KU LEUVEN

FACULTEIT ECONOMIE EN BEDRIJFSWETENSCHAPPEN

Bootstrap of residual processes in regression: to smooth or not to smooth?

Neumeyer N, Van Keilegom I.

KBI_1822

Biometrika (2017), **103**, 1, *pp.* 1–16
Printed in Great Britain Advance Access publication on 31 July 2016

Bootstrap of residual processes in regression: to smooth or not to smooth ?

BY N. NEUMEYER
Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany
neumeyer@math.uni-hamburg.de

AND I. VAN KEILEGOM
ORSTAT, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
ingrid.vankeilegom@kuleuven.be

SUMMARY

In this paper we consider regression models with centered errors, independent from the covariates. Given independent and identically distributed data and given an estimator of the regression function (which can be parametric or nonparametric of nature), we estimate the distribution of the error term by the empirical distribution of estimated residuals. To approximate the distribution of this estimator, Koul and Lahiri (1994) and Neumeyer (2009) proposed bootstrap procedures based on smoothing the residuals before drawing bootstrap samples. So far it has been an open question whether a classical non-smooth residual bootstrap is asymptotically valid in this context. In this paper we solve this open problem, and show that the non-smooth residual bootstrap is consistent. We illustrate this theoretical result by means of simulations, that show the accuracy of this bootstrap procedure for various models, testing procedures and sample sizes.

Some key words: Bootstrap; Empirical distribution function; Kernel smoothing; Linear regression; Location model; Nonparametric regression.

1. INTRODUCTION

Consider the model

$$Y = m(X) + \varepsilon, \tag{1}$$

where the response Y is univariate, the covariate X is of dimension $p \geq 1$, and the error term ε is independent of X. The regression function $m(\cdot)$ can be parametric (for instance linear) or nonparametric of nature, and the distribution F of ε is completely unknown, except that $E(\varepsilon) = 0$. The estimation of the distribution F has been the object of many papers in the literature, starting with the seminal papers of Durbin (1973), Loynes (1980) and Koul (1987) in the case where $m(\cdot)$ is parametric, whereas the nonparametric case has been studied by Van Keilegom and Akritas (1999), Akritas and Van Keilegom (2001) and Müller et al. (2004), among others.

The estimator of the error distribution has been shown to be very useful for testing hypotheses regarding several features of model (1), like for instance testing for the form of the regression function $m(\cdot)$ (Van Keilegom et al., 2008), comparing regression curves (Pardo-Fernández et al., 2007), testing independence between ε and X (Einmahl and Van Keilegom, 2008, and Racine and Van Keilegom, 2017), testing for a symmetric error distribution (Koul, 2002, Neumeyer and

Dette, 2007), among others. The idea in each of these papers is to compare an estimator of the error distribution obtained under the null hypothesis with an estimator that is not based on the null. Since the asymptotic distribution of the estimator of F has a complicated covariance structure, bootstrap procedures have been proposed to approximate the distribution of the estimator and the critical values of the tests.

Koul and Lahiri (1994) proposed a residual bootstrap for linear regression models, where the bootstrap residuals are drawn from a smoothed empirical distribution of the residuals. Neumeyer (2009) proposed a similar bootstrap procedure for nonparametric regression models. The reason why a smooth bootstrap was proposed is that the methods of proof in both papers require a smooth distribution of the bootstrap error. Smooth residual bootstrap procedures have been applied by De Angelis et al. (1993), Mora (2005), Pardo-Fernández et al. (2007), and Huskova and Meintanis (2009), among many others. An alternative bootstrap procedure for nonparametric regression was considered by Neumeyer (2008), where bootstrap residuals were drawn from the non-smoothed empirical distribution of the residuals, after which smoothing is applied on the empirical distribution of the bootstrap residuals. Further it has been shown that wild bootstrap in the context of residual-based procedures can only be applied for specific testing problems as testing for a symmetric error distribution (Neumeyer et al., 2005, Neumeyer and Dette, 2007), whereas it is not valid in general (as shown in the 2006 Ruhr-Universität Bochum habilitation thesis by N. Neumeyer). It has been an open question so far whether a classical non-smooth residual bootstrap is asymptotically valid in this context. In this paper we solve this open problem, and show that the non-smooth residual bootstrap is consistent when applied to residual processes. We will do this for the case of univariate nonparametric regression with random design and for multivariate linear regression with fixed design. Other models (nonparametric regression with fixed design, nonlinear or semiparametric regression,..) can be treated similarly. The question whether smooth bootstrap procedures should be preferred over non-smooth bootstrap procedures has been discussed in different contexts, see Silverman and Young (1987) and Hall et al. (1989).

The finite sample performance of the smooth and non-smooth residual bootstrap for residual processes has been studied by Neumeyer (2009). The paper shows that for small sample sizes using the classical residual bootstrap version of the residual empirical process in the nonparametric regression context yields too small quantiles. However, as we will show in this paper, this problem is diminished for larger sample sizes and it is not very relevant when applied to testing problems.

In this paper we consider bootstrap procedures that can be applied to obtain confidence bands for the error distribution or bootstrap versions of hypotheses tests based on residual empirical processes. Those have to be distinguished from bootstrap procedures in regression models for other purposes. First, bootstrap procedures for linear models have been considered by Efron (1979), Freedman (1981), and Wu (1986), among others, and can be applied for hypotheses testing or derivation of confidence sets for the regression parameter; see also Davison and Hinkley (1997) and the references given there. Second, there is a vast literature on bootstrap confidence sets for the regression function in nonparametric models. See Härdle and Bowman (1988), Härdle and Marron (1991), Neumann and Polzehl (1998) and Claeskens and Van Keilegom (2003). Third, several authors considered bootstrap procedures applied to hypotheses testing with test statistics that directly depend on the regression estimator. Among others, Härdle and Mammen (1993), Stute et al. (1998) and Delgado and González Manteiga (2001) proved validity of bootstrap procedures in the context of specific test statistics for nonparametric regression models (not depending on residual empirical processes).

2. NONPARAMETRIC REGRESSION

We start with the case of nonparametric regression with random design. The covariate is supposed to be one-dimensional. To estimate the regression function we use a kernel estimator based on Nadaraya-Watson weights :

$$\hat{m}(x) = \sum_{i=1}^{n} \frac{k_h(x - X_i)}{\sum_{j=1}^{n} k_h(x - X_j)} Y_i,$$

where k is a kernel density function, $k_h(\cdot) = k(\cdot/h)/h$ and $h = h_n$ is a positive bandwidth sequence converging to zero when n tends to infinity. Our main result is valid under the following regularity assumptions.

(A1) The univariate covariates $X_1, \ldots, X_n$ are independent and identically distributed on a compact support, say $[0, 1]$. They have a twice continuously differentiable density f_X that is bounded away from zero. The regression function m is twice continuously differentiable in $(0, 1)$.

(A2) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with distribution function F. They are centered and are independent of the covariates. F is twice continuously differentiable with strictly positive density f such that $\sup_{y \in \mathbb{R}} f(y) < \infty$ and $\sup_{y \in \mathbb{R}} |f'(y)| < \infty$. Further, $E(|\varepsilon_1|^{\upsilon}) < \infty$ for some $\upsilon \geq 7$.

(A3) k is a twice continuously differentiable symmetric density with compact support $[-1, 1]$, say, such that $\int uk(u)\,du = 0$ and $k(-1) = k(1) = 0$. The first derivative of k is of bounded variation.

(A4) h_n is a sequence of positive bandwidths such that $h_n \sim c_n n^{-1/3+\eta}$ with $4/(3 + 9\upsilon) < \eta < 1/12$, c_n is only of logarithmic rate, and υ is defined in (A2).

Under those assumptions one in particular has that $nh_n^4 = o(1)$ and it is possible to find some $\delta \in (0, 1/2)$ with

$$\frac{nh_n^{3+2\delta}}{\log(h_n^{-1})} \to \infty. \tag{2}$$

Let residuals be defined as $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, $i = 1, \ldots, n$. In Theorem 1 in Akritas and Van Keilegom (2001) it is shown that the residual process $n^{-1/2} \sum_{i=1}^{n} \{I(\hat{\varepsilon}_i \leq y) - F(y)\}$, $y \in \mathbb{R}$, converges weakly to a zero-mean Gaussian process $W(y)$ with covariance function given by

$$\text{cov}\{W(y_1), W(y_2)\} = E\Big[\{I(\varepsilon \leq y_1) + f(y_1)\varepsilon\}\{I(\varepsilon \leq y_2) + f(y_2)\varepsilon\}\Big], \tag{3}$$

where ε has distribution function F and density f.

Neumeyer (2009) studied a smooth bootstrap procedure to approximate the distribution of this residual process, and she showed that the smooth bootstrap 'works' in the sense that the limiting distribution of the bootstrapped residual process, conditional on the data, equals the process $W(y)$ defined above, in probability. We will study an alternative bootstrap procedure that has the advantage of not requiring smoothing of the residual distribution. For $i = 1, \ldots, n$, let $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - n^{-1} \sum_{j=1}^{n} \hat{\varepsilon}_j$, and let

$$\hat{F}_{0,n}(y) = n^{-1} \sum_{i=1}^{n} I(\tilde{\varepsilon}_i \leq y)$$

be the (non-smoothed) empirical distribution of the centered residuals. Then, we randomly draw bootstrap errors $\varepsilon_{0,1}^*, \ldots, \varepsilon_{0,n}^*$ with replacement from $\hat{F}_{0,n}$. Let $Y_i^* = \hat{m}(X_i) + \varepsilon_{0,i}^*$,

$i = 1, \dots, n$, and let $\hat{m}_0^*(\cdot)$ be the same as $\hat{m}(\cdot)$, except that we use the bootstrap data $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$. Define now

$$\hat{\varepsilon}_{0,i}^* = Y_i^* - \hat{m}_0^*(X_i) = \varepsilon_{0,i}^* + \hat{m}(X_i) - \hat{m}_0^*(X_i). \tag{4}$$

We are interested in the asymptotic behavior of the process $n^{1/2}(\hat{F}_{0,n}^* - \hat{F}_{0,n})$ with

$$\hat{F}_{0,n}^*(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{0,i}^* \le y) \tag{5}$$

and we will show below that it converges to the same limiting Gaussian process as the original residual process $n^{-1/2} \sum_{i=1}^n \{I(\hat{\varepsilon}_i \le y) - F(y)\}$, $y \in \mathbb{R}$, which means that smoothing of the residuals is not necessary to obtain a consistent bootstrap procedure.

In order to prove this result, we will use the results proved in Neumeyer (2009) to show that the difference between the smooth and the non-smooth bootstrap residual process is asymptotically negligible. To this end, we can write $\varepsilon_{0,i}^* = \hat{F}_{0,n}^{-1}(U_i)$, $i = 1, \dots, n$, where $U_1, \dots, U_n$ are independent random variables from a $U[0,1]$ distribution. Strictly speaking the U_i's form a triangular array $U_{1,n}, \dots, U_{n,n}$ of $U[0,1]$ variables, but since we are only interested in convergence in distribution of the bootstrap residual process (as opposed to convergence in probability or almost surely), we can work with $U_1, \dots, U_n$ without loss of generality.

We introduce the following notations : let $\varepsilon_{s,i}^* = \hat{F}_{s,n}^{-1}(U_i)$, where $\hat{F}_{s,n}(y) = \int \hat{F}_{0,n}(y - vs_n)\, dL(v)$ is the convolution of the distribution $\hat{F}_{0,n}(y - \cdot s_n)$ and the integrated kernel $L(\cdot) = \int_{-\infty}^{\cdot} \ell(u)\, du$, where ℓ is a kernel density function and s_n is a sequence of positive bandwidths controlling the smoothness of $\hat{F}_{s,n}$ such that $s_n \to 0$ for $n \to \infty$. Then, similarly to the definition of $\hat{\varepsilon}_{0,i}^*$ in (4), we define

$$\hat{\varepsilon}_{s,0,i}^* = \varepsilon_{s,i}^* + \hat{m}(X_i) - \hat{m}_0^*(X_i). \tag{6}$$

We then decompose the bootstrap residual process as follows :

$$\begin{aligned} n^{1/2}\{\hat{F}_{0,n}^*(y) - \hat{F}_{0,n}(y)\} &= n^{-1/2} \sum_{i=1}^{n} \{I(\hat{\varepsilon}_{0,i}^* \le y) - I(\hat{\varepsilon}_{s,0,i}^* \le y)\} \\ &\quad + n^{-1/2} \sum_{i=1}^{n} \{I(\hat{\varepsilon}_{s,0,i}^* \le y) - \hat{F}_{s,n}(y)\} \\ &\quad + n^{1/2}\{\hat{F}_{s,n}(y) - \hat{F}_{0,n}(y)\} \\ &= T_{n1}(y) + T_{n2}(y) + T_{n3}(y). \end{aligned} \tag{7}$$

In Lemmas S1.1 and S1.3 in the supplement we show that under assumptions (A1)–(A4) above and conditions (C1), (C2) (given in the supplement) concerning the choice of ℓ and s_n in the proof, the terms T_{n1} and T_{n3} are asymptotically negligible. For the proof of negligibility of T_{n1} two features of our construction are of utmost importance. On the one hand, in (6) the same function $\hat{m}_0^*$ needs to be used as in (4) (in contrast to (9) below). On the other hand, the same uniform random variables U_i need to be used to generate the bootstrap errors $\varepsilon_{0,i}^*$ and $\varepsilon_{s,i}^*$, $i = 1, \dots, n$. In this way the difference between the empirical distribution functions of $\hat{\varepsilon}_{0,1}^*, \dots, \hat{\varepsilon}_{0,n}^*$ and of $\hat{\varepsilon}_{s,0,1}^*, \dots, \hat{\varepsilon}_{s,0,n}^*$ can be bounded by the difference $E_n(f_1) - E_n(f_2)$, where E_n is an asymptotically equicontinuous empirical process indexed in a function class. Negligibility of T_{n1} then follows because the distance between the indices f_1 and f_2 can be bounded by the difference between $\hat{F}_{s,n}$ and $\hat{F}_{0,n}$, which is shown to be $o_{\text{pr}}(n^{-1/2})$. From the latter fact also negligibility of T_{n3} follows. Further in Lemma S1.2 we show that the process T_{n2} is asymptotically equivalent (in terms of weak convergence) to the smooth bootstrap residual process $n^{1/2}(\hat{F}^*_{s,n} - \hat{F}_{s,n})$ with

$$\hat{F}^*_{s,n}(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}^*_{s,i} \leq y), \tag{8}$$

where, in contrast to (6),

$$\hat{\varepsilon}^*_{s,i} = \varepsilon^*_{s,i} + \hat{m}(X_i) - \hat{m}^*_s(X_i) \tag{9}$$

with $\hat{m}^*_s$ defined as $\hat{m}$, but based on smoothed bootstrap data $(X_i, \hat{m}(X_i) + \varepsilon^*_{s,i})$, $i = 1, \dots, n$. Neumeyer (2009) showed weak convergence of the residual process based on the smooth residual bootstrap, $n^{1/2}(\hat{F}^*_{s,n} - \hat{F}_{s,n})$, to the Gaussian process defined in (3). Thus the main idea to prove Lemma S1.2 is to show that the estimator $\hat{m}^*_0$ has similar asymptotic properties as $\hat{m}^*_s$, such that the use of the different estimator does not make a difference in the proofs of fidi-convergence, tightness and the calculation of the asymptotic covariance. The three lemmas lead to the following main result regarding the validity of the non-smooth bootstrap residual process.

THEOREM 1. *Assume (A1)–(A4). Then, conditionally on the data $(X_1, Y_1), \dots, (X_n, Y_n)$, the process $n^{1/2}\{\hat{F}^*_{0,n}(y) - \hat{F}_{0,n}(y)\}$, $y \in \mathbb{R}$, converges weakly to the zero-mean Gaussian process $W(y)$, $y \in \mathbb{R}$, defined in (3), in probability.*

The proof is given in section S1 of the supplement.

The result of Theorem 1 can be applied to obtain confidence bands for the error distribution. It can further be used to approximate critical values for hypotheses tests in nonparametric regression models which are based on residual empirical processes such as tests for properties of the error distribution (see for instance Neumeyer and Dette, 2007, or Einmahl and Van Keilegom, 2008), or tests concerning the regression function (see for instance Pardo-Fernández et al., 2007, or Van Keilegom et al., 2008). The application of the bootstrap procedure needs to be modified in order to obtain data that fulfil the null hypothesis; see section 4 for examples and also section 5 in Neumeyer (2009).

Remark 1. If one aims at confidence sets for the regression function, one needs different kinds of bootstrap results. To demonstrate that both the non-smooth and the smooth residual bootstrap can be applied in this context as well note that under assumptions (A1)–(A4) (for fixed $0 < x < 1$), $(nh_n)^{1/2}\{\hat{m}(x) - m(x)\}$ converges in distribution to a centered normally distributed random variable Z with variance $E(\varepsilon_1^2) \int k^2(u)\, du \,/ f_X(x)$. No bias term appears because under our assumptions $nh_n^5 \to 0$. Under the same assumptions one obtains, conditionally on the data $(X_1, Y_1), \dots, (X_n, Y_n)$, weak convergence of the non-smooth residual bootstrap version $(nh_n)^{1/2}\{\hat{m}^*_0(x) - m^*(x)\}$ to Z, in probability, if the centering term is chosen as $m^*(x) = \sum_{i=1}^n k_h(X_i - x)\hat{m}(X_i) / \sum_{j=1}^n k_h(X_j - x)$. We provide a sketch of proof in the supplement (see subsection S1.1) to obtain the result under exactly our model and our assumptions, though similar results are well known in the literature. Choosing this centering is analogous to the approach of Härdle and Bowman (1988) who obtain confidence bands for the regression function with a residual bootstrap approach in the case of fixed design points and for the Priestley-Chao regression estimator. If one wants to replace the centering $m^*(x)$ by $\hat{m}(x)$, one should use a larger pilot bandwidth for $\hat{m}$ when constructing the bootstrap observations. This was demonstrated by Härdle and Marron (1991) for a wild bootstrap (for the Nadaraya-Watson estimator and random covariates) and the same reasoning applies for residual bootstrap. Cao-Abad and Gonzalez-Manteiga (1993) obtain similar results for a bootstrap procedure with smoothing in the explanatory variable. More recently, McMurry and Politis (2008) considered an alternative way of bias-correction using infinite order kernels for the Gasser-Müller estimator in the case of fixed design points. Concerning the smooth residual bootstrap, one obtains conditional weak convergence of $(nh_n)^{1/2}\{\hat{m}_s^*(x) - m^*(x)\}$ to Z, in probability, if $s_n \to 0$ and ℓ is a symmetric density with second moments. See the supplement (subsection S1.1) for a derivation of this result.

3. LINEAR MODEL

Consider independent observations from the linear model

$$Y_{ni} = x_{ni}^{\mathrm{T}}\beta + \varepsilon_{ni}, \quad i = 1, \ldots, n, \tag{10}$$

where $\beta \in \mathbb{R}^p$ denotes the unknown parameter and the errors ε_{ni} are assumed to be independent and identically distributed with $E(\varepsilon_{ni}) = 0$ and distribution function F. Throughout this section let $X_n \in \mathbb{R}^{n\times p}$ denote the design matrix in the linear model, where the vector $x_{ni}^{\mathrm{T}} = (x_{ni1}, \ldots, x_{nip})$ corresponds to the ith row of the matrix X_n and is not random. The design matrix is assumed to be of rank $p \le n$. We use the following regularity assumptions.

(AL1) The fixed design fulfils

a. $\max_{i=1,\ldots,n} x_{ni}^{\mathrm{T}}(X_n^{\mathrm{T}}X_n)^{-1}x_{ni} = O(n^{-1})$,
b. $\lim_{n\to\infty} n^{-1}X_n^{\mathrm{T}}X_n = \Sigma \in \mathbb{R}^{p\times p}$ with invertible Σ,
c. $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n x_{ni} = m \in \mathbb{R}^p$.

(AL2) The errors ε_{ni}, $i = 1, \ldots, n$, $n \in \mathbb{N}$, are independent and identically distributed with distribution function F and density f that is strictly positive, bounded, and continuously differentiable with bounded derivative on $\mathbb{R}$. Assume $E(|\varepsilon_{11}|^{\upsilon}) < \infty$ for some $\upsilon > 3$.

We consider the least squares estimator

$$\hat{\beta}_n = (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}Y_n = \beta + (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}\varepsilon_n \tag{11}$$

with the notations $Y_n = (Y_{n1}, \ldots, Y_{nn})^T, \varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$, and define residuals $\hat{\varepsilon}_{ni} = Y_{ni} - x_{ni}^{\mathrm{T}}\hat{\beta}_n$, $i = 1, \ldots, n$. Residual processes in linear models have been extensively studied by Koul (2002). It is shown there that, under assumptions (AL1) and (AL2), the process $n^{-1/2}\sum_{i=1}^n \{I(\hat{\varepsilon}_{ni} \le y) - F(y)\}$, $y \in \mathbb{R}$, converges weakly to a zero-mean Gaussian process $W(y)$ with covariance function

$$\begin{aligned}\mathrm{cov}\{W(y_1), W(y_2)\} = F(y_1 \wedge y_2) - F(y_1)F(y_2) + m^{\mathrm{T}}\Sigma^{-1}m\Big[f(y_1)f(y_2)\mathrm{var}(\varepsilon) \\ + f(y_1)E\{I(\varepsilon \le y_2)\varepsilon\} + f(y_2)E\{I(\varepsilon \le y_1)\varepsilon\}\Big],\end{aligned} \tag{12}$$

where ε has distribution function F and density f, and m and Σ are defined in (AL1).

For the bootstrap procedure we generate $\varepsilon_{0,i}^*$, $i = 1, \ldots, n$, from the distribution function

$$\hat{F}_{0,n}(y) = n^{-1}\sum_{i=1}^n I(\tilde{\varepsilon}_{ni} \le y) \tag{13}$$

with $\tilde{\varepsilon}_{ni} = \hat{\varepsilon}_{ni} - n^{-1}\sum_{j=1}^n \hat{\varepsilon}_{nj}$, $i = 1, \ldots, n$. The centering of residuals is not necessary when the covariate includes an intercept. Note also that in the bootstrap residuals we suppress the index n to match the notation in the nonparametric case. We now define bootstrap observations by

$$Y_{ni}^* = x_{ni}^{\mathrm{T}}\hat{\beta}_n + \varepsilon_{0,i}^* \quad (i = 1, \ldots, n),$$

and calculate estimated residuals from the bootstrap sample

$$\hat{\varepsilon}^*_{0,i} = Y^*_{ni} - x^{\mathrm{T}}_{ni}\hat{\beta}^*_{0,n} = \varepsilon^*_{0,i} + x^{\mathrm{T}}_{ni}(\hat{\beta}_n - \hat{\beta}^*_{0,n}), \tag{14}$$

where $\hat{\beta}^*_{0,n}$ is the least squares estimator

$$\hat{\beta}^*_{0,n} \;=\; (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}Y_n^* \;=\; \hat{\beta}_n + (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}\varepsilon^*_{0,n} \tag{15}$$

(with the notations $Y_n^* = (Y^*_{n1}, \ldots, Y^*_{nn})^{\mathrm{T}}$, $\varepsilon^*_{0,n} = (\varepsilon^*_{0,1}, \ldots, \varepsilon^*_{0,n})^{\mathrm{T}}$). We will show that the bootstrap residual process $n^{1/2}\{\hat{F}^*_{0,n}(y) - \hat{F}_{0,n}(y)\}$, with

$$\hat{F}^*_{0,n}(y) = n^{-1}\sum_{i=1}^{n} I(\hat{\varepsilon}^*_{0,i} \le y), \tag{16}$$

converges to the same limiting process $W(y)$, $y \in \mathbb{R}$, as the original residual process $n^{-1/2}\sum_{i=1}^{n}\{I(\hat{\varepsilon}_{ni} \le y) - F(y)\}$, $y \in \mathbb{R}$. Using the representations $\varepsilon^*_{0,i} = \hat{F}^{-1}_{0,n}(U_i)$, $\varepsilon^*_{s,i} = \hat{F}^{-1}_{s,n}(U_i)$, $i = 1, \ldots, n$, where $U_1, \ldots, U_n$ are independent and $U[0,1]$-distributed and $\hat{F}_{s,n}(y) = \int \hat{F}_{0,n}(y - vs_n)\, dL(v)$ is the smoothed empirical distribution function of the residuals, we have the same decomposition (7) as in the nonparametric case. In Lemmas S2.1 and S2.3 in the Appendix we show that under assumptions (AL1), (AL2) and conditions (CL1), (CL2) on the choice of s_n and L (given in the supplement) the terms T_{n1} and T_{n3} are asymptotically negligible. The idea of the proof is the same as in the nonparametric case. In particular one needs to use the same estimator $\hat{\beta}^*_{0,n}$ in (14) and in the definition of $\hat{\varepsilon}^*_{s,0,i} = \varepsilon^*_{s,i} + x^{\mathrm{T}}_{ni}(\hat{\beta}_n - \hat{\beta}^*_{0,n})$ with smooth bootstrap errors $\varepsilon^*_{s,i}$, and one needs to use the same uniform random variables U_i to define $\varepsilon^*_{0,i}$ and $\varepsilon^*_{s,i}$, $i = 1, \ldots, n$. Further we show in Lemma S2.2 that the limiting distribution of

$$T_{n2}(y) = n^{-1/2}\sum_{i=1}^{n}\left\{I(\hat{\varepsilon}^*_{s,0,i} \le y) - \hat{F}_{s,n}(y)\right\} \tag{17}$$

is the same as the limiting distribution of $n^{1/2}\{\hat{F}^*_{s,n}(y) - \hat{F}_{s,n}(y)\}$, with

$$\hat{F}^*_{s,n}(y) = n^{-1}\sum_{i=1}^{n} I(\hat{\varepsilon}^*_{s,i} \le y) \tag{18}$$

and $\hat{\varepsilon}^*_{s,i} = \varepsilon^*_{s,i} + x^{\mathrm{T}}_{ni}(\hat{\beta}_n - \hat{\beta}^*_{s,n})$, where $\hat{\beta}^*_{s,n} = \hat{\beta}_n + (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}\varepsilon^*_{s,n}$ and $\varepsilon^*_{s,n} = (\varepsilon^*_{s,1}, \ldots, \varepsilon^*_{s,n})^{\mathrm{T}}$. To show this we apply results from Koul and Lahiri (1994) and demonstrate that the use of the estimator $\hat{\beta}^*_{0,n}$ in the definition of $\hat{\varepsilon}^*_{s,0,i}$ (instead of $\hat{\beta}^*_{s,n}$ in the definition of $\hat{\varepsilon}^*_{s,i}$) does not change the asymptotic distribution. In this way we obtain the validity of the classical residual bootstrap.

THEOREM 2. *Assume (AL1), (AL2). Then, conditionally on the data $Y_{1n}, \ldots, Y_{nn}$, the process $n^{1/2}\{\hat{F}^*_{0,n}(y) - \hat{F}_{0,n}(y)\}$, $y \in \mathbb{R}$, converges weakly to the zero-mean Gaussian process $W(y)$, $y \in \mathbb{R}$, defined in* (12)*, in probability.*

The proof is given in section S2 of the supplement.

The result can be applied to obtain confidence bands for the error distribution or for hypotheses testing with procedures based on the residual empirical process, see for instance Koul (2002) or Neumeyer et al. (2007).

Remark 2. The same bootstrap procedures can be applied to mimic the distribution of $n^{1/2}(\hat\beta_n - \beta)$, which converges to a p-dimensional centered normal random variable Z with covariance matrix $E(\varepsilon_{11}^2)\Sigma^{-1}$. Freedman (1981) showed that, along almost all sequences $Y_{1n}, \ldots, Y_{nn}$, $n^{1/2}(\hat\beta^*_{0,n} - \hat\beta_n)$ converges in distribution to Z. His result holds under our (stronger) assumptions (AL1), (AL2). Thus the asymptotic distribution of $n^{1/2}(\hat\beta_n - \beta)$ is mimicked by the non-smooth residual bootstrap. Freedman (1981) further demonstrated that the residual bootstrap may fail if the residuals are not centered (see (13)). Concerning the smooth residual bootstrap, one obtains conditional weak convergence of $n^{1/2}(\hat\beta^*_{s,n} - \hat\beta_n)$ to Z, in probability, if $s_n \to 0$ and ℓ is a symmetric density with second moments. We will demonstrate this in the supplement (see subsection S2.1).

4. SIMULATIONS

In this section we will study the behavior of the smooth and the non-smooth residual bootstrap for a range of models, sample sizes and contexts. We start with an empirical study to assess the quality of bootstrap confidence bands for the error distribution. Next, we will consider the use of the residual bootstrap for approximating the critical values in two representative examples of hypothesis testing procedures based on residual processes: the case of testing for a symmetric error distribution, and the case of goodness-of-fit tests for the regression function.

4·1. *Confidence bands*

Consider model (1) in the nonparametric case and generate data with $m(x) = 2x$, where X follows a uniform distribution on $[0, 1]$, and $\varepsilon \sim N(0, 0{\cdot}25^2)$. In order to assess the quality of the smooth and the non-smooth bootstrap approximation, we calculate confidence bands for the error distribution $F(\cdot)$ by means of the two bootstraps. The bands are defined as $\hat F_{0,n}(\cdot) \pm d_{\alpha,n,0}$ for the non-smooth bootstrap, and $\hat F_{0,n}(\cdot) \pm d_{\alpha,n,s}$ for the smooth bootstrap, where $d_{\alpha,n,0}$ and $d_{\alpha,n,s}$ are respectively the quantiles of level $1-\alpha$ of the distribution of $\max_{1\le i\le n} |\hat F^*_{0,n}(\hat\varepsilon^*_{0,i}) - \hat F_{0,n}(\hat\varepsilon^*_{0,i})|$ and $\max_{1\le i\le n} |\hat F^*_{s,n}(\hat\varepsilon^*_{s,i}) - \hat F_{s,n}(\hat\varepsilon^*_{s,i})|$.

In order to verify whether the bootstrap approximation works well, we calculate the coverage and the average width of the confidence bands for several values of the sample size n and the confidence level $1-\alpha$. The results are in Table 1 and are based on 1000 simulation runs, and for each simulation 1000 bootstrap samples are generated. The bandwidth h_n is taken equal to $h_n = \hat\sigma_X n^{-0{\cdot}3}$ with $\hat\sigma_X$ the empirical standard deviation of $X_1, \ldots, X_n$, and the kernel k is the Gaussian kernel. For the smooth bootstrap, bootstrap errors $\varepsilon^*_{s,i}$ are generated from $\hat F_{s,n}(y) = \int \hat F_{0,n}(y - vs_n)\, dL(v)$, where L is the distribution of a standard normal and s_n is chosen by means of the cross-validation procedure proposed by Li et al. (2017). The latter paper studies the estimation of distribution functions by applying kernel smoothing on the empirical distribution, exactly in the same way as we do for obtaining our estimator $\hat F_{s,n}(\cdot)$. The bandwidth selector is included in the package 'np' in R, and is obtained from the function 'npudistbw'. This bandwidth satisfies regularity condition (CL2) in the supplement imposed on s_n, thanks to Theorem 3.2 in the latter paper.

The table shows that the smooth bootstrap leads to too small coverages, and the coverage does not improve when n increases. On the other hand, the non-smooth bootstrap leads to too large coverage probabilities for small values of n, but the coverage is close to the nominal level $1-\alpha$ when n equals 1000. So, the smooth bootstrap is anti-conservative and the non-smooth bootstrap tends to be conservative in this situation. A natural consequence of this tendency to under- and overestimate the coverage probability, is that the average width of the confidence bands obtained with the smooth bootstrap is smaller than of those obtained with the non-smooth bootstrap, and the difference in width is of the order of $10 - 20\%$.

n	Bootstrap	Coverage			Average width		
		90	95	99	90	95	99
50	Smooth	86·3	92·7	99·0	23·6	25·8	30·2
	Non-smooth	97·6	99·4	100	28·1	30·3	35·4
100	Smooth	86·6	92·7	98·9	17·0	18·5	21·6
	Non-smooth	96·4	98·8	998·	19·8	21·5	24·9
200	Smooth	86·9	93·1	98·2	12·2	13·3	15·4
	Non-smooth	95·8	97·8	99·8	13·9	15·1	17·4
500	Smooth	86·3	92·6	98·2	7·8	8·5	9·8
	Non-smooth	93·4	98·0	99·7	8·7	9·4	10·8
1000	Smooth	85·1	91·1	96·9	5·5	6·0	7·0
	Non-smooth	92·2	95·3	98·8	6·1	6·6	7·6

Table 1. *Coverage (in* $\%$*) and average width of confidence bands for* $F(\cdot)$ *for the smooth and non-smooth bootstrap in the nonparametric model and for* $1-\alpha = 0{\cdot}90, 0{\cdot}95$ *and* $0{\cdot}99$.

4·2. *Testing for a symmetric error distribution*

As already mentioned before, the residual bootstrap is very much used in hypothesis testing regarding various aspects of model (1). As a first illustration we consider a test for the symmetry of the error density in a linear regression model with fixed design. More precisely, consider the model $Y_{ni} = x_{ni}^{\mathrm{T}}\beta + \varepsilon_{ni}$, where $E(\varepsilon_{ni}) = 0$, and suppose we are interested in testing the following hypothesis regarding the distribution F of ε_{ni} :

$$H_0 : F(t) = 1 - F(-t) \text{ for all } t \in \mathbb{R}.$$

When the design is fixed and the regression function is linear, Koul (2002) considered a test for H_0 based on the residual process

$$\hat{F}_{0,n}(\cdot) - \hat{F}_{-0,n}(\cdot) = n^{-1}\sum_{i=1}^{n}\big\{I(\hat{\varepsilon}_{ni} \le \cdot) - I(-\hat{\varepsilon}_{ni} \le \cdot)\big\},$$

where $\hat{F}_{-0,n}$ is the empirical distribution of $-\hat{\varepsilon}_{n1}, \ldots, -\hat{\varepsilon}_{nn}$, and $\hat{\varepsilon}_{ni} = Y_{ni} - x_{ni}^{\mathrm{T}}\hat{\beta}_n$. Natural test statistics are the Kolmogorov-Smirnov and the Crámer-von Mises statistics :

$$T_{KS} = \sup_y |\hat{F}_{0,n}(y) - \hat{F}_{-0,n}(y)|, \quad T_{CM} = \int \big\{\hat{F}_{0,n}(y) - \hat{F}_{-0,n}(y)\big\}^2 d\hat{F}_{0,n}(y).$$

It is clear from the covariance function given in (12) that their asymptotic distribution is not easy to approximate, and that the residual bootstrap offers a valid alternative. We will compare the level and power of the two tests, using the smooth and the non-smooth bootstrap. The bootstrapped versions of T_{CM} (and similarly for T_{KS}) are given by

$$T^*_{CM,0} = n\int \big\{\hat{F}^*_{0,n}(y) - \hat{F}^*_{-0,n}(y)\big\}^2 d\hat{F}^*_{0,n}(y), \quad T^*_{CM,s} = n\int \big\{\hat{F}^*_{s,n}(y) - \hat{F}^*_{-s,n}(y)\big\}^2 d\hat{F}^*_{s,n}(y),$$

where for the non-smooth bootstrap, bootstrap errors $\varepsilon^*_{0,1}, \ldots, \varepsilon^*_{0,n}$ are drawn from $\{\hat{F}_{0,n}(\cdot) + \hat{F}_{-0,n}(\cdot)\}/2$, which is by construction a symmetric distribution, and for the smooth bootstrap we smooth this distribution using a Gaussian kernel and by choosing s_n by means of cross-validation as in the previous simulation study. The estimators $\hat{F}^*_{0,n}(\cdot)$ and $\hat{F}^*_{s,n}(\cdot)$ are defined as in (16) and (18), and $\hat{F}^*_{-0,n}(\cdot)$ and $\hat{F}^*_{-s,n}(\cdot)$ are defined accordingly. Finally, we reject H_0 if the observed value of T_{CM} exceeds the quantile of level $1-\alpha$ of the distribution of $T^*_{CM,0}$ or $T^*_{CM,s}$.

n	Test	$d=0$			$d=2$			$d=4$		
		2·5	5·0	10·0	2·5	5·0	10·0	2·5	5·0	10·0
50	$T^*_{KS,s}$	2·5	5·3	9·8	6·1	9·5	18·1	16·5	22·3	36·7
	$T^*_{KS,0}$	2·0	4·2	8·3	6·4	9·5	18·1	16·9	23·6	36·8
	$T^*_{CM,s}$	2·3	4·6	9·4	6·9	11·8	20·0	19·8	29·3	41·5
	$T^*_{CM,0}$	2·2	3·9	8·5	6·8	12·2	20·0	19·9	29·6	41·3
100	$T^*_{KS,s}$	1·8	4·7	9·4	11·2	16·8	27·0	36·3	46·1	61·1
	$T^*_{KS,0}$	1·5	3·9	8·1	11·6	18·2	27·3	38·0	48·0	61·3
	$T^*_{CM,s}$	1·9	4·1	8·7	13·3	21·7	31·2	43·3	56·5	67·3
	$T^*_{CM,0}$	1·8	3·9	8·1	13·0	20·6	30·3	42·2	55·5	66·5
200	$T^*_{KS,s}$	2·4	5·1	10·3	21·5	29·3	43·4	64·6	74·0	85·0
	$T^*_{KS,0}$	1·8	4·3	8·8	22·3	30·0	43·1	65·4	74·9	85·1
	$T^*_{CM,s}$	2·1	4·6	9·4	27·7	36·5	48·1	77·1	85·2	91·4
	$T^*_{CM,0}$	1·7	4·4	9·0	27·1	36·4	48·6	77·1	85·4	91·6
500	$T^*_{KS,s}$	3·0	5·8	10·9	50·5	61·9	73·8	97·7	98·9	99·8
	$T^*_{KS,0}$	2·7	5·0	9·9	51·2	61·8	73·7	97·7	98·9	99·8
	$T^*_{CM,s}$	2·8	5·0	10·7	60·5	72·2	80·7	99·7	100	100
	$T^*_{CM,0}$	2·7	5·0	10·3	60·0	71·3	80·8	99·6	100	100
1000	$T^*_{KS,s}$	2·8	4·7	10·4	81·1	89·8	94·5	100	100	100
	$T^*_{KS,0}$	2·5	4·4	9·2	81·6	90·1	94·5	100	100	100
	$T^*_{CM,s}$	2·4	4·7	9·4	89·8	94·9	97·6	100	100	100
	$T^*_{CM,0}$	2·3	4·5	9·2	90·5	94·9	97·6	100	100	100

Table 2. *Rejection probabilities (in %) of the test for symmetry in the linear model for several sample sizes n and for $\alpha = 0{\cdot}025, 0{\cdot}05$ and $0{\cdot}1$. Under the null we have a normal distribution ($d=0$), whereas under the alternative we have a skew normal distribution ($d=2$ and $d=4$). The power is obtained after calibrating the test in such a way that the size equals α, in order to ease comparison between the two bootstrap methods.*

Consider the model $Y_{ni} = 2x_{ni} + \varepsilon_{ni}$, where $x_{ni} = i/n$. We consider two error distributions under H_0. The first one is a normal distribution with mean zero and variance $0{\cdot}25^2$. Under the alternative we consider the skew normal distribution of Azzalini (1985), whose density is given by $2\phi(y)\Phi(dy)$, where ϕ and Φ are the density and distribution of the standard normal. More precisely, we let $d=2$ and $d=4$ and standardize these skew normal distributions so that they have mean zero and variance $0{\cdot}25^2$. When $d=0$ we find back the normal distribution. The second error distribution under H_0 is a Student-t distribution with 3 degrees of freedom, standardized in such a way that the variance equals $0{\cdot}25^2$. The asymptotic theory does not cover this case, but we like to know how sensitive the bootstrap methods are to the existence of moments of higher order. Under the alternative we consider a mixture of this Student-t distribution and a standard Gumbel distribution, again standardized to have mean zero and variance $0{\cdot}25^2$. The mixture proportions p are 1 (corresponding to H_0), $0{\cdot}75$ and $0{\cdot}50$.

The results, shown in Tables 2 and 3, are based on 2000 simulation runs, and for each simulated sample a total of 2000 bootstrap samples are generated. The power is obtained after calibrating the test in such a way that the size equals exactly $1-\alpha$, in order to ease comparison between the two bootstrap methods. The tables show that the Crámer-von Mises test outperforms the Kolmogorov-Smirnov test, and hence we focus on the former test. Table 2 shows that for the normal error distribution, the size is about right for the smooth bootstrap and a little bit too low for the non-smooth bootstrap. After correcting the critical value of the tests such that the rejection probabilities under H_0 equal α, we see that the smooth and the non-smooth bootstrap have almost identical power. The tabulated powers have a standard deviation of $\{\text{power}(1-\text{power})/2000\}^{1/2}$, which is bounded above by $1{\cdot}1\%$. Taking this standard deviation into account, we can conclude that there is no significant difference between the powers of the two tests.

Table 3 shows that when the distribution under H_0 is a Student-t distribution with 3 degrees of freedom, the size is in general too large for the smooth bootstrap, especially for small n, and about right for the non-smooth bootstrap. The level-adjusted power is again very similar for both types of bootstrap. However, the non-smooth bootstrap has the advantage that it does not depend on the selection of a bandwidth parameter and that it tends to be conservative in certain situations, whereas the smooth bootstrap has a tendency to be anti-conservative.

4·3. *Goodness-of-fit tests*

We end this section with a second application of the residual bootstrap, which concerns the use of residual processes and the residual bootstrap for testing the fit of a parametric model for the regression function m :

$$H_0 : m \in \mathcal{M} = \{m_\theta : \theta \in \Theta\},$$

where $\mathcal{M}$ is a class of parametric regression functions depending on a k-dimensional parameter space Θ. Van Keilegom et al. (2008) showed that testing H_0 is equivalent to testing whether the error distribution satisfies $F \equiv F_0$, where F_0 is the distribution of $Y - m_{\theta_0}(X)$ and θ_0 is the value of θ that minimizes $E[\{m(X) - m_\theta(X)\}^2]$. Consider the following Kolmogorov-Smirnov and Crámer-von Mises type test statistics :

$$T_{KS} = n^{1/2} \sup_y |\hat{F}_{0,n}(y) - \hat{F}_{\hat{\theta}}(y)|, \qquad T_{CM} = n \int \left\{\hat{F}_{0,n}(y) - \hat{F}_{\hat{\theta}}(y)\right\}^2 d\hat{F}_{\hat{\theta}}(y),$$

where $\hat{F}_{0,n}$ is as defined in Section 2 and $\hat{F}_\theta(y) = n^{-1}\sum_{i=1}^n I\{Y_i - \hat{m}_\theta(X_i) \le y\}$ with

$$\hat{m}_\theta(x) = \sum_{i=1}^n \frac{k_h(x - X_i)}{\sum_{j=1}^n k_h(x - X_j)} m_\theta(X_i)$$

for any θ, and $\hat{\theta}$ is the least squares estimator of θ. The critical values of these test statistics are approximated using our smooth and non-smooth residual bootstrap. More precisely, the bootstrapped versions of T_{CM} (and similarly for T_{KS}) are given by

$$T^*_{CM,0} = n \int \left\{\hat{F}^*_{0,n}(y) - \hat{F}^*_{0,\hat{\theta}^*_0}(y)\right\}^2 d\hat{F}^*_{0,\hat{\theta}^*_0}(y),$$

and

$$T^*_{CM,s} = n \int \left\{\hat{F}^*_{s,n}(y) - \hat{F}^*_{s,\hat{\theta}^*_s}(y)\right\}^2 d\hat{F}^*_{s,\hat{\theta}^*_s}(y),$$

n	Test	$p = 1$			$p = 0{\cdot}75$			$p = 0{\cdot}50$		
		2·5	5·0	10·0	2·5	5·0	10·0	2·5	5·0	10·0
50	$T^*_{KS,s}$	3·7	7·9	13·8	3·4	5·8	10·5	4·9	8·6	14·9
	$T^*_{KS,0}$	2·6	5·6	11·2	3·5	6·5	10·7	5·4	9·5	16·0
	$T^*_{CM,s}$	3·1	6·8	13·3	4·1	7·6	12·1	7·6	11·7	17·7
	$T^*_{CM,0}$	2·0	4·8	11·1	4·8	7·7	12·0	8·4	12·6	18·8
100	$T^*_{KS,s}$	3·7	7·2	12·9	3·9	6·9	13·0	8·4	13·5	23·7
	$T^*_{KS,0}$	2·7	5·5	10·0	3·6	7·0	13·3	8·2	13·6	23·9
	$T^*_{CM,s}$	3·9	6·5	13·0	4·1	7·6	13·4	9·9	17·0	27·0
	$T^*_{CM,0}$	2·8	5·3	10·7	4·2	7·5	13·6	10·0	17·2	27·6
200	$T^*_{KS,s}$	2·5	5·5	11·5	6·7	10·5	17·3	17·0	25·6	35·8
	$T^*_{KS,0}$	1·8	4·3	9·2	6·6	10·8	17·0	17·0	26·2	35·9
	$T^*_{CM,s}$	2·7	5·6	11·0	7·5	11·9	19·1	22·0	30·6	43·0
	$T^*_{CM,0}$	2·1	4·3	9·5	7·7	12·4	19·0	22·3	31·3	43·1
500	$T^*_{KS,s}$	2·5	4·8	10·7	10·7	17·9	25·3	42·2	54·6	64·1
	$T^*_{KS,0}$	2·2	4·2	9·7	10·5	17·4	25·5	42·1	54·3	64·3
	$T^*_{CM,s}$	2·4	4·9	10·7	14·2	20·4	29·5	52·2	62·4	71·1
	$T^*_{CM,0}$	2·1	4·3	9·9	13·9	20·5	30·0	52·2	62·6	71·9
1000	$T^*_{KS,s}$	2·6	5·6	10·5	18·9	27·3	40·0	75·7	83·9	90·6
	$T^*_{KS,0}$	2·0	4·8	9·5	19·2	27·1	40·2	76·1	83·0	90·6
	$T^*_{CM,s}$	2·5	5·0	10·9	23·7	34·0	44·2	83·7	89·4	93·8
	$T^*_{CM,0}$	2·3	4·8	9·9	24·0	33·8	44·9	84·3	89·4	93·9

Table 3. *Rejection probabilities (in %) of the test for symmetry in the linear model for several sample sizes n and for $\alpha = 0{\cdot}025, 0{\cdot}05$ and $0{\cdot}1$. Under the null we have a Student-t distribution with 3 degrees of freedom ($p = 1$), whereas under the alternative we have a mixture of a Student-t (3) and a Gumbel distribution ($p = 0{\cdot}75$ and $p = 0{\cdot}50$). The power is obtained after calibrating the test in such a way that the size equals α, in order to ease comparison between the two bootstrap methods.*

where $\hat\theta_0^*$ is the least squares estimator based on the bootstrap data $(X_i, Y_{0,i}^* = m_{\hat\theta}(X_i) + \varepsilon_{0,i}^*)$, $i = 1, \ldots, n$, $\hat F_{0,\theta}^*(y) = n^{-1}\sum_{i=1}^n I\{Y_{0,i}^* - \hat m_\theta(X_i) \le y\}$ for any θ, and similarly for $\hat\theta_s^*$ and $\hat F_{s,\theta}^*(y)$. We reject H_0 if the observed value of T_{CM} exceeds the quantile of level $1-\alpha$ of the distribution of $T_{CM,0}^*$ or $T_{CM,s}^*$.

We consider the model $m(x) = 2x$ and let $\mathcal{M} = \{x \to \theta x : \theta \in \Theta\}$, i.e. the null model is a linear model without intercept. The error term ε follows either a normal distribution or a Student-t distribution with 3 degrees of freedom, in both cases standardized in such a way that the variance equals $0{\cdot}25^2$. The covariate X has a uniform distribution on $[0, 1]$. The bandwidth h_n is taken equal to $h_n = \hat\sigma_X n^{-0{\cdot}3}$, and the kernel k is the Gaussian kernel. For the smooth bootstrap, we use a standard normal distribution and we select s_n via cross-validation as in the previous simulations. Under the alternative we consider the model $m(x) = 2x + ax^2$ for $a = 0{\cdot}25$ and $0{\cdot}5$. The rejection probabilities, given in Tables 4 and 5, are based on 2000 simulation runs, and for each simulation 2000 bootstrap samples are generated.

The tables show that the Crámer-von Mises test outperforms the Kolmogorov-Smirnov test, independently of the sample size, the type of bootstrap and the value of a (corresponding to the

n	Test	$a=0$			$a=0{\cdot}25$			$a=0{\cdot}5$		
		2·5	5·0	10·0	2·5	5·0	10·0	2·5	5·0	10·0
50	$T^*_{KS,s}$	3·9	6·3	13·3	3·3	6·4	11·6	9·3	15·2	24·1
	$T^*_{KS,0}$	2·3	4·8	10·7	3·2	6·9	12·5	10·1	16·3	25·8
	$T^*_{CM,s}$	2·4	4·9	9·7	5·6	10·1	17·9	16·8	24·8	36·6
	$T^*_{CM,0}$	1·5	3·8	8·0	5·8	11·0	18·1	17·5	25·7	36·7
100	$T^*_{KS,s}$	2·1	4·0	9·2	8·1	12·6	19·5	25·4	33·8	45·5
	$T^*_{KS,0}$	1·3	2·7	7·1	8·1	12·6	19·2	26·3	34·0	46·4
	$T^*_{CM,s}$	2·0	4·5	9·0	10·3	16·1	25·7	34·2	43·8	58·3
	$T^*_{CM,0}$	1·5	3·5	7·9	10·4	16·1	26·2	34·7	43·9	59·2
200	$T^*_{KS,s}$	2·1	4·5	9·3	12·5	20·9	32·1	48·9	60·9	73·7
	$T^*_{KS,0}$	1·5	3·3	7·6	14·1	20·9	32·3	52·2	61·5	74·2
	$T^*_{CM,s}$	2·2	4·2	8·5	19·2	29·8	42·5	63·6	74·9	84·3
	$T^*_{CM,0}$	1·7	3·6	7·3	19·8	29·3	43·0	64·9	74·8	84·8
500	$T^*_{KS,s}$	2·0	4·5	9·4	35·2	44·8	58·6	92·0	95·2	97·8
	$T^*_{KS,0}$	1·5	3·4	7·7	36·6	45·0	58·4	92·6	95·6	97·8
	$T^*_{CM,s}$	1·9	4·1	8·2	45·4	56·8	68·2	96·3	98·0	98·9
	$T^*_{CM,0}$	1·8	3·8	7·4	45·0	57·0	68·0	96·2	98·0	98·9
1000	$T^*_{KS,s}$	2·2	4·2	9·6	65·4	76·9	84·9	100	100	100
	$T^*_{KS,0}$	1·8	3·6	8·4	65·2	77·0	84·9	100	100	100
	$T^*_{CM,s}$	2·0	4·6	9·3	72·9	81·6	88·6	100	100	100
	$T^*_{CM,0}$	1·9	4·2	8·9	72·9	81·6	88·6	100	100	100

Table 4. *Rejection probabilities (in* $\%$*) of the goodness-of-fit test for several sample sizes and for* $\alpha = 0{\cdot}025, 0{\cdot}05$ *and* $0{\cdot}1$*, when the error term has a normal distribution. The regression function is* $m(x) = 2x + ax^2$ *and the null hypothesis corresponds to* $a = 0$*. The power is obtained after calibrating the test in such a way that the size equals* α*, in order to ease comparison between the two bootstrap methods.*

null or the alternative). Hence, we focus on the Crámer-von Mises test. Table 4 shows that when the error term has a normal distribution, both the smooth and the non-smooth bootstrap lead to conservative tests, although rejection probabilities are closer to the nominal level for the smooth bootstrap. After adjusting the critical value in such a way that the size equals α, we see that the size-adjusted powers under the two types of bootstrap are almost identical. A simple test of the equality of two proportions shows that there is no significant difference. When the error has a Student-t distribution (Table 5) the size is too large for the smooth bootstrap, and more or less equal to the nominal level for the non-smooth bootstrap. The size-adjusted powers are again very close.

Overall, when comparing the smooth bootstrap with the non-smooth bootstrap in the five contexts we have explored and for the different sample sizes and values of α we have considered, we can conclude that the smooth bootstrap has a tendency to be anti-conservative whereas the non-smooth bootstrap has a tendency to be conservative. As conservative tests are in general preferred over non-conservative tests (if we follow the general idea that we prefer to control the probability of committing a type I error), we conclude that the non-smooth bootstrap is preferable over the smooth bootstrap (for the testing problems under consideration and for confidence

n	Test	$a = 0$			$a = 0{\cdot}25$			$a = 0{\cdot}5$		
		2·5	5·0	10·0	2·5	5·0	10·0	2·5	5·0	10·0
50	$T^*_{KS,s}$	4·3	7·3	15·7	5·3	9·4	16·3	15·5	23·5	34·9
	$T^*_{KS,0}$	3·0	5·6	11·9	4·8	10·0	16·7	14·9	24·6	36·6
	$T^*_{CM,s}$	3·2	5·9	12·2	8·8	14·8	22·0	24·5	35·2	45·9
	$T^*_{CM,0}$	2·4	4·7	10·0	8·7	14·4	22·7	24·4	35·6	47·0
100	$T^*_{KS,s}$	4·2	7·7	13·2	10·1	15·7	24·2	31·3	41·7	54·9
	$T^*_{KS,0}$	2·9	5·8	10·8	10·5	15·8	24·5	32·3	42·4	55·5
	$T^*_{CM,s}$	3·4	6·4	12·9	14·2	21·4	29·9	42·5	53·0	64·2
	$T^*_{CM,0}$	2·5	5·2	10·8	14·7	21·5	30·0	43·1	53·1	64·4
200	$T^*_{KS,s}$	3·4	6·1	12·2	18·0	25·8	38·9	57·8	69·1	79·4
	$T^*_{KS,0}$	2·5	4·7	9·4	18·7	26·5	38·9	59·6	70·0	79·5
	$T^*_{CM,s}$	3·3	5·4	10·8	23·1	34·2	45·2	68·5	79·0	86·3
	$T^*_{CM,0}$	2·6	4·7	9·5	22·6	33·8	45·6	68·6	78·7	86·4
500	$T^*_{KS,s}$	3·6	6·8	12·6	34·6	48·6	61·8	92·2	95·7	97·7
	$T^*_{KS,0}$	3·1	5·5	10·9	33·8	48·2	62·4	91·7	95·6	97·8
	$T^*_{CM,s}$	3·4	6·2	11·9	44·0	54·9	67·4	95·3	97·1	98·4
	$T^*_{CM,0}$	3·1	5·8	10·7	43·8	55·0	67·6	95·2	97·1	98·5
1000	$T^*_{KS,s}$	3·0	6·4	12·0	67·2	78·5	86·0	99·4	99·7	99·8
	$T^*_{KS,0}$	2·4	5·6	11·0	69·2	78·5	85·9	99·5	99·7	99·8
	$T^*_{CM,s}$	3·2	6·2	11·5	73·8	82·6	89·8	99·6	99·8	99·9
	$T^*_{CM,0}$	3·0	5·5	10·6	73·9	82·6	89·7	99·6	99·8	99·9

Table 5. *Rejection probabilities (in %) of the goodness-of-fit test for several sample sizes and for* $\alpha = 0{\cdot}025, 0{\cdot}05$ *and* $0{\cdot}1$*, when the error term has a Student-t distribution with 3 degrees of freedom. The regression function is* $m(x) = 2x + ax^2$ *and the null hypothesis corresponds to* $a = 0$*. The power is obtained after calibrating the test in such a way that the size equals* α*, in order to ease comparison between the two bootstrap methods.*

bands for the error distribution). In addition, the smooth bootstrap depends on how we choose the smoothing parameter s_n (and the corresponding kernel), which is not the case with the non-smooth bootstrap.

5. DISCUSSION

In this paper we considered a regression model with parametric or fully nonparametric regression function, and with independence between the error term and the vector of covariates. We proposed a novel bootstrap procedure that can be used to approximate the distribution of an estimator of the error distribution under this model. The proposed bootstrap does not involve any smoothing, and hence it avoids the delicate choice of a bandwidth parameter, which existing bootstrap procedures suffer from. We showed the consistency of this bootstrap procedure, and applied it in three situations.

This paper is the first paper that studies the non-smoothed bootstrap of the error distribution. We restricted attention to the case of homoscedastic regression models, but in a second step it would be interesting to prove the consistency of the bootstrap when the error variance depends on one or several covariates. Other possible extensions of the model considered in this paper include the extension to semiparametric regression, nonparametric regression with more than one covariate, dependent data, missing or censored data, etc.

Acknowledgement

We are thankful to an associate editor and two referees for their constructive comments on the first version of the manuscript. N. Neumeyer acknowledges financial support from the DFG (Research Unit *Structural Inference in Statistics: Adaptation and Efficiency*). I. Van Keilegom acknowledges financial support from the European Research Council (Horizon 2020 / ERC grant), and from the Belgian government (Belgian Science Policy, IAP research network grant).

Supplementary material

Supplementary material available at *Biometrika* online includes the proofs of Theorems 1 and 2 and derivations concerning Remarks 1 and 2.

References

Akritas, M. G. & Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

Cao-Abad, R. & González Manteiga(1993). Bootstrap methods in regression smoothing. *J. Nonparametr. Statist.* **2**, 379–388.

Claeskens, G. & Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–1884.

Davison, A.C. & Hinkley, D.V. (1997). *Bootstrap methods and their applications*. Cambridge University Press.

Delgado, M.A. & González Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29**, 1469–1507.

De Angelis, D., Hall, P. & Young, G. A. (1993). Analytical and bootstrap approximations to estimator distributions in L^1 regression. *J. Amer. Statist. Assoc.* **88**, 1310–1316.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–26.

Einmahl, J. & Van Keilegom, I. (2008). Specification tests in nonparametric regression. *J. Econometrics* **143**, 88–102.

Freedman, D.A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218–1228.

Härdle, W. & Bowman, A.W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102–110.

Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–1947.

Härdle, W. & Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796.

Hall, P., DiCiccio, T.J. & Romano, J.P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17**, 692–704.

Huskova, M. & Meintanis, S.G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. *Kybernetika* **45**, 960–971.

Koul, H.L. (1987). Tests of goodness-of-fit in linear regression. Goodness-of-fit (Debrecen, 1984) *Colloq. Math. Soc. Janos Bolyai* **45**, 279–315, North-Holland, Amsterdam.

Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. Springer, New York, second ed.

Koul, H.L. & Lahiri, S.N. (1994). On bootstrapping M-estimated residual processes in multiple linear regression models. *J. Multiv. Anal.* **49**, 255–265.

Li, C., Li, H. & Racine, J.S. (2017). Cross-validated mixed-datatype bandwidth selection for nonparametric cumulative distribution/survivor functions. *Econometric Reviews* **36**, 970–987.

Loynes, R.M. (1980). The empirical sample distribution function of residuals from generalized regression. *Ann. Statist.* **8**, 285–298.

McMurry, T.L. & Politis, D.N (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statist. Probab. Lett.* **78**, 2463–2469.

MORA, J. (2005). Comparing distribution functions of errors in linear models: a nonparametric approach. *Statist. Probab. Lett.* **73**, 425–432.

MÜLLER, U.U., SCHICK, A. & WEFELMEYER, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Planning Inf.* **119**, 75–93.

NEUMANN, M.H. & POLZEHL, J. (1998). Simultaneous bootstrap confidence bands in nonparametric represson. *J. Nonparametr. Statist.* **9**, 307–333.

NEUMEYER, N. (2008). A bootstrap version of the residual-based smooth empirical distribution function. *J. Nonparametr. Statist.* **20**, 153–174.

NEUMEYER, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scand. J. Statist.* **36**, 204–228.

NEUMEYER, N., DETTE, H. & NAGEL, E.-R. (2005). A note on testing symmetry of the error distribution in linear regression models. *J. Nonparametr. Stat.* **17**, 697–715.

NEUMEYER, N. & DETTE, H. (2007). Testing for symmetric error distribution in nonparametric regression models. *Statist. Sinica* **17**, 775–795.

PARDO-FERNÁNDEZ, J.C., VAN KEILEGOM, I. & GONZÁLEZ-MANTEIGA, W. (2007). Testing for the equality of k regression curves. *Statist. Sinica* **17**, 1115–1137.

RACINE, J. & VAN KEILEGOM, I. (2017). A smooth nonparametric, multivariate, mixed-data location-scale test (submitted).

SILVERMAN, B.W. & YOUNG, G.A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74**, 469–479.

STUTE, W., GONZÁLEZ MANTEIGA, W. & PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statistic. Assoc.* **93**, 141–149.

VAN KEILEGOM, I. & AKRITAS, M.G (1999). Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745–1784.

VAN KEILEGOM, I., GONZÁLEZ-MANTEIGA, W. & SÁNCHEZ-SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17**, 401–415.

WU, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). *Ann. Statist.* **14**, 1261–1350.

[*Received* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be