*Biometrika* (2019), **106**, 2, *pp.* 385–400
*Printed in Great Britain*
doi: 10.1093/biomet/asz009
Advance Access publication 8 April 2019

# 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

*Research Centre for Operations Research and Business Statistics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium*

ingrid.vankeilegom@kuleuven.be

## Summary

In this paper we consider regression models with centred errors, independent of the covariates. Given independent and identically distributed data and given an estimator of the regression function, which can be parametric or nonparametric in nature, we estimate the distribution of the error term by the empirical distribution of estimated residuals. To approximate the distribution of this estimator, Koul & 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 as to whether a classical nonsmooth residual bootstrap is asymptotically valid in this context. Here we solve this open problem and show that the nonsmooth residual bootstrap is consistent. We illustrate the theoretical result by means of simulations, which demonstrate 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 $p \geqslant 1$, and the error term $\varepsilon$ is independent of $X$. The regression function $m(\cdot)$ can be parametric, for instance linear, or nonparametric in nature, and the distribution $F$ of $\varepsilon$ is completely unknown, except that $E(\varepsilon) = 0$. The estimation of the distribution $F$ has been the subject of many papers, starting with the seminal work of Durbin (1973), Loynes (1980) and Koul (1987) in the case where $m(\cdot)$ is parametric; the nonparametric case has been studied by Van Keilegom & Akritas (1999), Akritas & 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), such as testing for the form of the regression function

© 2019 Biometrika Trust

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

$m(\cdot)$ (Van Keilegom et al., 2008), comparing regression curves (Pardo-Fernández et al., 2007), testing independence between $\varepsilon$ and $X$ (Einmahl & Van Keilegom, 2008; Racine & Van Keilegom, 2019), or testing for a symmetric error distribution (Koul, 2002; Neumeyer & Dette, 2007). 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 & 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) considered 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), Huskova & Meintanis (2009), and others. An alternative bootstrap procedure for nonparametric regression was considered by Neumeyer (2008), where bootstrap residuals were drawn from the nonsmoothed empirical distribution of the residuals, after which smoothing is applied to the empirical distribution of the bootstrap residuals. Further, in the context of residual-based procedures, it has been shown that the wild bootstrap can be used for specific testing problems such as testing for a symmetric error distribution (Neumeyer et al., 2005; Neumeyer & 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 as to whether a classical nonsmooth residual bootstrap is asymptotically valid in this context. In the present paper we solve this open problem and show that the nonsmooth residual bootstrap is consistent when applied to residual processes. We will do this for the cases of univariate nonparametric regression with random design and of multivariate linear regression with fixed design. Other models, such as nonparametric regression with fixed design, and nonlinear or semiparametric regression, can be treated similarly. The question of whether smooth bootstrap procedures should be preferred over nonsmooth bootstrap procedures has been discussed in various contexts; see Silverman & Young (1987) and Hall et al. (1989).

The finite-sample performance of the smooth and nonsmooth residual bootstraps for residual processes has been studied by Neumeyer (2009) who showed that for small sample sizes, using the classical residual bootstrap version of the residual empirical process in the nonparametric regression context yields quantiles that are too small. However, as we will show here, this problem diminishes 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 used to obtain confidence bands for the error distribution or bootstrap versions of hypothesis tests based on residual empirical processes. These 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 used for hypothesis testing or derivation of confidence sets for the regression parameter; see also Davison & Hinkley (1997) and the references therein. Second, there is a vast literature on bootstrap confidence sets for the regression function in nonparametric models; see Härdle & Bowman (1988), Härdle & Marron (1991), Neumann & Polzehl (1998) and Claeskens & Van Keilegom (2003). Third, several authors have considered bootstrap procedures applied to hypothesis testing with test statistics that depend directly on the regression estimator. Among others, Härdle & Mammen (1993), Stute et al. (1998) and Delgado & González Manteiga (2001) proved validity of bootstrap procedures in the context of specific test statistics for nonparametric regression models that do not depend on residual empirical processes.

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

## 2. Nonparametric regression

We start with the case of nonparametric regression with random design. The covariate is assumed 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 that converges to zero as $n$ tends to infinity. Our main result is valid under the following regularity assumptions.

*Assumption* 1. 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 $(0, 1)$.

*Assumption* 2. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with distribution function $F$. They are centred and are independent of the covariates, and $F$ 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 $\upsilon \geqslant 7$.

*Assumption* 3. The function $k$ is a twice continuously differentiable symmetric density with compact support $[-1, 1]$, say, such that $\int uk(u)\,\mathrm{d}u = 0$ and $k(-1) = k(1) = 0$. The first derivative of $k$ is of bounded variation.

*Assumption* 4. The sequence $\{h_n\}$ consists of positive bandwidths such that $h_n \sim c_n n^{-1/3+\eta}$ with $4/(3 + 9\upsilon) < \eta < 1/12$, where $c_n$ is only of logarithmic rate and $\upsilon$ is as defined in Assumption 2.

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

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

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

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

where $\varepsilon$ has distribution function $F$ and density $f$.

Neumeyer (2009) studied a smooth bootstrap procedure for approximating the distribution of this residual process, and showed that using the smooth bootstrap, 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

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

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

$$\hat F_{0,n}(y) = n^{-1}\sum_{i=1}^n I(\tilde\varepsilon_i \leqslant y)$$

be the nonsmoothed empirical distribution of the centred residuals. Then we randomly draw bootstrap errors $\varepsilon^*_{0,1}, \ldots, \varepsilon^*_{0,n}$ with replacement from $\hat F_{0,n}$. Let $Y_i^* = \hat m(X_i) + \varepsilon^*_{0,i}$ $(i = 1, \ldots, n)$, and let $\hat m_0^*(\cdot)$ be the same as $\hat m(\cdot)$ except that we use the bootstrap data $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$. Now define

$$\hat\varepsilon^*_{0,i} = Y_i^* - \hat m_0^*(X_i) = \varepsilon^*_{0,i} + \hat m(X_i) - \hat m_0^*(X_i). \tag{4}$$

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

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

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

To prove this result, we will use the results in Neumeyer (2009) to show that the difference between the smooth and nonsmooth bootstrap residual processes is asymptotically negligible. To this end, we can write $\varepsilon^*_{0,i} = \hat F^{-1}_{0,n}(U_i)$ $(i = 1, \ldots, n)$, where $U_1, \ldots, U_n$ are independent random variables from a Un[0, 1] distribution. Strictly speaking, the $U_i$ form a triangular array $U_{1,n}, \ldots, U_{n,n}$ of Un[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 $U_1, \ldots, U_n$ without loss of generality.

We introduce the following notation: let $\varepsilon^*_{s,i} = \hat F^{-1}_{s,n}(U_i)$, where $\hat F_{s,n}(y) = \int \hat F_{0,n}(y - vs_n)\, \mathrm{d}L(v)$ is the convolution of the distribution $\hat F_{0,n}(y - \cdot s_n)$ and the integrated kernel $L(\cdot) = \int_{-\infty}^{\cdot} \ell(u)\, \mathrm{d}u$, with $\ell$ being a kernel density function and $s_n$ 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

$$\hat\varepsilon^*_{s,0,i} = \varepsilon^*_{s,i} + \hat m(X_i) - \hat m_0^*(X_i). \tag{5}$$

We then decompose the bootstrap residual process as follows:

$$\begin{aligned} n^{1/2}\{\hat F^*_{0,n}(y) - \hat F_{0,n}(y)\} &= n^{-1/2}\sum_{i=1}^n \{I(\hat\varepsilon^*_{0,i} \leqslant y) - I(\hat\varepsilon^*_{s,0,i} \leqslant y)\} \\ &\quad + n^{-1/2}\sum_{i=1}^n \{I(\hat\varepsilon^*_{s,0,i} \leqslant y) - \hat F_{s,n}(y)\} \\ &\quad + n^{1/2}\{\hat F_{s,n}(y) - \hat F_{0,n}(y)\} \\ &= T_{n1}(y) + T_{n2}(y) + T_{n3}(y). \end{aligned} \tag{6}$$

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

In the Supplementary Material we show that under Assumptions 1–4 above and conditions S1 and S2 in the Supplementary Material, concerning the choice of $\ell$ 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 (5) the same function $\hat{m}_0^*$ needs to be used as in (4), in contrast to (8) 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 $\varepsilon_{s,i}^*$ $(i = 1, \ldots, n)$. In this way the difference between the empirical distribution functions of $\hat{\varepsilon}_{0,1}^*, \ldots, \hat{\varepsilon}_{0,n}^*$ and $\hat{\varepsilon}_{s,0,1}^*, \ldots, \hat{\varepsilon}_{s,0,n}^*$ 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 $o_{\rm pr}(n^{-1/2})$. From the latter fact negligibility of $T_{n3}$ also follows. Further, in the Supplementary Material 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

$$\hat{F}_{s,n}^*(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{s,i}^* \leqslant y), \tag{7}$$

where, in contrast to (5),

$$\hat{\varepsilon}_{s,i}^* = \varepsilon_{s,i}^* + \hat{m}(X_i) - \hat{m}_s^*(X_i), \tag{8}$$

with $\hat{m}_s^*$ defined as $\hat{m}$ but based on smoothed bootstrap data $(X_i, \hat{m}(X_i) + \varepsilon_{s,i}^*)$ $(i = 1, \ldots, n)$. 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 in proving the lemma in the Supplementary Material is to show that the estimator $\hat{m}_0^*$ has similar asymptotic properties to $\hat{m}_s^*$, such that using the different estimator does not make a difference in the proofs of fidi-convergence and tightness and in the calculation of the asymptotic covariance. The three lemmas in the Supplementary Material lead to the following main result regarding the validity of the nonsmooth bootstrap residual process.

THEOREM 1. *Suppose that Assumptions 1–4 hold. Then, conditionally on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$, the process $n^{1/2}\{\hat{F}_{0,n}^*(y) - \hat{F}_{0,n}(y)\}$, $y \in \mathbb{R}$, converges weakly in probability to the zero-mean Gaussian process $W(y)$, $y \in \mathbb{R}$, defined in* (3).

The proof is given in the Supplementary Material.

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

*Remark* 1. If the aim is to obtain confidence sets for the regression function, one needs different kinds of bootstrap results. To demonstrate that both the nonsmooth and the smooth

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

residual bootstraps can be applied in this context as well, note that under Assumptions 1–4, for fixed $0 < x < 1$, $(nh_n)^{1/2}\{\hat{m}(x) - m(x)\}$ converges in distribution to a centred normally distributed random variable $Z$ with variance $E(\varepsilon_1^2)\int k^2(u)\,\mathrm{d}u/f_X(x)$. No bias term appears because under our assumptions $nh_n^5 \to 0$. Under the same assumptions one obtains, conditionally on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$, weak convergence in probability of the nonsmooth residual bootstrap version $(nh_n)^{1/2}\{\hat{m}_0^*(x) - m^*(x)\}$ to $Z$, if the centring term is chosen as $m^*(x) = \sum_{i=1}^n k_h(X_i - x)\hat{m}(X_i)/\sum_{j=1}^n k_h(X_j - x)$. In the Supplementary Material we provide a sketch of the proof to obtain the result under exactly our model and our assumptions, though similar results are well known in the literature. Choosing this centring is analogous to the approach of Härdle & Bowman (1988), who obtained 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 centring $m^*(x)$ by $\hat{m}(x)$, one should use a larger pilot bandwidth for $\hat{m}$ when constructing the bootstrap observations. This was demonstrated by Härdle & Marron (1991) for a wild bootstrap for the Nadaraya–Watson estimator and random covariates, and the same reasoning applies to the residual bootstrap. Cao-Abad & González-Manteiga (1993) obtained similar results for a bootstrap procedure with smoothing in the explanatory variable. More recently, McMurry & Politis (2008) considered an alternative method of bias correction by 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 $\ell$ is a symmetric density with second moments. See the Supplementary Material for a derivation of this result.

## 3. LINEAR MODEL

Consider independent observations from the linear model

$$Y_{ni} = x_{ni}^{\mathrm{T}}\beta + \varepsilon_{ni} \quad (i = 1, \ldots, n), \tag{9}$$

where $\beta \in \mathbb{R}^p$ denotes the unknown parameter and the errors $\varepsilon_{ni}$ 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 $i$th row of the matrix $X_n$ and is not random. The design matrix is assumed to be of rank $p \leqslant n$. We assume the following regularity conditions.

*Assumption* 5. The fixed design satisfies:

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

*Assumption* 6. The errors $\varepsilon_{ni}$ $(i = 1, \ldots, n;\, n \in \mathbb{N})$ 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 that $E(|\varepsilon_{11}|^{\upsilon}) < \infty$ for some $\upsilon > 3$.

We consider the least squares estimator

$$\hat{\beta}_n = (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}Y_n = \beta + (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}\varepsilon_n, \tag{10}$$

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

with the notation $Y_n = (Y_{n1}, \ldots, Y_{nn})^{\mathrm{T}}$ and $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^{\mathrm{T}}$, 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 in Koul (2002). It is shown there that, under Assumptions 5 and 6, the process $n^{-1/2}\sum_{i=1}^{n}\{I(\hat{\varepsilon}_{ni} \leqslant y) - F(y)\}$, $y \in \mathbb{R}$, converges weakly to a zero-mean Gaussian process $W(y)$ with covariance function

$$\begin{aligned}\mathrm{cov}\big\{W(y_1), W(y_2)\big\} = F(y_1 \wedge y_2) - F(y_1)F(y_2) + m^{\mathrm{T}}\Sigma^{-1}m\Big[&f(y_1)f(y_2)\mathrm{var}(\varepsilon) \\ &+ f(y_1)E\{I(\varepsilon \leqslant y_2)\varepsilon\} + f(y_2)E\{I(\varepsilon \leqslant y_1)\varepsilon\}\Big],\end{aligned} \tag{11}$$

where $\varepsilon$ has distribution function $F$ and density $f$, and $m$ and $\Sigma$ are as defined in Assumption 5.

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

$$\hat{F}_{0,n}(y) = n^{-1}\sum_{i=1}^{n} I(\tilde{\varepsilon}_{ni} \leqslant y) \tag{12}$$

with $\tilde{\varepsilon}_{ni} = \hat{\varepsilon}_{ni} - n^{-1}\sum_{j=1}^{n}\hat{\varepsilon}_{nj}$ $(i = 1, \ldots, n)$. The centring of residuals is not necessary when the covariate includes an intercept. 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,

$$\hat{\varepsilon}_{0,i}^{*} = Y_{ni}^{*} - x_{ni}^{\mathrm{T}}\hat{\beta}_{0,n}^{*} = \varepsilon_{0,i}^{*} + x_{ni}^{\mathrm{T}}(\hat{\beta}_n - \hat{\beta}_{0,n}^{*}), \tag{13}$$

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

$$\hat{\beta}_{0,n}^{*} = (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}Y_n^{*} = \hat{\beta}_n + (X_n^{\mathrm{T}}X_n)^{-1}X_n^{\mathrm{T}}\varepsilon_{0,n}^{*}, \tag{14}$$

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

$$\hat{F}_{0,n}^{*}(y) = n^{-1}\sum_{i=1}^{n} I(\hat{\varepsilon}_{0,i}^{*} \leqslant y), \tag{15}$$

converges to the same limiting process $W(y)$, $y \in \mathbb{R}$, as the original residual process $n^{-1/2}\sum_{i=1}^{n}\{I(\hat{\varepsilon}_{ni} \leqslant y) - F(y)\}$, $y \in \mathbb{R}$. Using the representations $\varepsilon_{0,i}^{*} = \hat{F}_{0,n}^{-1}(U_i)$ and $\varepsilon_{s,i}^{*} = \hat{F}_{s,n}^{-1}(U_i)$ $(i = 1, \ldots, n)$, where $U_1, \ldots, U_n$ are independent and Un[0, 1]-distributed and $\hat{F}_{s,n}(y) = \int \hat{F}_{0,n}(y - vs_n)\,\mathrm{d}L(v)$ is the smoothed empirical distribution function of the residuals, we have the same decomposition (6) as in the nonparametric case. In the Supplementary Material we show that under Assumptions 5 and 6 and the conditions in the Supplementary Material on the choice of $s_n$ and $L$, the terms $T_{n1}$ and $T_{n3}$ are asymptotically negligible. The idea of the proof is analogous to that in the nonparametric case. In particular, one needs to use the same estimator $\hat{\beta}_{0,n}^{*}$ in (13) and in the definition of $\hat{\varepsilon}_{s,0,i}^{*} = \varepsilon_{s,i}^{*} + x_{ni}^{\mathrm{T}}(\hat{\beta}_n - \hat{\beta}_{0,n}^{*})$ with smooth bootstrap errors $\varepsilon_{s,i}^{*}$,

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

and one needs to use the same uniform random variables $U_i$ to define $\varepsilon^*_{0,i}$ and $\varepsilon^*_{s,i}$ $(i = 1, \dots, n)$. We further show in the Supplementary Material that the limiting distribution of

$$T_{n2}(y) = n^{-1/2} \sum_{i=1}^{n} \{I(\hat{\varepsilon}^*_{s,0,i} \leqslant y) - \hat{F}_{s,n}(y)\} \tag{16}$$

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

$$\hat{F}^*_{s,n}(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}^*_{s,i} \leqslant y) \tag{17}$$

and $\hat{\varepsilon}^*_{s,i} = \varepsilon^*_{s,i} + x^{\mathrm{T}}_{ni}(\hat{\beta}_n - \hat{\beta}^*_{s,n})$, where $\hat{\beta}^*_{s,n} = \hat{\beta}_n + (X^{\mathrm{T}}_n X_n)^{-1} X^{\mathrm{T}}_n \varepsilon^*_{s,n}$ and $\varepsilon^*_{s,n} = (\varepsilon^*_{s,1}, \dots, \varepsilon^*_{s,n})^{\mathrm{T}}$. To show this we apply results from Koul & 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 that Assumptions 5 and 6 hold. Then, conditionally on the data $Y_{1n}, \dots, Y_{nn}$, the process $n^{1/2}\{\hat{F}^*_{0,n}(y) - \hat{F}_{0,n}(y)\}$, $y \in \mathbb{R}$, converges weakly in probability to the zero-mean Gaussian process $W(y)$, $y \in \mathbb{R}$, defined in (11).*

The proof is given in the Supplementary Material.

The result can be used to obtain confidence bands for the error distribution or for hypothesis testing with procedures based on the residual empirical process; see, for example, Koul (2002) or Neumeyer et al. (2005).

*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 centred normal random variable $Z$ with covariance matrix $E(\varepsilon^2_{11})\Sigma^{-1}$. Freedman (1981) showed that along almost all sequences $Y_{1n}, \dots, Y_{nn}$, $n^{1/2}(\hat{\beta}^*_{0,n} - \hat{\beta}_n)$ converges in distribution to $Z$. His result holds under our, stronger, Assumptions 5 and 6. Thus the asymptotic distribution of $n^{1/2}(\hat{\beta}_n - \beta)$ is mimicked by the nonsmooth residual bootstrap. Freedman (1981) further demonstrated that the residual bootstrap may fail if the residuals are not centred; see (12). 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 $\ell$ is a symmetric density with second moments. We will demonstrate this in the Supplementary Material.

## 4. SIMULATIONS

### 4.1. *Confidence bands*

We study the behaviour of the smooth and nonsmooth residual bootstraps 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.

Consider model (1) in the nonparametric case and generate data with $m(x) = 2x$, where $X$ follows a uniform distribution on $[0, 1]$ and $\varepsilon \sim N(0, 0.25^2)$. To assess the quality of the smooth and nonsmooth bootstrap approximations, 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 nonsmooth bootstrap, and $\hat{F}_{0,n}(\cdot) \pm d_{\alpha,n,s}$ for the smooth bootstrap, where $d_{\alpha,n,0}$ and

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

Table 1. *Coverage (%) and average width of confidence bands for $F(\cdot)$ for the smooth and nonsmooth bootstraps in the nonparametric model, with* $1-\alpha = 0.90, 0.95$ *and* $0.99$

| $n$ | Bootstrap | Coverage | | | Average width | | |
|---|---|---|---|---|---|---|---|
| | | 90 | 95 | 99 | 90 | 95 | 99 |
| 50 | Smooth | 86.3 | 92.7 | 99.0 | 23.6 | 25.8 | 30.2 |
| | Nonsmooth | 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 |
| | Nonsmooth | 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 |
| | Nonsmooth | 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 |
| | Nonsmooth | 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 |
| | Nonsmooth | 92.2 | 95.3 | 98.8 | 6.1 | 6.6 | 7.6 |

$d_{\alpha,n,s}$ are the level-$(1-\alpha)$ quantiles of the distributions of $\max_{1\leqslant i\leqslant n} |\hat{F}^*_{0,n}(\hat{\varepsilon}^*_{0,i}) - \hat{F}_{0,n}(\hat{\varepsilon}^*_{0,i})|$ and $\max_{1\leqslant i\leqslant n} |\hat{F}^*_{s,n}(\hat{\varepsilon}^*_{s,i}) - \hat{F}_{s,n}(\hat{\varepsilon}^*_{s,i})|$, respectively.

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, based on 1000 simulation runs, are shown in Table 1; for each simulation 1000 bootstrap samples are generated. The bandwidth $h_n$ is taken equal to $h_n = \hat{\sigma}_X n^{-0.3}$, where $\hat{\sigma}_X$ is 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 $\hat{F}_{s,n}(y) = \int \hat{F}_{0,n}(y - vs_n)\,\mathrm{d}L(v)$, where $L$ is the standard normal distribution and $s_n$ is chosen by means of the crossvalidation procedure proposed by Li et al. (2017). The latter paper studies the estimation of distribution functions by applying kernel smoothing to 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 `R` package `np` (R Development Core Team, 2019) and is obtained from the function `npudistbw`. This bandwidth satisfies a regularity condition imposed on $s_n$, given in the Supplementary Material, thanks to Theorem 3.2 of Li et al. (2017).

The table shows that the smooth bootstrap leads to coverages that are too small, and the coverage does not improve as $n$ increases. On the other hand, the nonsmooth 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 anticonservative while the nonsmooth bootstrap tends to be conservative in this situation. A natural consequence of this tendency to underestimate or overestimate the coverage probability is that the average width of the confidence bands obtained with the smooth bootstrap is smaller than that for the nonsmooth bootstrap, and the difference in width is of the order of 10–20%.

### 4.2. *Testing for a symmetric error distribution*

As already mentioned, the residual bootstrap is 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 $\varepsilon_{ni}$:

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

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

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

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

where $\hat{F}_{-0,n}(y) = n^{-1} \sum_{i=1}^{n} I(-\hat{\varepsilon}_{ni} < y)$ and $\hat{\varepsilon}_{ni} = Y_{ni} - x_{ni}^{\mathrm{T}} \hat{\beta}_n$. Natural test statistics are the Kolmogorov–Smirnov and Crámer–von Mises statistics,

$$T_{\mathrm{KS}} = n^{1/2} \sup_y \big| \hat{F}_{0,n}(y) - \hat{F}_{-0,n}(y) \big|, \quad T_{\mathrm{CM}} = n \int \big\{ \hat{F}_{0,n}(y) - \hat{F}_{-0,n}(y) \big\}^2 \, \mathrm{d}\hat{F}_{0,n}(y).$$

It is clear from the covariance function given in (11) that their asymptotic distributions are 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 nonsmooth bootstraps. The bootstrapped versions of $T_{\mathrm{CM}}$ are given by

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

$$T_{\mathrm{CM},s}^{*} = n \int \big\{ \hat{F}_{s,n}^{*}(y) - \hat{F}_{-s,n}^{*}(y) \big\}^2 \, \mathrm{d}\hat{F}_{s,n}^{*}(y);$$

bootstrapped versions of $T_{\mathrm{KS}}$ can be defined similarly. Here, for the nonsmooth 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 crossvalidation as in the previous simulation study. The estimators $\hat{F}_{0,n}^{*}(\cdot)$ and $\hat{F}_{s,n}^{*}(\cdot)$ are defined as in (15) and (17), 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 $T_{\mathrm{CM}}$ exceeds the level-$(1-\alpha)$ quantile of the distribution of $T_{\mathrm{CM},0}^{*}$ or $T_{\mathrm{CM},s}^{*}$.

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 $0.25^2$. Under the alternative we consider the skew-normal distribution of Azzalini (1985), whose density is given by $2\phi(y)\Phi(dy)$, where $\phi$ and $\Phi$ are the standard normal density and distribution, respectively. More precisely, we let $d = 2$ and $d = 4$ and standardize these skew-normal distributions so that they have mean zero and variance $0.25^2$. When $d = 0$ we recover the normal distribution. The second error distribution under $H_0$ is a Student $t$ distribution with three degrees of freedom, standardized in such a way that the variance equals $0.25^2$. The asymptotic theory does not cover this case, but we would 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 $0.25^2$. The mixture proportions $p$ are 1, 0.75 and 0.50, where $p = 1$ corresponds to $H_0$.

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 too low for the nonsmooth bootstrap. After correcting the critical value of the tests so that the rejection

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

Table 2. *Rejection probabilities* (%) *of the test for symmetry in the linear model for several sample sizes n and for* $\alpha = 0.025$, $0.05$ *and* $0.1$*: 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* $\alpha$*, in order to ease comparison between the two bootstrap methods*

| $n$ | 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 | $T^*_{\mathrm{KS},s}$ | 2.5 | 5.3 | 9.8 | 6.1 | 9.5 | 18.1 | 16.5 | 22.3 | 36.7 |
| | $T^*_{\mathrm{KS},0}$ | 2.0 | 4.2 | 8.3 | 6.4 | 9.5 | 18.1 | 16.9 | 23.6 | 36.8 |
| | $T^*_{\mathrm{CM},s}$ | 2.3 | 4.6 | 9.4 | 6.9 | 11.8 | 20.0 | 19.8 | 29.3 | 41.5 |
| | $T^*_{\mathrm{CM},0}$ | 2.2 | 3.9 | 8.5 | 6.8 | 12.2 | 20.0 | 19.9 | 29.6 | 41.3 |
| 100 | $T^*_{\mathrm{KS},s}$ | 1.8 | 4.7 | 9.4 | 11.2 | 16.8 | 27.0 | 36.3 | 46.1 | 61.1 |
| | $T^*_{\mathrm{KS},0}$ | 1.5 | 3.9 | 8.1 | 11.6 | 18.2 | 27.3 | 38.0 | 48.0 | 61.3 |
| | $T^*_{\mathrm{CM},s}$ | 1.9 | 4.1 | 8.7 | 13.3 | 21.7 | 31.2 | 43.3 | 56.5 | 67.3 |
| | $T^*_{\mathrm{CM},0}$ | 1.8 | 3.9 | 8.1 | 13.0 | 20.6 | 30.3 | 42.2 | 55.5 | 66.5 |
| 200 | $T^*_{\mathrm{KS},s}$ | 2.4 | 5.1 | 10.3 | 21.5 | 29.3 | 43.4 | 64.6 | 74.0 | 85.0 |
| | $T^*_{\mathrm{KS},0}$ | 1.8 | 4.3 | 8.8 | 22.3 | 30.0 | 43.1 | 65.4 | 74.9 | 85.1 |
| | $T^*_{\mathrm{CM},s}$ | 2.1 | 4.6 | 9.4 | 27.7 | 36.5 | 48.1 | 77.1 | 85.2 | 91.4 |
| | $T^*_{\mathrm{CM},0}$ | 1.7 | 4.4 | 9.0 | 27.1 | 36.4 | 48.6 | 77.1 | 85.4 | 91.6 |
| 500 | $T^*_{\mathrm{KS},s}$ | 3.0 | 5.8 | 10.9 | 50.5 | 61.9 | 73.8 | 97.7 | 98.9 | 99.8 |
| | $T^*_{\mathrm{KS},0}$ | 2.7 | 5.0 | 9.9 | 51.2 | 61.8 | 73.7 | 97.7 | 98.9 | 99.8 |
| | $T^*_{\mathrm{CM},s}$ | 2.8 | 5.0 | 10.7 | 60.5 | 72.2 | 80.7 | 99.7 | 100 | 100 |
| | $T^*_{\mathrm{CM},0}$ | 2.7 | 5.0 | 10.3 | 60.0 | 71.3 | 80.8 | 99.6 | 100 | 100 |
| 1000 | $T^*_{\mathrm{KS},s}$ | 2.8 | 4.7 | 10.4 | 81.1 | 89.8 | 94.5 | 100 | 100 | 100 |
| | $T^*_{\mathrm{KS},0}$ | 2.5 | 4.4 | 9.2 | 81.6 | 90.1 | 94.5 | 100 | 100 | 100 |
| | $T^*_{\mathrm{CM},s}$ | 2.4 | 4.7 | 9.4 | 89.8 | 94.9 | 97.6 | 100 | 100 | 100 |
| | $T^*_{\mathrm{CM},0}$ | 2.3 | 4.5 | 9.2 | 90.5 | 94.9 | 97.6 | 100 | 100 | 100 |

probabilities under $H_0$ equal $\alpha$, we see that the smooth and nonsmooth bootstraps 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 1.1%. 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 three degrees of freedom, the size is in general too large for the smooth bootstrap, especially for small $n$, and is about right for the nonsmooth bootstrap. The level-adjusted power is again very similar for the two types of bootstrap. However, the nonsmooth bootstrap has the advantage that it does not depend on the selection of a bandwidth parameter and tends to be conservative in certain situations, whereas the smooth bootstrap has a tendency to be anticonservative.

### 4.3. *Goodness-of-fit tests*

We now test 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 $\Theta$. 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 $\theta_0$ is the value of $\theta$ that minimizes $E[\{m(X) - m_\theta(X)\}^2]$. Consider the following Kolmogorov–Smirnov and

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

Table 3. *Rejection probabilities (%) of the test for symmetry in the linear model for several sample sizes n and for* $\alpha = 0.025$, $0.05$ *and* $0.1$*: under the null we have a Student t distribution with three degrees of freedom* ($p = 1$)*, whereas under the alternative we have a mixture of a Student* $t(3)$ *and a Gumbel distribution* ($p = 0.75$ *and* $p = 0.50$)*; the power is obtained after calibrating the test in such a way that the size equals* $\alpha$*, in order to ease comparison between the two bootstrap methods*

| $n$ | Test | | $p=1$ | | | $p=0.75$ | | | $p=0.50$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 |
| 50 | $T^*_{\mathrm{KS},s}$ | 3.7 | 7.9 | 13.8 | 3.4 | 5.8 | 10.5 | 4.9 | 8.6 | 14.9 |
| | $T^*_{\mathrm{KS},0}$ | 2.6 | 5.6 | 11.2 | 3.5 | 6.5 | 10.7 | 5.4 | 9.5 | 16.0 |
| | $T^*_{\mathrm{CM},s}$ | 3.1 | 6.8 | 13.3 | 4.1 | 7.6 | 12.1 | 7.6 | 11.7 | 17.7 |
| | $T^*_{\mathrm{CM},0}$ | 2.0 | 4.8 | 11.1 | 4.8 | 7.7 | 12.0 | 8.4 | 12.6 | 18.8 |
| 100 | $T^*_{\mathrm{KS},s}$ | 3.7 | 7.2 | 12.9 | 3.9 | 6.9 | 13.0 | 8.4 | 13.5 | 23.7 |
| | $T^*_{\mathrm{KS},0}$ | 2.7 | 5.5 | 10.0 | 3.6 | 7.0 | 13.3 | 8.2 | 13.6 | 23.9 |
| | $T^*_{\mathrm{CM},s}$ | 3.9 | 6.5 | 13.0 | 4.1 | 7.6 | 13.4 | 9.9 | 17.0 | 27.0 |
| | $T^*_{\mathrm{CM},0}$ | 2.8 | 5.3 | 10.7 | 4.2 | 7.5 | 13.6 | 10.0 | 17.2 | 27.6 |
| 200 | $T^*_{\mathrm{KS},s}$ | 2.5 | 5.5 | 11.5 | 6.7 | 10.5 | 17.3 | 17.0 | 25.6 | 35.8 |
| | $T^*_{\mathrm{KS},0}$ | 1.8 | 4.3 | 9.2 | 6.6 | 10.8 | 17.0 | 17.0 | 26.2 | 35.9 |
| | $T^*_{\mathrm{CM},s}$ | 2.7 | 5.6 | 11.0 | 7.5 | 11.9 | 19.1 | 22.0 | 30.6 | 43.0 |
| | $T^*_{\mathrm{CM},0}$ | 2.1 | 4.3 | 9.5 | 7.7 | 12.4 | 19.0 | 22.3 | 31.3 | 43.1 |
| 500 | $T^*_{\mathrm{KS},s}$ | 2.5 | 4.8 | 10.7 | 10.7 | 17.9 | 25.3 | 42.2 | 54.6 | 64.1 |
| | $T^*_{\mathrm{KS},0}$ | 2.2 | 4.2 | 9.7 | 10.5 | 17.4 | 25.5 | 42.1 | 54.3 | 64.3 |
| | $T^*_{\mathrm{CM},s}$ | 2.4 | 4.9 | 10.7 | 14.2 | 20.4 | 29.5 | 52.2 | 62.4 | 71.1 |
| | $T^*_{\mathrm{CM},0}$ | 2.1 | 4.3 | 9.9 | 13.9 | 20.5 | 30.0 | 52.2 | 62.6 | 71.9 |
| 1000 | $T^*_{\mathrm{KS},s}$ | 2.6 | 5.6 | 10.5 | 18.9 | 27.3 | 40.0 | 75.7 | 83.9 | 90.6 |
| | $T^*_{\mathrm{KS},0}$ | 2.0 | 4.8 | 9.5 | 19.2 | 27.1 | 40.2 | 76.1 | 83.0 | 90.6 |
| | $T^*_{\mathrm{CM},s}$ | 2.5 | 5.0 | 10.9 | 23.7 | 34.0 | 44.2 | 83.7 | 89.4 | 93.8 |
| | $T^*_{\mathrm{CM},0}$ | 2.3 | 4.8 | 9.9 | 24.0 | 33.8 | 44.9 | 84.3 | 89.4 | 93.9 |

Crámer–von Mises-type test statistics:

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

where $\hat{F}_{0,n}$ is as defined in § 2 and $\hat{F}_\theta(y) = n^{-1}\sum_{i=1}^n I\{Y_i - \hat{m}_\theta(X_i) \leqslant y\}$ 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 $\theta$, and $\hat{\theta}$ is the least squares estimator of $\theta$. The critical values of these test statistics are approximated using our smooth and nonsmooth residual bootstraps. More precisely, the bootstrapped versions of $T_{\mathrm{CM}}$ are given by

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

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

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

Table 4. *Rejection probabilities* (%) *of the goodness-of-fit test for several sample sizes and for* $\alpha = 0.025$, $0.05$ *and* $0.1$, *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* $\alpha$, *in order to ease comparison between the two bootstrap methods*

| $n$ | Test | $a=0$ | | | $a=0.25$ | | | $a=0.5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 |
| 50 | $T^*_{\mathrm{KS},s}$ | 3.9 | 6.3 | 13.3 | 3.3 | 6.4 | 11.6 | 9.3 | 15.2 | 24.1 |
| | $T^*_{\mathrm{KS},0}$ | 2.3 | 4.8 | 10.7 | 3.2 | 6.9 | 12.5 | 10.1 | 16.3 | 25.8 |
| | $T^*_{\mathrm{CM},s}$ | 2.4 | 4.9 | 9.7 | 5.6 | 10.1 | 17.9 | 16.8 | 24.8 | 36.6 |
| | $T^*_{\mathrm{CM},0}$ | 1.5 | 3.8 | 8.0 | 5.8 | 11.0 | 18.1 | 17.5 | 25.7 | 36.7 |
| 100 | $T^*_{\mathrm{KS},s}$ | 2.1 | 4.0 | 9.2 | 8.1 | 12.6 | 19.5 | 25.4 | 33.8 | 45.5 |
| | $T^*_{\mathrm{KS},0}$ | 1.3 | 2.7 | 7.1 | 8.1 | 12.6 | 19.2 | 26.3 | 34.0 | 46.4 |
| | $T^*_{\mathrm{CM},s}$ | 2.0 | 4.5 | 9.0 | 10.3 | 16.1 | 25.7 | 34.2 | 43.8 | 58.3 |
| | $T^*_{\mathrm{CM},0}$ | 1.5 | 3.5 | 7.9 | 10.4 | 16.1 | 26.2 | 34.7 | 43.9 | 59.2 |
| 200 | $T^*_{\mathrm{KS},s}$ | 2.1 | 4.5 | 9.3 | 12.5 | 20.9 | 32.1 | 48.9 | 60.9 | 73.7 |
| | $T^*_{\mathrm{KS},0}$ | 1.5 | 3.3 | 7.6 | 14.1 | 20.9 | 32.3 | 52.2 | 61.5 | 74.2 |
| | $T^*_{\mathrm{CM},s}$ | 2.2 | 4.2 | 8.5 | 19.2 | 29.8 | 42.5 | 63.6 | 74.9 | 84.3 |
| | $T^*_{\mathrm{CM},0}$ | 1.7 | 3.6 | 7.3 | 19.8 | 29.3 | 43.0 | 64.9 | 74.8 | 84.8 |
| 500 | $T^*_{\mathrm{KS},s}$ | 2.0 | 4.5 | 9.4 | 35.2 | 44.8 | 58.6 | 92.0 | 95.2 | 97.8 |
| | $T^*_{\mathrm{KS},0}$ | 1.5 | 3.4 | 7.7 | 36.6 | 45.0 | 58.4 | 92.6 | 95.6 | 97.8 |
| | $T^*_{\mathrm{CM},s}$ | 1.9 | 4.1 | 8.2 | 45.4 | 56.8 | 68.2 | 96.3 | 98.0 | 98.9 |
| | $T^*_{\mathrm{CM},0}$ | 1.8 | 3.8 | 7.4 | 45.0 | 57.0 | 68.0 | 96.2 | 98.0 | 98.9 |
| 1000 | $T^*_{\mathrm{KS},s}$ | 2.2 | 4.2 | 9.6 | 65.4 | 76.9 | 84.9 | 100 | 100 | 100 |
| | $T^*_{\mathrm{KS},0}$ | 1.8 | 3.6 | 8.4 | 65.2 | 77.0 | 84.9 | 100 | 100 | 100 |
| | $T^*_{\mathrm{CM},s}$ | 2.0 | 4.6 | 9.3 | 72.9 | 81.6 | 88.6 | 100 | 100 | 100 |
| | $T^*_{\mathrm{CM},0}$ | 1.9 | 4.2 | 8.9 | 72.9 | 81.6 | 88.6 | 100 | 100 | 100 |

bootstrapped versions of $T_{\mathrm{KS}}$ can be defined similarly. Here, $\hat{\theta}^*_0$ is the least squares estimator based on the bootstrap data $(X_i, Y^*_{0,i} = m_{\hat{\theta}}(X_i) + \varepsilon^*_{0,i})$ $(i = 1, \ldots, n)$, $\hat{F}^*_{0,\theta}(y) = n^{-1} \sum_{i=1}^n I\{Y^*_{0,i} - \hat{m}_\theta(X_i) \leqslant y\}$ for any $\theta$, and similarly for $\hat{\theta}^*_s$ and $\hat{F}^*_{s,\theta}(y)$. We reject $H_0$ if the observed value of $T_{\mathrm{CM}}$ exceeds the level-$(1-\alpha)$ quantile of the distribution of $T^*_{\mathrm{CM},0}$ or $T^*_{\mathrm{CM},s}$.

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 $\varepsilon$ follows either a normal distribution or a Student $t$ distribution with three degrees of freedom, in both cases standardized in such a way that the variance equals $0.25^2$. The covariate $X$ has a uniform distribution on $[0, 1]$. The bandwidth $h_n$ is taken to be $h_n = \hat{\sigma}_X n^{-0.3}$, and the kernel $k$ is the Gaussian kernel. For the smooth bootstrap, we use a standard normal distribution and select $s_n$ via crossvalidation as in the previous simulations. Under the alternative we consider the model $m(x) = 2x + ax^2$ for $a = 0.25$ and $0.5$. 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 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 nonsmooth bootstraps 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 $\alpha$, we see that the size-adjusted powers under the two types of bootstrap are almost identical. A simple test of the

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

Table 5. *Rejection probabilities* (%) *of the goodness-of-fit test for several sample sizes and for* $\alpha = 0.025$, $0.05$ *and* $0.1$, *when the error term has a Student t distribution with three 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* $\alpha$, *in order to ease comparison between the two bootstrap methods*

| $n$ | Test | $a=0$ | | | $a=0.25$ | | | $a=0.5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 | 2.5 | 5.0 | 10.0 |
| 50 | $T^*_{\mathrm{KS},s}$ | 4.3 | 7.3 | 15.7 | 5.3 | 9.4 | 16.3 | 15.5 | 23.5 | 34.9 |
| | $T^*_{\mathrm{KS},0}$ | 3.0 | 5.6 | 11.9 | 4.8 | 10.0 | 16.7 | 14.9 | 24.6 | 36.6 |
| | $T^*_{\mathrm{CM},s}$ | 3.2 | 5.9 | 12.2 | 8.8 | 14.8 | 22.0 | 24.5 | 35.2 | 45.9 |
| | $T^*_{\mathrm{CM},0}$ | 2.4 | 4.7 | 10.0 | 8.7 | 14.4 | 22.7 | 24.4 | 35.6 | 47.0 |
| 100 | $T^*_{\mathrm{KS},s}$ | 4.2 | 7.7 | 13.2 | 10.1 | 15.7 | 24.2 | 31.3 | 41.7 | 54.9 |
| | $T^*_{\mathrm{KS},0}$ | 2.9 | 5.8 | 10.8 | 10.5 | 15.8 | 24.5 | 32.3 | 42.4 | 55.5 |
| | $T^*_{\mathrm{CM},s}$ | 3.4 | 6.4 | 12.9 | 14.2 | 21.4 | 29.9 | 42.5 | 53.0 | 64.2 |
| | $T^*_{\mathrm{CM},0}$ | 2.5 | 5.2 | 10.8 | 14.7 | 21.5 | 30.0 | 43.1 | 53.1 | 64.4 |
| 200 | $T^*_{\mathrm{KS},s}$ | 3.4 | 6.1 | 12.2 | 18.0 | 25.8 | 38.9 | 57.8 | 69.1 | 79.4 |
| | $T^*_{\mathrm{KS},0}$ | 2.5 | 4.7 | 9.4 | 18.7 | 26.5 | 38.9 | 59.6 | 70.0 | 79.5 |
| | $T^*_{\mathrm{CM},s}$ | 3.3 | 5.4 | 10.8 | 23.1 | 34.2 | 45.2 | 68.5 | 79.0 | 86.3 |
| | $T^*_{\mathrm{CM},0}$ | 2.6 | 4.7 | 9.5 | 22.6 | 33.8 | 45.6 | 68.6 | 78.7 | 86.4 |
| 500 | $T^*_{\mathrm{KS},s}$ | 3.6 | 6.8 | 12.6 | 34.6 | 48.6 | 61.8 | 92.2 | 95.7 | 97.7 |
| | $T^*_{\mathrm{KS},0}$ | 3.1 | 5.5 | 10.9 | 33.8 | 48.2 | 62.4 | 91.7 | 95.6 | 97.8 |
| | $T^*_{\mathrm{CM},s}$ | 3.4 | 6.2 | 11.9 | 44.0 | 54.9 | 67.4 | 95.3 | 97.1 | 98.4 |
| | $T^*_{\mathrm{CM},0}$ | 3.1 | 5.8 | 10.7 | 43.8 | 55.0 | 67.6 | 95.2 | 97.1 | 98.5 |
| 1000 | $T^*_{\mathrm{KS},s}$ | 3.0 | 6.4 | 12.0 | 67.2 | 78.5 | 86.0 | 99.4 | 99.7 | 99.8 |
| | $T^*_{\mathrm{KS},0}$ | 2.4 | 5.6 | 11.0 | 69.2 | 78.5 | 85.9 | 99.5 | 99.7 | 99.8 |
| | $T^*_{\mathrm{CM},s}$ | 3.2 | 6.2 | 11.5 | 73.8 | 82.6 | 89.8 | 99.6 | 99.8 | 99.9 |
| | $T^*_{\mathrm{CM},0}$ | 3.0 | 5.5 | 10.6 | 73.9 | 82.6 | 89.7 | 99.6 | 99.8 | 99.9 |

equality of two proportions shows that there is no significant difference. When the error has a Student $t$ distribution, see Table 5, the size is too large for the smooth bootstrap, but is more or less equal to the nominal level for the nonsmooth bootstrap. The size-adjusted powers are again very close.

Overall, upon comparing the smooth bootstrap with the nonsmooth bootstrap in the five contexts and for the different sample sizes and values of $\alpha$, we can conclude that the smooth bootstrap has a tendency to be anticonservative whereas the nonsmooth bootstrap has a tendency to be conservative. As conservative tests are in general preferred over nonconservative tests, if we follow the general idea that we prefer to control the probability of committing a Type I error, we conclude that the nonsmooth bootstrap is preferable to the smooth bootstrap in the testing problems under consideration and in obtaining confidence bands for the error distribution. In addition, the smooth bootstrap depends on how we choose the smoothing parameter $s_n$ and the corresponding kernel, but this is not the case for the nonsmooth bootstrap.

## 5. DISCUSSION

To the best of our knowledge, this paper is the first work that establishes consistency of the nonsmoothed bootstrap of the error distribution. We have restricted attention to the case of homoscedastic regression models, but in a next step it would be interesting to prove the consistency of the bootstrap when the error variance depends on one or several covariates. Other possible generalizations of the model considered in this paper include extensions to semiparametric regression, nonparametric regression with more than one covariate, dependent data, and missing or censored data.

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

## Acknowledgement

We thank an associate editor and two referees for their constructive comments on the first version of the manuscript. Neumeyer acknowledges financial support from the Deutsche Forschungsgemeinschaft. Van Keilegom was supported by the European Research Council and the Belgian government.

## 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**, 549–67.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–8.

Cao-Abad, R. & González-Manteiga, W. (1993). Bootstrap methods in regression smoothing. *J. Nonparam. Statist.* **2**, 379–88.

Claeskens, G. & Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–84.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

De Angelis, D., Hall, P. & Young, G. A. (1993). Analytical and bootstrap approximations to estimator distributions in $L^1$ regression. *J. Am. Statist. Assoc.* **88**, 1310–6.

Delgado, M. A. & González Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29**, 1469–507.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–90.

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. Economet.* **143**, 88–102.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218–28.

Härdle, W. & Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Am. Statist. Assoc.* **83**, 102–10.

Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–47.

Härdle, W. & Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–96.

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**, 960–71.

Koul, H. L. (1987). Tests of goodness-of-fit in linear regression. In *Goodness-of-fit (Debrecen, 1984)*, vol. 45 of *Colloquia Mathematica Societatis János Bolyai*. Amsterdam: North-Holland, pp. 279–315.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. New York: Springer, 2nd ed.

Koul, H. L. & Lahiri, S. N. (1994). On bootstrapping $M$-estimated residual processes in multiple linear regression models. *J. Mult. Anal.* **49**, 255–65.

Li, C., Li, H. & Racine, J. S. (2017). Cross-validated mixed-datatype bandwidth selection for nonparametric cumulative distribution/survivor functions. *Economet. Rev.* **36**, 970–87.

Loynes, R. M. (1980). The empirical sample distribution function of residuals from generalized regression. *Ann. Statist.* **8**, 285–98.

McMurry, T. L. & Politis, D. N (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statist. Prob. Lett.* **78**, 2463–9.

Mora, J. (2005). Comparing distribution functions of errors in linear models: A nonparametric approach. *Statist. Prob. Lett.* **73**, 425–32.

Müller, U. U., Schick, A. & Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plan. Infer.* **119**, 75–93.

Neumann, M. H. & Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparam. Statist.* **9**, 307–33.

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019

NEUMEYER, N. (2008). A bootstrap version of the residual-based smooth empirical distribution function. *J. Nonparam. Statist.* **20**, 153–74.

NEUMEYER, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scand. J. Statist.* **36**, 204–28.

NEUMEYER, N. & DETTE, H. (2007). Testing for symmetric error distribution in nonparametric regression models. *Statist. Sinica* **17**, 775–95.

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

PARDO-FERNÁNDEZ, J. C., VAN KEILEGOM, I. & GONZÁLEZ-MANTEIGA, W. (2007). Testing for the equality of $k$ regression curves. *Statist. Sinica* **17**, 1115–37.

R DEVELOPMENT CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

RACINE, J. S. & VAN KEILEGOM, I. (2019). A smooth nonparametric, multivariate, mixed-data location-scale test. *J. Bus. Econ. Statist.* to appear.

SILVERMAN, B. W. & YOUNG, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74**, 469–79.

STUTE, W., GONZÁLEZ MANTEIGA, W. & PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Am. Statist. Assoc.* **93**, 141–9.

VAN KEILEGOM, I. & AKRITAS, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745–84.

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**, 401–15.

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with Discussions). *Ann. Statist.* **14**, 1261–350.

[*Received on* 7 *December* 2017. *Editorial decision on* 6 *November* 2018]

Downloaded from https://academic.oup.com/biomet/article-abstract/106/2/385/5431311 by Universita di Trento - Biblioteca user on 07 September 2019