A NOTE ON CONFIDENCE INTERVALS FOR DEBLURRED IMAGES

Michał Biel and Zbigniew Szkutnik

Communicated by Mirosław Pawlak

Abstract. We consider pointwise asymptotic confidence intervals for images that are blurred and observed in additive white noise. This amounts to solving a stochastic inverse problem with a convolution operator. Under suitably modified assumptions, we fill some apparent gaps in the proofs published in [N. Bissantz, M. Birke, *Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators*, J. Multivariate Anal. 100 (2009), 2364–2375]. In particular, this leads to modified bootstrap confidence intervals with much better finite-sample behaviour than the original ones, the validity of which is, in our opinion, questionable. Some simulation results that support our claims and illustrate the behaviour of the confidence intervals are also presented.

Keywords: inverse problems, confidence intervals, convolution, deblurring.

Mathematics Subject Classification: 62G08, 62G15, 62G20.

1. INTRODUCTION

Many practical problems in science, medical imaging, astronomy etc. can be formulated as stochastic inverse problems, i.e. problems with indirect and noisy observations. The distribution of observable data then depends on a parameter, say g, related to the object of real interest, say θ, through an operator equation $g = \mathcal{K}\theta$. Typically, θ and g are elements of same function spaces and $\mathcal{K}^{-1}$ is unbounded, which makes the problem ill-posed in the Hadamard sense and some sort of regularization or smoothing becomes necessary. If $\mathcal{K}$ is the identity operator, the problem becomes a direct one, although the distinction is not sharp and some problems, e.g. density estimation, may be viewed both as direct and indirect problems (cf. [12]). A good review of non-stochastic inverse problems and related analytical and numerical techniques is given, e.g., in [13]. For a general treatment of estimation in stochastic inverse problems we refer to [7, 9, 16]. A recent discussion of adaptivity issues in inverse problems can be found in [8].

Construction of pointwise confidence intervals and/or uniform confidence bands is the most informative way of quantifying the accuracy of estimation in problems of function estimation. The estimators typically depend on regularization/smoothing parameters. For theoretical construction of confidence regions, those parameters are usually assumed to be chosen a priori, i.e. not to depend on data. The point estimates are then often linear, which paves the way to finding their asymptotic distributions and the corresponding confidence intervals via an application of a suitable central limit theorem and some sort of bias correction or undersmoothing, which makes the bias negligible. The constructions of uniform confidence regions are usually based on limit theorems for the sup-norm of an appropriately centered functional estimator.

Construction of confidence bands in direct problems of function estimation started in 1973 with the seminal paper by Bickel and Rosenblatt [1], who constructed confidence bands for density estimated from an i.i.d. sample, and continued in several further developments, as summarized, e.g., in [14, Ch. 5.1.3] and [6]. The latter paper was also the first step towards the construction of confidence bands in inverse problems and was followed in recent years by several similar works [2, 4, 5, 10, 11, 15, 18, 19].

In typical cases, the widths rates of the pointwise and the uniform confidence intervals differ only by a logarithmic factor (see, e.g., [18]), which makes studying both of them interesting. The uniform confidence bands typically need larger sample sizes to work reasonably, because of the slow convergence to the asymptotic extremal distributions. This, and also the dependence of asymptotic distributions on unknown parameters that have to be estimated, make the alternative bootstrap constructions attractive.

In this paper, we reconsider the construction of pointwise confidence intervals in an inverse regression model originally studied in [3] and correct some deficiencies of the original article.

With $\mathbf{r} = (r_1, \ldots, r_d) \in \{-n, \ldots, n\}^d$ and a positive sequence $a_n \to 0$, consider a d-dimensional grid $\mathbf{z_r} = (r_1/(na_n), \ldots, r_d/(na_n))$ and corresponding measurements $Y_\mathbf{r}$ of a scalar function of d variables. The measurements may represent, e.g., the intensity of a d-dimensional image in pixels centered at $\mathbf{z_r}$. Assume that the true image, say $\theta(\mathbf{z})$, is blurred by convolution with a known function Ψ, and that the resulting function $g(\mathbf{z}) = (\Psi * \theta)(\mathbf{z}) = \int_{\mathbb{R}^d} \Psi(\mathbf{z} - \mathbf{t})\theta(\mathbf{t})d\mathbf{t}$ is measured with additive white noise, i.e.,

$$Y_\mathbf{r} = g(\mathbf{z_r}) + \epsilon_\mathbf{r}, \qquad (1.1)$$

with i.i.d. $\epsilon_\mathbf{r}$'s such that $\mathbb{E}(\epsilon_\mathbf{r}) = 0$ and $\mathbb{E}(\epsilon_\mathbf{r}^2) = \sigma^2$. The problem of estimating θ and its derivatives in this inverse regression model with convolution operator was studied in [3], along with a discussion of the background literature. Notice that the design grid expands as $n \to \infty$ and, if $na_n \to \infty$, it becomes asymptotically dense in $\mathbb{R}^d$. As discussed in [3], this allows for estimation of θ without assuming its periodicity and compactness of its support.

Let $\theta^{(\mathbf{j})}(\mathbf{z}) = \partial^j \theta(\mathbf{z})/(\partial z_1^{j_1} \ldots \partial z_d^{j_d})$, with $\mathbf{z} = (z_1, \ldots, z_d)$, $\mathbf{j} = (j_1, \ldots, j_d)$ and $j = \sum_{k=1}^d j_k$. A smoothed spectral-cut-off estimator of $\theta^{(\mathbf{j})}$ of the form

$$\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) = \sum_{\mathbf{r} \in \{-n,\ldots,n\}^d} \frac{1}{n^d h^{j+d} a_n^d} K_n^{(\mathbf{j})}\big((\mathbf{z} - \mathbf{z_r})/h\big) Y_\mathbf{r} \qquad (1.2)$$

was proposed in [3], where

$$K_n^{(\mathbf{j})}(\mathbf{z}) = \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} (-i\boldsymbol{\omega})^{\mathbf{j}} \exp(-i\langle \boldsymbol{\omega}, \mathbf{z}\rangle) \frac{\mathcal{F}k(\boldsymbol{\omega})}{\mathcal{F}\Psi(\boldsymbol{\omega}/h)} \mathrm{d}\boldsymbol{\omega},$$

h is a scalar smoothing parameter, $\mathcal{F}$ denotes the Fourier transform operator, $k(\cdot)$ is a suitable flat-top kernel, and $\boldsymbol{\omega}^{\mathbf{j}} = \prod_k \omega_k^{j_k}$, with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$. Asymptotic normality of $\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})$ was proved in [3], and pointwise asymptotic confidence intervals for $\theta^{(\mathbf{j})}(\mathbf{z})$ were constructed. Bootstrap confidence intervals were studied as well, and recommended because they usually perform in finite samples better than the asymptotic ones. However, there seem to be some gaps in the proofs presented in [3]. In Section 2, we fill the gaps under slightly strengthened assumptions. In case of the bootstrap confidence intervals, this also results in a different form of the intervals. Results of a simulation study are presented in Section 3. In particular, we show a rather peculiar finite sample behaviour of the originally proposed bootstrap intervals, and a much better performance of the corrected ones.

2. MAIN RESULTS

Various subsets of the following list of regularity assumptions will be used in the sequel. The interpretation of Assumptions 1-4 has been discussed in [3].

(A1) $\mathcal{F}k$ is symmetric and supported on $[-1,1]^d$, $\mathcal{F}k(\boldsymbol{\omega}) = 1$ for $\boldsymbol{\omega} \in [-b,b]^d$ with some $b > 0$, and $|\mathcal{F}k(\boldsymbol{\omega})| \leq 1$ for all $\boldsymbol{\omega} \in [-1,1]^d$.

(A2) $\mathcal{F}\Psi(\boldsymbol{\omega})\|\boldsymbol{\omega}\|^{\beta} \to C$, as $\boldsymbol{\omega} \to \infty$, for some positive β and some complex, nonzero C.

(A3) $\int_{\mathbb{R}^d} |\mathcal{F}\theta(\boldsymbol{\omega})|\|\boldsymbol{\omega}\|^{s-1}\mathrm{d}\boldsymbol{\omega} < \infty$ for some $s > p + 1$.

(A4) $\int_{\mathbb{R}^d} |g(\mathbf{z})|\|\mathbf{z}\|^r\mathrm{d}\mathbf{z} < \infty$.

(A5) g has bounded partial derivatives of order one.

2.1. CORRECTED ASSUMPTIONS FOR ASYMPTOTIC NORMALITY

The first important gap in [3] seems to be in the proof of Theorem 1, which concerns asymptotic normality of $\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})$ and states that under Assumptions 1 and 2 and with

$$w_{\mathbf{j},\mathbf{r},n}(\mathbf{z}) = \frac{1}{n^d h^{j+d} a_n^d} K_n^{(\mathbf{j})}\big((\mathbf{z} - \mathbf{z_r})/h\big)$$

one has

$$\frac{\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) - \mathbb{E}\big(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})\big)}{\sigma\sqrt{\sum_{\mathbf{r}} w_{\mathbf{j},\mathbf{r},n}^2(\mathbf{z})}} \xrightarrow{\mathrm{d}} Z \sim \mathrm{N}(0,1), \tag{2.1}$$

if $nha_n \to \infty$. We see the need for correction in the second and third line on page 2372 in [3], where the sum in the denominator is interpreted as the midpoint quadrature for

an integral. It is unclear, how the midpoint quadrature error can give $O((nha_n)^{-d})$, as Bissantz and Birke claim without giving any supporting argument (h is also missing there, but this seems to be a misprint). Standard first-order Taylor expansion of the integrand $F_1(\mathbf{s}) := \left[K_n^{(\mathbf{j})}((\mathbf{z} - \mathbf{s})/h) \right]^2$, along with the fact that under Assumptions 1 and 2 all its first-order derivatives can be shown to be $O(h^{-2\beta-1})$, give the quadrature error related term of the order $O(n^{-1}h^{-d-2\beta-1}a_n^{-d-1})$. This is obtained by first showing in a standard way that

$$
\left| \int\limits_{A_\mathbf{r}} F_1(\mathbf{s}) \mathrm{d}\mathbf{s} - \frac{1}{(na_n)^d} F_1(\mathbf{z_r}) \right| = O\left(\frac{1}{h^{2\beta+1}} \frac{1}{(na_n)^{d+1}} \right),
$$

where $A_\mathbf{r}$ is the cube centered at $\mathbf{z_r}$ with edge $(na_n)^{-1}$, then summing over the grid points $\mathbf{z_r}$, and a suitable change of integration variable. Different order of approximation results in a stronger condition to be imposed in order to obtain the asymptotic equivalence in the fourth line on page 2372 in [3]. One needs for that $nh^{d+1}a_n^{d+1} \to \infty$, rather than $nha_n \to \infty$, as assumed in [3].

Midpoint quadrature is also used in the proof of Lemma 3 in [3] that gives an estimate of the order of magnitude of the bias. Bissantz and Birke claim to have the quadrature error of order $O(n^{-d}a_n^{-d}h^{-j-d})$ in the third line of formula (5) in the proof on page 2372. This again seems questionable, so we propose to use the same error estimation method as before, which gives the quadrature error $O(n^{-1}a_n^{-d-1}h^{-1})$ in the third line and, finally, $O(n^{-1}a_n^{-d-1}h^{-j-d-1})$ in the fourth line in formula (5) in [3]. Consequently, one needs to strengthen the original assumptions even more significantly than before. Not only has the condition $n^d h^{\beta+s+d-1}a_n^d \to \infty$ to be strengthened to $nh^{\beta+s+d}a_n^{d+1} \to \infty$, but also, in order to again estimate the quadrature error via Taylor expansion of the integrand $F_2(\mathbf{y}) := g(\mathbf{y}) \exp(i\langle \boldsymbol{\omega}, \mathbf{y} \rangle / h)$, one has to assume that all first-order derivatives of the function g in (1.1) are bounded.

It is known (see, e.g., [17]) that midpoint quadratures may perform extremely well, with rates of convergence exponential in mesh size, when applied to analytic and fastly decaying integrands. However, out of the two integrands F_1 and F_2 used in the proofs, only F_1 is guaranteed to be analytic, as a band-limited function, and an application of the results from [17] does not seem possible, also because F_1 does not satisfy the remaining conditions imposed in [17].

To summarize:

– under assumptions (A1) and (A2), the asymptotic normality given in (2.1) holds true, if $h \to 0$, $a_n \to 0$ and $nh^{d+1}a_n^{d+1} \to \infty$ (this is stronger than $nha_n \to \infty$, assumed in [3]);

– under assumptions (A1)–(A5), bias $\left(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) \right) = o(h^{s-j-1})$, if $nh^{\beta+s+d}a_n^{d+1} \to \infty$ (this is stronger than $n^d h^{\beta+s+d-1}a_n^d \to \infty$, assumed in [3]).

Under these strengthened conditions, the construction of asymptotic confidence intervals, as described in Corollary 4 in [3], remains valid.

2.2. CORRECTIONS TO THE CONSTRUCTION OF BOOTSTRAP CONFIDENCE INTERVALS

A more essential gap in [3] apparently occurs in the bootstrap part of the paper and results in a seemingly faulty form of the bootstrap confidence intervals. To explain that point, let us recall the residual bootstrap approach proposed in [3]. First, the function $g = \Psi * \theta$ is estimated with $\hat{g}_{n,\tilde{h}} = \Psi * \hat{\theta}_{n,\tilde{h}}$, where $\hat{\theta}_{n,\tilde{h}}$ is the estimator (1.2) constructed with an auxiliary bandwidth $\tilde{h}$. Then, a set of centered residuals $\hat{\epsilon}_{\mathbf{r}} - \bar{\hat{\epsilon}}$, $\mathbf{r} \in \{-n, \ldots, n\}^d$ is constructed, where $\hat{\epsilon}_{\mathbf{r}} = Y_{\mathbf{r}} - \hat{g}_{n,\tilde{h}}(\mathbf{z}_{\mathbf{r}})$ and $\bar{\hat{\epsilon}}$ is the mean of $\hat{\epsilon}_{\mathbf{r}}$'s. Bootstrap data is generated as $Y_{\mathbf{r}}^* = \hat{g}_{n,\tilde{h}}(\mathbf{z}_{\mathbf{r}}) + \epsilon_{\mathbf{r}}^*$, with $\epsilon_{\mathbf{r}}^*$ drawn with replacement from the set of centered residuals. The bootstrap estimator is then defined as

$$\hat{\theta}_{n,h}^{(\mathbf{j})*}(\mathbf{z}) = \sum_{\mathbf{r} \in \{-n, \ldots, n\}^d} \frac{1}{n^d h^{j+d} a_n^d} K_n^{(\mathbf{j})}\big((\mathbf{z} - \mathbf{z}_{\mathbf{r}})/h\big) Y_{\mathbf{r}}^*, \tag{2.2}$$

and its bootstrap distribution is used to construct a confidence interval for $\theta^{(\mathbf{j})}(\mathbf{z})$. Typically, as suggested in [3], $\tilde{h}$ is taken larger than h.

The properties of the bootstrap confidence intervals are studied under the following additional conditions, of which (A8) was missing in [3].

(A6) $\mathcal{F}\hat{\theta}_{n,\tilde{h}}$ exists and satisfies

$$\int_{\mathbb{R}^d} |\mathcal{F}\theta(\boldsymbol{\omega}) - \mathcal{F}\hat{\theta}_{n,\tilde{h}}| \, \|\boldsymbol{\omega}\|^{s-1} \mathrm{d}\boldsymbol{\omega} = o_P(1).$$

(A7) $\int_{\mathbb{R}^d} \left|\hat{g}_{n,\tilde{h}}(\mathbf{z})\right| \|\mathbf{z}\|^r \mathrm{d}\mathbf{z} < \infty$ for some $r > 0$.
(A8) Almost surely, $\hat{g}_{n,\tilde{h}}$ have bounded partial derivatives of order one.

The conclusion of Theorem 5 in [3] concerns the asymptotics of the bootstrap distribution of $\hat{\theta}_{n,h}^{(\mathbf{j})*} - \mathbb{E}_*\big(\hat{\theta}_{n,h}^{(\mathbf{j})*}\big)$, where $\mathbb{E}_*$ denotes the expectation with respect to the bootstrap distribution, all conditionally on the initial sample $\{Y_{\mathbf{r}}\}$, and gives the consistency of the residual bootstrap. It should be noted that only assumptions (A1) and (A2) are used in the proof of this result, and that assumptions (A3), (A4) and (A6), (A7), imposed in the statement of Theorem 5, are only needed in the construction of bootstrap confidence intervals in the paragraph that follows Theorem 5. Also, since condition (4) from page 2366 in [3] is used in the proof on page 2374, one has to additionally assume that $nh^{d+1}a_n^{d+1} \to \infty$. Because of reasons discussed above in Section 2.1, the weaker condition $nha_n \to \infty$ does not seem sufficient.

Bootstrap confidence intervals proposed in [3] have the form

$$\left(2\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) - \vartheta_{n,1-\alpha/2}^*(\mathbf{z}), 2\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) - \vartheta_{n,\alpha/2}^*(\mathbf{z})\right), \tag{2.3}$$

where $\vartheta_{n,\gamma}^*(\mathbf{z})$ is the γ-quantile of the bootstrap distribution of $\hat{\theta}_{n,h}^*(\mathbf{z})$. For the asymptotic validity of such intervals, i.e., for the coverage probability to be $1 - \alpha + o_P(1)$, one would need

$$\left[\mathbb{E}\big(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})\big) - \theta^{(\mathbf{j})}(\mathbf{z})\right] - \left[\mathbb{E}_*\big(\hat{\theta}_{n,h}^{(\mathbf{j})*}(\mathbf{z})\big) - \hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})\right] = o_P\left((na_n)^{-d/2} h^{-\beta-j-d/2}\right) \tag{2.4}$$

which is asserted, with $d = 1$ and without detailed proof, on page 2374 in [3]. It does not seem, however, to be true in general (unless $\tilde{h} = h$).

On the other hand, it can be proved that, with s defined in (A6), if

$$n^d h^{2s+2\beta+d-2} a_n^d = o(1), \quad nh^{\beta+s+d} a_n^{d+1} \to \infty,$$

and (A5) and (A8) hold true, then

$$\left[\mathbb{E}\big(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})\big) - \theta^{(\mathbf{j})}(\mathbf{z})\right] - \left[\mathbb{E}_*\big(\hat{\theta}_{n,h}^{(\mathbf{j})\,*}(\mathbf{z})\big) - \hat{\theta}_{n,\tilde{h}}^{(\mathbf{j})}(\mathbf{z})\right] = o_P\left(h^{s-j-1}\right) \qquad (2.5)$$

and this shows the validity of the bootstrap confidence interval of the form

$$\left(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) + \hat{\theta}_{n,\tilde{h}}^{(\mathbf{j})}(\mathbf{z}) - \vartheta_{n,1-\alpha/2}^*(\mathbf{z}), \hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z}) + \hat{\theta}_{n,\tilde{h}}^{(\mathbf{j})}(\mathbf{z}) - \vartheta_{n,\alpha/2}^*(\mathbf{z})\right), \qquad (2.6)$$

because

$$h^{s-j-1} = o(n^{-d/2} h^{-\beta-j-d/2} a_n^{-d/2}).$$

To see the details, write

$$\hat{\theta}_{n,\tilde{h}}^{(\mathbf{j})}(\mathbf{z}) - \theta^{(\mathbf{j})}(\mathbf{z}) = A + B + C_1,$$

where

$$A = \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} (-i\boldsymbol{\omega})^{\mathbf{j}} \exp\left[-\frac{i\langle \boldsymbol{\omega}, \mathbf{z}\rangle}{h}\right] \left(1 - \mathcal{F}k(\boldsymbol{\omega})\right) \left[\mathcal{F}\hat{\theta}_{n,\tilde{h}}(\boldsymbol{\omega}/h) - \mathcal{F}\theta(\boldsymbol{\omega}/h)\right] \mathrm{d}\boldsymbol{\omega},$$

$$B = \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} q(\boldsymbol{\omega}, \mathbf{z}) \int\limits_{\mathbb{R}^d \setminus D_n} \exp\left[\frac{i\langle \boldsymbol{\omega}, \mathbf{y}\rangle}{h}\right] (\hat{g}_{n,\tilde{h}}(\mathbf{y}) - g(\mathbf{y}))\mathrm{d}\mathbf{y}\mathrm{d}\boldsymbol{\omega},$$

$$C_1 = \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} q(\boldsymbol{\omega}, \mathbf{z}) \int\limits_{D_n} \exp\left[\frac{i\langle \boldsymbol{\omega}, \mathbf{y}\rangle}{h}\right] (\hat{g}_{n,\tilde{h}}(\mathbf{y}) - g(\mathbf{y}))\mathrm{d}\mathbf{y}\mathrm{d}\boldsymbol{\omega},$$

with

$$q(\boldsymbol{\omega}, \mathbf{z}) = (-i\boldsymbol{\omega})^{\mathbf{j}} \exp\left[-\frac{i\langle \boldsymbol{\omega}, \mathbf{z}\rangle}{h}\right] \frac{\mathcal{F}k(\boldsymbol{\omega})}{\mathcal{F}\Psi(\boldsymbol{\omega}/h)}$$

and

$$D_n = \left[-\frac{1}{a_n} - \frac{1}{2na_n}, \frac{1}{a_n} + \frac{1}{2na_n}\right]^d.$$

In the same way as in the proof of Lemma 3 in [3], but using (A6)–(A8), it can be proved that $A = o_P(h^{s-j-1})$ and $B = o_P(h^{s-j-1})$. (Notice that (A6) sets some restrictions on the order of magnitude of the auxiliary bandwidth $\tilde{h}$.) Further, rewrite the difference $C_2 := \mathbb{E}_*\big(\hat{\theta}_{n,h}^{(\mathbf{j})\,*}(\mathbf{z})\big) - \mathbb{E}\big(\hat{\theta}_{n,h}^{(\mathbf{j})}(\mathbf{z})\big)$ as

$$C_2 = \frac{1}{(2\pi)^d h^{j+d}} \int\limits_{\mathbb{R}^d} q(\boldsymbol{\omega}, \mathbf{z}) \left[\sum_{\mathbf{r} \in \{-n,\dots,n\}} \frac{1}{n^d a_n^d} \exp\left(\frac{i\langle \boldsymbol{\omega}, \mathbf{z_r}\rangle}{h}\right)(\hat{g}_{n,\tilde{h}}(\mathbf{z_r}) - g(\mathbf{z_r}))\right]\mathrm{d}\boldsymbol{\omega}.$$

It is now easy to prove that

$$C_1 - C_2 = O_P\left(\frac{1}{nh^{\beta+j+d+1}a_n^{d+1}}\right) = o_P\left(h^{s-j-1}\right)$$

using Taylor series expansion and exploiting (A5) and (A8). This reasoning fails, however, for the original construction in [3], because of difficulties with bounding of $C_1 - C_2$.

To summarize:

– under assumptions (A1)–(A8), the bootstrap confidence intervals (2.6) are asymptotically valid, if $nh^{d+1}a_n^{d+1} \to \infty$, $n^d h^{2s+2\beta+d-2}a_n^d \to 0$ and $nh^{\beta+s+d}a_n^{d+1} \to \infty$.

Notice that the original confidence intervals (2.3) and our confidence intervals (2.6) coincide, if the auxiliary bandwidth $\tilde{h}$ equals the main bandwidth h. Otherwise, the two intervals are of equal length but they differ in centering.

Finally, even if condition (2.4) were true, our confidence intervals (2.6) based on condition (2.5) perform much better in finite samples than the original intervals (2.3), as will be seen in our simulation study.

3. SIMULATION RESULTS

Following [3], we used in simulations for $d = 1$ two true functions:

$$\theta_1(x) = \exp[-(x-1.1)^2/(2 \cdot 0.64)]$$

and

$$\theta_2(x) = \exp[-(x-0.2)^2/(2 \cdot 0.09)] + 1.2\exp[-(x-0.85)^2/(2 \cdot 0.04)],$$

and the Laplace convolution kernel $\Psi(x) = 1.5\exp(-3|x|)$. The sinc kernel was used in the definition of the estimator. For each setup, 200 data sets were generated from model (1.1) with Gaussian noise and confidence intervals were constructed for selected values of the x-argument. The empirical coverages were computed as fractions. For each data sample, 400 bootstrap samples were generated and the quantiles of the bootstrap distribution were estimated with empirical quantiles. As in [3], the sampling distribution for the residuals was constructed from the observations satisfying $|\mathbf{z_k}| < 1/a_n - 2.01h$. All computations were performed in the R environment ver. 3.3.1. The convolutions were numerically computed using the R function *fft*() and the Gaussian noise was generated by means of the R function *rnorm*().

In order to cross-check our implementation with [3], we started with sample size $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$. No indication was given in [3] how the secondary bandwidth $\tilde{h}$ was chosen in simulations. With $\tilde{h} = h$, i.e. in the case in which the intervals (2.3) and (2.6) coincide, we obtained a very good agreement for $\theta_1(\cdot)$, cf. Figure 1 with Figure 1 in [3]. We thus conjecture that $\tilde{h} = h$ was used in simulations in [3], which would make the original intervals identical with ours and valid, thus explaining their reasonable behaviour in the simulations described in [3].
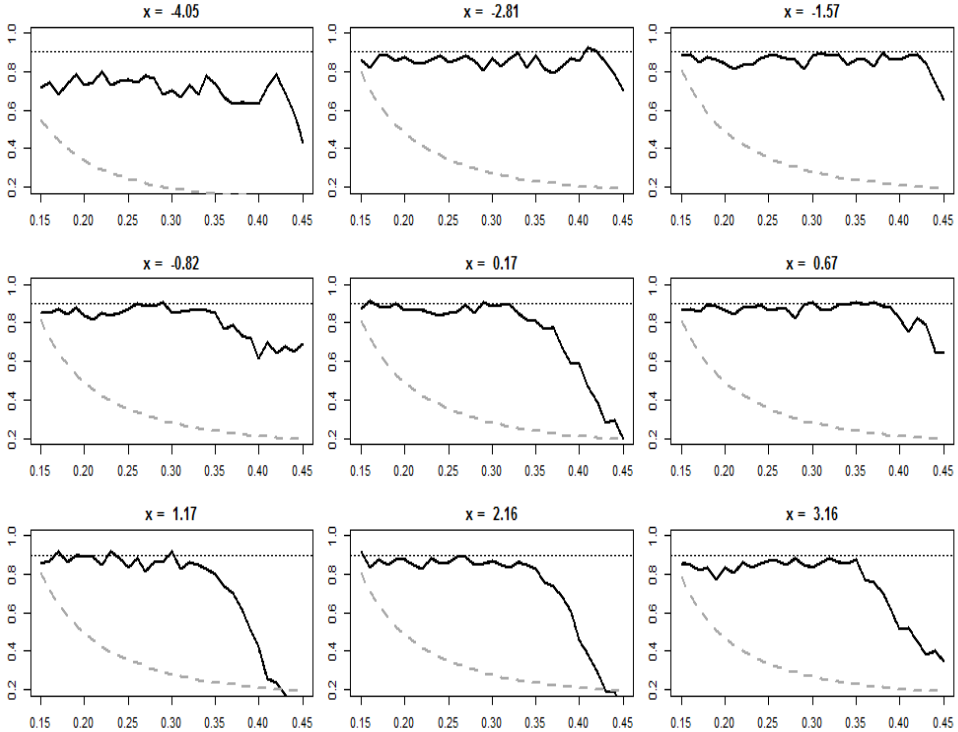
Fig. 1. Simulated coverage probabilities (solid) and interval length (dashed, multiplied by a factor of 3) for 90% nominal coverage probability (dotted), for selected values of x as functions of the bandwidth h, for the true Gaussian function $\theta_1(x)$ and for $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$. The auxiliary bandwidth $\tilde{h}$ was taken equal to the main bandwidth h

For $\theta_2(\cdot)$, our Figure 2 and Figure 2 in [3] visibly differ. As Figure 3 in [3] clearly shows a function different from that given as $\theta_2(\cdot)$ in p. 2368, it is not clear which of those functions was actually used in [3], and we do not pursue that question any further.

Next, to compare the global behaviour of both types of confidence intervals, we ran the simulation with $\tilde{h} = 2h$ and averaged the empirical coverages over 16 x-values, as in Table 1 in [3]. As clearly seen in Figure 3 and Figure 4, our modified intervals keep the nominal confidence level in a long range of h-values and uniformly in h dominate the intervals defined in [3]. Moreover, the empirical coverages of the latter are significantly and uniformly too low, which, indeed, confirms our doubts about their formal validity. A comparison of Figures 3 and 4 also shows, that smaller bandwidth should be used for less regular functions, which is an expected conclusion.
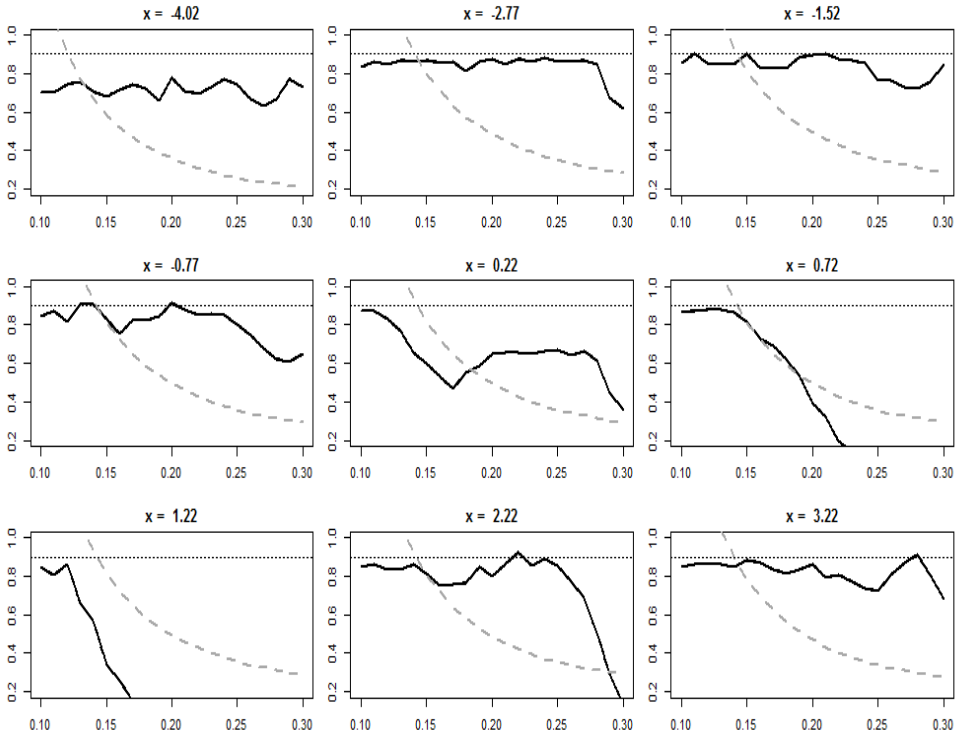
Fig. 2. Same as in Figure 1, but for the true bimodal function $\theta_2(x)$

Figure 5 illustrates the effect of various choices of the multiplier c in the relation $\tilde{h} = ch$. Larger c generally allows for more smoothing (i.e., larger h without losing the coverage probability), and more smoothing makes the intervals shorter. On the other hand, however, the larger c, the longer the intervals for a given h. In effect, as illustrated in Figure 5 for the function $\theta_1(\cdot)$, $c = 1.5$ may be a reasonable choice in the sense that it may produce the shortest intervals, provided h is optimally chosen as the maximum value that still gives the assumed coverage probability. The difference with respect to $c = 1$ is not large, however, at least for the example presented in Figure 5.

The data driven choice of h is a much more critical and difficult issue. An L_∞-motivated approach was advocated in [3]. It consists in computing the estimators $\hat{\theta}_{n,h}^{(j)}(\mathbf{z})$ for bandwidths from an equidistantly spaced grid and "choosing among these the largest bandwidth, where the supremum of the differences between the estimators for two adjacent bandwidth steps exceeds a certain threshold". This method was introduced in [6] and the chosen bandwidth is believed to approximate that minimizing the L_∞ distance between the estimator and the true function.
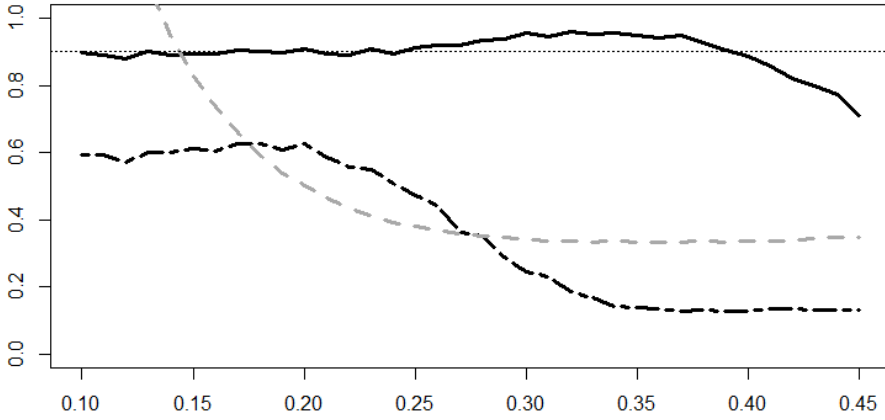
Fig. 3. Simulated coverage probabilities for original (dotdash) and new (solid) confidence intervals, and interval length (dashed, multiplied by a factor of 3) for 90% nominal coverage probability (dotted) averaged over 16 equidistant values of x approximately covering the interval [-1,3] as functions of the bandwidth h, for the true Gaussian function $\theta_1(\cdot)$ and for $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$. The auxiliary bandwidth $\tilde{h} = 2h$ was used in this case
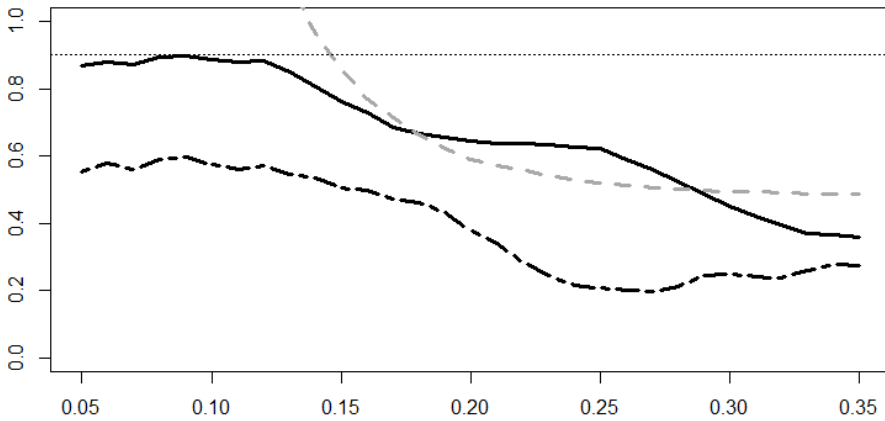

Fig. 4. Same as in Figure 3, but for the true bimodal function $\theta_2(\cdot)$
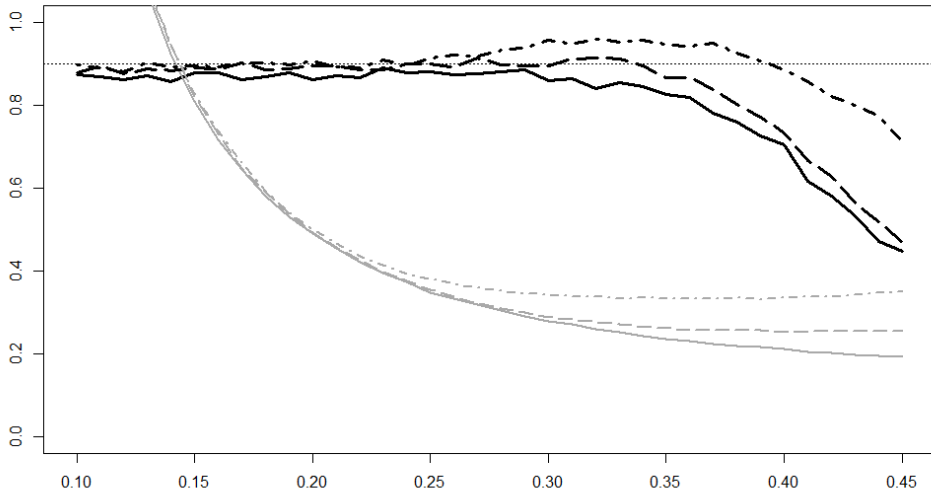
Fig. 5. Simulated coverage probabilities (black) and interval length (grey, multiplied by a factor of 3) for the new intervals, for 90% nominal coverage probability (dotted), averaged over 16 equidistant values of x approximately covering the interval [-1,3] as functions of the bandwidth h, for the true Gaussian function $\theta_1(\cdot)$ and for $2n + 1 = 201$, $\sigma = 0.1$ and $a_n = 0.25$. Various choices of the auxiliary bandwidth correspond to various line patterns: $\tilde{h} = h$ (solid), $\tilde{h} = 1.5h$ (dashed) and $\tilde{h} = 2h$ (dotdash)

According to our experience, that L_∞ optimal bandwidth h could be too large when confidence intervals are constructed. For instance, the L_∞ optimal h for the function θ_2 ($n = 100$, $\sigma = 0.1$, $a_n = 0.25$) varies between 0.15 and 0.18 and the resulting actual coverage probability is about 60-70%, with the nominal 90%. Contrary to the claims in [3], it seems that undersmoothing might sometimes be necessary.

Our simulation experience with the data driven choice of h was rather pessimistic. The method suggested in [3] did not reliably approximate the L_∞ optimal bandwidth and the obtained bandwidths were typically too small for the unimodal function θ_1 and much too large for the bimodal function θ_2, when the confidence intervals were constructed. Neither experimenting with the grid step and range for h, nor attempts to adjust other parameters of the search algorithm, or even the algorithm itself (e.g. employing local L_∞ norms for seeking different bandwidths for different points) brought any significant improvement.

We thus illustrate the performance of the confidence intervals with subjectively selected bandwidth. For the Gaussian function θ_1 and for a typical data set, the original bootstrap confidence intervals from [3] are compared with our corrected version in Figure 6, with the bandwidth $h = 0.33$ (cf. Figure 5). Analogous comparison for the bimodal function θ_2 is given in Figure 7 with $h = 0.12$. In both cases, $\tilde{h} = 1.5h$ was used.
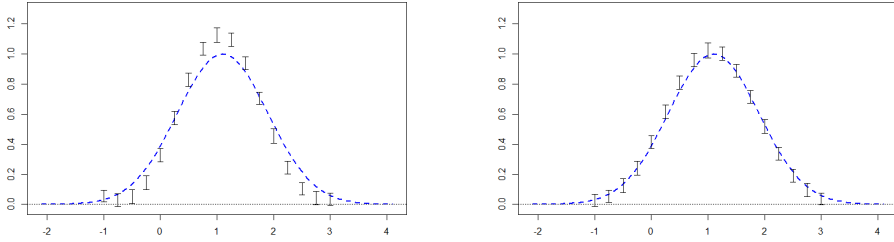
Fig. 6. Nominal 90% bootstrap confidence intervals originally constructed in [3] (left panel) and our corrected intervals (right panel) for a typical sample from the unimodal function θ_1 (dashed) with $n = 100, \sigma = 0.1, a_n = 0.25$, $h = 0.33$ and $\tilde{h} = 1.5h$
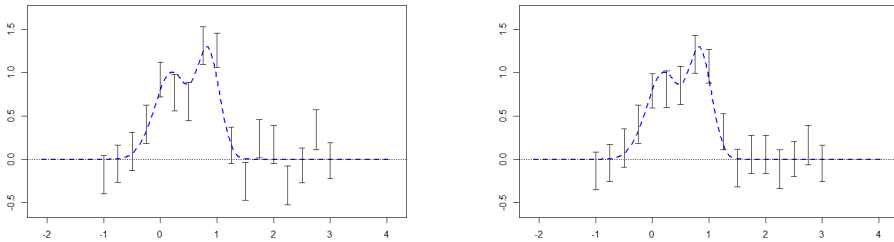

Fig. 7. Same as in Figure 6, but for the bimodal function θ_2 and with $h = 0.12$

REFERENCES

[1] P.J. Bickel, M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist. **1** (1973), 1071–1095.

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

[3] N. Bissantz, M. Birke, *Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators*, J. Multivariate Anal. **100** (2009), 2364–2375.

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

[5] N. Bissantz, H. Holzmann, K. Proksch, *Confidence regions for images observed under the Radon transform*, J. Multivariate Anal. **128** (2014), 86–107.

[6] N. Bissantz, L. Dümbgen, H. Holzman, A. Munk, *Non-parametric confidence bands in deconvolution density estimation*, J. Roy. Statist. Soc. B **69** (2007), 483–506.

[7] N. Bissantz, T. Hohage, A. Munk, F.H. Ruymgaart, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM. J. Numer. Anal. **45** (2007), 2610–2636.

[8] G. Blanchard, M. Hoffman, M. Reiss, *Optimal adaptation for early stopping in statistical inverse problems*, SIAM/ASA J. Uncertain. Quantif. **6** (2018), 1043–1075.

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

[10] B. Ćmiel, Z. Szkutnik, J. Wojdyła, *Asymptotic confidence bands in the Spektor––Lord–Willis problem via kernel estimation of intensity derivative*, Electron. J. Stat. **12** (2018), 194–223.

[11] A. Delaigle, P. Hall, F. Jamshidi, *Confidence bands in non-parametric error-in-variables regression*, J. Roy. Statist. Soc. B **77** (2015), 149–169.

[12] A.K. Dey, F.H. Ruymgaart, *Direct density estimation as an ill-posed inverse estimation problem*, Stat. Neerl. **53** (1999), 309–326.

[13] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Mathematics and Its Applications, vol. 375, Kluwer Academic, Dordrecht, 1996.

[14] E. Giné, R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York, 2016.

[15] K. Lounici, R. Nickl, *Global uniform risk bounds for wavelet deconvolution estimator*, Ann. Statist. **39** (2011), 201–231.

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

[17] D.T.P. Nguyen, D. Nuyens, *Multivariate integration over $\mathbb{R}^s$ with exponential rate of convergence*, J. Comput. Appl. Math. **315** (2016), 327–342.

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

[19] J. Wojdyła, Z. Szkutnik, *Nonparametric confidence bands in Wicksell's problem*, Statist. Sinica **28** (2018), 93–113.

Michał Biel
michal.biel@agh.edu.pl

Faculty of Applied Mathematics
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow, Poland

Zbigniew Szkutnik (corresponding author)
szkutnik@agh.edu.pl

Faculty of Applied Mathematics
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow, Poland