^c(h_n)$, it follows that $|\Lambda(h_n k)| \le 1$, which implies both statements $|\Lambda(h_n k) - 1| \le 2$ and $|k| > Mh_n^{-1}$ hold. The first term in the right-hand side of (5.4) is therefore bounded by

$$2 \sum_{k\in I^c(h_n)} |\Theta(k)| \le 2h_n^s M^{-s} \sum_{k=-\infty}^{\infty} |k|^s |\Theta(k)|. \tag{5.5}$$

This implies the first term in (5.4) is of the order $O(h_n^s)$, uniformly in $x \in [-1/2,\, 1/2]$.

We now turn to the second term in (5.4). It follows from Assumptions 1 and 2 for the series in this term to be bounded by

$$h_n^{-1}\Big[\min_{k\in\{z\in\mathbb{Z}:\,|z|\le\Gamma\}} |\Psi(k)|\Big]^{-1}\Big\{h_n \sum_{\omega\in h_n\mathbb{Z}} |\Lambda(\omega)|\Big\} + h_n^{-b-1}C_\Psi^{-1}\Big\{h_n \sum_{\omega\in h_n\mathbb{Z}} |\omega|^b|\Lambda(\omega)|\Big\},$$

which is easily seen to be of the order $O(h_n^{-b-1})$. The additional factor of h_n^{-1} appears in the bound above because we have a shrinkage of $|k|$ by h_n. This implies $\sum_{k=-\infty}^{\infty}\{|\Lambda(h_n k)|/|\Psi(k)|\}$ is of the order $O(h_n^{-b-1})$. Now we only need to consider the term $\max_{k\in\mathbb{Z}} |\hat{R}(k) - R(k)|$. The assumptions of Lemma 3 are satisfied. It then follows for $\max_{k\in\mathbb{Z}} |\hat{R}(k) - R(k)| = O(n^{-1})$. Hence, the second term in (5.4) is of the order $O((nh_n^{b+1})^{-1})$, uniformly in $x \in [-1/2,\, 1/2]$. Combining the results above, we have that (5.4) is of the order $O(h_n^s + (nh_n^{b+1})^{-1})$, uniformly in $x \in [-1/2,\, 1/2]$, and the assertion of Lemma 1 follows. □

We are now prepared to state the proof of Lemma 2.

Proof of Lemma 2. Without loss of generality we can assume that $n \ge 3$. Our argument is similar to the arguments found in Masry (1993), who gives related results for an errors-in-variables model. We will employ truncation as follows. Let the stabilizing sequence $\{\eta_n\}_{n\ge3}$ satisfy $\eta_n = O((nh_n^{2b+1})^{-1/2}\log^{1/2}(n))$ and the truncation sequence $\{t_n\}_{n\ge3}$ satisfy $t_n = O((n\log(n)(\log\log(n))^{1+\delta})^{1/\kappa})$, with $\delta > 0$. Write $K_j = E^{1/\kappa}[|Y_j|^\kappa]$. We can decompose $\hat{\theta}(x) - E[\hat{\theta}(x)]$ into the sum of $D_1(x) = \hat{\theta}(x) - \hat{\theta}^t(x)$, $D_2(x) = E[\hat{\theta}^t(x)] - E[\hat{\theta}(x)]$ and $D_3(x) = \hat{\theta}^t(x) - E[\hat{\theta}^t(x)]$, where

$$\hat{\theta}^t(x) = \frac{1}{2n+1}\sum_{j=-n}^{n} Y_j \mathbf{1}\big[|Y_j| \le K_j t_n\big] W_{j,h_n}(x), \qquad x \in [-1/2,\, 1/2].$$

Beginning with $D_1(x)$, it follows along the same lines as the arguments in the proof of Lemma 2.1 of Masry (1993) for $\sup_{x\in[-1/2,\,1/2]} |D_1(x)| = o(\eta_n)$, almost surely. Turning our

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

attention now to $D_2(x)$, it is easy to show $\sup_{x\in[-1/2,\,1/2]}|W_{j,h_n}(x)|$ is bounded by the series $\sum_{k=-\infty}^{\infty}\{|\Lambda(h_nk)|/|\Psi(k)|\}$, and we have already shown this series is of the order $O(h_n^{-b-1})$ in the proof of Lemma 1. Hence, we have that $\sup_{x\in[-1/2,\,1/2]}|W_{j,h_n}(x)| = O(h_n^{-b-1})$. It follows that we can find an appropriate constant $C>0$ such that we can bound $\sup_{x\in[-1/2,\,1/2]}|D_2(x)|$ by

$$Ch_n^{-b-1}\frac{1}{2n+1}\sum_{j=-n}^{n}E\Big[|Y_j|\mathbf{1}\big[|Y_j|>K_jt_n\big]\Big]. \tag{5.6}$$

Since $\kappa>1$, writing $M_K=\max_{j=-n,\ldots,n}K_j$, we can apply Markov's inequality to obtain

$$\max_{j=-n,\ldots,n}E[|Y_j|\mathbf{1}[|Y_j|>K_jt_n]] = \max_{j=-n,\ldots,n}\int_0^\infty P\big(|Y_j|>\max\{s,\,K_jt_n\}\big)\,ds \le \frac{\kappa}{\kappa-1}M_Kt_n^{1-\kappa}.$$

Enlarging the constant C in (5.6) implies $\sup_{x\in[-1/2,\,1/2]}|D_2(x)|\le Ch_n^{-b-1}t_n^{1-\kappa}=o(\eta_n)$.

To continue, we will require an additional result. For any $u,v\in[-1/2,\,1/2]$, we can repeat the arguments in the proof of Lemma 3 to see that

$$\Big|W_{j,h_n}(u)-W_{j,h_n}(v)\Big| \le \big|u-v\big|2^{3/2}\pi\sum_{k=-\infty}^{\infty}|k|\frac{|\Lambda(h_nk)|}{|\Psi(k)|}.$$

Hence, arguing as in the proof of Lemma 1 we can find an appropriate constant $C>0$ such that

$$\big|W_{j,h_n}(u)-W_{j,h_n}(v)\big|\le Ch_n^{-b-2}|u-v|,\qquad u,v\in[-1/2,\,1/2]. \tag{5.7}$$

Now we consider $D_3(x)$. Let $\{s_n\}_{n\ge3}$ be a sequence satisfying $s_n=O(h_n^{b+2}\eta_nt_n^{-1})=o(1)$ such that, when we shatter the interval $[-1/2,\,1/2]$ into s_n^{-1} many fragments of the form $(x_i,\,x_{i+1}]$, with the first fragment defined to be $[-1/2,\,x_2]=\{-1/2\}\cup(-1/2,\,x_2]$, our fragments satisfy $\max_{i=1,\ldots,s_n^{-1}}|x_{i+1}-x_i|\le s_n$. For any $x\in[-1/2,\,1/2]$, there is exactly one fragment $(x_{i'},\,x_{i'+1}]$ that contains x, and on this interval we can write

$$D_3(x)=D_{4,i}(x)-D_{5,i}(x)+D_{6,i},$$

where $D_{4,i}(x)=\hat\theta^t(x)-\hat\theta^t(x_{i'})$, $D_{5,i}(x)=E[\hat\theta^t(x)]-E[\hat\theta^t(x_{i'})]$ and $D_{6,i}=\hat\theta^t(x_{i'})-E[\hat\theta^t(x_{i'})]$. It follows that $\sup_{x\in[-1/2,\,1/2]}|D_3(x)|$ is bounded by

$$\max_{i=1,\ldots,s_n^{-1}}\sup_{x\in(x_i,\,x_{i+1}]}\big|D_{4,i}(x)\big|+\max_{i=1,\ldots,s_n^{-1}}\sup_{x\in(x_i,\,x_{i+1}]}\big|D_{5,i}(x)\big|+\max_{i=1,\ldots,s_n^{-1}}\big|D_{6,i}\big|.$$

Hence, to show the result $\sup_{x\in[-1/2,\,1/2]}|D_3(x)|=O(\eta_n)$, almost surely, we will instead show that each of the following statements hold:

$$\max_{i=1,\ldots,s_n^{-1}}\sup_{x\in(x_i,\,x_{i+1}]}|D_{4,i}(x)|=O(\eta_n),\quad\text{a.s.,} \tag{5.8}$$

$$\max_{i=1,\ldots,s_n^{-1}}\sup_{x\in(x_i,\,x_{i+1}]}|D_{5,i}(x)|=O(\eta_n) \tag{5.9}$$

and

$$\max_{i=1,\ldots,s_n^{-1}}|D_{6,i}|=O(\eta_n),\quad\text{a.s..} \tag{5.10}$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Beginning with (5.8), fix an arbitrary interval $(x_i,\, x_{i+1}]$. On this interval $D_{4,i}(x)$ is equal to

$$\frac{1}{2n+1}\sum_{j=-n}^{n} Y_j \mathbf{1}\big[|Y_j| \le K_j t_n\big]\big\{W_{j,h_n}(x) - W_{j,h_n}(x_i)\big\}, \qquad x \in (x_i,\, x_{i+1}].$$

It follows from (5.7) that we can find an appropriate constant $C > 0$ for the inequality $\sup_{x\in(x_i,\,x_{i+1}]} |D_{4,i}(x)| \le C t_n h_n^{-b-2} s_n$ to hold, almost surely, independent of i. Therefore, by construction of $\{s_n\}_{n\ge 3}$, we find that (5.8) holds. Observing that $D_{5,i}(x) = E[D_{4,i}(x)]$, we have that (5.9) holds as well.

To see the final statement (5.10) holds, define the random variables $U_{j,i} = \{Y_j \mathbf{1}[|Y_j| \le K_j t_n] - E[Y_j \mathbf{1}[|Y_j| \le K_j t_n]]\} W_{j,h_n}(x_i)$, $j = -n, \ldots, n$. Standard arguments can then be used to show that $U_{-n,i}, \ldots, U_{n,i}$ are independent, and each have mean equal to zero, variance bounded by $C_1 h_n^{-2b-1}$ and bounded in absolute value by $C_2 t_n h_n^{-b-1}$, where $C_1 > 0$ and $C_2 > 0$ are appropriately chosen constants and both bounds are independent of both j and i. Applying Bernstein's Inequality (see, for example, Lemma 2.2.11 in van der Vaart and Wellner, 1996), we can find an appropriate constant $C > 0$ and obtain

$$P\Big(\max_{i=1,\ldots,s_n^{-1}} |D_{6,i}| > \eta_n\Big) \le 2 s_n^{-1} \exp\Big(-C \frac{n\eta_n^2}{h_n^{-2b-1} + t_n h_n^{-b-1}\eta_n}\Big). \tag{5.11}$$

In light of the fact that $t_n h_n^{-b-1}\eta_n = o(h_n^{-2b-1})$, which holds since $\kappa > 2 + 1/b$, we can enlarge C for the right-hand side of (5.11) to be further bounded by a positive constant multiplied by

$$h_n^{-3/2} n^{(1/2)+(1/\kappa)-C} \log^{-(1/2-1/\kappa)}(n)\big(\log\log(n)\big)^{(1+\delta)/\kappa},$$

which is summable provided we take $C > (3/2)(1+1/(2b+1)) + 1/\kappa$, where $1/(2b+1)$ accounts for the expansion of h_n^{-1}; i.e. $(n^{1/(2b+1)} h_n)^{-3/2} \to 0$, as $n \to \infty$. It then follows by the Borel-Cantelli lemma that (5.10) holds. This completes the proof. □

We can now state the proof of Theorem 1 from Section 2.1:

Proof of Theorem 1. The first two assertions follow immediately from the results of Lemma 1 and Lemma 2 in combination with our choice of regularizing sequence as discussed in Section 2.1. This means we only need to show the last assertion. Let us begin by calculating the Fourier coefficients $\{\hat{\Theta}(\xi)\}_{\xi\in\mathbb{Z}}$ of $\hat{\theta}$:

$$\begin{aligned}\hat{\Theta}(\xi) &= \int_{-1/2}^{1/2} \hat{\theta}(x) e^{-i2\pi\xi x}\, dx = \sum_{k=-\infty}^{\infty} \frac{\Lambda(h_n k)}{\Psi(k)} \hat{R}(k) \int_{-1/2}^{1/2} e^{i2\pi(k-\xi)x}\, dx \\ &= \Lambda(h_n\xi)\Theta(\xi) + \left\{ E\big[\hat{R}(\xi)\big] - R(\xi) + \frac{1}{2n+1}\sum_{j=-n}^{n} \varepsilon_j e^{-i2\pi\xi x_j} \right\} \frac{\Lambda(h_n\xi)}{\Psi(\xi)},\end{aligned}$$

where we have used the orthonormality of the basis $\{\exp(i2\pi kx) : x \in [-1/2,\, 1/2]\}_{k\in\mathbb{Z}}$ in the final equality. The definition of $\mathscr{R}_{s-1/2}$ requires that we show the series condition

$$\sum_{\xi=-\infty}^{\infty} |\xi|^{s-1/2} \big|\hat{\Theta}(\xi)\big| < \infty \tag{5.12}$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

is satisfied. We can see that $|\hat{\Theta}(\xi)|$ is bounded by

$$|\Theta(\xi)| + \left\{ \max_{k\in\mathbb{Z}} \left| E\big[\hat{R}(k)\big] - R(k) \right| + \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \varepsilon_j e^{i2\pi\xi x_j} \right| \right\} \frac{|\Lambda(h_n\xi)|}{|\Psi(\xi)|}. \tag{5.13}$$

Observing that $\theta \in \mathscr{R}_s$ implies $\theta \in \mathscr{R}_{s-1/2}$, we have $\sum_{\xi=-\infty}^{\infty} |\xi|^{s-1/2} |\Theta(\xi)| < \infty$. Hence, we only need to verify the series condition (5.12) stated for the last term in (5.13) holds.

The assumptions of Lemma 3 are satisfied, which gives $\max_{k\in\mathbb{Z}} |E[\hat{R}(k)] - R(k)| = O(n^{-1})$. Additionally, the map $x \mapsto \exp(-i2\pi kx)$ is confined to the unit circle in the complex plane. A standard argument then shows

$$\max_{k\in\mathbb{Z}} \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \varepsilon_j e^{-i2\pi k x_j} \right| = O\big(n^{-1/2} \log^{1/2}(n)\big), \qquad \text{a.s.}$$

Finally, in the proof of Lemma 1, we have shown $\sum_{k=-\infty}^{\infty} \{|\Lambda(h_n k)|/|\Psi(k)|\} = O(h_n^{-b-1})$, and similar lines of argument can be used to show $\sum_{k=-\infty}^{\infty} \{|k|^{s-1/2} |\Lambda(h_n k)|/|\Psi(k)|\} = O(h_n^{-s-b-1/2})$ with the assumption $\int_{-\infty}^{\infty} |u|^{s+b-1/2} |\Lambda(u)|\, du < \infty$. Together, these results imply

$$\left\{ \max_{k\in\mathbb{Z}} \left| E\big[\hat{R}(k)\big] - R(k) \right| + \max_{k\in\mathbb{Z}} \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \varepsilon_j e^{-i2\pi k x_j} \right| \right\} \sum_{\xi=-\infty}^{\infty} |\xi|^{s-1/2} \frac{|\Lambda(h_n\xi)|}{|\Psi(\xi)|}$$

is of order $O(1 + (n\log(n))^{-1/2}) = O(1)$, almost surely. Hence, the series condition (5.12) stated for the last term in (5.13) holds. It follows that $\hat{\theta} - \theta \in \mathscr{R}_{s-1/2}$, almost surely, for large enough n. Combining this statement with the first assertion then proves the third assertion. □

Nickl and Pötscher (2007) study classes of functions of Besov- and Sobolev-type. These authors derive results concerning the bracketing metric entropy and the related central limit theorems of these spaces using weighted norms. We can see the results of their Corollary 4 on bracketing numbers for weighted Sobolev spaces can be extended to our case (see Section 3.3.4 on page 189). We summarize this result in the following proposition:

PROPOSITION 2. *For the function space $\mathscr{R}_{s,1}$, with $s > 1/2$, a finite constant $C > 0$ exists such that*

$$\log N_{[]}\big(\epsilon, \mathscr{R}_{s,1}, \|\cdot\|_\infty\big) \le C\epsilon^{-1/s}, \qquad \epsilon > 0,$$

where $N_{[]}(\epsilon, \mathscr{R}_{s,1}, \|\cdot\|_\infty)$ is the number of brackets of length ϵ required to cover the metric space $(\mathscr{R}_{s,1}, \|\cdot\|_\infty)$.

In light of the results on the estimator $\hat{\theta}$, we can now state a result on the modulus of continuity relating $\hat{\mathbb{F}}(t)$ to $(2n+1)^{-1} \sum_{j=-n}^{n} \mathbf{1}[\varepsilon_j \le t]$. Using results on Donsker classes of functions, we can show this modulus of continuity holds up to a negligible term of order $o_P(n^{-1/2})$. The proof of this result follows along the same lines as the proof of Lemma A.1 in Van Keilegom and Akritas (1999) and, therefore, it is omitted (see also Neumeyer, 2009).

LEMMA 4. *Let the assumptions of Theorem 1 be satisfied with $s > 3/2$. In addition, assume that F admits a bounded Lebesgue density function f. Then $\sup_{t\in\mathbb{R}} |M_n(t)| = o_P(n^{-1/2})$, where*

$$M_n(t) = \frac{1}{2n+1} \sum_{j=-n}^{n} \mathbf{1}\big[\varepsilon_j \le t + \big[K\big(\hat{\theta} - \theta\big)\big](x_j)\big] - \int_{-1/2}^{1/2} F\big(t + \big[K\big(\hat{\theta} - \theta\big)\big](x)\big)\, dx$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$$- \frac{1}{2n+1} \sum_{j=-n}^{n} \mathbf{1}[\varepsilon_j \leq t] + F(t).$$

Direct regression estimators typically allow for appropriate expansions into averages of the model errors up to some negligible remainder term. This representation motivates the term $\varepsilon f(t)$ in the expansion of the empirical distribution function of the these model residuals. In the following result, we provide a similar expansion for the indirect regression estimator $\hat{\theta}$, and we show this expansion holds up to a negligible term of order $o_P(n^{-1/2})$. Hence, we can immediately see that our indirect regression function estimator $\hat{\theta}$ and typical direct regression function estimators share this property. This combined with the modulus of continuity result above implies that our residual-based empirical distribution function behaves similarly to that in the usual direct estimation setting (see, for example, Müller, Schick and Wefelmeyer, 2007, who construct expansions for many residual-based empirical distribution functions based on direct regression function estimators).

PROPOSITION 3. *Let the assumptions of Lemma 1 be satisfied, and assume that* $E[\varepsilon_j^2] < \infty$, $j = -n, \ldots, n$. *Let the regularizing sequence* $\{h_n\}_{n\geq 1}$ *satisfy* $h_n^{s+b+1} = o(n^{-1/2})$ *and* $(nh_n)^{-1} = o(n^{-1/2})$. *Then*

$$\left| \int_{-1/2}^{1/2} \big[K\big(\hat{\theta} - \theta\big)\big](x)\, dx - \frac{1}{2n+1} \sum_{j=-n}^{n} \varepsilon_j \right| = o_P(n^{-1/2}).$$

PROOF. Note that $\hat{R}(k) - E[\hat{R}(k)] = (2n+1)^{-1} \sum_{j=-n}^{n} \varepsilon_j \exp(-i2\pi k x_j)$. We can write $1 = \int_{-1/2}^{1/2} \sum_{k=-\infty}^{\infty} e^{i2\pi k x}\, dx$ so that we can bound the left-hand side of the assertion by $S_1 + S_2 + S_3$, where

$$S_1 = \left| \frac{1}{2n+1} \sum_{j=-n}^{n} \varepsilon_j \int_{-1/2}^{1/2} \left\{ \sum_{k=-\infty}^{\infty} \big\{\Lambda(h_n k) - 1\big\} e^{i2\pi k(x - x_j)} \right\} dx \right|,$$

$$S_2 = \sum_{k=-\infty}^{\infty} \big|\Lambda(h_n k) - 1\big| \big|R(k)\big| \left| \int_{-1/2}^{1/2} e^{i2\pi k x}\, dx \right|.$$

and

$$S_3 = \left[\max_{k \in \mathbb{Z}} \Big| E\big[\hat{R}(k)\big] - R(k) \Big| \right] \sum_{k=-\infty}^{\infty} |\Lambda(h_n k)|$$

The assertion then follows, if we show $S_1 = o_P(n^{-1/2})$, $S_2 = o(n^{-1/2})$ and $S_3 = o(n^{-1/2})$.

We can see that it follows for $S_1 = o_P(n^{-1/2})$ from Assumption 2 and the fact that

$$\frac{1}{2n+1} \sum_{j=-n}^{n} \left\{ \int_{-1/2}^{1/2} \left\{ \sum_{k=-\infty}^{\infty} \big\{\Lambda(h_n k) - 1\big\} e^{i2\pi k(x - x_j)} \right\} dx \right\}^2 = o(1).$$

To show that $S_2 = o(n^{-1/2})$, recall the convolution theorem for Fourier transformation implies that $|R(k)| = |\Theta(k)||\Psi(k)|$. The integral term in S_2 is bounded by a positive constant C multiplied by $|k|^{-1}$. Combining this fact with the constant C_Ψ^* from Assumption 1 and that $|\Lambda(h_n k) - 1| \leq 2$ shows that we can enlarge C such that S_2 is bounded by

$$C h_n^{s+b+1} \sum_{k=-\infty}^{\infty} |k|^s |\Theta(k)| = O\big(h_n^{s+b+1}\big) = o\big(n^{-1/2}\big).$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Finally, we consider the last remainder term S_3. The assumptions of Lemma 3 are satisfied, which gives $\max_{k\in\mathbb{Z}} |E[\hat{R}(k)] - R(k)| = O(n^{-1})$. Similar lines of argument to those in the proof of Lemma 1 shows the series term in S_3 is of the order $O(h_n^{-1})$. Combining these facts, we have that S_3 is of the order $O((nh_n)^{-1}) = o(n^{-1/2})$, which concludes the proof. □

Combining the results above, we can now state the proof of Theorem 2.

Proof of Theorem 2. Recall $M_n(t)$ from Lemma 4. A straightforward calculation shows that

$$\frac{1}{2n+1}\sum_{j=-n}^{n}\Big\{\mathbf{1}\big[\hat{\varepsilon}_j \le t\big] - \mathbf{1}\big[\varepsilon_j \le t\big] - \varepsilon_j f(t)\Big\} = M_n(t) + H_n(t) + L_n(t),$$

where

$$H_n(t) = \int_{-1/2}^{1/2} F\big(t + \big[K\big(\hat{\theta}-\theta\big)\big](x)\big)\,dx - F(t) - f(t)\int_{-1/2}^{1/2}\big[K\big(\hat{\theta}-\theta\big)\big](x)\,dx$$

and

$$L_n(t) = f(t)\Big\{\int_{-1/2}^{1/2}\big[K\big(\hat{\theta}-\theta\big)\big](x)\,dx - \frac{1}{2n+1}\sum_{j=-n}^{n}\varepsilon_j\Big\}.$$

The assumptions of Lemma 4 are satisfied, which implies $\sup_{t\in\mathbb{R}}|M_n(t)| = o_P(n^{-1/2})$. Hence, the assertion follows from showing $\sup_{t\in\mathbb{R}}|H_n(t)| = o_P(n^{-1/2})$ and $\sup_{t\in\mathbb{R}}|L_n(t)| = o_P(n^{-1/2})$.

Beginning with $H_n(t)$, writing $C_{f,\gamma}$ for the Hölder constant of f with exponent γ, we have

$$H_n(t) = \int_{-1/2}^{1/2}\big[K\big(\hat{\theta}-\theta\big)\big](x)\int_0^1\Big\{f\big(t + s\big[K\big(\hat{\theta}-\theta\big)\big](x)\big) - f(t)\Big\}\,ds\,dx$$

so that $\sup_{t\in\mathbb{R}}|H_n(t)|$ is bounded by

$$\frac{C_{f,\gamma}}{1+\gamma}\Big[\sup_{x\in[-1/2,\,1/2]}\Big|\hat{\theta}(x) - \theta(x)\Big|\Big]^{1+\gamma}.$$

The assumptions of Theorem 1 are satisfied, which implies the second term in the bound above is $o(n^{-1/2})$, almost surely. It then follows that $\sup_{t\in\mathbb{R}}|H_n(t)| = o_P(n^{-1/2})$.

Now we will consider $L_n(t)$. Since f is bounded, we have that $\sup_{t\in\mathbb{R}}|L_n(t)|$ is bounded by

$$\sup_{t\in\mathbb{R}}|f(t)|\Big|\int_{-1/2}^{1/2}\big[K\big(\hat{\theta}-\theta\big)\big](x)\,dx - \frac{1}{2n+1}\sum_{j=-n}^{n}\varepsilon_j\Big|.$$

The parameter sequence $\{h_n\}_{n\ge1}$ satisfies (2.3), and we have both

$$h_n^{s+b+1} = O\big(n^{-1/2-1/(4s+4b+2)}\log^{(s+b+1)/(2s+2b+1)}(n)\big) = o\big(n^{-1/2}\big)$$

and

$$\big(nh_n\big)^{-1} = O\big(n^{-(2s+2b)/(2s+2b+1)}\log^{-1/(2s+2b+1)}(n)\big) = o\big(n^{-1/2}\big),$$

which follows from the fact that $s + b > 1/2$. The assumptions of Proposition 3 are satisfied, which implies the second term in the bound above is $o_P(n^{-1/2})$. This shows that $\sup_{t\in\mathbb{R}}|L_n(t)| = o_P(n^{-1/2})$, and, hence, the assertion of Theorem 2 holds. □

Here we provide a short proof of Proposition 1.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

PROOF OF PROPOSITION 1. Beginning with the first assertion, we can write the integrated variance of $\hat{\theta}$ as

$$\int_{-1/2}^{1/2} E\Big[\big\{\hat{\theta}(x) - E\big[\hat{\theta}(x)\big]\big\}^2\Big]\, dx = \frac{\sigma^2}{2n+1} \sum_{k=-\infty}^{\infty} \frac{\Lambda^2(h_n k)}{\Psi^2(k)}.$$

Repeating the arguments in the proof of Lemma 1 in Section 5 then shows

$$\sum_{k=-\infty}^{\infty} \frac{\Lambda^2(h_n k)}{\Psi^2(k)} = O\big(h_n^{-2b-1}\big).$$

Therefore, we can specify $C_\Lambda > 0$ for the first assertion to hold. The second assertion follows directly by an application of Lemma 1. □

Now returning to the discussion in Section 3.1, the choice of scaling sequence $\{c_n\}_{n\geq 1}$ used for the contaminates $c_n U_j$, $j = -n, \ldots, n$, in the smooth bootstrap, always satisfies $(nc_n)^{-1}\log(n) = o(1)$, and, hence, we can apply Theorem A of Silverman (1978) in combination with the Hölder continuity of w and the results of Theorem 1, or the combination of the results from Lemma 1 and Lemma 2, to show that f_n^* is strongly consistent for f, uniformly over the entire real line. The result only holds when the density function f is Hölder with exponent $2/3 < \gamma \leq 1$, the density function w is also chosen to be Hölder continuous with similar smoothness, and the smoothness index s of the function space $\mathscr{R}_s$ satisfies $s > (1+\gamma)(2b+1)/(3\gamma-2)$, which is larger than the lower bound on s required by the second statement of Theorem 1. The additional smoothness in θ is required due to the fact that residuals are used in the estimator f_n^* rather than the model errors. Under these conditions, further technical but standard arguments similar to those used to prove related results in Neumeyer (2009) can be used to prove the following result.

PROPOSITION 4. *Let the assumptions of Theorem 3 be satisfied, but now requiring the density f to be Hölder continuous with exponent $2/3 < \gamma \leq 1$ and choosing the density w also to be Hölder continuous with the same exponent. Additionally, assume the smoothness index s of the function space $\mathscr{R}_s$ satisfies $s > (1+\gamma)(2b+1)/(3\gamma-2)$. Then*

$$\sup_{t\in\mathbb{R}} \Big|f_n^*(t) - f(t)\Big| = o_P(1),$$

$$\sup_{t\in\mathbb{R}} \Big|F_n^*(t) - F(t)\Big| = o_P(1)$$

and

$$\sup_{t\in\mathbb{R}} \Big|E^*\big[\varepsilon^* \mathbf{1}\big[\varepsilon^* \leq t\big]\big] - E\big[\varepsilon \mathbf{1}[\varepsilon \leq t]\big]\Big| = o_P(1).$$

We omit proof of the following result because it is proven in exactly the same manner as Theorem 1.

PROPOSITION 5. *Let the assumptions of Theorem 1 be satisfied. Choose the regularizing sequence $\{g_n\}_{n\geq 1}$ according to (2.3) and let the scaling sequence $\{c_n\}_{n\geq 1}$ satisfy $c_n = O(n^{-\alpha})$, with $0 < \alpha < 1/2 + 1/\kappa$. Then, P^*-almost surely, we have*

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

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

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

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

and, for large enough n,

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

References

[1] Adorf, H.M. (1995). Hubble Space Telescope image reconstruction in its fourth year. *Inverse Problems* **11**, 639-653.
[2] Bertero, M., Boccacci, P., Desiderà, G. and Vicidomini, G. (2009). Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems* **25**, 123006.
[3] Birke, M., Bissantz, N. and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems* **26**, 115020.
[4] Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24**, 034009.
[5] Bissantz, N. and Holzmann, H. (2013). Asymptotics for spectral regularization estimators in statistical inverse problems. *Computat. Statist.* **28**, 435-453.
[6] Bissantz, N., Mair, B. and Munk, A. (2006). A multi-scale stopping criterion for MLEM reconstructions in PET. *IEEE Nuclear Science Symposium Conference Record* **6**, 3376-3379.
[7] Bissantz, N., Mair, B. and Munk, A. (2008). A statistical stopping rule for MLEM reconstructions in PET. *IEEE Nuclear Science Symposium Conference Record* **8**, 4198-4200.
[8] Blanchard, G., Hoffmann, M. and Reiß, M. (2016). Optimal adaptation for early stopping in statistical inverse problems. arXiv:1606.07702v1.
[9] Blanchard, G. and Mathé, P. (2012). Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems* **28**, 115011.
[10] Cao, R. (1993). Bootstrapping the mean integrated squared error. *J. Multivariate Anal.* **45**, 137-160.
[11] Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Ann. Statist.* **28**, 630-647.
[12] Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems* **24**, 034004.
[13] Cavalier, L. and Golubev, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34**, 1653-1677.
[14] Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123**, 323-354.
[15] Davies, P.L. and Meise, M. (2008). Approximating data with weighted smoothing splines. *J. Nonparametr. Stat.* **20**, 207-228.
[16] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.
[17] Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5**, 907-925.
[18] 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.
[19] Hall, P. and Horowitz, J. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Ann. Statist.* **33**, 2904-2929.
[20] 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.
[21] Mair, B.A. and Ruymgaart, F.H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56**, 1424-1444.
[22] Marteau, C. and Mathé, P. (2014). General regularization schemes for signal detection in inverse problems. *Math. Methods Statist.* **23**, 176-200.
[23] Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37**, 1105-1115.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

[24] Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary processes. *J. Nonparametr. Stat.* **3**, 13-36.
[25] Mathé, P. and Pereverzev, S.V. (2006). Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comp.* **75**, 1913-1929.
[26] 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.
[27] Müller, U.U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions* **25**, 1-18.
[28] Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scand. J. Stat.* **36**, 204-228.
[29] Nickl, R. and Pötscher, B.M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20**, 177-199.
[30] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.
[31] 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.
[32] Proksch, K., Bissantz, N. and Dette, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli* **21**, 144-175.
[33] Rochet, P. (2013). Adaptive hard-thresholding for linear inverse problems. *ESAIM Probab. Stat.* **17**, 485-499.
[34] Silverman, B.W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**, 177-184.
[35] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Vol. 26. CRC press.
[36] van der Vaart, A.W. and Wellner J.A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York.
[37] Van Keilegom, I. and Akritas, M.G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745-1784.