# 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, deconvolution function estimator, indirect nonparametric regression, 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. Hence, important quantities that are of interest to the study are not directly available for statistical inference, but images of these quantities under some transformation such as a convolution can be used instead. These so–called inverse problems frequently occur, e.g. in signal detection or biological and medical imaging. A common example is the reconstruction of astronomical images, where the connection between the true image and the observable image is at least approximately given by convolution–type operators (see Adorf, 1995 or Bertero et al., 2009). Another typical example occurs in reconstruction of medical images like those obtained from Positron Emission Tomography. Here the connection between the true image and the observations involves the Radon Transform (see Cavalier, 2000).

In this article we consider an inverse regression model, i.e. observing a signal of interest from indirect observations

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

Corresponding author: Justin Chown (justin.chown@ruhr-uni-bochum.de)
*Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl für Stochastik, 44780 Bochum, DE.*

where $K$ is an operator specifying convolution of the true underlying regression $\theta$ with a point spread function $\psi$, i.e.

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

The resulting function $r$ can be viewed as a blurred regression function. We will assume that $\psi$ is known and behaves like a probability density function on the interval $[-1/2,\, 1/2]$, i.e. $\psi$ is positive–valued on the interval $[-1/2,\, 1/2]$ and integrates to one so that $K1 = 1$. However, we will only assume that $\theta$ is known to be smooth. The covariates $x_j$ in model (1.1) are uniformly distributed design points in the interval $[-1/2,\, 1/2]$, i.e. $x_j = j/2n$, $j = -n, \ldots, n$. The errors $\varepsilon_j$ 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 $\theta$ is identifiable, and later we will require further assumptions for our results to hold.

Statistical inverse problems have received a great amount of attention on construction of estimators for various densities and indirect regression models. Masry (1991) constructs an estimator of a multivariate density function in an errors-in-variables model. Later, Masry (1993) investigates an estimator of a multivariate regression function in a similar errors-in-variables model. Mair and Ruymgaart (1996) consider estimates of the indirect regression function using a regularized inversion technique and show their constructions satisfy minimax optimality. Politis and Romano (1999) estimate a multivariate density function by introducing the flat-top kernel approach, and in a highly constrained model they show their technique is root–$n$ consistent. Cavalier and Tsybakov (2002) consider estimating the indirect regression function in a heteroskedastic model and find their approach satisfies minimax optimality. Bissantz and Holzmann (2008) construct confidence intervals and confidence bands for a variety of statistical inverse problems. Later, Bissantz and Holzmann (2013) investigate similar problems using spectral regularization. Birke, Bissantz and Holzmann (2010) construct uniform confidence bands for the indirect regression function in a univariate regression model, and this technique is extended to the case multiple covariates by Proksch, Bissantz and Dette (2015).

All of these estimators depend on some kind of regularization parameter. This quantity is similar to the bandwidth 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. Popular approaches for choosing this regularization parameter are based on multiscale and related methods (see, for example, Bissantz, Mair and Munk, 2006, Bissantz, Mair and Munk, 2008, Davies and Meise, 2008, González-Manteiga, Martinez-Miranda and Pérez-González, 2004 and Hotz et al., 2012). From a different perspective, 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 estimates generated from a sequence of regularization parameters. In the case of iterative estimation procedures this is the problem of finding a stopping iteration. In this article, we provide a statistical methodology for selecting a best fitting (most feasible) regression estimate from a sequence of function estimates based on observations from model (1.1) using the resulting model residuals.

Many statistical procedures are residual–based, which requires studying the distribution function $F$ of the model errors, which is in general unknown and must be estimated. We estimate $F$ using the empirical distribution function of the model residuals:

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

where $\hat{\theta}$ is a suitable estimate of $\theta$, which depends on a regularization parameter. There are many results in the literature on signal deconvolution problems that motivate our work on $\hat{\theta}$. In addition, there are many results in the literature on residual–based empirical distribution functions in direct regression models; for example, consistency and asymptotic optimality.

To the best of our knowledge, very little attention has been paid to the analysis of residuals from indirect regression modeling. Our work is new in the sense that we show the empirical distribution function of the residuals $\hat{\mathbb{F}}$ behaves similarly in the indirect regression model as it does in the usual nonparametric regression model, which has broad implications for the construction of residual–based tests for indirect regression models. We then use these results to develop a valid smooth bootstrap technique, which uses the residuals from the indirect regression model (1.1), to find an optimal regularization parameter for the estimator $\hat{\theta}$.

The article is organized as follows. Some notation and the estimation method are introduced in Section 2. We present our main results in Section 3, where we characterize the crucial technical properties of the indirect regression function estimator $\hat{\theta}$ investigated in this paper and the resulting uniform expansion of the residual–based empirical distribution function $\hat{\mathbb{F}}$. In Section 3.1, 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). We also develop a data–driven approach for selecting the regularization parameter for the estimator $\hat{\theta}$ using a smooth bootstrap of the model residuals in the spirit of Neumeyer (2009). We conclude the article with a numerical study in Section 4, which indicates good finite sample performance of the proposed regularization. Many of the technical details used in the proofs of our results are given 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

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

In order to construct an estimator for the function $\theta$ we will need to restrict $\theta$ to a smooth class of functions from $\mathcal{L}_2([-1/2,\, 1/2])$. This means we only consider functions $q$ 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 $q^{(i)}$, $1 \le i \le d$, 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 $\phi$ with support $[-1/2,\, 1/2]$ that have evaluations of $\phi$, $(d^{i'}\phi)/(dx^{i'})$, $i' = 1, \ldots, i$, at $1/2$ and $-1/2$ equal to zero. We can define the space of functions $\mathscr{R}_d$ as

$$\mathscr{R}_d = \Big\{ q \in \mathcal{L}_2([-1/2,\, 1/2]) \,:\, q^{(1)}, \ldots, q^{(d)} \in \mathcal{L}_2([-1/2,\, 1/2]) \Big\},$$

and a norm for these functions is given by

$$\|q\|_d^2 = \sum_{i=0}^{d} \int_{-1/2}^{1/2} \big| q^{(i)}(x) \big|^2 \, dx,$$

writing $q^{(0)}$ for $q$. Using the Plancherel identity, this norm has an equivalent representation, for an appropriate constant $C > 0$,

$$\|q\|_d^2 = C \sum_{k=-\infty}^{\infty} \big(1 + k^2\big)^d |\rho(k)|^2,$$

where $\{\rho(k)\}_{k \in \mathbb{Z}}$ are the Fourier coefficients of $q$, i.e.

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

Replacing $d$ with a positive real number motivates considering smoothness orders $s > 0$, where $\mathscr{R}_s$ now becomes a Sobolev space of smoothness $s$, i.e.

$$\mathscr{R}_s = \Big\{ q \in \mathcal{L}_2([-1/2,\, 1/2]) \,:\, \sum_{k=-\infty}^{\infty} \big(1 + k^2\big)^s |\rho(k)|^2 < \infty \Big\}.$$

Note, whenever $\theta \in \mathscr{R}_s$ we have $r = K\theta \in \mathscr{R}_s$, and the characteristic series in the definition of $\mathscr{R}_s$ defines a restriction on the Fourier coefficients $\{R(k)\}_{k \in \mathbb{Z}}$ of the blurred regression $r$, which are defined similarly to the Fourier coefficients $\{\rho(k)\}_{k \in \mathbb{Z}}$ above. This has particular advantages. For example, suppose we wanted to check whether or not a specific function belongs to $\mathscr{R}_s$. Using this norm, we only need to calculate the Fourier coefficients of our function and check whether or not the series condition in the definition of $\mathscr{R}_s$ holds. This is in contrast to Hölder spaces, where checking whether or not a function belongs to the space is often more difficult because it typically involves direct calculation of derivatives and proving statements in the supremum norm.

Much of the research in the area of deconvolution problems has focused into two important cases. The first case is that of the so–called ordinarily smooth point spread functions, and the second case is that of the so–called super smooth point spread functions. The first case means assuming the Fourier coefficients $\{\Psi(k)\}_{k \in \mathbb{Z}}$ of $\psi$, which are defined similarly to the Fourier coefficients $\{\rho(k)\}_{k\in\mathbb{Z}}$ above, decay at a polynomic rate: there are constants $b > 0$ and $C_\Psi > 0$ such that $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$. Under this assumption, we can construct an estimator $\hat{\theta}$ for $\theta$ 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. While the second case means assuming the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ of $\psi$ decay at an exponential rate: there are constants $b_0 \in \mathbb{R}$, $b > 0$, $C > 0$ and $C_\Psi > 0$ such that $|k|^{-b_0}\exp(C|k|^b)|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$. 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 point spread functions $\psi$.

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

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

The double integral in the right-hand side of (2.1) is an average, and, therefore, we can estimate it using the empirical average from our data $(x_j, Y_j)$, $j = -n, \ldots, n$, to obtain

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

To estimate $\theta$ defined by the equation $r(\cdot) = [K\theta](\cdot)$, we will make use of the Fourier coefficients $\{R(k)\}_{k\in\mathbb{Z}}$ of $r$, which are unknown because $\theta$ is not specified, and the Fourier coefficients $\{\Psi(k)\}_{k\in\mathbb{Z}}$ of $\psi$, which are known because $\psi$ is specified. Throughout this article, we will assume $\{\Psi(k)\}_{k\in\mathbb{Z}}$ is bounded away from zero in absolute value on any bounded region $\mathcal{Z} \subset \mathbb{Z}$. This implies the Fourier inversion operator involving $\Psi^{-1}$ is well–behaved (see, for example, the discussion on preconditioning on page 1425 of Mair and Ruymgaart, 1996). Since the 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 estimate $\hat{R}$ for $R$, we need to control the random fluctuations that occur at high frequency spectra, i.e. large values of $|k|$. Politis and Romano (1999) introduce spectral smoothing to control these fluctuations in higher frequencies. The idea is to utilize lower frequencies, where $\hat{R}$ is well–behaved, and taper down the contributions of higher frequencies, where $\hat{R}$ is not well–behaved, by using a regularizing sequence to control the length of the window of acceptable frequencies based on the amount of data available.

To continue, we will introduce some notation. Let $\{h_n\}_{n\geq 1}$ be a regularizing sequence that satisfies $h_n \to 0$, as $n \to \infty$, and $M > 0$ is a constant chosen to control the amount of high–frequency smoothing applied. Now we consider smoothing kernel functions similar to those used in typical nonparametric function estimators by restricting our choice of smoothing kernel based on its Fourier transform. We will require our smoothing kernel to have a Fourier transform $\Lambda$, which itself does not depend on $h_n$, that satisfies the representation $\Lambda(h_n k) = \lambda(k)$, $k \in \mathbb{Z}$, for some function $\lambda$, which in general does depend on $h_n$. This representation means we only need to consider the shrinkage $h_n k$ of $k$. As a consequence, we will also consider the shrinkage of the Fourier frequency domain $\mathbb{Z}$ into $h_n\mathbb{Z}$, where $h_n\mathbb{Z}$ denotes the shrinkage of the integers $\mathbb{Z}$ by $h_n$. We will require that $\Lambda$ additionally satisfies the following general assumption:

ASSUMPTION 1. *There is an integer $M > 0$ specifying the region $I = \{z \in \mathbb{Z} : |z| \leq M\}$ such that $\Lambda(k) = 1$, for $k \in I$, $|\Lambda(k)| \leq 1$, for $k \notin I$, and $\Lambda$ satisfies $\sum_{k=-\infty}^{\infty} |k|^b |\Lambda(k)| < \infty$, where $b > 0$ is a constant.*

Note, we will require additional assumptions on $\Lambda$ to obtain specific rates of convergence. However, our assumptions are less restrictive than those of Politis and Romano (1999). Since our rates of convergence are impacted by the ill–posedness of the inverse problem, we only require $\Lambda$ to be equal to one in a neighborhood containing the zeroth Fourier frequency. The additional summability requirements are used to obtain our explicit rates of convergence. For example, we cannot achieve a bias of the order $h_n^s$, when $\theta \in \mathscr{R}_s$, which is achievable in the direct estimation setting. Instead, we can only obtain a bias of order $h_n^{s-b}$, where $b$ is the degree of ill–posedness. However, our formulation has the advantage that it is still comparable to the so–called "superkernels" that give the order $h_n^s$ (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).

We can then estimate $\theta$ using a kernel smoother:

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

where the weights $W_{j,h_n}$ are defined by

$$W_{j,h_n}(x) = \sum_{k=-\infty}^{\infty} \frac{\lambda(k)}{\Psi(k)} e^{i2\pi k(x-x_j)} = \sum_{\omega\in h_n\mathbb{Z}} \frac{\Lambda(\omega)}{\Psi(\omega/h_n)} \exp\left(i2\pi\omega \frac{x-x_j}{h_n}\right).$$

The smoothing kernel $W_{j,h_n}$ is sometimes called a deconvolution kernel (see, for example, Birke, Bissantz and Holzmann, 2010). In the following section we will investigate the asymptotic properties of the empirical process from the residuals $\hat{\varepsilon}_j = Y_j - K\hat{\theta}(x_j)$ obtained from the estimator $\hat{\theta}$.

## 3. Main results

Our first result specifies the asymptotic order of the bias of $\hat{\theta}$ and is proved in Section 5.

LEMMA 1. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$. Assume that $0 < b < s$ and $C_\psi > 0$ are constants such that $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$, and Assumption 1 is satisfied for this $b$. Then, for any regularizing sequence $\{h_n\}_{n\geq 1}$ satisfying $h_n \to 0$ and $nh_n^b \to \infty$, as $n \to \infty$, we have*

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

Next we consider the consistency of $\hat{\theta}$. The asymptotic order of the bias of $\hat{\theta}$ is impacted by the degree of ill-posedness of the inverse problem, and we will see this detrimental effect in the asymptotic order of consistency as well. In the following result we give the asymptotic order of the strong uniform consistency of $\hat{\theta}$, which is also proved in Section 5.

LEMMA 2. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$, and suppose there exists $0 < b < s$ such that $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$, with $C_\Psi > 0$ a constant. Let Assumption 1 hold for this $b$, with $\Lambda$ additionally satisfying $\sum_{k=-\infty}^{\infty} |k|^{b+1}|\Lambda(k)| < \infty$. Assume the random variables $Y_{-n}, \dots, Y_n$ have a finite absolute moment of order $\kappa > 2$. In addition, let the regularizing sequence $\{h_n\}_{n\geq 1}$ satisfy $h_n \to 0$ such that $(nh_n^{2b})^{-1/2}\log^{1/2}(n) \to 0$, as $n \to \infty$. Then*

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

Using the results of Lemma 1 and Lemma 2, we can obtain a uniform rate of convergence of the estimator $\hat{\theta}$ for $\theta$ by choosing a regularizing sequence $\{h_n\}_{n\geq 1}$ that balances the asymptotic orders of both the bias and consistency: $\{h_n\}_{n\geq 1}$ is chosen to satisfy $h_n^{s-b} = O((nh_n^{2b})^{-1/2}\log^{1/2}(n))$. This implies choosing

$$h_n = O(n^{-1/(2s)}\log^{1/(2s)}(n)), \tag{3.1}$$

and we have both $nh_n^{2b} \to \infty$ and $(nh_n^{2b})^{-1/2}\log^{1/2}(n) \to 0$, as $n \to \infty$. In addition, we have $(nh_n^b)^{-1} = o(h_n^{s-b})$ so the bias has order $O(h_n^{s-b})$. We can also see that

$$(n^{-(s-b)/(2s)}\log^{(s-b)/(2s)}(n))^{1+\gamma} = o(n^{-1/2})$$

whenever $\gamma > b/(s-b)$, and we can restrict $0 < \gamma \leq 1$ by assuming that $s > 2b$. In the following result we give the uniform rate of convergence of $\hat{\theta}$ for $\theta$. The proof is complicated and can be found in Section 5.

THEOREM 1. *Let $\theta \in \mathscr{R}_s$, with $s \geq 1$, and suppose there exists $0 < b < s$ such that $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$, with $C_\Psi > 0$ a constant. Let Assumption 1 hold for this $b$, with $\Lambda$ additionally satisfying $\sum_{k=-\infty}^{\infty} |k|^{b+1}|\Lambda(k)| < \infty$. Assume the random variables $Y_{-n}, \dots, Y_n$ have a finite absolute moment of order $\kappa > 2$. Finally, let the regularizing sequence $\{h_n\}_{n\geq 1}$ satisfy (3.1). Then*

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

*If, additionally, $s > 2b$, then we have, for every $b/(s-b) < \gamma \leq 1$,*

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

*Let $\tau = \max\{1, b\}$. If $\Lambda$ satisfies $\sum_{k=-\infty}^{\infty} |k|^{s+\tau}|\Lambda(k)| < \infty$, then, for large enough $n$,*

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

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

The results on $\hat{\theta}$ above guarantee our model residuals are well–behaved so that we can study the limiting behavior of the empirical distribution function $\hat{\mathbb{F}}$. We arrive at our main result: the uniform expansion of the residual–based empirical distribution function. The proof of this result requires further technical arguments. Therefore, we have placed this proof and its supporting results in Section 5. Also note, the uniform expansion of $\hat{\mathbb{F}}$ implies that $\hat{\mathbb{F}}$ satisfies a functional central limit theorem.

Theorem 2. *Suppose there are constants $b > 0$ and $C_\Psi > 0$ such that $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$, and Assumption 1 is satisfied for this $b$. Let $\theta \in \mathscr{R}_s$, with $s > \max\{2b, 1\}$. In addition, let $\tau = \max\{1, b\}$ and suppose $\Lambda$ satisfies $\sum_{k=-\infty}^{\infty} |k|^{s+\tau}|\Lambda(k)| < \infty$. Assume the distribution function $F$ admits a bounded Lebesgue density function $f$ that is Hölder continuous with exponent $b/(s-b) < \gamma \leq 1$, and let $\varepsilon_{-n}, \ldots, \varepsilon_n$ have a finite absolute moment of order $\kappa > 2$. Finally, let the regularizing sequence $\{h_n\}_{n\geq 1}$ satisfy (3.1). Then*

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

Remark 1. In light of the fact that $r(\cdot) = [K\theta](\cdot)$ is nonparametric because $\theta$ is nonparametric, we can see that model (1.1) is a type of nonparametric regression. The estimator $\hat{\mathbb{F}}$ has influence function $\mathbf{1}[\varepsilon \leq t] - F(t) - \varepsilon f(t)$. Hence, 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).

Remark 2. The set $\{\exp(i2\pi kx) : x \in [-1/2, 1/2]\}_{k\in\mathbb{Z}}$ is an orthonormal basis for the class of functions $\mathcal{L}_2([-1/2, 1/2])$ and the design space has unit volume. If either the design space is a compact set of volume $d_1 > 0$ or the corresponding basis vectors have squared length, in the corresponding $\mathcal{L}_2$–norm, $d_2 > 0$, then the estimator $\hat{\mathbb{F}}$ has influence function equal to $\mathbf{1}[\varepsilon \leq t] - d_1F(t) - d_1d_2\varepsilon f(t)$, which may no longer be efficient for estimating $F$. To avoid the possible inefficiency of this approach, it is recommended to use an affine transformation mapping the design space into the interval $[-1/2, 1/2]$, where the orthonormal basis $\{\exp(i2\pi kx) : x \in [-1/2, 1/2]\}_{k\in\mathbb{Z}}$ can be used in the indirect regression estimator $\hat{\theta}$.

**3.1. Asymptotically optimal regularization parameter selection.** We now consider the problem of choosing an appropriate sequence of regularization parameters $\{h_n\}_{n\geq 1}$ that is used in the estimator $\hat{\theta}$. Theorem 1 suggests a practical choice of regularization would be a scheme that minimizes the integrated mean squared error (IMSE) of $\hat{\theta}$. Our formulation of the deconvolution kernel causes the regularizing sequence to interact only with the Fourier frequencies considered in the estimator $\hat{\theta}$, which follows from our representation $\Lambda(h_nk) = \lambda(k)$. Hence, we will not be able to determine an exact sequence that minimizes the IMSE of $\hat{\theta}$. In the following result, we give asymptotic bounds on the integrated variance and the integrated squared bias of the estimator $\hat{\theta}$ that lead to an appropriate choice of regularization that approximately minimizes the IMSE of $\hat{\theta}$.

PROPOSITION 1. *Let Assumption 1 hold. Assume that $\theta \in \mathscr{R}_s$, with $s > 2b$, and $E[\varepsilon_{-n}^2] = \ldots = E[\varepsilon_n^2] = \sigma^2$. Then, for any regularizing sequence $\{h_n\}_{n\geq 1}$ satisfying $h_n \to 0$ such that $nh_n^s \to \infty$, as $n \to \infty$, there are constants $C_\Lambda > 0$ and $C_R > 0$ such that*

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

*and*

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

PROOF. Beginning with the first assertion, we can write $\{\hat{\theta}(x) - E[\hat{\theta}(x)]\}^2$ as

$$\begin{aligned}
&\sum_{k=-\infty}^{\infty} \frac{\lambda^2(k)}{\Psi^2(k)} \bigg\{\frac{1}{(2n+1)^2} \sum_{j=-n}^{n} \varepsilon_j^2\bigg\} + \sum_{k=-\infty}^{\infty} \frac{\lambda^2(k)}{\Psi^2(k)} \bigg\{\frac{1}{(2n+1)^2} \sum_{j\neq l} \varepsilon_j \varepsilon_l e^{i2\pi k(x_l - x_j)}\bigg\} \\
&+ \sum_{k\neq\xi} \frac{\lambda(k)}{\Psi(k)} \frac{\lambda(\xi)}{\Psi(\xi)} \bigg\{\frac{1}{(2n+1)^2} \sum_{j=-n}^{n} \varepsilon_j^2\bigg\} e^{i2\pi(k-\xi)x} \\
&+ \sum_{k\neq\xi} \frac{\lambda(k)}{\Psi(k)} \frac{\lambda(\xi)}{\Psi(\xi)} \bigg\{\frac{1}{(2n+1)^2} \sum_{j\neq l} \varepsilon_j \varepsilon_l e^{i2\pi\{k(x-x_j) - \xi(x-x_l)\}}\bigg\}
\end{aligned}$$

so that

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

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

$$\sum_{k=-\infty}^{\infty} \frac{\lambda^2(k)}{\Psi^2(k)} \leq O(h_n^{-2b}),$$

and, therefore, we can specify $C_\Lambda > 0$ for the first assertion to hold.

Turning our attention to the second assertion, let $I^c(h_n) = \{z \in \mathbb{Z} : z > Mh_n^{-1}\}$. We can write $\{E[\hat{\theta}(x)] - \theta(x)\}^2$ as

$$\begin{aligned}
&\sum_{k\in I^c(h_n)} \big\{\lambda(k) - 1\big\}^2 \frac{R^2(k)}{\Psi^2(k)} + 2\sum_{k\in I^c(h_n)} \frac{\lambda(k)\{\lambda(k)-1\}}{\Psi(k)} \frac{R(k)}{\Psi(k)} \Big\{E\big[\hat{R}(-k)\big] - R(-k)\Big\} \\
&+ \sum_{k=-\infty}^{\infty} \frac{\lambda^2(k)}{\Psi^2(k)} \Big\{E\big[\hat{R}(k)\big] - R(k)\Big\}\Big\{E\big[\hat{R}(-k)\big] - R(-k)\Big\} \\
&+ \sum_{\{k,\xi\in I^c(h_n)\,:\,k\neq\xi\}} \frac{\lambda(k)-1}{\Psi(k)} \frac{\lambda(\xi)-1}{\Psi(\xi)} R(k)R(\xi) e^{i2\pi(k-\xi)x}
\end{aligned}$$

$$+ 2\sum_{k\neq\xi} \frac{\lambda(k)-1}{\Psi(k)}\frac{\lambda(\xi)}{\Psi(\xi)} R(k)\Big\{E\big[\hat R(-\xi)\big] - R(-\xi)\Big\} e^{i2\pi(k-\xi)x}$$
$$+ \sum_{k\neq\xi} \frac{\lambda(k)}{\Psi(k)}\frac{\lambda(\xi)}{\Psi(\xi)} \Big\{E\big[\hat R(k)\big] - R(k)\Big\}\Big\{E\big[\hat R(-\xi)\big] - R(-\xi)\Big\} e^{i2\pi(k-\xi)x}$$

so that

$$\begin{aligned}\int_{-1/2}^{1/2} \Big\{E\big[\hat\theta(x)\big] - \theta(x)\Big\}^2 dx = &\sum_{k\in I^c(h_n)} \big\{\lambda(k)-1\big\}^2 \frac{R^2(k)}{\Psi^2(k)} \\ &+ 2\sum_{k\in I^c(h_n)} \frac{\lambda(k)\{\lambda(k)-1\}}{\Psi(k)}\frac{R(k)}{\Psi(k)}\Big\{E\big[\hat R(-k)\big] - R(-k)\Big\} \\ &+ \sum_{k=-\infty}^{\infty} \frac{\lambda^2(k)}{\Psi^2(k)}\Big\{E\big[\hat R(k)\big] - R(k)\Big\}\Big\{E\big[\hat R(-k)\big] - R(-k)\Big\}.\end{aligned}$$

The assumptions of Lemma 3 in Section 5 are satisfied. An application of this result shows both the second term in the display above is bounded in absolute value by

$$O(n^{-1}h_n^{s-2b}) = o(h_n^{2(s-b)})$$

and the third term in the same display is also bounded in absolute value by

$$O((nh_n^b)^{-2}) = o(h_n^{2(s-b)}).$$

Again, repeating the arguments in the proof of Lemma 1 in Section 5 shows the first term in the display above to be bounded by $O(h_n^{2(s-b)})$, and, therefore, we can specify $C_R > 0$ for the second assertion to hold. □

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

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

Using a suitable estimate $\hat\sigma^2$ for $\sigma^2$ then leads to a rule of thumb in the spirit of Silverman (1986):

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

REMARK 4. Our recommended optimal regularizing sequence depends on the constants $s$, $C_\Lambda$ and $C_R$, which may be unknown. Specifically, the ratio $C_\Lambda/C_R$ can be viewed as a measure of how well the Fourier coefficients of the smoothing kernel function $\{\Lambda(k)\}_{k\in\mathbb{Z}}$ are controlling the expansion of the Fourier inversion operator to the amount of the Fourier expansion of $\theta$ that is ignored by the estimate $\hat\theta$. In applications, a suitable approximation of the ratio $C_\Lambda/C_R$ may be obtainable. When this is not appropriate, a numerical search routine, via bootstrap or cross–validation, is then recommended to find a suitable regularizing sequence.

**3.2. Smooth bootstrap of residuals.** Computational approaches for automated, or 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 of 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 bootstrap estimate of the IMSE of the estimator $\hat{\theta}$, which requires mirroring the restrictions given by Theorem 2 on model (1.1) in the bootstrap scheme. This technique allows some functionals from the original data–generating process to have equivalent representations in the bootstrap process with similar properties, which motivates our use of it to estimate the IMSE of $\hat{\theta}$. Throughout this section, we will describe the stochastic properties of our random quantities using $P^*$–outer measure, which, for a single bootstrap response $Y^*$, reduces to 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 $\varepsilon^*$ 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 $\varepsilon^*$ satisfies having a mean equal to zero, independence, a finite moment of order $\kappa > 2$ and a common distribution function $F_n^*$ that admits a bounded Lebesgue density function $f_n^*$ that is Hölder continuous. The first requirement is satisfied merely by centering our original model residuals:

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

Turning our attention to the next constraint, we can see that conditioning on the original sample $\mathbb{D}$ and selecting from $\tilde{\varepsilon}_{-n}, \ldots, \tilde{\varepsilon}_n$ completely at random and with replacement satisfies independence, in the sense of $P^*$–outer measure. 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 $\kappa > 2$ 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, and later we will impose requirements on $\{c_n\}_{n \geq 1}$ that are appropriate to form a bootstrap version of our indirect regression function estimator as Neumeyer (2009) does with her nonparametric regression function estimator. Consequently,

$\varepsilon_j^*$ has the common distribution function

$$F_n^*(t) = P^*(\varepsilon_j^* \leq t) = \frac{1}{(2n+1)c_n} \sum_{j=-n}^{n} \int_{-\infty}^{t} w\left(\frac{u - \tilde{\varepsilon}_j}{c_n}\right) du, \qquad t \in \mathbb{R}, \tag{3.2}$$

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 estimate 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

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

Following the observations of Neumeyer (2009), we need to choose $\{c_n\}_{n\geq 1}$ such that our bootstrap indirect regression estimator $\hat{\theta}^*$ satisfies similar properties as $\hat{\theta}$ given in Theorem 1, where $\hat{\theta}^*$ by $\hat{\theta}$ is defined in (2.2), where $Y_j^*$ replaces $Y_j$ and a regularizing sequence $\{g_n\}_{n\geq 1}$ replaces the regularizing sequence $\{h_n\}_{n\geq 1}$. When the assumptions of Theorem 1 hold, repeating the arguments in Section 5, using our bootstrap data, shows that we only need to choose $\{c_n\}_{n\geq 1}$ to satisfy $c_n = O(n^{-\alpha})$, for any $0 < \alpha < 1/4 - 1/(2\kappa)$, for the associated results of Theorem 1 to hold for $\hat{\theta}^*$. Consequently, $f_n^*$ is also uniformly consistent for $f$ with our choice of $\{c_n\}_{n\geq 1}$; see Theorem A in Silverman (1978), which permits a wide variety of density functions $w$ to be chosen including the standard normal density. For example, if the contaminants $U_j$ satisfy a finite moment of order larger than 10, then we can simply use $c_n = O(n^{-1/5})$. This implies that we can use normally distributed contaminates $U_j$ and $c_n = O(n^{-1/5})$. We summarize these results in the following corollary. For brevity, we omit its proof because it is proven in exactly the same manner as Theorem 1 and its supporting results (see Section 5).

COROLLARY 1. *Let the assumptions of Theorem 1 be satisfied. Choose the regularizing sequence $\{g_n\}_{n\geq 1}$ to satisfy $g_n = O(n^{-1/(2s)} \log^{1/(2s)}(n))$ and the scaling sequence $\{c_n\}_{n\geq 1}$ to satisfy $c_n = O(n^{-\alpha})$, for any $0 < \alpha < 1/4 - 1/(2\kappa)$. Then, $P^*$–outer almost surely, we have*

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

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

*for every $b/(s-b) < \gamma \leq 1$, and, for large enough $n$,*

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

Remark 5. Following the discussion on pages 207-209 in Neumeyer (2009), validity of the proposed smooth bootstrap of the model residuals is obtained as follows. Define

$$R_n(t) = (2n+1)^{-1/2} \sum_{j=-n}^{n} \Big\{ \mathbf{1}\big[\hat{\varepsilon}_j \le t\big] - F(t) \Big\}$$

and its smooth bootstrap analogue

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

where $\hat{\varepsilon}_j^* = Y_j^* - [K\hat{\theta}^*](x_j)$ is a residual obtained in the smooth bootstrap sample. The analogous results of Theorem 2 for $\hat{\mathbb{F}}^*(t) = (2n+1)^{-1}\sum_{j=1}^{n} \mathbf{1}[\hat{\varepsilon}_j^* \le t]$ can then be obtained using Corollary 1. This result combined with the uniform consistency of $f_n^*$ for $f$ then implies the limiting distributions of $R_n(t)$ and $R_n^*(t)$ are given by the same Gaussian process, which has continuous sample paths.

It then follows for statistics $T_n = \mu(R_n)$ and their smooth bootstrap version $T_n^* = \mu(R_n^*)$ obtained from continuous functionals $\mu$ satisfy the following consistency property. Define $c_{n,\alpha}^*$ by $P^*(T_n^* \le c_{n,\alpha}^*) = \alpha$. Combining the continuity of the functional $\mu$ and the continuous sample paths of the resulting Gaussian process with the continuous mapping theorem we obtain

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

which characterizes the validity of the proposed smooth bootstrap of the model residuals.

Now we turn our attention to choosing the regularizing sequence $\{g_{n,opt}\}_{n\ge1}$ that minimizes the IMSE between $\hat{\theta}^*$ and $\hat{\theta}$, conditionally on the observed data $\mathbb{D}$. The IMSE of $\hat{\theta}$, which we want to minimize with respect to the regularizing sequence, is given by

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

Following Cao (1993), we will arbitrarily choose the original regularizing sequence $\{h_n\}_{n\ge1}$ according to Theorem 1 as a pilot sequence to form an initial and consistent estimate $\hat{\theta}_{h_n}$, which we can plug–in for the unknown function $\theta$ in (3.3) (also a reasonable approximation to the rule-of-thumb in Remark 3 can be used). This leads to an analogous form of (3.3) in $P^*$–outer measure, which is given by

$$IMSE^*\big(\hat{\theta}^*\big) = \int_{-1/2}^{1/2} E^*\Big[\big\{\hat{\theta}^*(x) - \hat{\theta}_{h_n}(x)\big\}^2\Big]\,dx = E^*\bigg[\int_{-1/2}^{1/2} \big\{\hat{\theta}^*(x) - \hat{\theta}_{h_n}(x)\big\}^2\,dx\bigg]. \tag{3.4}$$

Since $\hat{\theta}_{h_n}$ satisfies (2.2) and $\hat{\theta}^*$ also satisfies (2.2), with $Y_j^*$ in place of $Y_j$, the expected values on the far right–hand sides of (3.3) and (3.4) are averages taken with respect to the distribution functions $F$ and $F_n^*$ from (3.2), respectively. We can then use standard arguments to show

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

Hence, we obtain $IMSE^*(\hat{\theta}^*) = IMSE(\hat{\theta}) + o_p(1)$. This implies our bootstrap analogue of IMSE is an effective predictor of the true IMSE.

It follows that we can choose $\{g_{n,opt}\}_{n\geq 1}$ such that

$$g_{n,opt} = \underset{g_n \in (0,\,\hbar]}{\arg\min}\, E^*\left[\int_{-1/2}^{1/2} \left\{\hat{\theta}^*(x) - \hat{\theta}_{h_n}(x)\right\}^2 dx\right], \tag{3.5}$$

where $\hbar > 0$ is an appropriate constant such that $h_{n,opt} \in (0,\,\hbar)$. The outer expectation $E^*$ can be approximated using the usual Monte Carlo approach and we can minimize this criterion using a standard grid search.

Consider the Fourier coefficients $\{\Lambda(k)\}_{k\in\mathbb{Z}}$ used in the estimators $\hat{\theta}$ and $\hat{\theta}^*$. Working only with the Fourier coefficients $\{\Lambda(k)\}_{k\in\mathbb{Z}}$ means viewing $\Lambda$ as a mapping from $\mathbb{Z}$ to $[-1,\,1]$. However, we can also view $\Lambda$ as a mapping from $\mathbb{R}$ to $[-1,\,1]$ because we plug–in $h_n k$ for $k$ to form the estimator $\hat{\theta}$ (also we plug–in $g_n k$ for $k$ to form the estimator $\hat{\theta}^*$). It is then easy to see that imposing standard smoothness assumptions on $\Lambda$, viewed as a mapping from $\mathbb{R}$ to $[-1,\,1]$, leads to the desired consistency property between the smooth bootstrap selected optimal regularizing sequence $\{g_{n,opt}\}_{n\geq 1}$ defined by (3.5), which minimizes (3.4), and the desired optimal regularizing sequence $\{h_{n,opt}\}_{n\geq 1}$, which minimizes (3.3). We summarize these observations in the following remark.

Remark 6. From the discussion above, we expect $h_{n,opt} = C_{opt} n^{-1/(2s)}$, where $C_{opt} > 0$ is an appropriate constant. We can restrict our choice of smoothing kernel such that its Fourier transform $\Lambda$ allows for $\{h_{n,opt}\}_{n\geq 1}$ to be the unique minimizer of (3.3). Let $\{g_{n,opt}\}_{n\geq 1}$ satisfy (3.5). Since $IMSE^*(\hat{\theta}^*)$ is consistent for $IMSE(\hat{\theta})$, we have the desired $g_{n,opt} = h_{n,opt} + o_p(1)$.

## 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 our simulations, we chose the regression function

$$\theta(x) = 3e^{-20x^2}, \qquad x \in [-1/2,\,1/2],$$

and the point spread function $\psi$ is taken as the Laplace density restricted to the interval $[-1/2,\,1/2]$ with a mean of zero and a scale of $1/10$, which satisfies the ordinary smoothness assumption with $b = 2$. 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 in estimation of the function $\theta$. Finally, the model errors are randomly generated from a normal distribution with mean zero and scale $2/3$. Our simulations consider samples of sizes 51, 101, 201 and 501, i.e. $n$ is taken as 25, 50, 100 and 250.

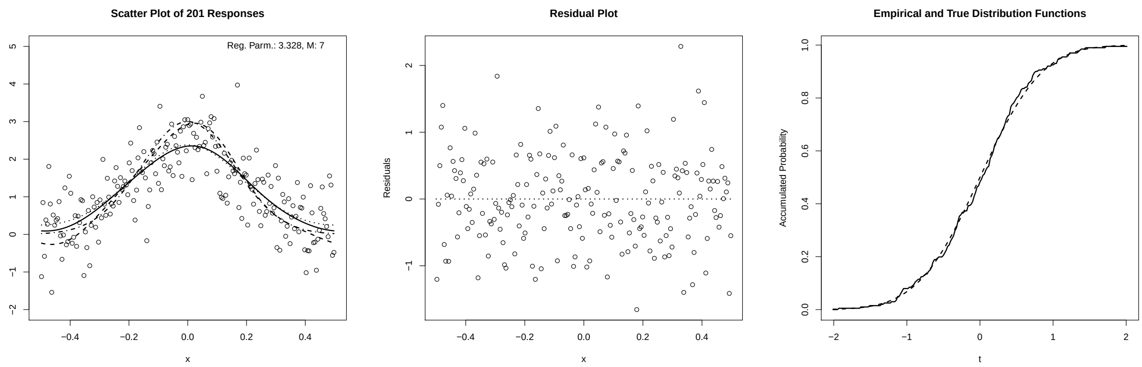


FIGURE 1. *From left to right: A scatter plot of the data overlaid with the fitted blurred regression (solid), estimated regression (dashed), true blurred regression function (dotted) and the true regression function (dot-dashed); A scatter plot of the model residuals overlaid with a line at zero; A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

We work with the smoothing kernel that has Fourier coefficients satisfying

$$\Lambda(k) = \begin{cases} 1, & \text{if } |k| \leq 7 \\ |k|^{-8}, & \text{if } 7 < |k| \leq n \\ 0, & \text{otherwise.} \end{cases}$$

In order to select an appropriate regularization parameter for the function estimator $\hat{\theta}$, we work with the pilot sequence $h_n = 5n^{-1/9} \log^{1/9}(n)$. We have used standard normally distributed contaminates $U_j$ and, following Silverman's rule for selecting a bandwidth in kernel density estimation, we work with 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 by using the pilot sequence $h_n$ to estimate $\theta$. Using 200 smooth bootstrap replications to construct a suitable approximation of the IMSE of $\hat{\theta}$ for each of 500 equally spaced candidate regularization parameters in (0, 8], we then choose the optimal regularization parameter $g_{n,opt}$ as the grid point that minimizes this approximate IMSE, which we then use to construct the function estimate $\hat{\theta}$ (see our discussion on the proposed smooth bootstrap approximation of $IMSE(\hat{\theta})$ in subsection 3.2).

The assumptions of Theorem 2, and Corollary 1, are satisfied for the choices made above. Figure 1 displays the results of our indirect regression estimator for a typical data set 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 $\theta$ and $r$. Turning our attention to the scatter plot of the residuals, we can see the indirect regression estimator $\hat{\theta}$ 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 residuals. Finally, the remaining plot of the

| $n$ \ $t$ | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| 51 | 0.0152 (0.0034) | 0.1723 (0.0502) | 0.0129 (0.0842) | -0.1693 (0.0434) | -0.0159 (0.0036) |
| 101 | 0.0154 (0.0028) | 0.2078 (0.0704) | 0.0220 (0.0826) | -0.2312 (0.0749) | -0.0195 (0.0032) |
| 201 | -0.0019 (0.0013) | -0.0400 (0.0418) | -0.0001 (0.0886) | 0.0365 (0.0427) | 0.0027 (0.0012) |
| 501 | -0.0026 (0.0012) | -0.0149 (0.0461) | 0.0000 (0.0957) | 0.0244 (0.0477) | 0.0019 (0.0014) |

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* $2$.

| $t$ \ $n$ | 51 | 101 | 201 | 501 | $\infty$ |
|---|---|---|---|---|---|
| -2 | 0.0036 | 0.0030 | 0.0013 | 0.0013 | 0.0014 |
| -1 | 0.0799 | 0.1136 | 0.0434 | 0.0463 | 0.0460 |
| 0 | 0.0844 | 0.0831 | 0.0886 | 0.0957 | 0.0908 |
| 1 | 0.0721 | 0.1284 | 0.0441 | 0.0483 | 0.0460 |
| 2 | 0.0039 | 0.0036 | 0.0012 | 0.0014 | 0.0014 |

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* $2$.

| 51 | 101 | 201 | 501 | $\infty$ |
|---|---|---|---|---|
| 0.2614 | 0.3595 | 0.1926 | 0.1858 | 0.1889 |

TABLE 3. *Asymptotic integrated mean squared error of* $(2n+1)^{1/2}\{\hat{\mathbb{F}} - F\}$ *by sample size.*

distribution functions shows the empirical distribution function of the residuals $\hat{\mathbb{F}}$ matches very closely to the true error distribution function $F$.

Turning our attention to the numerical summaries of the estimator $\hat{\mathbb{F}}$, we can plainly see this estimator is performing well. In Table 1, we have calculated the simulated asymptotic biases and variances of $\hat{\mathbb{F}}$ at the points $-2$, $-1$, 0, 1 and 2. The simulated asymptotic biases are calculated by computing the simulated biases of $\hat{\mathbb{F}}$ and multiplying these by the square–root of the corresponding sample size, 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 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 $\infty$ 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.

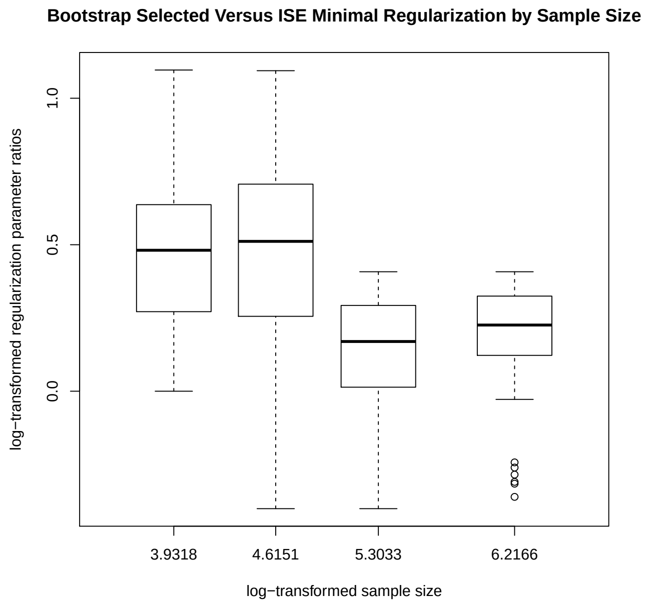


Figure 2. *Boxplots of log-transformed ratios of regularization parameters by log-transformed sample size.*

| Regularization | 51 | 101 | 201 | 501 |
|---|---|---|---|---|
| Bootstrap | 0.3208 | 0.2812 | 0.0540 | 0.0295 |
| Best | 0.1593 | 0.0933 | 0.0536 | 0.0276 |

Table 4. *Integrated mean squared error of $\hat{\theta}$ by sample size. 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.*

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 in estimating $F$ 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 $\hat{\theta}$.

The results concerning our indirect regression estimator are interesting. From Remark 6, we can see that our optimal regularization parameter depends on both the sample size and the smoothness index $s$ of the function class used to approximate $\theta$. In addition to finding an 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 estimate $\hat{\theta}$ and the function $\theta$. In general, this methodology is not available in applications, but we expect it to produce the best resulting estimator of $\theta$ with respect to the IMSE.

In Figure 2 we give boxplots of the log-transformed ratios of the 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 0, which we expect to continue as the sample size increases. This confirms the consistency between the two regularization techniques mentioned in Remark 6. It appears that with increasing sample size both the bootstrap selection methodology and the ISE selection methodology choose regularizations that result in maintaining the smaller Fourier frequencies used in the estimator $\hat{\theta}$ until enough data is available to incorporate larger frequencies, i.e. the smoothness index $s$ appears to be automatically selected. This is particularly convenient because the smoothness index $s$ is in general unknown and very important for building an optimal indirect regression function estimator.

We have also numerically measured the performance of the estimator $\hat{\theta}$ by simulating the IMSE using both regularization techniques. The results are given in Table 4. We can see the estimator $\hat{\theta}$ using each regularization parameter has IMSE decaying to zero as the sample size increases, and both IMSE values appear to be very close at the larger sample sizes 201 and 501, which also confirms the conjecture of consistency between the two regularization techniques given in Remark 6. In summary, we find the residual–based empirical distribution function is performing well in estimating the distribution function of the errors, and the proposed smooth bootstrap methodology for selecting the regularization parameter used in the indirect regression estimate provides a useful and convenient tool for precise indirect regression function estimation.

**Acknowledgements** 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).

## 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 $r \in \mathscr{R}_s$, with $s \geq 1$. Then*

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

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 $s = 1$. Without any loss of generality, we can assume that $|r(0)| < \infty$. We can write

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

$$= \frac{1}{2n+1} \sum_{j=1}^{n} r(x_j) e^{-i2\pi k x_j} + \frac{r(0)}{2n+1} + \frac{1}{2n+1} \sum_{j=-n}^{-1} r(x_j) e^{-i2\pi k x_j}.$$

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

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

so that $R(k)$ is equal to

(5.2)

$$\sum_{j=1}^{n} \int_{(2j-1)/(4n+2)}^{(2j+1)/(4n+2)} r(x) e^{-i2\pi k x}\, dx + \int_{-1/(4n+2)}^{1/(4n+2)} r(x) e^{-i2\pi k x}\, dx + \sum_{j=-n}^{-1} \int_{(2j-1)/(4n+2)}^{(2j+1)/(4n+2)} r(x) e^{-i2\pi k x}\, dx$$

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

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

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

Since $x_j = j/(2n)$, we have $v/(2n+1) + j/(2n+1) = x_j + (v - x_j)/(2n+1)$. Using the far right–hand sides of (5.1) and (5.2), we find 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

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

$R_2(k)$ is equal to

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

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

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

and $R_5(k)$ is equal to

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

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

Beginning with $R_1(k)$, it follows from $r \in \mathscr{R}_1$ that we can find an appropriate constant $C > 0$ such that

$$\int_{-1/2}^{1/2} \left| r(x_j) - r\left( x_j + \frac{v - x_j}{2n+1} \right) \right| dv \le C n^{-1} \left\{ \int_{-1/2}^{1/2} (v - x_j)^2 \, dv \right\}^{1/2}.$$

Therefore, we can bound $R_1(k)$ by

$$C n^{-2} \sum_{j=1}^{n} \left\{ \int_{-1/2}^{1/2} (v - x_j)^2 \, dv \right\}^{1/2},$$

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

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

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

$$\frac{1}{2n+1}\sum_{j=1}^{n}\bigg|\int_{-1/2}^{1/2}\bigg\{r(x_j)\exp\big(-i2\pi k x_j\big) - r\bigg(x_j+\frac{v-x_j}{2n+1}\bigg)\exp\bigg(-i2\pi k\bigg(x_j+\frac{v-x_j}{2n+1}\bigg)\bigg)\bigg\}\,dv\bigg|. \tag{5.3}$$

We can use the Fourier inversion formula to write

$$\begin{aligned} r(x_j)\exp\big(-i2\pi k x_j\big) - r\bigg(x_j+\frac{v-x_j}{2n+1}\bigg)\exp\bigg(-i2\pi k\bigg(x_j+\frac{v-x_j}{2n+1}\bigg)\bigg) \\ = \sum_{\xi=-\infty}^{\infty} R(\xi)\bigg\{\exp\big(i2\pi(\xi-k)x_j\big) - \exp\bigg(i2\pi(\xi-k)\bigg(x_j+\frac{v-x_j}{2n+1}\bigg)\bigg)\bigg\}, \end{aligned} \tag{5.4}$$

and we can choose $w_j(v) \in (\min\{x_j, x_j+(v-x_j)/(2n+1)\}, \max\{x_j, x_j+(v-x_j)/(2n+1)\})$ for the right–hand side of (5.4) to be equal to

$$\frac{i2\pi(v-x_j)}{2n+1}\sum_{|\xi-k|>0} R(\xi)(\xi-k)\exp\big(i2\pi(\xi-k)w_j(v)\big).$$

Since $r \in \mathscr{R}_1$, 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 (5.3) to be further bounded by

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

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 arguments similar to that for showing $\max_{k\in\mathbb{Z}} R_1(k) = O(n^{-1})$ above, 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 3:

Proof of Lemma 1. We begin with the decomposition

$$\begin{aligned} E\big[\hat\theta(x)\big] &= E\bigg[\frac{1}{2n+1}\sum_{j=-n}^{n} Y_j W_{j,h_n}(x)\bigg] \\ &= \frac{1}{2n+1}\sum_{j=-n}^{n} r(x_j)\bigg\{\sum_{k=-\infty}^{\infty}\frac{\lambda(k)}{\Psi(k)}\exp\big(i2\pi k(x-x_j)\big)\bigg\} \\ &= \sum_{k=-\infty}^{\infty}\frac{\lambda(k)}{\Psi(k)}R(k)\exp(i2\pi kx) + \sum_{k=-\infty}^{\infty}\frac{\lambda(k)}{\Psi(k)}\Big\{E\big[\hat R(k)\big] - R(k)\Big\}\exp(i2\pi kx) \end{aligned}$$

so that

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

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

$$\sum_{k=-\infty}^{\infty} \frac{|\lambda(k)-1|}{|\Psi(k)|}|R(k)| + \max_{k\in\mathbb{Z}}\Big|E\big[\hat{R}(k)\big] - R(k)\Big| \sum_{k=-\infty}^{\infty} \frac{|\lambda(k)|}{|\Psi(k)|}. \tag{5.5}$$

Following the representation $\Lambda(h_n k) = \lambda(k)$, we can partition $\mathbb{Z}$ into $I(h_n) \cup I^c(h_n)$, where $I(h_n) = \{z \in \mathbb{Z} : h_n|z| \le M\} = \{z \in \mathbb{Z} : |z| \le Mh_n^{-1}\}$. For every $k \in I^c(h_n)$, we have $|\Lambda(h_n k)| \le 1$, which implies $|\Lambda(h_n k) - 1| \le 2$, and the first term in the right–hand side of (5.5) is bounded by

$$2\sum_{k\in I^c(h_n)} \frac{|R(k)|}{|\Psi(k)|} \le \frac{2h_n^{s-b}}{M^{s-b}} \sum_{k=-\infty}^{\infty} \frac{|k|^s|R(k)|}{|k|^b|\Psi(k)|}. \tag{5.6}$$

To continue, let $\epsilon > 0$ be arbitrary. Since we have $|k|^b|\Psi(k)| \to C_\Psi$, as $|k| \to \infty$, it follows that we can find a constant $\Gamma > 0$ such that $|k|^b|\Psi(k)| > C_\Psi/2$, for every $|k| > \Gamma$. Hence, the fraction in the series in (5.6) is bounded by

$$\begin{cases} |k|^{s-b}|R(k)|\big/\big[\min_{k\in\{z\in\mathbb{Z}\,:\,|z|\le\Gamma\}}|\Psi(k)|\big], & \text{if } |k| \le \Gamma, \\ 2|k|^s|R(k)|/C_\Psi, & \text{if } |k| > \Gamma. \end{cases}$$

Therefore, for any $\epsilon > 0$, the series in (5.6) is bounded by

$$\Big[\min_{k\in\{z\in\mathbb{Z}\,:\,|z|\le\Gamma\}} |\Psi(k)|\Big]^{-1} \sum_{k=-\infty}^{\infty} |k|^{s-b}|R(k)| + \frac{2}{C_\Psi}\sum_{k=-\infty}^{\infty} |k|^s|R(k)|,$$

which is finite. This implies the first term in (5.5) is of order $O(h_n^{s-b})$, uniformly in $x \in [-1/2, 1/2]$.

We now turn to the second term in (5.5). It follows along the same lines as the arguments in the previous paragraph for the series in this term to be bounded by

$$\Big[\min_{k\in\{z\in\mathbb{Z}\,:\,|z|\le\Gamma\}} |\Psi(k)|\Big]^{-1} \sum_{\omega\in h_n\mathbb{Z}} |\Lambda(\omega)| + \frac{2h_n^{-b}}{C_\Psi}\sum_{\omega\in h_n\mathbb{Z}} |\omega|^b|\Lambda(\omega)|,$$

where $\Gamma$ is given above. The factor $h_n^{-b}$ appears in the bound above because we have used the representation $\Lambda(h_n k) = \lambda(k)$, which leads to shrinking $|k|$ by $h_n$. 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 (5.5) is of order $O((nh_n^b)^{-1})$, uniformly in $x \in [-1/2, 1/2]$. Combining the results above, we have that (5.5) is of order $O(h_n^{s-b} + (nh_n^b)^{-1})$, uniformly in $x \in [-1/2, 1/2]$, and the assertion of Lemma 1 follows. □

Proof of Lemma 2. Without loss of generality we can assume that $n \geq 3$. 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 $\{\eta_n\}_{n\geq 3}$ satisfy $\eta_n = O((nh_n^{2b})^{-1/2}\log^{1/2}(n))$ and the truncation sequence $\{t_n\}_{n\geq 3}$ 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

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

Beginning with $D_1(x)$, we can write this term as

$$\frac{1}{2n+1}\sum_{j=-n}^{n} Y_j \mathbf{1}\big[|Y_j| > K_j t_n\big] W_{j,h_n}(x)$$

so that $\sup_{x\in[-1/2,\,1/2]} |D_1(x)|$ is bounded by

$$\max_{j\in\{-n,\ldots,n\}} \sup_{x\in[-1/2,\,1/2]} |W_{j,h_n}(x)| \frac{1}{2n+1}\sum_{j=-n}^{n} |Y_j| \mathbf{1}\big[|Y_j| > K_j t_n\big]. \tag{5.7}$$

We have that $\max_{j\in\{-n,\ldots,n\}} \sup_{x\in[-1/2,\,1/2]} |W_{j,h_n}(x)|$ is bounded by $\sum_{k=-\infty}^{\infty}\{|\lambda(k)|/|\Psi(k)|\}$, and in the proof of Lemma 1 we have already shown this series is of order $O(h_n^{-b})$. Turning our attention to the indicator function in (5.7), we can use Markov's inequality to obtain $P(|Y_j| > K_j t_n) \leq t_n^{-\kappa}$. Since $\delta > 0$, we have

$$\sum_{n=3}^{\infty} t_n^{-\kappa} = \sum_{n=3}^{\infty} \frac{1}{n\log(n)(\log\log(n))^{1+\delta}} < \infty.$$

It then follows by the Borel-Cantelli lemma for the event $\{|Y_j| \leq K_j t_n\}$ to occur infinitely often. Since $\{t_n\}_{n\geq 3}$ is increasing, we have, for large enough $n$, $|Y_j| \leq K_j t_n$, almost surely. Finally, since $h_n^{-b} < \infty$, for all $n$ finite, we can conclude that (5.7) is equal to zero, for large enough $n$, almost surely. It then follows for $\sup_{x\in[-1/2,\,1/2]} |D_1(x)| = o(\eta_n)$, almost surely.

We now turn our attention to $D_2(x)$. We have already shown that $\sup_{x\in[-1/2,\,1/2]} |W_{j,h_n}(x)| = O(h_n^{-b})$, and 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

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

Since $\kappa > 1$, writing $M_K = \max_{j=-n,\ldots,n} K_j$, we have

$$\max_{j=-n,\ldots,n} E[|Y_j|\mathbf{1}[|Y_j| > K_j t_n]] = \max_{j=-n,\ldots,n} \int_{K_j t_n}^{\infty} P\big(|Y_j| > s\big)\, ds \leq \frac{M_K}{\kappa - 1} t_n^{1-\kappa}.$$

This implies that we can enlarge $C$ such that $\sup_{x\in[-1/2,\,1/2]} |D_2(x)| \leq Ch_n^{-b} t_n^{1-\kappa} = o(\eta_n)$.

To continue, we will require an additional result. For any $u, v \in [-1/2,\, 1/2]$, we can find $w_j \in (\min\{u - x_j, v - x_j\}, \max\{u - x_j, v - x_j\})$ such that

$$W_{j,h_n}(u) - W_{j,h_n}(v) = i\pi(u - v) \sum_{k=-\infty}^{\infty} \frac{k\lambda(k)}{\Psi(k)} \exp\big(i\pi k w_j\big).$$

Following the arguments in the proof of Lemma 1, we can bound $|W_{j,h_n}(u) - W_{j,h_n}(v)|$ by the product of $|u - v|$ and

$$\pi\bigg[\min_{k\in\{z\in\mathbb{Z}:\,|z|\le\Gamma\}} |\Psi(k)|\bigg]^{-1} h_n^{-1} \sum_{\omega\in h_n\mathbb{Z}} |\omega||\Lambda(\omega)| + \frac{2\pi}{C_\Psi} h_n^{-b-1} \sum_{\omega\in h_n\mathbb{Z}} |\omega|^{b+1}|\Lambda(\omega)|,$$

where $\epsilon > 0$ is arbitrarily chosen and $\Gamma > 0$ is a constant that satisfies the condition that, for all $|k| > \Gamma$, $|k|^b|\Psi(k)| > (C_\Psi/2)$. This shows that we can find an appropriate constant $C > 0$ such that

$$\big|W_{j,h_n}(u) - W_{j,h_n}(v)\big| \le C h_n^{-b-1}|u - v|, \qquad u, v \in [-1/2,\, 1/2]. \tag{5.8}$$

Now we consider $D_3(x)$. Let $\{s_n\}_{n\ge3}$ be a sequence satisfying $s_n = O(h_n^{b+1}\eta_n t_n^{-1}) = o(1)$ such that, when we shatter the interval $[-1/2,\, 1/2]$ into $s_n^{-1}$ many fragments of the form $[x_i,\, x_{i+1}]$, 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

$$D_3(x) = D_4(x) - D_5(x) + D_6(x_{i'}),$$

where $D_4(x) = \hat\theta^t(x) - \hat\theta^t(x_{i'})$, $D_5(x) = E[\hat\theta^t(x)] - E[\hat\theta^t(x_{i'})]$ and $D_6(x_{i'}) = \hat\theta^t(x_{i'}) - E[\hat\theta^t(x_{i'})]$. Then we have

$$\begin{aligned}\sup_{x\in[-1/2,\,1/2]} |D_3(x)| &= \max_{i=1,\ldots,s_n^{-1}} \sup_{x\in[x_i,\,x_{i+1}]} |D_3(x)| \\ &\le \max_{i=1,\ldots,s_n^{-1}} \sup_{x\in[x_i,\,x_{i+1}]} \big|D_4(x)\big| + \max_{i=1,\ldots,s_n^{-1}} \sup_{x\in[x_i,\,x_{i+1}]} \big|D_5(x)\big| + \max_{i=1,\ldots,s_n^{-1}} \big|D_6(x_i)\big|.\end{aligned}$$

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:

$$\max_{j=1,\ldots,s_n^{-1}} \sup_{x\in[x_j,\,x_{j+1}]} |D_4(x)| = O(\eta_n), \quad \text{a.s.}, \tag{5.9}$$

$$\max_{j=1,\ldots,s_n^{-1}} \sup_{x\in[x_j,\,x_{j+1}]} |D_5(x)| = O(\eta_n) \tag{5.10}$$

and

$$\max_{j=1,\ldots,s_n^{-1}} |D_6(x_j)| = O(\eta_n), \quad \text{a.s..} \tag{5.11}$$

Beginning with (5.9), fix an arbitrary interval $[x_i,\, x_{i+1}]$. On this interval $D_4(x)$ 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 (5.8) that we can find an appropriate constant $C > 0$ for the inequality $\sup_{x\in[x_i,\, x_{i+1}]} |D_4(x)| \le C t_n h_n^{-b-1} s_n$ to hold, almost surely, independent of $i$. Therefore, by construction of $\{s_n\}_{n\ge3}$, we find that (5.9) holds. Observing that $D_5(x) = E[D_4(x)]$, we have that (5.10) holds as well.

To see the final statement (5.11) holds, define the random variables $U_j(x_i) = \{Y_j \mathbf{1}[|Y_j| \le K_j t_n] - E[Y_j \mathbf{1}[|Y_j| \le K_j t_n]]\} W_{j,h_n}(x_i)$, $j = -n, \ldots, n$. It then follows that $U_{-n}, \ldots, U_n$ are independent, and each have mean equal to zero, variance bounded by $C_1 h_n^{-2b}$ and bounded in absolute value by $C_2 t_n h_n^{-b}$, where $C_1 > 0$ and $C_2 > 0$ are appropriately chosen constants and both bounds are independent of $j$. 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

$$P\Big( \max_{i=1,\ldots,s_n^{-1}} \big|D_6(x_i)\big| > \eta_n \Big) \le 2 s_n^{-1} \exp\Big( - C \frac{n\eta_n^2}{h_n^{-2b} + t_n h_n^{-b} \eta_n} \Big). \tag{5.12}$$

In light of the fact that $t_n h_n^{-b} \eta_n = o(h_n^{-2b})$, we can enlarge $C$ for the right–hand side of (5.12) to be further bounded by a positive constant multiplied by

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

which is summable provided we take $C > (3\kappa+2)/(2\kappa) + 1/(2b)$, where $1/(2b)$ accounts for the expansion of $h_n^{-1}$; i.e. $(n^{1/(2b)} h_n)^{-1} \to 0$, as $n \to \infty$. It then follows by the Borel–Cantelli lemma that (5.11) holds, which also concludes the proof of Lemma 2. □

We can now state the proof of Theorem 1 from Section 3:

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 3. 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 $\hat{\theta}$:

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

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 $\mathscr{R}_s$ requires that we show the series condition

$$\sum_{\xi=-\infty}^{\infty} \big(1+\xi^2\big)^s \hat{\Theta}^2(\xi) < \infty \tag{5.13}$$

is satisfied. For any real numbers $a$, $b$ and $c \geq 0$, we have the inequality $|a+b|^{1+c} \leq 2^c(|a|^{1+c} + |b|^{1+c})$. Applying this inequality twice, we can see that $\hat{\Theta}^2(\xi)$ is bounded by

$$2\frac{R^2(\xi)}{\Psi^2(\xi)} + 4\left[\max_{k\in\mathbb{Z}}\left|E\big[\hat{R}(k)\big] - R(k)\right|\right]^2 \frac{\lambda^2(\xi)}{\Psi^2(\xi)} + 4\frac{\lambda^2(\xi)}{\Psi^2(\xi)}\left\{\frac{1}{2n+1}\sum_{j=-n}^{n}\varepsilon_j e^{i2\pi\xi x_j}\right\}^2. \tag{5.14}$$

Observing that $\theta \in \mathscr{R}_s$, we have $\sum_{\xi=-\infty}^{\infty}(1+\xi^2)^s\{R^2(\xi)/\Psi^2(\xi)\} < \infty$. Hence, we only need to verify the series condition (5.13) stated for the last two terms in (5.14) holds.

Similar lines of arguments for showing the result $\sum_{k=-\infty}^{\infty}\{|\lambda(k)|/|\Psi(k)|\} = O(h_n^{-b})$ in the proof of Lemma 1 give $\sum_{\xi=-\infty}^{\infty}(1+\xi^2)^s\{\lambda^2(\xi)/\Psi^2(\xi)\} = O(h_n^{-2b})$. Since the assumptions of Lemma 1 are satisfied, we have $\max_{k\in\mathbb{Z}}|E[\hat{R}(k)] - R(k)| = O(n^{-1})$. This implies

$$4\left[\max_{k\in\mathbb{Z}}\left|E\big[\hat{R}(k)\big] - R(k)\right|\right]^2 \sum_{\xi=-\infty}^{\infty}\big(1+\xi^2\big)^s\frac{\lambda^2(\xi)}{\Psi^2(\xi)} = O\big((nh_n^b)^{-2}\big) = o(1).$$

Hence, the series condition (5.13) stated for the second term in (5.14) holds.

Since $\exp(-i2\pi kx)$ is confined to the unit circle in the complex plane, a standard argument gives

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

Turning our attention to the third term of (5.14), we have

$$4\left[\max_{k\in\mathbb{Z}}\left|\frac{1}{2n+1}\sum_{j=-n}^{n}\varepsilon_j e^{-i2\pi kx_j}\right|\right]^2 \sum_{\xi=-\infty}^{\infty}\big(1+\xi^2\big)^s\frac{\lambda^2(\xi)}{\Psi^2(\xi)} = O\big(n^{-1}\log(n)h_n^{-2b}\big), \qquad \text{a.s.}$$

Since $n^{-1}\log(n)h_n^{-2b} = O(n^{-(s-b)/s}\log^{(s-b)/s}(n)) = o(1)$, we can see the series condition (5.13) stated for the third term of (5.14) is satisfied, almost surely, for large enough $n$. Combining these results shows that (5.13) holds, i.e.

$$\sum_{\xi=-\infty}^{\infty}\big(1+\xi^2\big)^s\hat{\Theta}^2(\xi) < \infty,$$

almost surely, for large enough $n$. It follows that $\hat{\theta} - \theta \in \mathscr{R}_s$, almost surely, for large enough $n$. Combining this statement with the first assertion 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 in weighted norms. Since our space $\mathscr{R}_s$ is a collection of functions with compact support on the interval $[-1/2, 1/2]$, we can see the results of their Corollary 4 on bracketing numbers for weighted Sobolev spaces immediately apply to our case by repeating the steps in the proof of their Corollary 2 for Besov-type functions of bounded support, i.e. our function space $\mathscr{R}_{s,1}$ is the unit ball of the metric space $(\mathscr{R}_s, \|\cdot\|_\infty)$ and $\mathscr{R}_{s,1}$ can be viewed as a restriction of a larger weighted Sobolev space of similar type, where the weighting function is now defined to be equal to 1 on the interval $[-1/2,\,1/2]$. We can summarize this result in the following corollary:

Corollary 2 (Corollary 4 of Nickl and Pötscher, 2007). *For the function space $\mathscr{R}_{s,1}$, with $s > 1/2$, there is a constant $C > 0$ 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 $\epsilon$ 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 $o_P(n^{-1/2})$.

Lemma 4. *Let the assumptions of Theorem 1 be satisfied with $s > 1$. 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*

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

Proof. This argument is similar to the proof of Lemma A.1 of Van Keilegom and Akritas (1999), who prove a similar result for a direct regression model. We will begin by showing the class of functions

$$\mathfrak{F} = \left\{(x,\varepsilon) \mapsto \mathbf{1}\big[\varepsilon \le t + [Kq](x)\big] - \int_{-1/2}^{1/2} F\big(t + [Kq](x)\big)\,dx \;:\; t\in\mathbb{R},\, q \in \mathscr{R}_{s,1}\right\}$$

is $\mu \times F$–Donsker, where $\mu$ is the Lebesgue measure on the interval $[-1/2,\,1/2]$. We will then use this property to prove the assertion. To show $\mathfrak{F}$ is $\mu\times F$–Donsker, we need to show Dudley's entropy integral condition,

$$\int_0^\infty \sqrt{\log N_{[]}\big(\epsilon,\, \mathfrak{F},\, L_2(\mu\times F)\big)}\,d\epsilon < \infty,$$

is satisfied (see, for example, Theorem 2.5.6 in van der Vaart and Wellner, 1996), where we write $N_{[]}(\epsilon,\, \mathfrak{F},\, L_2(\mu\times F))$ for the number of brackets of length $\epsilon$ required to cover $(\mathfrak{F}, L_2(\mu\times F))$ and we write $L_2(\mu\times F)$ for the $L_2$–norm with respect to the product measure $\mu\times F$. Since $\mathfrak{F}$ is composed of a sum of elements, we only need to show the result for the simpler class $\mathfrak{F}_1 = \{(x,\varepsilon)\mapsto \mathbf{1}[\varepsilon \le t + [Kq](x)],\, :\, t\in\mathbb{R},\, q\in\mathscr{R}_{s,1}\}$ because the proof for the second class is almost the same and, therefore, omitted.

The assumptions of Corollary 2 are satisfied. Hence, it follows for there to be a constant $C > 0$ such that $n_q = N_{[]}(\epsilon^2/(2\|f\|_\infty),\, \mathscr{R}_{s,1},\, \|\cdot\|_\infty) \le \exp(C\epsilon^{-2/s})$. Let $\varrho_{l,1} \le \varrho_{u,1},\ldots,\varrho_{l,n_q} \le \varrho_{u,n_q}$ be the $n_q$ brackets that cover $(\mathscr{R}_{s,1}, \|\cdot\|_\infty)$. Now write $F_{l,i}(t) = F(t + [K\varrho_{l,i}](x))$ and $F_{u,i}(t) = F(t + [K\varrho_{u,i}](x))$, for each $i = 1,\ldots,n_q$. Observing that $F_{l,i}$ and $F_{u,i}$ are probability measures, we can shatter $\mathbb{R} \cup \{-\infty, \infty\}$ into $O(\epsilon^{-2})$ many fragments of the form $[t_{l,i,j_1}, t_{l,i,j_1+1}]$ such that $\max_{j_1=1,\dots,O(\epsilon^{-2})} |F_{l,i}(t_{l,i,j_1+1}) - F_{l,i}(t_{l,i,j_1})| \le \epsilon^2/4$, and, separately, we can construct a similar shattering of $\mathbb{R} \cup \{-\infty, \infty\}$ using $F_{u,i}$ obtaining fragments of the form $[t_{u,i,j_2}, t_{u,i,j_2+1}]$, $j_2 = 1, \dots, O(\epsilon^{-2})$. It then follows that $t \in \mathbb{R}$ is bracketed by $t^-_{l,i,j_1} \le t^+_{u,i,j_2}$, where $t^-_{l,i,j_1}$ is the largest $t_{l,i,j_1}$ that is less than or equal to $t$ and $t^+_{u,i,j_2}$ is the smallest $t_{u,i,j_2}$ that is greater than or equal to $t$.

We will now show our brackets for $\mathfrak{F}_1$ are given by

$$\mathbf{1}\big[\varepsilon \le t^-_{l,i,j_1} + \big[K\varrho_{l,i}\big](x)\big] \le \mathbf{1}\big[\varepsilon \le t + [Kq](x)\big] \le \mathbf{1}\big[\varepsilon \le t^+_{u,i,j_2} + \big[K\varrho_{u,i}\big](x)\big].$$

The squared length of our proposed brackets is

$$\int_{-1/2}^{1/2} \Big\{ F_{u,i}\big(t^+_{u,i,j_2}\big) - F_{l,i}\big(t^-_{l,i,j_1}\big) \Big\}\, dx,$$

which is bounded by

$$\int_{-1/2}^{1/2} \Big\{ F_{u,i}(t) - F_{l,i}(t) \Big\}\, dx + \frac{\epsilon^2}{2}. \tag{5.15}$$

Observing that $F$ has a bounded Lebesgue density $f$, the integral in (5.15) is bounded by

$$\|f\|_\infty \int_{-1/2}^{1/2} \big[K\big(\varrho_{u,i} - \varrho_{l,i}\big)\big](x)\, dx \le \|f\|_\infty \|\varrho_{u,i} - \varrho_{l,i}\|_\infty \le \frac{\epsilon^2}{2},$$

where we have used that $K1 = 1$ and our construction of the bracket $\varrho_{l,i} \le \varrho_{u,i}$. This implies (5.15) is bounded by $\epsilon^2$, and, therefore, our proposed brackets for $(\mathfrak{F}_1, L_2(\mu \times F))$ have $L_2(\mu \times F)$–length no greater than $\epsilon$ as required.

When $0 < \epsilon < 1$, it then follows that we need at most $O(\epsilon^{-2} \exp(C\epsilon^{-2/s}))$ many brackets to cover $(\mathfrak{F}_1, L_2(\mu \times F))$, and, when $\epsilon \ge 1$, only one bracket is required. This implies that we can find appropriate constants $C_1$ and $C_2$ such that

$$\int_0^\infty \sqrt{\log N_{[]}\big(\epsilon, \mathfrak{F}_1, L_2(\mu \times F)\big)}\, d\epsilon = \int_0^1 \sqrt{\log N_{[]}\big(\epsilon, \mathfrak{F}_1, L_2(\mu \times F)\big)}\, d\epsilon \le C_1 + C_2 \frac{s}{s-1}.$$

Since $s > 1$, the bound above is finite and so Dudley's entropy integral condition is satisfied. This shows the class $\mathfrak{F}_1$ is $\mu \times F$–Donsker, and, therefore, $\mathfrak{F}$ is also $\mu \times F$–Donsker.

By Corollary 2.3.12 of van der Vaart and Wellner (1996), $\mathfrak{F}$ is Donsker implies empirical processes indexed by $\mathfrak{F}$ are asymptotically equicontinuous in the sense that, for every $\eta > 0$,

$$\lim_{\alpha \downarrow 0} \limsup_{n\to\infty} P\bigg( \sup_{\{\mathfrak{f}_1,\mathfrak{f}_2 \in \mathfrak{F}\,:\, \mathrm{Var}[\mathfrak{f}_1 - \mathfrak{f}_2] < \alpha\}} (2n+1)^{-1/2} \bigg| \sum_{j=-n}^{n} \Big\{ \mathfrak{f}_1(x_j, \varepsilon_j) - \mathfrak{f}_2(x_j, \varepsilon_j) \Big\} \bigg| > \eta \bigg) = 0. \tag{5.16}$$

Since the assumptions of Theorem 1 are satisfied, we have that $\hat\theta - \theta \in \mathscr{R}_{s,1}$, almost surely, for large enough $n$. Respectively using $\hat\theta - \theta$ and the zero function in place of $q$, the difference $\mathfrak{f}_1(x_j, \varepsilon_j) - \mathfrak{f}_2(x_j, \varepsilon_j)$ now becomes $\mathbf{1}[\varepsilon_j \le t + [K(\hat\theta - \theta)](x_j)] - \int_{-1/2}^{1/2} F(t + [K(\hat\theta-\theta)](x))\, dx - \mathbf{1}[\varepsilon_j \le t] + F(t)$, which, for large enough $n$, belongs to $\mathfrak{F}$ almost surely. Therefore, we only need to check the variance condition under the supremum in (5.16) to finish proving the assertion.

To fix the function $\hat\theta$, we condition on the observed data $(x_j, Y_j)$, $j = -n, \ldots, n$, and the variance condition becomes

$$\begin{aligned}
&\int_{-1/2}^{1/2}\int_{-\infty}^{\infty}\Big\{\mathbf{1}\big[v \le t + \big[K\big(\hat\theta-\theta\big)\big](x)\big] - \int_{-1/2}^{1/2} F\big(t + \big[K\big(\hat\theta-\theta\big)\big](x)\big)\,dx \\
&\qquad\qquad - \mathbf{1}[v \le t] + F(t)\Big\}^2 F(dv)\,dx \\
&= \int_{-1/2}^{1/2}\bigg\{F\bigg(\max\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg) - F\bigg(\min\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg)\bigg\}\,dx \\
&\quad - \bigg\{\int_{-1/2}^{1/2}\bigg\{F\bigg(\max\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg) - F\bigg(\min\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg)\bigg\}\,dx\bigg\}^2,
\end{aligned}$$

which is bounded by

$$\begin{aligned}
&\sup_{x\in[-1/2,\,1/2]}\ \sup_{t\in\mathbb{R}}\bigg|F\bigg(\max\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg) - F\bigg(\min\Big\{t,\, t + \big[K\big(\hat\theta-\theta\big)\big](x)\Big\}\bigg)\bigg| \\
&\le \|f\|_\infty \big\|\hat\theta-\theta\big\|_\infty.
\end{aligned}$$

Also by Theorem 1, we have that $\|\hat\theta-\theta\|_\infty = o(1)$, almost surely, and so it follows for the bound above to be $o(1)$, almost surely. Hence, the variance condition in (5.16) is satisfied. The assertion is then implied by the equicontinuity of empirical processes indexed by the restriction of $\mathfrak{F}$ to those elements in $\mathfrak{F}$ corresponding to $\hat\theta-\theta$ and the zero function. $\square$

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 2. *Let the assumptions of Lemma 1 be satisfied, and assume that $E[\varepsilon_j^2] < \infty$, $j = -n, \ldots, n$. In addition, let the regularizing sequence $\{h_n\}_{n\ge1}$ satisfy $h_n^{s+1} = o(n^{-1/2})$. Then*

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

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 $1 = \int_{-1/2}^{1/2}\sum_{k=-\infty}^{\infty} e^{i2\pi kx}\,dx$ so that we can bound the left–hand side of the assertion by $S_1+S_2+S_3$, where

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

$$S_2 = \left[ \max_{k \in \mathbb{Z}} \left| E\big[\hat{R}(k)\big] - R(k) \right| \right] \sum_{k=-\infty}^{\infty} |\lambda(k)|$$

and

$$S_3 = \sum_{k=-\infty}^{\infty} |\lambda(k) - 1||R(k)| \left| \int_{-1/2}^{1/2} e^{i2\pi kx}\, dx \right|.$$

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 $S_1 = o_p(n^{-1/2})$, if we can show

$$\frac{1}{2n+1} \sum_{j=-n}^{n} \left\{ \int_{-1/2}^{1/2} \left\{ \sum_{k=-\infty}^{\infty} \big\{\lambda(k) - 1\big\} e^{i2\pi k(x - x_j)} \right\} dx \right\}^2 = o(1). \tag{5.17}$$

Since $|\lambda(k) - 1| \le 2$, it follows that

$$\left| \int_{-1/2}^{1/2} \left\{ \sum_{k=-\infty}^{\infty} \big\{\lambda(k) - 1\big\} e^{-i2\pi k(x - x_j)} \right\} dx \right| \le 2,$$

and we have $\lambda(k) - 1 = 0$ when $k \in I(h_n) = \{z \in \mathbb{Z} : z \le M h_n^{-1}\}$. Hence, the sum is only indexed by $k \in I^c(h_n) = \{z \in \mathbb{Z} : z > M h_n^{-1}\}$, which is asymptotically empty. This implies (5.17) holds.

Now we consider the remainder term $S_2$. The assumptions of Lemma 3 are satisfied and so the first term of $S_2$ is of order $O(n^{-1})$. This and the absolute summability of $\lambda$ yield that $S_2 = o(n^{-1/2})$.

Finally, for the term $S_3$, we can assume that $k \in I^c(h_n)$ for $R_3$, which does not include $k = 0$, and the integral term in this quantity is bounded by $(2/\pi)|k|^{-1}$. Since $\theta \in \mathscr{R}_s$, it follows that we can find an appropriate constant $C > 0$ for the inequality $|R(k)| \le C|k|^{-s}$ to hold, and we can enlarge $C$ such that $R_3$ is bounded by

$$C h_n^{s+1} \sum_{\omega \in h_n \mathbb{Z}} |\omega|^{-(s+1)}.$$

This shows that $R_3 = O(h_n^{s+1}) = o(n^{-1/2})$. □

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 in Section 5 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 $\gamma$, 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

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

It follows that $h_n^{s+1} = O(n^{-(1/2)-1/(2s)}\log^{(s+1)/(2s)}(n)) = o(n^{-1/2})$, and, hence, the assumptions of Proposition 2 in Section 5 are satisfied, which implies the second term in the bound above is $o_p(n^{-1/2})$. This shows that $\sup_{t\in\mathbb{R}}|L_n(t)| = o_p(n^{-1/2})$. The assertion of Theorem 2 then follows. $\square$

## 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 (26pp.).

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

[4] Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems*, 24: 034009 (17pp.).

[5] Bissantz, N. and Holzmann, H. (2013). Asymptotics for spectral regularization estimators in statistical inverse problems. *Computational Statistics*, 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] Cao, R. (1993). Bootstrapping the mean integrated squared error. *Journal of Multivariate Analysis*, 45: 137-160.

[9] Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Annals of Statistics*, 28: 630-647.

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

[11] Davies, P.L. and Meise, M. (2008). Approximating data with weighted smoothing splines. *Journal of Nonparametric Statistics*, 20: 207-228.

[12] González-Manteiga, W., Martinez-Miranda, M.D. and Pérez-González, A. (2004). The choice of smoothing parameter in nonparametric regression through wild bootstrap. *Computational Statistics and Data Analysis*, 47: 487-515.

[13] Hotz, T., Marnitz, P., Stichtenoth, R., Davies, L., Kabluchko, Z. and Munk, A. (2012). Locally adaptive image denoising by a statistical multiresolution criterion. *Computational Statistics and Data Analysis*, 56: 543-558.

[14] Mair, B.A. and Ruymgaart, F.H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56: 1424-1444.

[15] Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory*, 37: 1105-1115.

[16] Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary processes. *Journal of Nonparametric Statstics*, 3: 13-36.

[17] Müller, U.U., Schick, A. and Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *Journal of Statistical Planning and Inference*, 119: 75-93.

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

[19] Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics*, 36: 204-228.

[20] Nickl, R. and Pötscher, B.M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *Journal of Theoretical Probability*, 20: 177-199.

[21] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33: 1065-1076.

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

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

[24] Silverman, B.W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 6: 177-184.

[25] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Vol. 26. CRC press.

[26] 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.

[27] Van Keilegom, I. and Akritas, M.G. (1999). Transfer of tail information in censored regression models. *Annals of Statistics*, 27: 1745-1784.