Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 21–57

ON AUTOMATIC KERNEL DENSITY ESTIMATE BASED TESTS FOR GOODNESS-OF-FIT

CARLOS TENREIRO

ABSTRACT: Although estimation and testing are different statistical problems, if we want to use a test statistic based on the Parzen-Rosenblatt estimator to test the hypothesis that the underlying density function f is a member of a location-scale family of probability density functions, it may be found reasonable to choose the smoothing parameter in such a way that the kernel density estimator is an effective estimator of f irrespective of which of the null or the alternative hypothesis is true. In this paper we address this question by considering the well-known Bickel-Rosenblatt test statistics which are based on the quadratic distance between the nonparametric kernel estimator and two parametric estimators of f under the null hypothesis. For each one of these test statistics we describe their asymptotic behaviours for a general data-dependent smoothing parameter, and we state their limiting gaussian null distribution and the consistency of the associated goodness-of-fit test procedures for location-scale families. In order to compare the finite sample power performance of the Bickel-Rosenblatt tests based on a null hypothesis-based bandwidth selector with other bandwidth selector methods existing in the literature, a simulation study for the normal, logistic and Gumbel null location-scale models is included in this work.

KEYWORDS: Kernel density estimator; Goodness-of-fit tests; Bickel-Rosenblatt tests; Bandwidth selection.
AMS SUBJECT CLASSIFICATION (2020): 62G10, 62G07.

1. Introduction

Given $X_1, \ldots, X_n$ independent and identically distributed real-valued random variables from an absolutely continuous distribution with continuous density function f, it is well known that the unknown density function f may be estimated by using the Parzen-Rosenblatt estimator (Rosenblatt, 1956, Parzen, 1962) defined, for $x \in \mathbb{R}$, by

$$f_h(x) := \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

where $K_h(\cdot) = K(\cdot/h)/h$, for $h > 0$, with K a kernel in $\mathbb{R}$, that is, K is a bounded and symmetric probability density function, and the bandwidth

Received December 21, 2021.

$h = h_n$ is a sequence of strictly positive real numbers converging to zero as n tends to infinity, which we always assume along this paper (see Devroye and Györfi, 1985, Silverman, 1986, Bosq and Lecoutre, 1987, Wand and Jones, 1995, Simonoff, 1996, and Tsybakov, 2009, for general reviews on density estimation).

Other than the estimation of the underlying probability density function, the kernel density estimator can also be used for testing the null hypothesis

$$H_0: \; f \in \mathscr{F}_0, \tag{1}$$

where $\mathscr{F}_0$ is a parametric family of density functions, against a general alternative hypothesis. This idea was first explored in Bickel and Rosenblatt (1973) who considered, among other, two test statistics based on the L^2 distance between the nonparametric estimator f_h and two parametric estimators of f under the null hypothesis. Focusing our attention on the case where $\mathscr{F}_0$ is a location-scale family, that is,

$$\mathscr{F}_0 = \big\{ g(\cdot;\theta_1,\theta_2) : \theta_1 \in \mathbb{R}, \theta_2 > 0 \big\}, \tag{2}$$

with $g(x;\theta_1,\theta_2) = f_0((x-\theta_1)/\theta_2)/\theta_2$, and f_0 is a known probability density function on $\mathbb{R}$, the Bickel-Rosenblatt test statistics we are interested in are given by

$$I_n(h) = I_n(X_1,\ldots,X_n;h) := nh \int \{ f_h(x) - K_h * g(x;\hat\theta_1,\hat\theta_2)\}^2 dx, \tag{3}$$

and

$$J_n(h) = J_n(X_1,\ldots,X_n;h) := nh \int \{ f_h(x) - g(x;\hat\theta_1,\hat\theta_2)\}^2 dx, \tag{4}$$

where the integrals are over $\mathbb{R}$ with respect to the Lebesgue measure, $\hat\theta_k$, for $k = 1, 2$, are consistent estimators of θ_k under H_0, and $*$ denotes the convolution operator. The theoretical properties of goodness-of-fit tests based on $I_n(h)$ and $J_n(h)$ were first studied by Bickel and Rosenblatt (1973) in the univariate case by using strong approximation techniques for empirical processes, and by Rosenblatt (1975) in the multivariate case, by using a Poissonization of sample size technique. However, a full description of their asymptotic behaviour was later provided in Fan (1994) by using the fact, first noticed in Hall (1984), that central limit theorems for the integrated squared error of kernel density estimators can be derived by using a central limit theorem for degenerate U-statistics with variable kernels (see Ghosh and Huang, 1991, Fan, 1998, Gouriéroux and Tenreiro, 2001 and Cao and Lugosi, 2005, for other works on goodness-of-fit tests based on the kernel density estimator).

Taking into account that the class $\mathscr{F}_0$ is closed with respect to affine transformations, some authors argue that any reasonable statistic $T_n = T_n(X_1, \ldots, X_n)$ for testing H_0 should be location-scale invariant, that is, it should satisfy the equality

$$T_n(a + bX_1, \ldots, a + bX_n) = T_n(X_1, \ldots, X_n),$$

for each $a \in \mathbb{R}$ and $b > 0$ (see Henze, 2002, p. 469, Ebner and Henze, 2020, p. 847). As we can easily see, this invariance property does not hold for the functionals $I_n(h)$ and $J_n(h)$ whenever we take for h a deterministic bandwidth, even when $\hat{\theta}_1$ is location-scale equivariant and $\hat{\theta}_2$ is scale equivariant, that is,

$$\hat{\theta}_1(a + bX_1, \ldots, a + bX_n) = a + b\,\hat{\theta}_1(X_1, \ldots, X_n)$$

and

$$\hat{\theta}_2(a + bX_1, \ldots, a + bX_n) = b\,\hat{\theta}_2(X_1, \ldots, X_n),$$

for each $a \in \mathbb{R}$ and $b > 0$. However, if we further assume that $h = \hat{h}(X_1, \ldots, X_n)$ depends on the observations and is scale equivariant, then $I_n(\hat{h})$ and $J_n(\hat{h})$ are location-scale invariant test statistics. This invariance property follows easily from the representations

$$I_n(\hat{h}) = n(\hat{h}/\hat{\theta}_2) \int \{\tilde{f}_{\hat{h}/\hat{\theta}_2}(y) - K_{\hat{h}/\hat{\theta}_2} * f_0(y)\}^2 dy, \tag{5}$$

and

$$J_n(\hat{h}) = n(\hat{h}/\hat{\theta}_2) \int \{\tilde{f}_{\hat{h}/\hat{\theta}_2}(y) - f_0(y)\}^2 dy, \tag{6}$$

where

$$\tilde{f}_h(y) = \frac{1}{n} \sum_{i=1}^{n} K_h(y - Y_{n,i}),$$

is the kernel estimator with kernel K and smoothing parameter h, based on the so-called scaled residuals $Y_{n,j} = (X_j - \hat{\theta}_1)/\hat{\theta}_2$, $j = 1, \ldots, n$. When $\hat{h}$ takes the form $\hat{h} = \hat{\theta}_2 h$ with h a deterministic bandwidth, the statistic $I_n(\hat{h})$ is considered in Bowman (1992) (see also Fan, 1994, pp. 332–336) and the theoretical properties of the goodness-of-fit test based on $I_n(\hat{h})$ are described in Tenreiro (2007) in the case where θ_1 and θ_2 are, respectively, the mean and the standard deviation of $g(\cdot; \theta_1, \theta_2)$, and $\hat{\theta}_1 = \bar{X}_n$ and $\hat{\theta}_2 = S_n$, where

$\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ is the sample mean and $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is the sample variance. Moreover, Bowman (1992, p. 3) also suggests to take for the deterministic bandwidth h the asymptotic optimal bandwidth, in the sense of the mean integrated square error, for estimating the null density f_0. In this case, we have

$$h = h_1 = h_1(f_0; K, n) = c_K\, R(f_0'')^{-1/5} n^{-1/5}, \tag{7}$$

with

$$c_K = R(K)^{1/5}\mu_2(K)^{-2/5} \tag{8}$$

(see Bosq and Lecoutre, 1987, pp. 78–83 and Wand and Jones, 1995, pp. 19–23), where $R(\varphi) = \int \varphi(x)^2 dx$ and $\mu_2(\varphi) = \int x^2\varphi(x)dx$, for an arbitrary real-valued measurable function φ, which leads to consider for $\hat{h}$ the null hypothesis based bandwidth selector

$$\hat{h}_{H_0} = \hat{\theta}_2 h_1(f_0; K, n). \tag{9}$$

In the case of testing an hypothesis of normality, that is, $f_0 = \phi$, where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, $x \in \mathbb{R}$, is the standard Gaussian density, and taking $K = \phi$ and $\hat{\theta}_2 = S_n$, this leads to the data-dependent bandwidth

$$\hat{h}_{H_0} = (4/3)^{1/5} S_n n^{-1/5}. \tag{10}$$

This approach, also considered in Bowman and Foster (1993, p. 535) for testing a multivariate hypothesis of normality, was first suggested with apparent good results by Henze and Zirkler (1990, p. 3600; see also Ebner and Henze, 2020) and the corresponding theoretical properties of the test statistic $I_n(\hat{h})$ first established in Gürtler (2000).

From an estimation perspective, the choice of the bandwidth is crucial to the performance of the kernel density estimator, this being one of the most studied topics in kernel density estimation, and several data-based approaches have been proposed for selecting h (see Wand and Jones, 1995, pp. 58–89, and also Tenreiro, 2017, p. 3440, where more recent bandwidth selection methods are mentioned). Although estimation and testing are different statistical problems, if we want to test H_0 through a test statistic based on the kernel density estimator, it may be found reasonable to select the smoothing parameter in such a way that f_h is an effective estimator of the underlying density f irrespective of which of the null or the alternative hypothesis is true, a property that is not fulfilled by the automatic bandwidth selector (9). Despite some scepticism have been expressed about this approach by Bowman (1992, p. 3), mainly due to the extra source of variation introduced into the null distribution of the test statistic by the considered bandwidth selector, in this paper we intend to address this issue deeply by considering the situation where the data-dependent smoothing parameter $\hat{h}$ satisfies the relative consistency condition

$$\frac{\hat{h}}{h_0} - 1 = o_p(1), \tag{11}$$

where $h_0 = h_0(f; K, n)$ is the exact optimal bandwidth in the sense that it minimizes the kernel density estimator mean integrated square error, that is,

$$h_0 = \operatorname*{argmin}_{h>0} \mathrm{MISE}(f; K, n, h), \tag{12}$$

where

$$\mathrm{MISE}(f; K, n, h) = \mathrm{E}\biggl(\int \{f_h(x) - f(x)\}^2 dx \biggr).$$

For a square integrable density f, the existence of this exact optimal bandwidth for all $n \in \mathbb{N}$ can be established whenever the kernel K is continuous at zero with $R(K) < 2K(0)$ (see Chacón et al., 2007). Classical data-based bandwidth selectors such as the least squares cross-validation bandwidth or the two-stage direct plug-in bandwidth selector based on $h_1 = h_1(f; K, n)$, which are both described in Wand and Jones (1995, pp. 63–65, 71–72), are scale equivariant and satisfy (11).

The remainder of this work is organised as follows. In Section 2 we describe the asymptotic behaviour of the Bickel-Rosenblatt test statistics $I_n(\hat{h})$ and $J_n(\hat{h})$ with $\hat{h} = \hat{h}_n(X_1, \ldots, X_n)$ a general data-dependent smoothing parameter. In a univariate context these results extend those obtained by Fan (1994), Gürtler (2000) and Tenreiro (2007). The limiting null distribution and the consistency of the considered Bickel-Rosenblatt tests for location-scale families are stated in Section 3. In Section 4 we conduct a simulation study to compare the finite sample power performance of the Bickel-Rosenblatt tests based on the null hypothesis-based bandwidth selector $\hat{h}_{H_0}$ with other scale equivariant bandwidth selectors $\hat{h}$ satisfying condition (11). We consider the cases of the normal, logistic and Gumbel null location-scale models. Although $\hat{h}_{H_0}$ does not satisfy this relative consistency condition unless $f \in \mathscr{F}_0$, we conclude that the tests based on it, especially those based on I_n, are in general more powerful than, or at least as powerful as, those based on the considered bandwidth selectors that satisfy such condition. Some other data-driven bandwidths inspired in the methods considered in Cao and Van Keilegom (2006), Martínez-Camblor et al. (2008), and Martínez-Camblor and Uña-Álvarez (2013) in the context of smooth tests for the k-sample problem are adapted to our context and included in the simulation study. These last bandwidth selectors, which can be computed by resampling, take the general form $\hat{\lambda}\hat{h}$, where $\hat{h}$ is a scale equivariant bandwidth selector (e.g. $\hat{h} = \hat{h}_{H_0}$) and $\hat{\lambda}$ is a data-driven tuning parameter selector taking values in a finite set of tuning parameters Λ (e.g. $\Lambda = \{0.5, 0.75, 1, 1.5, 2\}$). Nevertheless, none of these bandwidth selectors have shown to be preferable to $\hat{h}_{H_0}$. Section 5 includes a brief summary and some conclusions. For convenience of exposition the proofs are deferred to Section 6 and some of the simulation results are displayed after the bibliography.

2. Test statistics asymptotic behaviour

In this section we are interested in the asymptotic behaviour of the Bickel-Rosenblatt test statistics $I_n(\hat{h})$ and $J_n(\hat{h})$ given by (3) and (4), respectively, where $\hat{h} = \hat{h}_n(X_1, \ldots, X_n)$ is a general data-dependent smoothing parameter. In a univariate framework the results presented here extend those obtained by Fan (1994), Gürtler (2000) and Tenreiro (2007).

2.1. Asymptotic behaviour of $I_n(\hat{h})$. In order to describe the asymptotic behaviour of the integrated square error $I_n(\hat{h})$ we consider the following assumptions on the underlying probability density function f, the parametric family $\mathscr{F}_0$ given by (2), the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, the kernel K and the data-dependent bandwidth $\hat{h}$. We denote by $\mathscr{F}$ an appropriate set of probability density functions on $\mathbb{R}$ that contains $\mathscr{F}_0$ and to which the underlying probability density function f belongs, and by L^r, for $r \in [1, \infty]$, the normed vector space of measurable functions $\varphi : \mathbb{R} \to \mathbb{R}$ for which $||\varphi||_r < \infty$, where $||\varphi||_r := \left(\int |\varphi(x)|^r dx\right)^{1/r} < \infty$ for $r \in [1, \infty[$, and $||\varphi||_\infty = \inf\{c \geq 0 : |\varphi(x)| \leq c \text{ for almost every } x\}$.

Assumption (D). $f \in L^\infty$, for all $f \in \mathscr{F}$.

Assumption (F). f_0 is two times continuously differentiable with $f_0 \in L^\infty$, $f_0', y \mapsto y f_0'(y) \in L^2 \cap L^r$ and $f_0'', y \mapsto y^2 f_0''(y) \in L^r$, for some $r \in]2, \infty]$.

Assumption (P). For all $f \in \mathscr{F}$ there exist $\theta_1(f) \in \mathbb{R}$ and $\theta_2(f) > 0$ such that $\hat{\theta}_k \xrightarrow{p} \theta_k(f)$, for $k = 1, 2$. Moreover, if $f = g(\cdot; \theta_1, \theta_2)$, for some $\theta_1 \in \mathbb{R}$

and $\theta_2 > 0$ (i.e. $f \in \mathscr{F}_0$), we assume that

$$\sqrt{n}\big(\hat{\theta}_k - \theta_k\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_k(X_i; \theta_1, \theta_2) + o_p(1),$$

where $\psi_k(\cdot; \theta_1, \theta_2)$ is a real function depending on θ_1 and θ_2, with $\mathrm{E}_f(\psi_k(X; \theta_1, \theta_2)) = 0$ and $\mathrm{E}_f(\psi_k(X; \theta_1, \theta_2)^2) < \infty$, for $k = 1, 2$.

Assumption (K). The kernel K belongs to $\mathscr{K}^\omega$, for some $\omega \in \{2, 3, \dots\}$, where $\mathscr{K}^\omega$ is the set of real-valued functions K on $\mathbb{R}$ with continuous derivatives up to order ω such that $\lim_{|u|\to\infty} uK(u) = 0$, for which there exists $\eta \in]0, 1[$, such that the real-valued functions $K^{\ell,\eta}$ defined, for $u \in \mathbb{R}$, by $K^{\ell,\eta}(u) = |u|^\ell \sup_{|h-1|\leq\eta} |K^{(\ell)}(u/h)|$, are bounded and integrable on $\mathbb{R}$ for $\ell = 0, 1, \dots, \omega$.

The standard Gaussian kernel $K = \phi$ belongs to $\mathscr{K}^\omega$ for all ω, and every kernel with compact support with continuous derivatives up to order ω belongs to $\mathscr{K}^\omega$.

Assumption (B). For all $f \in \mathscr{F}$, there exists a deterministic sequence $(h_n(f)) = (h(f))$ of strictly positive real numbers satisfying $h(f) \to 0$ and $nh(f) \to \infty$, as $n \to \infty$, such that

$$\xi_n := \frac{\hat{h}}{h(f)} - 1 = o_p(1).$$

As mentioned before, under some conditions on f and K, assumption (B) is fulfilled by the least squares cross-validation bandwidth and by the two-stage direct plug-in bandwidth selector with $h(f) = h_0$, where h_0 is given by (12). Of course, in these cases assumption (B) is also fulfilled with $h(f) = h_1$, where h_1 is given by (7), as h_0 and h_1 are asymptotically equivalent (see Hall and Marron, 1991, p. 159). From a density estimation point of view, the distinction between bandwidth selectors is usually based on the rate of convergence to zero of the relative error ξ_n. For example, we have $\xi_n = O_p\big(n^{-1/10}\big)$ for the least squares cross-validation bandwidth (see Scott and Terrel, 1987, Hall and Marron, 1987), and $\xi_n = O_p\big(n^{-5/14}\big)$ for the two-stage direct plug-in bandwidth selector (see Tenreiro, 2003). A better order of convergence is achieved by the smoothed cross-validation method of Hall et al. (1992) and by the plug-in method of Hall et al. (1991) for which we have $\xi_n = O_p\big(n^{-1/2}\big)$. Note that these rates of convergence are not directly comparable since the conditions imposed to f in each case are not necessarily the same. A different situation occurs when $\hat{h}$ is the well known normal scale bandwidth selector defined by $\hat{h} = (8\sqrt{\pi}/3)^{1/5} c_K n^{-1/5} \hat{\sigma}$, where c_K is given by (8) and $\hat{\sigma}$ is a consistent estimator of the standard deviation σ_f of f (see Wand and Jones, 1995, p. 60). Although this bandwidth selector satisfies assumption (B) with $h(f) = (8\sqrt{\pi}/3)^{1/5} c_K n^{-1/5} \sigma_f$, and we have $\xi_n = O_p\big(n^{-1/2}\big)$ whenever the scale estimator is such that $\hat{\sigma} - \sigma_f = O_p\big(n^{-1/2}\big)$, the normal scale bandwidth selector does not fulfil the relative consistency condition (11).

In the next result, which proof is given in Section 6.1, we describe the asymptotic behaviour of the Bickel-Rosenblatt test statistic $I_n(\hat{h})$ given by (3) where $\hat{h} = \hat{h}_n(X_1, \ldots, X_n)$ is a general data-dependent smoothing parameter. Recall that $R(\varphi) = \int \varphi(x)^2 dx$ for an arbitrary real-valued measurable function φ.

Theorem 1. *Under assumptions (D), (F), (P), (K) and (B), let us assume that*

$$h(f)^{-1/2} \xi_n^2 + n h(f)^{1/2} \xi_n^{\omega} = o_p(1). \tag{13}$$

a) If the null hypothesis is true then

$$\nu_f^{-1} h(f)^{-1/2} \big(I_n(\hat{h}) - R(K)\big) \xrightarrow{d} N(0,1),$$

where

$$\nu_f^2 = 2R(K * K) R(f).$$

b) If the alternative hypothesis is true then

$$(nh(f))^{-1} \big(I_n(\hat{h}) - R(K)\big) \xrightarrow{p} R\big(f - g(\cdot; \theta_1(f), \theta_2(f))\big).$$

Remark 1. If $h(f) = cn^{-1/5}(1 + o(1))$ and $\xi_n = O_p\big(n^{-\alpha}\big)$, for some $c > 0$ and $0 < \alpha \leq 1/2$, condition (13) is satisfied whenever $\alpha > \max(1/20, 9/(10\omega))$. Therefore, it holds for the least squares cross-validation bandwidth selector whenever $\omega \geq 10$, and for the two-stage direct plug-in bandwidth selector if $\omega \geq 3$.

2.2. Asymptotic behaviour of $J_n(\hat{h})$. In order to describe the asymptotic behaviour of the integrated square error $J_n(\hat{h})$ some additional assumptions are needed.

Assumption (D'). For all $f \in \mathscr{F}$, f is two times continuously differentiable on $\mathbb{R}$ with $f'' \in L^\infty \cap L^2$.

Assumption (F'). f_0 is such that $f_0'' \in L^\infty \cap L^s$, with $1/r + 1/s = 1$ and $r \in]2, \infty]$ is given in assumption (F).

Assumption (K'). The functions $u \mapsto u^2 K^{\ell,\epsilon}(u)$, for $\ell = 0, 1, \ldots, \omega$, where $K^{\ell,\epsilon}$ is defined in assumption (K), are bounded and integrable on $\mathbb{R}$.

Assumption (B'). For all $f \in \mathscr{F}$, the deterministic sequence $h(f)$ is such that $nh(f)^5 \to \lambda_f$, as $n \to \infty$, for some $\lambda_f \in]0, \infty[$.

Note that if $h(f) = h_0$, where h_0 is given by (12), then $\lambda_f = c_K^5 R(f'')^{-1}$, where c_K is given in (8).

In the next result, which proof is given in Section 6.2, we describe the asymptotic behaviour of the Bickel-Rosenblatt test statistic $J_n(\hat{h})$ given by (4) where $\hat{h} = \hat{h}_n(X_1, \ldots, X_n)$ is a general data-dependent smoothing parameter.

Theorem 2. *Under assumptions (D), (D'), (F), (F'), (P), (K), (K'), (B), (B'), let us assume that*

$$h(f)^{-1/2}\xi_n + nh^{1/2}\xi_n^\omega = o_p(1). \tag{14}$$

a) If the null hypothesis is true then

$$\upsilon_f^{-1} h(f)^{-1/2}\big(J_n(\hat{h}) - R(K) - c_n(f;K)\big) \xrightarrow{d} N(0,1)$$

where

$$c_n(f;K) = nh(f)R(K_{h(f)} * f - f),$$

and

$$\upsilon_f^2 = 2R(K*K)R(f) + \lambda_f \mu_2(K)^2 \mathrm{Var}_f(\varphi_f(X))$$

with

$$\varphi_f(u) = f''(u) - \sum_k \psi_k(u;\theta_1(f),\theta_2(f)) \int \bar{f}''(x) \frac{\partial g}{\partial \theta_k}(x;\theta_1(f),\theta_2(f))dx,$$

where $\bar{f}(x) = f(-x)$, for $x \in \mathbb{R}$.
b) If the alternative hypothesis is true then

$$(nh(f))^{-1}\big(J_n(\hat{h}) - R(K) - c_n(f;K)\big) \xrightarrow{p} R\big(f - g(\cdot;\theta_1(f),\theta_2(f))\big).$$

Remark 2. Under the conditions of Remark 1, condition (14) holds if $\alpha > \max(1/10, 9/(10\omega))$. Therefore, it is not fulfilled by the least squares cross-validation bandwidth selector, and it holds for the two-stage direct plug-in bandwidth selector whenever $\omega \geq 3$.

3. Bickel-Rosenblatt tests for location-scale families

Under the assumptions of Theorems 1 and 2, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are location-scale and scale equivariant estimators of θ_1 and θ_2, respectively, and the deterministic sequence $h(f)$ is scale equivariant (that is, $h(g) = bh(f)$, where $g(\cdot) = f((\cdot - a)/b)/b$, for all $a \in \mathbb{R}$ and $b > 0$), a property that is satisfied by the exact optimal bandwidth (12), we can easily conclude that $\nu_f^{-1} h(f)^{-1/2} = \nu_{f_0}^{-1} h(f_0)^{-1/2}$, $\upsilon_f^{-1} h(f)^{-1/2} = \upsilon_{f_0}^{-1} h(f_0)^{-1/2}$ and $c_n(f;K) = c_n(f_0;K)$. Therefore, from Theorems 1 and 2 we deduce that the tests based on the critical regions

$$C_n(I_n(\hat{h}),\alpha) = \big\{\nu_{f_0}^{-1} h(f_0)^{-1/2}\big(I_n(\hat{h}) - R(K)\big) > \Phi^{-1}(1-\alpha)\big\}$$

and

$$C_n(J_n(\hat{h}),\alpha) = \big\{\upsilon_{f_0}^{-1} h(f_0)^{-1/2}\big(J_n(\hat{h}) - R(K) - c_n(f_0;K)\big) > \Phi^{-1}(1-\alpha)\big\},$$

where $\alpha \in]0,1[$ and $\Phi^{-1}(1-\alpha)$ is the quantile of order $1-\alpha$ of the standard normal distribution, are asymptotically of level α and consistent to test $f \in \mathscr{F}_0$ against $f \in \mathscr{F} \setminus \mathscr{F}_0$, that is, $P_f\big(C_n(T_n,\alpha)\big) \to \alpha$, for all $f \in \mathscr{F}_0$, and $P_f\big(C_n(T_n,\alpha)\big) \to 1$, for all $f \in \mathscr{F} \setminus \mathscr{F}_0$, where $T_n = T_n(X_1,\ldots,X_n)$ stands for either $I_n(X_1,\ldots,X_n;\hat{h}(X_1,\ldots,X_n))$ or $J_n(X_1,\ldots,X_n;\,\hat{h}(X_1,\ldots,X_n))$.

Such as in the case where $\hat{h}$ is deterministic (see Fan, 1995, p. 372), some simulation results reveal that the asymptotic normal distribution provides a poor approximation to the finite sample distributions of $I_n(\hat{h})$ and $J_n(\hat{h})$ under the null hypothesis, which implies large differences between the true level and the nominal level of the tests based on the previous critical regions. This fact is illustrated in Table 1 where type I error estimates based on 20,000 simulations under the null hypothesis are shown for the normality tests based on the previous critical regions with $K = \phi$ and $\hat{h} = \hat{h}_{H_0}$ given by (10).

In order to circumvent this problem, the standard strategy (see Fan, 1995, pp. 372–373) is to consider instead the test defined by the critical region

$$\mathscr{C}(T_n,\alpha) = \big\{T_n > q(T_n^*,\alpha)\big\},$$

where $T_n = T_n(X_1,\ldots,X_n)$ stands for either $I_n(X_1,\ldots,X_n;\hat{h}(X_1,\ldots,X_n))$ or $J_n(X_1,\ldots,X_n;\,\hat{h}(X_1,\ldots,X_n))$, and $q(T_n^*,\alpha) = q(T_n^*,\alpha;X_1,\ldots,X_n)$ denotes the quantile of order $1-\alpha$ of the distribution of the random variable T_n^* defined as follows:

1) use the original sample $X_1,\ldots,X_n$ to compute $\hat{\theta}_1$ and $\hat{\theta}_2$;

	$I_n(\hat{h}_{H_0})$			$J_n(\hat{h}_{H_0})$		
level α	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^2$	$n = 10^3$	$n = 10^4$
0.01	0.000	0.002	0.006	0.000	0.002	0.008
0.05	0.001	0.006	0.018	0.001	0.008	0.020
0.10	0.001	0.011	0.030	0.002	0.015	0.035

TABLE 1. *Type I error estimates for the normality tests based on the critical regions $C_n(I_n(\hat{h}),\alpha)$ and $C_n(J_n(\hat{h}),\alpha)$, with $K = \phi$ and $\hat{h} = \hat{h}_{H_0}$, for nominal significant levels $\alpha = 0.1, 0.05, 0.01$ and sample sizes $n = 10^k$, $k = 2, 3, 4$. The number of replications for each case is $20,000$.*

2) draw a random sample $U_1, \ldots, U_n$ from the distribution f_0 and define the bootstrap sample by $X^*_{n,i} = \hat{\theta}_1 + \hat{\theta}_2 U_i$, for $i = 1, \ldots, n$;

3) use the bootstrap sample to compute $T_n(X^*_{n,1}, \ldots, X^*_{n,n})$ and call it T^*_n.

Of course, if the test statistic T_n is location-scale invariant, which occurs if we further assume that $\hat{h}$ is scale equivariant, the quantile $q(T^*_n, \alpha)$, which does not depend on the observed sample, is the quantile of order $1 - \alpha$ of T_n under H_0, we denote by $q(T_n, \alpha)$. This quantile is assumed to be a known quantity as is can be well approximated by repeating steps 2) and 3) for a large number of times. As stated in the next result, which proof is presented in Section 6.3, in this important case the test based on the critical region $\mathscr{C}(T_n, \alpha)$, has a level of significance at most equal to α for each sample size n, and is consistent to test $f \in \mathscr{F}_0$ against $f \in \mathscr{F} \setminus \mathscr{F}_0$.

Theorem 3. *Under the assumptions of Theorems 1 or 2, let us assume that $\hat{\theta}_1$ and $\hat{\theta}_2$ are location-scale and scale equivariant estimators of θ_1 and θ_2, respectively. If the bandwidth selector $\hat{h}$ is scale equivariant, then the test statistic T_n, where T_n stands for either $I_n(\hat{h})$ or $J_n(\hat{h})$, is location-scale invariant, and the test based on the critical region*

$$\mathscr{C}(T_n, \alpha) = \big\{T_n > q(T_n, \alpha)\big\}, \tag{15}$$

where $\alpha \in]0, 1[$, is such that

$$P_f\big(\mathscr{C}(T_n, \alpha)\big) \leq \alpha, \text{ for all } f \in \mathscr{F}_0,$$

and

$$P_f\big(\mathscr{C}(T_n, \alpha)\big) \to 1, \text{ for all } f \in \mathscr{F} \setminus \mathscr{F}_0.$$

4. Finite sample results

In this section we conduct a simulation study to compare the finite sample power performance of goodness-of-fit tests based on the critical regions (15) for several choices of the scale equivariant bandwidth selector $\hat{h}$. More precisely, we intend to compare the null hypothesis based bandwidth selector $\hat{h}_{H_0}$ proposed by Bowman (1992) given by (9), with other scale equivariant bandwidth selectors $\hat{h}$ satisfying the relative consistency condition (11), for which it is expected, at least from an asymptotic point of view, that the kernel estimator $f_{\hat{h}}$ is an effective estimator of the underlying density f irrespective of which of the null or the alternative hypothesis is true. To this end, besides $\hat{h}_{H_0}$ three other automatic and scale equivariant bandwidth selectors are considered in our study. They are the least squares cross-validation bandwidth selector $\hat{h}_{\mathrm{CV}}$, the two-stage direct plug-in bandwidth selector $\hat{h}_{\mathrm{PI}}$ (see Wand and Jones, 1995, pp. 63–65, 71–72), and also a modified version of the bandwidth selector proposed in Chacón and Tenreiro (2013), where the cross-validation function is replaced by the weighted cross-validation function with $\gamma = 0.5$ (for the definition of the weighted cross-validation function, see Tenreiro, 2017, p. 3440). Under some conditions on f, $\hat{h}_{\mathrm{CT}}$ fulfils assumption (B) with $h(f) = h_0$ and $\xi_n = O_p\big(n^{-5/14}\big)$ (see Chacón and Tenreiro, 2013, Theorem 3.1, p. 2207). The power results observed in our simulation study for the bandwidths $\hat{h}_{\mathrm{CV}}$, $\hat{h}_{\mathrm{PI}}$ and $\hat{h}_{\mathrm{CT}}$ reveal that this latter bandwidth presents a good overall performance for a wide range of alternative density features, which is relevant for real data situations where there is usually little prior information on the alternative density shape. For this reason, and because no essential feature is lost, hereafter we confine to the results obtained by the bandwidths $\hat{h}_{H_0}$ and $\hat{h}_{\mathrm{CT}}$.

From the representations (5) and (6), and taking for K the standard normal density, which we always assume from now on, the test statistics $I_n(\hat{h})$ and $J_n(\hat{h})$ can be evaluated from the equalities

$$I_n(\hat{h}) = \frac{\tilde{h}}{n} \sum_{i,j=1}^{n} Q(Y_{n,i}, Y_{n,j}; \tilde{h})$$

and

$$J_n(\hat{h}) = \frac{\tilde{h}}{n} \sum_{i,j=1}^{n} R(Y_{n,i}, Y_{n,j}; \tilde{h}),$$

where for $u, v \in \mathbb{R}$ and $h > 0$,

$$Q(u, v; h) = \phi_{\sqrt{2}h}(u - v) - \phi_{\sqrt{2}h} * f_0(u) - \phi_{\sqrt{2}h} * f_0(v) + \phi_{\sqrt{2}h} * \bar{f}_0 * f_0(0)$$

and

$$R(u, v; h) = \phi_{\sqrt{2}h}(u - v) - \phi_h * f_0(u) - \phi_h * f_0(v) + \bar{f}_0 * f_0(0),$$

with $\tilde{h} = \hat{h}/\hat{\theta}_2$, $Y_{n,j} = (X_j - \hat{\theta}_1)/\hat{\theta}_2$, $j = 1, \ldots, n$, and $\bar{f}_0(u) = f_0(-u)$, for $u \in \mathbb{R}$. Taking into account the convolution properties of the Gaussian densities (see Wand and Jones, 1995, pp. 177–180), the calculation of $I_n(\hat{h})$ and $J_n(\hat{h})$ is especially simple for the normality test in which case no numerical integration is required. In this case, we have

$$I_n(\hat{h}) = n\tilde{h} \sum_{k,l=1}^{n+1} w_k \phi_{(\beta_k^2+\beta_l^2)^{1/2}}(\alpha_k - \alpha_l) w_l$$

and

$$J_n(\hat{h}) = n\tilde{h} \sum_{k,l=1}^{n+1} w_k \phi_{(\gamma_k^2+\gamma_l^2)^{1/2}}(\alpha_k - \alpha_l) w_l,$$

where $w = (\frac{1}{n}, \ldots, \frac{1}{n}, -1)$, $\alpha = (Y_{n,1}, \ldots, Y_{n,n}, 0)$, $\beta = \big(\tilde{h}, \ldots, \tilde{h}, (\tilde{h}^2 + 1)^{\frac{1}{2}}\big)$ and $\gamma = (\tilde{h}, \ldots, \tilde{h}, 1)$.

We start the study on the finite sample performance of the tests based on the critical regions (15) for nominal levels $\alpha = 0.1, 0.05, 0.01$, by considering the test of normality in which case the null model is given by (2) with $f_0 = \phi$, and we take $\hat{\theta}_1 = \bar{X}_n$ and $\hat{\theta}_2 = S_n$ the maximum likelihood estimators of θ_1 and θ_2 under H_0. As the test statistics $I_n(\hat{h})$ and $J_n(\hat{h})$ for $\hat{h} = \hat{h}_{H_0}$ and $\hat{h} = \hat{h}_{\mathrm{CT}}$ are invariant under null hypothesis (see Theorem 3), the quantiles of order $1-\alpha$ in the critical regions (15) are estimated by performing 100,000 simulations under the null hypothesis. We consider alternative distributions from a well-known set of normal mixture densities considered in Marron and Wand (1992) which is often used in the context of kernel density estimation. This set is very rich, containing densities with a wide variety of features, such as kurtosis, skewness, multimodality, etc. The densities of the considered alternatives jointly with the density of the normal distribution with the same mean and variance are shown in Figure 1. The densities are identified as in Marron and Wand (1992) and the values for the parameters of this set of normal mixture densities are given in the Table 1 of the same article. For the nominal level $\alpha = 0.05$ and sample sizes $n = 20, 50, 80$ we report in Table

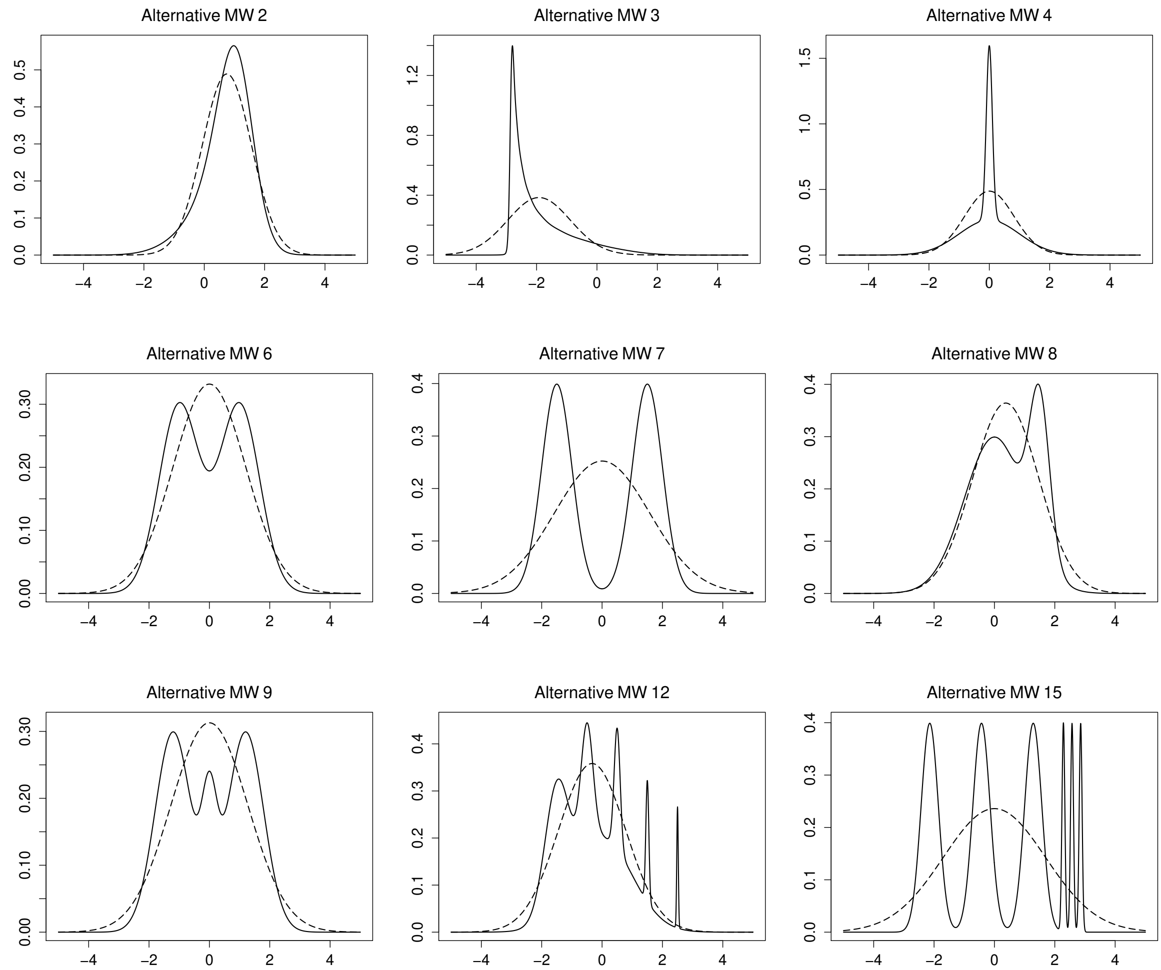

FIGURE 1. *Probability density functions of alternatives from the Marron and Wand (1992) set of normal mixture densities (solid) and the probability function of the Gaussian distribution with the same mean and variance of the considered alternative (dashed).*

2 the power estimates based on 10,000 samples from the considered set of alternative densities. All the simulations in this work were carried out using programs written in the R language (R Development Core Team, 2019).

Taking into account some simulation experiments, not presented here to save space, to estimate the mean integrated squared error of the kernel density estimator for each one of the bandwidths $\hat{h}_{H_0}$ and $\hat{h}_{\mathrm{CT}}$, we can conclude that the kernel estimator based on $\hat{h}_{H_0}$ performs better than that based on $\hat{h}_{\mathrm{CT}}$ for the normal mixture densities $2, 6, 8, 9$ and 12 (for the considered sample sizes). This may explain the results shown in Table 2 where the

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
			Alternative MW 2			
$\hat{h}_{H_0}$	0.18	0.34	0.49	0.08	0.16	0.26
$\hat{h}_{\mathrm{CT}}$	0.13	0.26	0.39	0.11	0.19	0.27
			Alternative MW 3			
$\hat{h}_{H_0}$	0.86	1.00	1.00	0.74	1.00	1.00
$\hat{h}_{\mathrm{CT}}$	0.81	1.00	1.00	0.80	1.00	1.00
			Alternative MW 4			
$\hat{h}_{H_0}$	0.32	0.72	0.93	0.07	0.28	0.63
$\hat{h}_{\mathrm{CT}}$	0.57	0.93	0.99	0.54	0.90	0.99
			Alternative MW 6			
$\hat{h}_{H_0}$	0.16	0.51	0.76	0.32	0.70	0.89
$\hat{h}_{\mathrm{CT}}$	0.13	0.41	0.66	0.20	0.61	0.86
			Alternative MW 7			
$\hat{h}_{H_0}$	0.97	1.00	1.00	1.00	1.00	1.00
$\hat{h}_{\mathrm{CT}}$	0.96	1.00	1.00	0.99	1.00	1.00
			Alternative MW 8			
$\hat{h}_{H_0}$	0.13	0.38	0.59	0.22	0.51	0.72
$\hat{h}_{\mathrm{CT}}$	0.12	0.33	0.53	0.15	0.44	0.66
			Alternative MW 9			
$\hat{h}_{H_0}$	0.22	0.67	0.89	0.42	0.84	0.96
$\hat{h}_{\mathrm{CT}}$	0.18	0.55	0.81	0.26	0.77	0.95
			Alternative MW 12			
$\hat{h}_{H_0}$	0.10	0.21	0.33	0.12	0.25	0.38
$\hat{h}_{\mathrm{CT}}$	0.09	0.20	0.35	0.10	0.24	0.39
			Alternative MW 15			
$\hat{h}_{H_0}$	0.23	0.72	0.95	0.44	0.89	0.99
$\hat{h}_{\mathrm{CT}}$	0.64	1.00	1.00	0.65	1.00	1.00

TABLE 2. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the normal distribution based on $I_n(\hat{h})$ and $J_n(\hat{h})$ with $\hat{h} = \hat{h}_{H_0}$ and $\hat{h} = \hat{h}_{\mathrm{CT}}$, for some alternatives from the Marron and Wand (1992) set of normal mixture densities. The power estimates are based on $10,000$ samples from the considered alternatives.*

tests based on $\hat{h}_{H_0}$ perform generally better than those based on $\hat{h}_{\mathrm{CT}}$ for alternatives $2, 6, 8$ and 9, and they perform similarly for alternative 12. The opposite situation occurs for the remaining four normal mixtures where the kernel density estimator based on $\hat{h}_{\mathrm{CT}}$ performs much better than that based on $\hat{h}_{H_0}$. However, only for the normal mixtures 4 and 15 the tests based on

$\hat{h}_{\mathrm{CT}}$ perform clearly better than those based on $\hat{h}_{H_0}$. For densities 3 and 7 the tests perform similarly. As the considered alternative densities are far from the null hypothesis density family in shape, we can conclude that even a low performing bandwidth selector from a density estimation point of view is good enough to detect such alternatives. In this situation, estimation and testing demand different answers regarding bandwidth selection. The results presented in Table 2 for the skewed unimodal density 2 also deserve an additional comment. This is an interesting case because density 2 is not far from the normal density in shape, and we may expect that $\hat{h}_{H_0}$, as based on the null density family, may reach good power results for alternative densities which are not far from the null density model in shape. The simulations results observed for density 2 support this idea. The results presented in Table 2 also show different performances for the tests based on the test statistics $I_n(\hat{h})$ and $J_n(\hat{h})$ no matter which bandwidth is used. The statistic $J_n(\hat{h})$ seems to be more effective in detecting multimodal alternatives, whereas $I_n(\hat{h})$ shows in general a better performance in the detection of unimodal alternatives.

Based on the previous conclusions, we have good reasons to believe that $\hat{h}_{H_0}$ may reach a good power performance for wide sets of alternative distributions. In order to examine in detail this question, other than the goodness-of-fit test for the normal distribution we also consider two other null location-scale models. They are the logistic model where $f_0(x) = (\exp(-x/2) + \exp(x/2))^{-2}$, for $x \in \mathbb{R}$, and the Gumbel extreme value model where $f_0(x) = \exp(-x - \exp(-x))$, for $x \in \mathbb{R}$. For this latter family of distributions we take for $\hat{\theta}_1$ and $\hat{\theta}_2$ the maximum likelihood estimators of θ_1 and θ_2, which satisfy

$$\hat{\theta}_2 = \bar{X}_n - \frac{\sum_{j=1}^n X_j \exp(-X_j/\hat{\theta}_2)}{\sum_{j=1}^n \exp(-X_j/\hat{\theta}_2)}$$

and

$$\hat{\theta}_1 = -\hat{\theta}_2 \log\left(n^{-1}\sum_{j=1}^n X_j \exp(-X_j/\hat{\theta}_2)\right).$$

In the case of the goodness-of-fit test for the logistic distribution we use the moment estimators $\hat{\theta}_1 = \bar{X}_n$ and $\hat{\theta}_2 = \sqrt{3}\, S_n/\pi$ which are simpler to evaluate and nearly as efficient as the maximum likelihood estimators (see Johnson et al., 1995, pp. 127–130). Similarly to the goodness-of-fit test for the normal distribution, we are under the assumptions of Theorem 3 and the tests based on the critical regions (15) are implemented as explained before.

For comparison proposes, besides the bandwidth selectors $\hat{h}_{H_0}$ and $\hat{h}_{\mathrm{CT}}$, we consider in this study other bandwidth selectors which are based on the common principle that the bandwidth should be tuned in order to improve the power performance of the test. In order to implement this idea, we consider the set of scale equivariant bandwidths based on $\hat{h}$, where $\hat{h}$ stands for $\hat{h}_{H_0}$ or $\hat{h}_{\mathrm{CT}}$, given by

$$\hat{h}_\lambda(X_1, \ldots, X_n) = \lambda \hat{h}(X_1, \ldots, X_n), \ \lambda \in \Lambda,$$

where Λ is a finite set of strictly positive real numbers that will act as tuning parameters. Besides the value $\lambda = 1$ associated to the reference bandwidth $\hat{h}$, this set is meant to include tuning parameters smaller and larger than the unit. If we denote by $T_{n,\lambda}(X_1, \ldots, X_n)$ one the statistics $I_n(X_1, \ldots, X_n; \hat{h}_\lambda(X_1, \ldots, X_n))$ or $J_n(X_1, \ldots, X_n; \hat{h}_\lambda(X_1, \ldots, X_n))$, from the scale equivariant property of $\hat{h}$ we know that $T_{n,\lambda}$ is location-scale invariant, and therefore the null distribution of $T_{n,\lambda}$ does not depend on $f \in \mathscr{F}_0$, where $\mathscr{F}_0$ is given by (2). Therefore, the tests with critical regions

$$\mathscr{C}(T_{n,\lambda}, \alpha) = \{T_{n,\lambda} > q(T_{n,\lambda}, \alpha)\}, \ \lambda \in \Lambda, \tag{16}$$

where $q(T_{n,\lambda}, \alpha)$ denotes the quantile of order $1 - \alpha$ of $T_{n,\lambda}$ under H_0, have levels of significance at most equal to α. As before, we assumed that these quantiles are known quantities as they can be well approximated by simulating under the null hypothesis for a large number of times (100,000 replications under the null hypothesis are used). The power properties of each one of the previous test procedures depend on λ which is the reason that its choice is usually crucial to obtain a performing test procedure. In order to make such a choice, we need to define a suitable location-scale invariant measurable function taking values in Λ, $\hat{\lambda} = \hat{\lambda}(X_1, \ldots, X_n)$, called tuning parameter selector, on the basis of which we can consider a test procedure based on the critical region

$$\mathscr{C}(T_{n,\hat{\lambda}}, \alpha) = \{T_{n,\hat{\lambda}} > q(T_{n,\hat{\lambda}}, \alpha)\}, \tag{17}$$

where $q(T_{n,\hat{\lambda}}, \alpha)$ denotes the quantile of order $1 - \alpha$ of $T_{n,\hat{\lambda}}$ under H_0. This test has a level of significance at most equal to α for each sample size n.

In order to define efective methods for selecting the tuning parameter $\lambda \in \Lambda$, we will adapt to our situation three methods considered in Cao and Van Keilegom (2006), Martínez-Camblor et al. (2008), and Martínez-Camblor and Uña-Álvarez (2013) in the context of smooth tests for the k-sample problem.

Given the level α of the test, and a sample $X_1, \ldots, X_n$ from f, the first tuning parameter selector we consider, we denote by $\hat{\lambda}_1 = \hat{\lambda}_1(X_1, \ldots, X_n; \alpha)$, was originally used in Cao and Van Keilegom (2006, p. 69) and is defined as the value in Λ that maximises the smooth bootstrap power, that is,

$$\hat{\lambda}_1 = \operatorname*{argmax}_{\lambda \in \Lambda} \frac{1}{B_1} \sum_{k=1}^{B_1} I\big(T_{n,\lambda}(X^*_{k,1}, \ldots, X^*_{k,n}) > q(T_{n,\lambda}, \alpha)\big), \tag{18}$$

with

$$X^*_{k,j} = X_{U_{(k-1)n+j}} + \hat{h}_{\mathrm{CT}}(X_1, \ldots, X_n) N_{(k-1)n+j},$$

for $k = 1, \ldots, B_1$ and $j = 1, \ldots, n$, where N_l, for $l = 1, \ldots, nB_1$, are independent copies of the standard normal distribution, and U_l, for $l = 1, \ldots, nB_1$, are independent copies of the discrete uniform distribution on $\{1, \ldots, n\}$; that is, for each $k = 1, \ldots, B_1$, $X^*_{k1}, \ldots, X^*_{kn}$ is generated by resampling from the Parzen-Rosenblatt estimator with Gaussian kernel and smoothing parameter $\hat{h}_{\mathrm{CT}}(X_1, \ldots, X_n)$. As expressed by the notation $\hat{\lambda}_1(X_1, \ldots, X_n; \alpha)$, note that $\hat{\lambda}_1$ depends on the considered level α.

The second method for selecting λ we consider is based on the observation that given the values $T_{n,\lambda}(X_1, \ldots, X_n)$ of the test statistics for the observed sample $X_1, \ldots, X_n$, more evidence against the null hypothesis is obtained for smaller p-values. Therefore, to construct a powerful test it makes sense to minimise the bootstrap p-value along $\lambda \in \Lambda$, an idea that was used in Martínez-Camblor et al. (2008, pp. 4014–4015); see also Martínez-Camblor and Uña-Álvarez (2009). Hence, we denote by $\hat{\lambda}_2 = \hat{\lambda}_2(X_1, \ldots, X_n)$ the tuning parameter selector given by

$$\hat{\lambda}_2 = \operatorname*{argmin}_{\lambda \in \Lambda} \frac{1}{B_0} \sum_{j=1}^{B_0} I\big(T_{n,\lambda}(X_{0,(j-1)n+1}, \ldots, X_{0,jn}) \geq T_{n,\lambda}(X_1, \ldots, X_n)\big), \tag{19}$$

where $X_{0,l}$, for $l = 1, \ldots, nB_0$, are independent copies of the random variable X_0 with density f_0.

The last method for selecting λ we consider was introduced in Martínez-Camblor and Uña-Álvarez (2013, p. 273) and is based on the idea that λ should be chosen in order to maximise the discrimination capability, between the null and the alternative hypotheses, of the diagnostic variable $T_{n,\lambda}$ expressed by the area under the ROC curve associated to it. As this area is given by $P(T^0_{n,\lambda} < T^1_{n,\lambda})$, where $T^0_{n,\lambda}$ and $T^1_{n,\lambda}$ are independent random variable with the null and the alternative distributions of $T_{n,\lambda}$, respectively (see

Krzanowski and Hand, 2009, pp. 26–28), we consider the tuning parameter selector $\hat{\lambda}_3 = \hat{\lambda}_3(X_1, \ldots, X_n)$ defined by

$$\hat{\lambda}_3 = \underset{\lambda \in \Lambda}{\operatorname{argmax}} \frac{1}{B_0 B_1} \sum_{j=1}^{B_0} \sum_{k=1}^{B_1} I\big(T_{n,\lambda}(X_{0,(j-1)n+1}, \ldots, X_{0,jn}) < T_{n,\lambda}(X^*_{k,1}, \ldots, X^*_{k,n})\big), \qquad (20)$$

where $X_{0,l}$, for $l = 1, \ldots, nB_0$, and $X^*_{k,j}$, for $k = 1, \ldots, B_1$ and $j = 1, \ldots, n$, are defined as before.

Taking into account that, conditionally on the sequences $N = (N_l)$, $U = (U_l)$ and $X_0 = (X_{0,l})$, the previous tuning parameter selectors are location-scale invariants, we conclude that the tests based on the critical region (17), where $\hat{\lambda}$ stands for either $\hat{\lambda}_1$, $\hat{\lambda}_2$ or $\hat{\lambda}_3$, have levels of significance at most equal to α for each sample size n (conditionally on N, U and X_0). In the practical implementation of these tests we always take $\Lambda = \{0.5, 0.75, 1, 1.5, 2\}$. For the normality goodness-of-fit tests we take $B_0 = B_1 = 200$, and the quantiles $q(T_{n,\hat{\lambda}}, \alpha)$ are estimated by performing 100,000 simulations under the null hypothesis. For the goodness-of-fit tests for the logistic and Gumbel distributions we take $B_0 = B_1 = 100$, and the quantiles are estimated by performing 50,000 simulations under the null hypothesis, because the evaluation of the corresponding test statistics is more time-consuming than in the normal case.

For the alternatives 4 and 8 from the Marron and Wand (1992) set of normal mixture densities, Tables 3 and 4 present power estimates, at level $\alpha = 0.05$, for the normality goodness-of-fit tests based on the critical regions (16) with $\hat{h}_\lambda = \lambda \hat{h}$, $\lambda = 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 5$, and (17) with $\Lambda = \{0.5, 0.75, 1, 1.5, 2\}$, where $\hat{h} = \hat{h}_{H_0}, \hat{h}_{\mathrm{CT}}$. As mentioned before, for all samples sizes we see that the empirical power depends on λ. However, these two alternatives reveal different situations. For alternative 8 the best power results are in general observed for values of λ close or even equal to 1, and therefore the tests based on $\hat{\lambda}_j \hat{h}$, for $j = 1, 2, 3$, are not expected to be more powerful than those based on the bandwidth selector $\hat{h}$. The figures in both tables support this idea. A similar situation occurs for alternative 4 and bandwidth $\hat{h}_{\mathrm{CT}}$. However, when the bandwidth $\hat{h}_{H_0}$ is used for alternative 4, an alternative for which the kernel estimator based on $\hat{h}_{H_0}$ performs poorly from a density estimation point of view, we see that it is highly advisable to use a tuning parameter smaller than 1, which may explain the good results obtained by the tuning parameters selectors $\hat{\lambda}_2$ and $\hat{\lambda}_3$ for the test based on I_n and by $\hat{\lambda}_1$ for the test based on J_n.

For $\alpha = 0.01, 0.05, 0.1$, and sample sizes $n = 20, 50, 80$, we present in Tables 5, 9 and 10, estimates of the nominal levels of significance for the goodness-of-fit tests for the normal, logistic and Gumbel distributions, respectively, based on $I_n(\hat{h})$ and $J_n(\hat{h})$ for the different bandwidth selectors $\hat{h}$ based on $\hat{h}_{H_0}$ and $\hat{h}_{\mathrm{CT}}$. They are based on $20,000$ simulations under the null hypothesis. These results indicate that all the tests have an effective level of significance very close to α. With some few exceptions, the estimated levels are inside the approximate 95% confidence interval for the preassigned nominal levels.

Although a larger set of alternative distributions, usually considered in power studies for testing the normal, logistic and Gumbel distributions was considered in our study (see Epps and Pulley, 1983, Meintanis, 2004, Epps, 2005, Romão et al., 2010), we limit ourselves to present in Tables 6–8 (normal distribution), Tables 11–13 (logistic distribution) and Tables 14–16 (Gumbel distribution) the empirical power results for some of these alternatives. The first seven alternatives are from the following location-scale families: uniform, exponential, Laplace, Cauchy, normal, logistic and Gumbel. The remaining six alternatives are from the following families of distributions: Student, lognormal, Tukey, gamma, Weibull and beta. For the exact definition of the distributions included in these tables, see Epps (2005). We limit ourselves to present here the results obtained for the nominal level $\alpha = 0.05$ and sample sizes $n = 20, 50, 80$. However, similar conclusions can be drawn for the nominal levels $\alpha = 0.1, 0.01$ also considered in our study. For comparison purposes, we have include in the previous tables power estimates for the classical Anderson-Darling (1954) goodness-of-fit test which is based on a weighted quadratic distance between the empirical distribution function and a parametric estimator of the distribution function of f under the null hypothesis (see Stephens, 1986, and the references therein, for goodness-of-fit tests based on the empirical distribution function). In order to implement this test, the quantiles of order $1 - \alpha$ of the Anderson-Darling test statistic A^2 are estimated by performing 100,000 simulations under the null hypothesis. In the case of the goodness-of-fit test for the normal distribution we also include in our simulation study the highly recommended Shapiro-Wilk (1965) test SW implemented by the R-function `shapiro.test`. For all the tests

	$I_n(\lambda\hat{h}_{H_0})$			$J_n(\lambda\hat{h}_{H_0})$		
λ	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
	Alternative MW 4					
0.25	0.54	0.94	0.99	0.54	0.93	0.99
0.50	0.46	0.89	0.99	0.42	0.87	0.98
0.75	0.37	0.81	0.97	0.24	0.68	0.93
1.00	0.32	0.72	0.93	0.07	0.28	0.63
1.25	0.28	0.63	0.87	0.02	0.02	0.06
1.50	0.26	0.56	0.81	0.01	0.00	0.00
1.75	0.24	0.49	0.74	0.01	0.00	0.00
2.00	0.23	0.43	0.68	0.01	0.00	0.00
3.00	0.20	0.31	0.46	0.00	0.00	0.00
5.00	0.19	0.23	0.29	0.00	0.00	0.00
$\hat{\lambda}_1$	0.35	0.75	0.94	0.21	0.50	0.74
$\hat{\lambda}_2$	0.44	0.90	0.98	0.02	0.07	0.09
$\hat{\lambda}_3$	0.49	0.90	0.99	0.02	0.14	0.34
	Alternative MW 8					
0.25	0.15	0.33	0.49	0.15	0.33	0.49
0.50	0.17	0.40	0.59	0.18	0.43	0.62
0.75	0.16	0.40	0.61	0.20	0.49	0.69
1.00	0.13	0.38	0.59	0.22	0.51	0.72
1.25	0.11	0.35	0.56	0.23	0.51	0.71
1.50	0.08	0.31	0.52	0.23	0.49	0.68
1.75	0.07	0.27	0.49	0.22	0.46	0.65
2.00	0.07	0.23	0.44	0.21	0.44	0.61
3.00	0.06	0.15	0.29	0.21	0.39	0.53
5.00	0.05	0.12	0.21	0.20	0.37	0.48
$\hat{\lambda}_1$	0.13	0.35	0.56	0.17	0.41	0.61
$\hat{\lambda}_2$	0.16	0.37	0.54	0.14	0.10	0.08
$\hat{\lambda}_3$	0.14	0.24	0.39	0.11	0.08	0.13

TABLE 3. *Power estimates, at level $\alpha = 0.05$, for the normality goodness-of-fit tests based on $I_n(\lambda\hat{h}_{H_0})$ and $J_n(\lambda\hat{h}_{H_0})$, with $\lambda = 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 5$, and $I_n(\hat{\lambda}_j\hat{h}_{H_0})$ and $J_n(\hat{\lambda}_j\hat{h}_{H_0})$, $j = 1, 2, 3$, with $\Lambda = \{0.5, 0.75, 1, 1.5, 2\}$, for alternatives 4 and 8 from the Marron and Wand (1992) set of normal mixture densities. The number of replications used is $10,000$.*

included in the study, the power estimates are based on 10,000 samples from the considered alternatives.

Although none of the considered tests present uniformly better results for the considered set of alternative distributions, the main conclusion that can be drawn from this study is that the tests based on $\hat{h}_{H_0}$ present in fact a good

λ	$I_n(\lambda\hat{h}_{\mathrm{CT}})$ $n=20$	$n=50$	$n=80$	$J_n(\lambda\hat{h}_{\mathrm{CT}})$ $n=20$	$n=50$	$n=80$
	Alternative MW 4					
0.25	0.50	0.92	0.99	0.50	0.92	0.99
0.50	0.55	0.93	0.99	0.54	0.93	0.99
0.75	0.56	0.93	0.99	0.54	0.92	0.99
1.00	0.57	0.93	0.99	0.54	0.90	0.99
1.25	0.57	0.93	0.99	0.37	0.78	0.95
1.50	0.58	0.93	0.99	0.13	0.45	0.76
1.75	0.58	0.93	0.99	0.02	0.12	0.34
2.00	0.58	0.93	0.99	0.01	0.01	0.05
3.00	0.58	0.92	0.99	0.00	0.00	0.00
5.00	0.57	0.91	0.99	0.00	0.00	0.00
$\hat{\lambda}_1$	0.56	0.93	0.99	0.48	0.86	0.98
$\hat{\lambda}_2$	0.48	0.91	0.99	0.01	0.03	0.11
$\hat{\lambda}_3$	0.47	0.87	0.98	0.10	0.86	0.91
	Alternative MW 8					
0.25	0.14	0.31	0.48	0.15	0.32	0.48
0.50	0.15	0.34	0.53	0.15	0.36	0.55
0.75	0.13	0.34	0.54	0.15	0.40	0.61
1.00	0.12	0.33	0.53	0.15	0.44	0.66
1.25	0.11	0.31	0.51	0.08	0.28	0.51
1.50	0.10	0.29	0.49	0.04	0.04	0.11
1.75	0.09	0.27	0.47	0.03	0.02	0.02
2.00	0.08	0.25	0.45	0.03	0.02	0.01
3.00	0.07	0.20	0.37	0.03	0.02	0.01
5.00	0.07	0.16	0.30	0.03	0.02	0.01
$\hat{\lambda}_1$	0.12	0.33	0.53	0.12	0.33	0.51
$\hat{\lambda}_2$	0.15	0.35	0.51	0.02	0.01	0.00
$\hat{\lambda}_3$	0.16	0.30	0.46	0.04	0.22	0.17

TABLE 4. *Power estimates, at level* $\alpha = 0.05$*, for the normality goodness-of-fit tests based on* $I_n(\lambda\hat{h}_{\mathrm{CT}})$ *and* $J_n(\lambda\hat{h}_{\mathrm{CT}})$*, with* $\lambda = 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 5$*, and* $I_n(\hat{\lambda}_j\hat{h}_{\mathrm{CT}})$ *and* $J_n(\hat{\lambda}_j\hat{h}_{\mathrm{CT}})$*,* $j = 1, 2, 3$*, with* $\Lambda = \{0.5, 0.75, 1, 1.5, 2\}$*, for alternatives 4 and 8 from the Marron and Wand (1992) set of normal mixture densities. The number of replications used is* $10,000$*.*

overall performance for a wide set of alternative distributions. Regarding the two tests based on $\hat{h}_{H_0}$, our preference goes to the test based on the test statistic $I_n(\hat{h}_{H_0})$. This test is in general more powerful than, or at least as powerful as, the tests based on $\hat{h}_{\mathrm{CT}}$, and also proves to be quite competitive against the Anderson-Darling test, although slightly less performing than

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
			$\alpha=0.01$			
$\hat{h}_{H_0}$	0.009	0.010	0.011	0.009	0.010	0.011
$\hat{\lambda}_1\hat{h}_{H_0}$	0.010	0.010	0.009	0.009	0.010	0.010
$\hat{\lambda}_2\hat{h}_{H_0}$	0.010	0.010	0.010	0.010	0.010	0.010
$\hat{\lambda}_3\hat{h}_{H_0}$	0.011	0.010	0.011	0.011	0.010	0.011
$\hat{h}_{\mathrm{CT}}$	0.010	0.011	0.010	0.010	0.011	0.011
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.010	0.011	0.010	0.009	0.011	0.010
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.010	0.011	0.010	0.010	0.010	0.011
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.011	0.011	0.011	0.010	0.010	0.009
			$\alpha=0.05$			
$\hat{h}_{H_0}$	0.049	0.051	0.049	0.047	0.049	0.050
$\hat{\lambda}_1\hat{h}_{H_0}$	0.051	0.050	0.048	0.050	0.048	0.049
$\hat{\lambda}_2\hat{h}_{H_0}$	0.049	0.050	0.049	0.050	0.051	0.048
$\hat{\lambda}_3\hat{h}_{H_0}$	0.050	0.049	0.049	0.050	0.053	0.050
$\hat{h}_{\mathrm{CT}}$	0.051	0.050	0.050	0.049	0.050	0.050
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.051	0.050	0.050	0.051	0.049	0.050
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.049	0.050	0.047	0.051	0.049	0.052
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.049	0.049	0.049	0.051	0.051	0.050
			$\alpha=0.10$			
$\hat{h}_{H_0}$	0.101	0.103	0.098	0.099	0.101	0.100
$\hat{\lambda}_1\hat{h}_{H_0}$	0.100	0.101	0.096	0.099	0.101	0.103
$\hat{\lambda}_2\hat{h}_{H_0}$	0.099	0.100	0.096	0.099	0.101	0.100
$\hat{\lambda}_3\hat{h}_{H_0}$	0.102	0.101	0.099	0.098	0.103	0.102
$\hat{h}_{\mathrm{CT}}$	0.100	0.101	0.097	0.100	0.099	0.099
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.100	0.100	0.096	0.101	0.102	0.101
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.102	0.099	0.096	0.102	0.102	0.102
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.100	0.101	0.098	0.103	0.099	0.100

TABLE 5. *Estimates of the nominal levels of significance, for a preassigned level α, for the goodness-of-fit tests for the* normal distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors. The number of replications for each case is $20,000$.*

the Shapiro-Wilk test for normality. However, no matter the considered null hypothesis model, for some of the considered alternatives, such as the light-tailed alternatives uniform and beta, the test based on $J_n(\hat{h}_{H_0})$ shows to be more powerful than that based on $I_n(\hat{h}_{H_0})$. Finally, note that the new bandwidth selectors $\hat{\lambda}_j\hat{h}_{H_0}$ or $\hat{\lambda}_j\hat{h}_{\mathrm{CT}}$, for $j=1,2,3$, which are much more time-consuming to compute than $\hat{h}_{H_0}$ or $\hat{h}_{\mathrm{CT}}$, do not reveal in general any

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
			Uniform			
$\hat{h}_{H_0}$	0.19	0.62	0.87	0.39	0.83	0.97
$\hat{\lambda}_1\hat{h}_{H_0}$	0.18	0.59	0.85	0.29	0.73	0.93
$\hat{\lambda}_2\hat{h}_{H_0}$	0.21	0.52	0.76	0.30	0.23	0.11
$\hat{\lambda}_3\hat{h}_{H_0}$	0.19	0.40	0.67	0.22	0.33	0.71
$\hat{h}_{\mathrm{CT}}$	0.16	0.49	0.77	0.23	0.77	0.95
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.16	0.49	0.77	0.17	0.59	0.87
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.20	0.49	0.76	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.21	0.47	0.73	0.02	0.38	0.41
A^2	0.21	0.61	0.88	0.21	0.61	0.88
SW	0.21	0.74	0.97	0.21	0.74	0.97
			Exponential			
$\hat{h}_{H_0}$	0.76	0.99	1.00	0.58	0.97	1.00
$\hat{\lambda}_1\hat{h}_{H_0}$	0.70	0.98	1.00	0.49	0.94	1.00
$\hat{\lambda}_2\hat{h}_{H_0}$	0.49	0.96	1.00	0.02	0.07	0.40
$\hat{\lambda}_3\hat{h}_{H_0}$	0.26	0.56	0.93	0.08	0.30	0.63
$\hat{h}_{\mathrm{CT}}$	0.66	0.98	1.00	0.64	0.98	1.00
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.65	0.98	1.00	0.57	0.96	1.00
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.55	0.95	1.00	0.01	0.07	0.23
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.45	0.89	0.99	0.17	0.94	0.98
A^2	0.78	1.00	1.00	0.78	1.00	1.00
SW	0.83	1.00	1.00	0.83	1.00	1.00
			Laplace			
$\hat{h}_{H_0}$	0.24	0.51	0.71	0.05	0.11	0.24
$\hat{\lambda}_1\hat{h}_{H_0}$	0.22	0.47	0.68	0.08	0.17	0.28
$\hat{\lambda}_2\hat{h}_{H_0}$	0.15	0.40	0.61	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{H_0}$	0.15	0.30	0.45	0.01	0.01	0.01
$\hat{h}_{\mathrm{CT}}$	0.24	0.50	0.69	0.19	0.37	0.53
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.23	0.48	0.68	0.15	0.29	0.42
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.14	0.33	0.51	0.02	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.12	0.26	0.43	0.03	0.22	0.15
A^2	0.24	0.52	0.72	0.24	0.52	0.72
SW	0.25	0.52	0.71	0.25	0.52	0.71
			Cauchy			
$\hat{h}_{H_0}$	0.87	1.00	1.00	0.65	0.97	1.00
$\hat{\lambda}_1\hat{h}_{H_0}$	0.86	1.00	1.00	0.69	0.97	1.00
$\hat{\lambda}_2\hat{h}_{H_0}$	0.79	0.99	1.00	0.20	0.75	0.95
$\hat{\lambda}_3\hat{h}_{H_0}$	0.66	0.98	1.00	0.55	0.92	0.99
$\hat{h}_{\mathrm{CT}}$	0.87	1.00	1.00	0.85	0.99	1.00
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.87	1.00	1.00	0.82	0.99	1.00
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.80	0.99	1.00	0.38	0.82	0.96
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.78	0.99	1.00	0.58	0.99	1.00
A^2	0.88	1.00	1.00	0.88	1.00	1.00
SW	0.87	1.00	1.00	0.87	1.00	1.00

TABLE 6. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* normal distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2 and SW. The power estimates are based on $10,000$ samples from the considered alternatives.*

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n = 20$	$n = 50$	$n = 80$	$n = 20$	$n = 50$	$n = 80$
			Logistic			
$\hat{h}_{H_0}$	0.10	0.13	0.17	0.03	0.02	0.02
$\hat{\lambda}_1\hat{h}_{H_0}$	0.09	0.12	0.15	0.04	0.04	0.03
$\hat{\lambda}_2\hat{h}_{H_0}$	0.06	0.09	0.11	0.02	0.01	0.01
$\hat{\lambda}_3\hat{h}_{H_0}$	0.06	0.07	0.08	0.03	0.01	0.00
$\hat{h}_{\mathrm{CT}}$	0.09	0.12	0.15	0.07	0.07	0.07
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.08	0.11	0.15	0.06	0.05	0.05
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.06	0.07	0.08	0.04	0.02	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.06	0.05	0.06	0.04	0.04	0.01
A^2	0.10	0.14	0.19	0.10	0.14	0.19
SW	0.11	0.19	0.26	0.11	0.19	0.26
			Gumbel			
$\hat{h}_{H_0}$	0.28	0.57	0.76	0.14	0.36	0.55
$\hat{\lambda}_1\hat{h}_{H_0}$	0.25	0.51	0.70	0.12	0.27	0.45
$\hat{\lambda}_2\hat{h}_{H_0}$	0.12	0.31	0.57	0.03	0.01	0.00
$\hat{\lambda}_3\hat{h}_{H_0}$	0.08	0.06	0.13	0.03	0.01	0.01
$\hat{h}_{\mathrm{CT}}$	0.19	0.43	0.65	0.17	0.36	0.53
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.18	0.42	0.62	0.13	0.26	0.40
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.12	0.27	0.44	0.02	0.01	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.10	0.17	0.31	0.03	0.19	0.15
A^2	0.27	0.60	0.81	0.27	0.60	0.81
SW	0.31	0.69	0.89	0.31	0.69	0.89
			Student(6)			
$\hat{h}_{H_0}$	0.13	0.20	0.27	0.04	0.04	0.05
$\hat{\lambda}_1\hat{h}_{H_0}$	0.12	0.18	0.25	0.06	0.05	0.06
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.12	0.18	0.02	0.01	0.01
$\hat{\lambda}_3\hat{h}_{H_0}$	0.07	0.08	0.10	0.02	0.01	0.01
$\hat{h}_{\mathrm{CT}}$	0.11	0.17	0.23	0.09	0.11	0.12
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.11	0.17	0.22	0.08	0.08	0.08
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.07	0.10	0.13	0.03	0.02	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.06	0.07	0.10	0.03	0.06	0.02
A^2	0.13	0.22	0.30	0.13	0.22	0.30
SW	0.15	0.28	0.39	0.15	0.28	0.39
			Lognormal(0.5)			
$\hat{h}_{H_0}$	0.48	0.85	0.97	0.28	0.69	0.90
$\hat{\lambda}_1\hat{h}_{H_0}$	0.43	0.80	0.94	0.23	0.59	0.84
$\hat{\lambda}_2\hat{h}_{H_0}$	0.23	0.64	0.90	0.03	0.009	0.04
$\hat{\lambda}_3\hat{h}_{H_0}$	0.11	0.16	0.47	0.04	0.04	0.09
$\hat{h}_{\mathrm{CT}}$	0.35	0.75	0.93	0.32	0.70	0.89
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.34	0.74	0.92	0.26	0.59	0.81
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.24	0.59	0.83	0.01	0.006	0.02
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.18	0.45	0.73	0.04	0.50	0.55
A^2	0.47	0.87	0.98	0.47	0.87	0.98
SW	0.52	0.93	0.99	0.52	0.93	0.99

TABLE 7. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* normal distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2 and SW. The power estimates are based on $10,000$ samples from the considered alternatives.*

$\hat{h}$	$I_n(\hat{h})$ $n=20$	$n=50$	$n=80$	$J_n(\hat{h})$ $n=20$	$n=50$	$n=80$
	Tukey(5)					
$\hat{h}_{H_0}$	0.08	0.20	0.37	0.03	0.05	0.11
$\hat{\lambda}_1\hat{h}_{H_0}$	0.10	0.23	0.40	0.07	0.14	0.22
$\hat{\lambda}_2\hat{h}_{H_0}$	0.15	0.38	0.58	0.02	0.01	0.01
$\hat{\lambda}_3\hat{h}_{H_0}$	0.17	0.44	0.64	0.02	0.02	0.01
$\hat{h}_{\mathrm{CT}}$	0.17	0.37	0.56	0.15	0.30	0.45
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.16	0.37	0.56	0.12	0.22	0.35
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.14	0.34	0.51	0.03	0.01	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.14	0.38	0.57	0.03	0.19	0.11
A^2	0.10	0.24	0.41	0.10	0.24	0.41
SW	0.07	0.12	0.26	0.07	0.12	0.26
	Gamma(4,1)					
$\hat{h}_{H_0}$	0.27	0.55	0.75	0.15	0.38	0.60
$\hat{\lambda}_1\hat{h}_{H_0}$	0.24	0.49	0.69	0.12	0.29	0.48
$\hat{\lambda}_2\hat{h}_{H_0}$	0.11	0.30	0.55	0.04	0.01	0.01
$\hat{\lambda}_3\hat{h}_{H_0}$	0.08	0.06	0.10	0.04	0.01	0.01
$\hat{h}_{\mathrm{CT}}$	0.18	0.42	0.64	0.17	0.36	0.56
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.17	0.40	0.61	0.12	0.26	0.41
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.12	0.26	0.41	0.02	0.01	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.09	0.15	0.27	0.02	0.18	0.13
A^2	0.26	0.58	0.81	0.26	0.58	0.81
SW	0.30	0.69	0.90	0.30	0.69	0.90
	Weibul(2,1)					
$\hat{h}_{H_0}$	0.14	0.29	0.45	0.11	0.25	0.39
$\hat{\lambda}_1\hat{h}_{H_0}$	0.12	0.26	0.40	0.09	0.18	0.28
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.15	0.26	0.07	0.05	0.03
$\hat{\lambda}_3\hat{h}_{H_0}$	0.06	0.05	0.05	0.07	0.04	0.03
$\hat{h}_{\mathrm{CT}}$	0.10	0.19	0.32	0.10	0.20	0.33
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.09	0.19	0.31	0.08	0.14	0.21
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.08	0.13	0.18	0.03	0.02	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.07	0.07	0.10	0.03	0.08	0.04
A^2	0.13	0.31	0.50	0.13	0.31	0.50
SW	0.15	0.41	0.66	0.15	0.41	0.66
	Beta(2,3)					
$\hat{h}_{H_0}$	0.08	0.21	0.33	0.14	0.32	0.47
$\hat{\lambda}_1\hat{h}_{H_0}$	0.08	0.19	0.30	0.11	0.24	0.37
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.13	0.20	0.12	0.18	0.20
$\hat{\lambda}_3\hat{h}_{H_0}$	0.08	0.07	0.10	0.11	0.15	0.23
$\hat{h}_{\mathrm{CT}}$	0.07	0.13	0.22	0.08	0.24	0.42
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.07	0.13	0.22	0.08	0.17	0.27
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.08	0.13	0.18	0.04	0.03	0.02
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.08	0.10	0.15	0.04	0.08	0.07
A^2	0.09	0.19	0.32	0.09	0.19	0.32
SW	0.07	0.20	0.39	0.07	0.20	0.39

TABLE 8. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* **normal distribution** *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2 and SW. The power estimates are based on $10,000$ samples from the considered alternatives.*

special advantage over these simple to compute bandwidths, the exception being the Tukey(5) alternative distribution for the normal and the logistic models. As some simulation experiments reveal (not presented here), the extra source of variation they introduce into the null hypothesis distribution of the associated test statistics, especially those based on J_n, may explain the observed results.

5. Conclusions

The choice of the bandwidth is crucial to the performance of the Parzen-Rosenblatt estimator and several automatic bandwidth selectors considered in the literature satisfy the relative consistency condition (11). This is not the case of the null hypothesis based bandwidth selector $\hat{h}_{H_0}$ that only satisfies this condition when the null hypothesis is true. However, if we want to use the Bickel-Rosenblatt test statistics to test the hypothesis that the underlying density function f is a member of a location-scale family of probability density functions, the finite sample results presented in this paper support the idea that the tests based on $\hat{h}_{H_0}$ present a good overall performance for a wide set of alternative distributions. These tests are in general more powerful than, or at least as powerful as, those based on data-dependent smoothing parameters $\hat{h}$ that satisfy the relative consistency condition irrespective of which of the null or the alternative hypothesis is true, as well as those inspired on existing data-driven bandwidths for smooth tests for the k-sample problem which can be computed by resampling.

6. Proofs

6.1. Proof of Theorem 1. Consider the expansion

$$\begin{aligned}(n\hat{h})^{-1}I_n(\hat{h}) &= \int\{f_{\hat{h}}(x) - K_{\hat{h}} * f(x)\}^2dx \\ &\quad + \int\{K_{\hat{h}} * f(x) - K_{\hat{h}} * g(x;\hat{\theta}_1,\hat{\theta}_2)\}^2dx \\ &\quad + 2\int\{f_{\hat{h}}(x) - K_{\hat{h}} * f(x)\}\{K_{\hat{h}} * f(x) - K_{\hat{h}} * g(x;\hat{\theta}_1,\hat{\theta}_2)\}dx \\ &=: I_{n,1} + I_{n,2} + 2I_{n,3}. \end{aligned} \tag{21}$$

In order to establish the asymptotic behaviour of each one of the previous terms, we use the approach of Tenreiro (2001), which is based on the Taylor expansion

$$K_h(u) := W(u,h) = \sum_{\ell=0}^{\omega-1}(h-1)^\ell K^{\partial(\ell)}(u) + (h-1)^\omega K^{\partial(\omega)}(u,h),$$

where $u \in \mathbb{R}$, $h > 0$,

$$K^{\partial(\ell)}(u) := \frac{1}{\ell!}\frac{\partial^\ell W}{\partial h^\ell}(u,1),\ \ell = 0,\dots,\omega-1,$$

and

$$K^{\partial(\omega)}(u,h) := \frac{1}{(\omega-1)!}\int_0^1 (1-t)^{\omega-1}\frac{\partial^\omega W}{\partial h^\omega}(u,1+t(h-1))dt.$$

Note that, from assumption (K) the functions $K^{\partial(\ell)}$ are bounded and integrable on $\mathbb{R}$, for $\ell = 1,\dots,\omega-1$, and there exists $\eta \in]0,1[$ such that the function $K^{\partial(\omega),\eta}(u) := \sup_{|h-1|\le\eta}|K^{\partial(\omega)}(u,h)|$, is bounded and integrable on $\mathbb{R}$. From the previous Taylor expansion we deduce the following expansions for $f_{\hat h}$, $K_{\hat h} * f$ and $K_{\hat h} * g(\cdot;\hat\theta_1,\hat\theta_2)$, that play a crucial role in what follows. For $x \in \mathbb{R}$ and denoting by h the deterministic bandwidth $h(f)$ given in assumption (B), we have

$$f_{\hat h}(x) = \sum_{\ell=0}^{\omega-1}\xi_n^\ell \frac{1}{n}\sum_{i=1}^n K_h^{\partial(\ell)}(x-X_i) + \xi_n^\omega \frac{1}{n}\sum_{i=1}^n K_h^{\partial(\omega)}(x-X_i,\hat h), \tag{22}$$

$$K_{\hat h} * f(x) = \sum_{\ell=0}^{\omega-1}\xi_n^\ell K_h^{\partial(\ell)} * f(x) + \xi_n^\omega K_h^{\partial(\omega)}(\cdot,\hat h) * f(x), \tag{23}$$

and

$$K_{\hat h} * g(x;\hat\theta_1,\hat\theta_2) = \sum_{\ell=0}^{\omega-1}\xi_n^\ell K_h^{\partial(\ell)} * g(x;\hat\theta_1,\hat\theta_2) + \xi_n^\omega K_h^{\partial(\omega)}(\cdot,\hat h) * g(x;\hat\theta_1,\hat\theta_2), \tag{24}$$

where $K_h^{\partial(\ell)}(u) = K_h^{\partial(\ell)}(u/h)/h$ and $K_h^{\partial(\omega)}(u,\hat h) = K_h^{\partial(\omega)}(u/h,\hat h/h)/h$. Moreover, for $|\hat h/h - 1| \le \eta$ we have $|K_h^{\partial(\omega)}(u,\hat h)| \le K_h^{\partial(\omega),\eta}(u)$, for $u \in \mathbb{R}$.

Each one of the terms in (21) is studied in the following propositions. We denote by h the deterministic sequence $h(f)$ given in assumption (B).

Proposition 1. *We have*

$$I_{n,1} = (1-\xi_n)\frac{1}{nh}R(K) + \frac{1}{nh^{1/2}}U_n(1+o_p(1)) + O_p\big(n^{-1}h^{-1/2}\xi_n + (nh)^{-1}\xi_n^2 + \xi_n^\omega\big),$$

where U_n given by (25) is asymptotically normal with zero mean and variance $2R(K{*}K)R(f)$.

Proof: Using equalities (22) and (23), and assumptions (D), (K) and (B), from Proposition 2 of Tenreiro (2001, p. 290) we have

$$I_{n,1} = \int \{f_h(x) - K_h{*}f(x)\}^2 dx - \xi_n \frac{1}{nh} R(K) + O_p\big(n^{-1}h^{-1/2}\xi_n + (nh)^{-1}\xi_n^2 + \xi_n^\omega\big).$$

Moreover, using degenerated U-statistics techniques (see Hall, 1984, Tenreiro, 1997) we have

$$\int \{f_h(x) - K_h * f(x)\}^2 dx = \frac{1}{nh} R(K) + \frac{1}{nh^{1/2}} U_n (1 + o_p(1)),$$

with

$$U_n = \frac{2}{n} \sum_{1 \le i < j \le n} q_n(X_i, X_j), \tag{25}$$

$$q_n(u,v) = h^{1/2} \int \big\{K_h(x-u) - K_h{*}f(x)\big\}\big\{K_h(x-v) - K_h{*}f(x)\big\} dx,$$

and U_n is asymptotically normal with zero mean and variance equal to $2R(K{*}K)R(f)$. ■

Proposition 2. *We have*

$$I_{n,2} = R\big(f - g(\cdot;\theta_1(f),\theta_2(f))\big) + o_p(1).$$

Moreover, under the null hypothesis we have

$$I_{n,2} = O_p\big(n^{-1}\big).$$

Proof: From (23) and (24) we have

$$\begin{aligned} I_{n,2} = &\sum_{\ell,\ell'=0}^{\omega-1} \xi_n^{\ell+\ell'} \int K_h^{\partial(\ell)} * \hat\delta_n(x) K_h^{\partial(\ell')} * \hat\delta_n(x) dx \\ &+ 2\sum_{\ell=0}^{\omega-1} \xi_n^{\omega+\ell} \int K_h^{\partial(\ell)} * \hat\delta_n(x) K_h^{\partial(\omega)}(\cdot,\hat h) * \hat\delta_n(x) dx \\ &+ \xi_n^{2\omega} \int \big(K_h^{\partial(\omega)}(\cdot,\hat h) * \hat\delta_n(x)\big)^2 dx, \end{aligned}$$

where $\hat{\delta}_n(x) = f(x) - g(x;\hat{\theta}_1,\hat{\theta}_2)$. Moreover,

$$\left|\int K_h^{\partial(\ell)} * \hat{\delta}_n(x) K_h^{\partial(\ell')} * \hat{\delta}_n(x)dx\right| \leq ||K^{\partial(\ell)}||_1 ||K^{\partial(\ell')}||_1 ||\hat{\delta}_n||_2^2,$$

and for all $\epsilon \in]0,\eta[$ and for $|\hat{h}/h - 1| \leq \epsilon$ we have

$$\left|\int K_h^{\partial(\ell)} * \hat{\delta}_n(x) K_h^{\partial(\omega)}(\cdot,\hat{h}) * \hat{\delta}_n(x)dx\right| \leq ||K^{\partial(\ell)}||_1 ||K^{\partial(\omega),\eta}||_1 ||\hat{\delta}_n||_2^2$$

and

$$\left|\int \big(K_h^{\partial(\omega)}(\cdot,\hat{h}) * \hat{\delta}_n(x)\big)^2 dx\right| \leq ||K^{\partial(\omega),\eta}||_1^2 ||\hat{\delta}_n||_2^2.$$

Therefore, from assumption (B) we can write

$$I_{n,2} = R\big(K_h * \hat{\delta}_n\big) + O_p\big(||\hat{\delta}_n||_2^2 \xi_n\big). \tag{26}$$

On the other hand, from assumption (F) the function $(\theta_1,\theta_2) \mapsto g(x;\theta_1,\theta_2)$ has continuous first-order partial derivatives, and the functions $(\theta_1,\theta_2) \mapsto \big|\big|\frac{\partial g}{\partial \theta_k}(\cdot;\theta_1,\theta_2)\big|\big|_2$ are locally bounded on $\mathbb{R}\times]0,+\infty[$ for $k = 1,2$. Therefore, for each $x \in \mathbb{R}$, a Taylor expansion of $g(x;\hat{\theta}_1,\hat{\theta}_2)$ at the point $(\theta_1(f),\theta_2(f))$ leads to

$$\hat{\delta}_n(x) = f(x) - g(x;\hat{\theta}_1,\hat{\theta}_2) = f(x) - g(x;\theta_1(f),\theta_2(f)) + u_n(x), \tag{27}$$

where

$$||u_n||_2 = O_p\big(|\hat{\theta}_1 - \theta_1(f)| + |\hat{\theta}_2 - \theta_2(f)|\big). \tag{28}$$

The first part of the stated result follows now from (26) and the following convergence that can be established from standard arguments as h tends to zero, when n tends to infinity:

$$R\big(K_h*(f - g(\cdot;\theta_1(f),\theta_2(f)))\big) = R\big(f - g(\cdot;\theta_1(f),\theta_2(f))\big) + o(1).$$

Finally, taking into account that $\hat{\delta}_n = u_n$ under the null hypothesis, where $||u_n||_2 = O_p(n^{-1/2})$ from assumption (P), we deduce that $I_{n,2} = O_p(n^{-1})$ under the null hypothesis. ∎

To establish the order of convergence of $I_{n,3}$ we need the following lemma. Note that we are always assuming that $\hat{h}$ satisfies assumption (B).

Lemma 1. *Let φ be a real-valued function defined on $\mathbb{R}\times]0,+\infty[$, and assume that there exists $\eta \in]0,1[$ such that the function $\varphi^\eta(u) = \sup_{|h-1|\leq\eta}|\varphi(u,h)|$ is bounded and integrable.*

a) If $\gamma_n : \mathbb{R} \mapsto \mathbb{R}$ *is such* $||\gamma_n||_2 = O(1)$ *then*

$$\frac{1}{n}\sum_{i=1}^{n}\int \big\{\varphi_h(x - X_i) - \varphi_h * f(x)\big\}\gamma_n(x)dx = O_p\big(n^{-1/2}\big).$$

b) If $\gamma_n : \mathbb{R} \mapsto \mathbb{R}$ *is such* $||\gamma_n||_r = O(1)$*, for some* $r \in [1, \infty]$*, then*

$$\frac{1}{n}\sum_{i=1}^{n}\int |\varphi_h(x - X_i, \hat{h}) - \varphi_h(\cdot, \hat{h}) * f(x)|\gamma_n(x)dx = O_p(1).$$

c) If $\tilde{\gamma}_n = \tilde{\gamma}_n(\cdot; X_1, \ldots, X_n) : \mathbb{R} \mapsto \mathbb{R}$ *is such that* $||\tilde{\gamma}_n||_r = O_p(1)$*, for some* $r \in [1, \infty]$*, then*

$$\frac{1}{n}\sum_{i=1}^{n}\int |\varphi_h(x - X_i, \hat{h}) - \varphi_h(\cdot, \hat{h}) * f(x)|\tilde{\gamma}_n(x)dx = O_p\big(h^{-1/r}\big).$$

Proof: Write $S_{n,a}$, $S_{n,b}$ and $S_{n,c}$ for the sums considered in each one of the parts a), b) and c). The order of convergence stated in part a) follows from the inequalities

$$\begin{aligned}
\mathrm{E}(S^2_{n,a}) &\leq \frac{1}{n}\mathrm{E}\bigg(\int \varphi_h(x - X_i)\gamma_n(x)dx\bigg)^2 \\
&\leq \frac{1}{n}\iint \varphi(u)^2\gamma_n(z + uh)^2 f(z)dudz \\
&\leq \frac{1}{n}||f||_\infty||\varphi||_2^2||\gamma_n||_2^2.
\end{aligned}$$

In order to establish parts b) and c), it is enough to note that for all $\epsilon \in]0, \eta[$ and for $|\hat{h}/h - 1| \leq \epsilon$ we have

$$|S_{n,b}| \leq \frac{1}{n}\sum_{i=1}^{n}\bigg\{\int \varphi_h^\epsilon(x - X_i)|\gamma_n|(x)dx + \int \varphi_h^\epsilon * f(x)|\gamma_n|(x)dx\bigg\} =: S^\epsilon_{n,b},$$

and

$$|S_{n,c}| \leq \frac{1}{n}\sum_{i=1}^{n}\bigg\{\int \varphi_h^\epsilon(x - X_i)|\tilde{\gamma}_n|(x)dx + \int \varphi_h^\epsilon * f(x)|\tilde{\gamma}_n|(x)dx\bigg\} =: S^\epsilon_{n,c},$$

where

$$\mathrm{E}(S^\epsilon_{n,b}) \leq 2\int \varphi_h^\epsilon * f(x)|\gamma_n|(x)dx \leq 2||f||_\infty||\varphi^\epsilon||_s||\gamma_n||_r,$$

and

$$S^{\epsilon}_{n,c} \leq 2h^{-1/r}||\tilde{\gamma}_n||_r||\varphi^{\epsilon}||_s,$$

with $1/r + 1/s = 1$. Therefore, $S^{\epsilon}_{n,b} = O_p(1)$ and $S^{\epsilon}_{n,c} = O_p\big(h^{-1/r}\big)$ which implies the stated results as $\hat{h}/h - 1 = o_p(1)$. ■

Proposition 3. *We have*

$$I_{n,3} = O_p\big((nh)^{-1/2}\big).$$

Moreover, under the null hypothesis we have

$$I_{n,3} = O_p\big(n^{-1}h^{-1/r} + n^{-1/2}\xi^{\omega}_n\big),$$

where $r \in]2,\infty]$ is given in assumption (F).

Proof: The first statement follows from Propositions 1 and 2 since $|I_{n,3}| \leq I^{1/2}_{n,1}I^{1/2}_{n,2}$. On the other hand, from (22), (23) and (24) we have

$$\begin{aligned}
&I_{n,3}\\
&= \sum_{\ell,\ell'=0}^{\omega-1} \xi_n^{\ell+\ell'} \frac{1}{n}\sum_{i=1}^n \int \big\{K_h^{\partial(\ell)}(x-X_i) - K_h^{\partial(\ell)} * f(x)\big\} K_h^{\partial(\ell')} * \hat{\delta}_n(x)dx\\
&\quad + \sum_{\ell=0}^{\omega-1} \xi_n^{\omega+\ell} \frac{1}{n}\sum_{i=1}^n \big\{K_h^{\partial(\ell)}(x-X_i) - K_h^{\partial(\ell)} * f(x)\big\} K_h^{\partial(\omega)}(\cdot,\hat{h}) * \hat{\delta}_n(x)dx\\
&\quad + \sum_{\ell=0}^{\omega-1} \xi_n^{\omega+\ell} \frac{1}{n}\sum_{i=1}^n \int \big\{K_h^{\partial(\omega)}(x-X_i,\hat{h}) - K_h^{\partial(\omega)}(\cdot,\hat{h}) * f(x)\big\} K_h^{\partial(\ell)} * \hat{\delta}_n(x)dx\\
&\quad + \xi_n^{2\omega} \frac{1}{n}\sum_{i=1}^n \int \big\{K_h^{\partial(\omega)}(x-X_i,\hat{h}) - K_h^{\partial(\omega)}(\cdot,\hat{h}) * f(x)\big\} K_h^{\partial(\omega)}(\cdot,\hat{h}) * \hat{\delta}_n(x)dx\\
&\quad = I^1_{n,3} + I^2_{n,3} + I^3_{n,3} + I^4_{n,3}.
\end{aligned}$$

where $\hat{\delta}_n(x) = f(x) - g(x;\hat{\theta}_1,\hat{\theta}_2)$. From assumption (F), the function $(\theta_1,\theta_2) \mapsto g(x;\theta_1,\theta_2)$ has continuous second-order partial derivatives, and for some $r \in]2,\infty]$ the functions $(\theta_1,\theta_2) \mapsto \big|\big|\frac{\partial^2 g}{\partial\theta_k\partial\theta_l}(\cdot;\theta_1,\theta_2)\big|\big|_r$ are locally bounded on $\mathbb{R}\times]0,+\infty[$, for $k,l = 1,2$. Therefore, under the null hypothesis a Taylor expansion of $g(x;\hat{\theta}_1,\hat{\theta}_2)$ at the point $(\theta_1(f),\theta_2(f))$ leads to

$$\hat{\delta}_n(x) = -\sum_k (\hat{\theta}_k - \theta_k(f)) \frac{\partial g}{\partial\theta_k}(x;\theta_1(f),\theta_2(f)) + v_n(x), \tag{29}$$

for $x \in \mathbb{R}$, where from assumption (P)

$$||v_n||_r = O_p\big((\hat\theta_1 - \theta_1(f))^2 + (\hat\theta_2 - \theta_2(f))^2\big) = O_p\big(n^{-1}\big).$$

Therefore, from Lemma 1 we get

$$I^1_{n,3} = O_p\big(n^{-1}h^{-1/r}\big),\ I^2_{n,3} = O_p\big((n^{-1/2} + n^{-1}h^{-1/r})\xi_n^\omega\big),$$

$$I^3_{n,3} = O_p\big((n^{-1/2} + n^{-1}h^{-1/r})\xi_n^\omega\big) \text{ and } I^4_{n,3} = O_p\big((n^{-1/2} + n^{-1}h^{-1/r})\xi_n^{2\omega}\big),$$

which completes the proof. ■

We can now conclude the proof of Theorem 1. As $\xi_n = o_p(1)$ and $h \to 0$, as $n \to \infty$, from Proposition 1 we have

$$I_{n,1} = O_p\big((nh)^{-1} + \xi_n^\omega\big).$$

Therefore, from expansion (21) and Propositions 2 and 3, we get

$$(nh)^{-1}\big(I_n(\hat h) - R(K)\big) = R\big(f - g(\cdot;\theta_1(f),\theta_2(f))\big) + O_p\big((nh)^{-1/2} + \xi_n^\omega\big),$$

which completes the proof of part b) as $nh \to \infty$, when $n \to \infty$. Moreover, under the null hypothesis from Propositions 1, 2 and 3 we also have

$$h^{-1/2}\big(I_n(\hat h) - R(K)\big) = U_n + O_p\big(h^{-1/2}\xi_n^2 + h^{1/2-1/r} + nh^{1/2}\xi_n^\omega\big) + o_p(1).$$

Taking into account hypothesis (13), this completes the proof of part a) as $r > 2$ and U_n is asymptotically normal with zero mean and variance equal to $\nu_f^2 = 2R(K*K)R(f)$. ■

6.2. Proof of Theorem 2. Let us consider the expansion

$$\begin{aligned}(n\hat h)^{-1}J_n(\hat h) &= \int \{f_{\hat h}(x) - f(x)\}^2 dx \\ &\quad + \int \{f(x) - g(x;\hat\theta_1,\hat\theta_2)\}^2 dx \\ &\quad + 2\int \{f_{\hat h}(x) - f(x)\}\{f(x) - g(x;\hat\theta_1,\hat\theta_2)\} dx \\ &=: J_{n,1} + J_{n,2} + 2J_{n,3}. \end{aligned} \tag{30}$$

Each one of these terms will be studied in the following propositions. As before, we denote by h the deterministic sequence $h(f)$ which existence is assured by assumption (B).

Proposition 4. *We have*

$$J_{n,1} = \frac{1}{nh}R(K) + R(K_h * f - f) + \frac{1}{nh^{1/2}}U_n(1 + o_p(1)) + \frac{1}{\sqrt{n}h^{-2}}V_n + O_p\big(((nh)^{-1} + h^4)\xi_n + \xi_n^\omega\big),$$

where U_n is defined in Proposition 1 and V_n given by (31) is asymptotically normal with zero mean and variance $\mu_2(K)^2\mathrm{Var}_f(f''(X_1))$.

Proof: Taking into account equality (22) and assumptions (D), (D'), (K), (K') and (B), from Lemma 1 of Tenreiro (2001, p. 286) we have

$$J_{n,1} = \int \{f_h(x) - f(x)\}^2 dx + O_p\big(((nh)^{-1} + h^4)\xi_n + \xi_n^\omega\big).$$

Using degenerated U-statistics techniques (see Hall, 1984) we known that

$$\int \{f_h(x) - f(x)\}^2 dx = \frac{1}{nh}R(K) + R(K_h * f - f) + \frac{1}{nh^{1/2}}U_n(1 + o_p(1)) + \frac{1}{\sqrt{n}h^{-2}}V_n,$$

with U_n given by (25) and

$$V_n := \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int \{K_h(x - X_i) - K_h * f(x)\}h^{-2}\{K_h * f(x) - f(x)\}dx, \quad (31)$$

with

$$h^{-2}\{K_h * f(x) - f(x)\} = \iint_0^1 (1 - t)u^2K(u)f''(x - tuh)dudt, \qquad (32)$$

is asymptotically normal with zero mean and variance equal to $\mu_2(K)^2$ $\mathrm{Var}_f(f''(X_1))$. ∎

Proposition 5. *We have*

$$J_{n,2} = R\big(f - g(\cdot; \theta_1(f), \theta_2(f))\big) + o_p(1).$$

Moreover, under the null hypothesis we have

$$J_{n,2} = O_p\big(n^{-1}\big).$$

Proof: It follows straightforwardly from (27) and (28). ∎

Proposition 6. *We have*

$$J_{n,3} = O_p\big((nh)^{-1/2}\big).$$

Moreover, under the null hypothesis we have

$$J_{n,3} = -\frac{1}{\sqrt{n}h^{-2}}\big(W_n + o_p(1)\big) + O_p\big(n^{-1}h^{-1/r} + n^{-1/2}\xi_n^{\omega}\big),$$

where W_n is given by (36).

Proof: The first statement follows from Propositions 4 and 5 because $|J_{n,3}| \leq J_{n,1}^{1/2} J_{n,2}^{1/2}$ and $R(K_h * f - f) = O(h^4)$. Write

$$\begin{aligned} J_{n,3} &= \int \{f_{\hat{h}}(x) - K_{\hat{h}} * f(x)\}\hat{\delta}_n(x)dx + \int \{K_{\hat{h}} * f(x) - f(x)\}\hat{\delta}_n(x)dx \\ &=: J_{n,3}^1 + J_{n,3}^2, \end{aligned} \tag{33}$$

where $\hat{\delta}_n(x) = f(x) - g(x; \hat{\theta}_1, \hat{\theta}_2)$. From (22) and (23) we have

$$\begin{aligned} J_{n,3}^1 = &\sum_{\ell=0}^{\omega-1} \xi_n^{\ell} \frac{1}{n} \sum_{i=1}^{n} \int \big\{K_h^{\partial(\ell)}(x - X_i) - K_h^{\partial(\ell)} * f(x)\big\}\hat{\delta}_n(x)dx \\ &+ \xi_n^{\omega} \frac{1}{n} \sum_{i=1}^{n} \int \big\{K_h^{\partial(\omega)}(x - X_i, \hat{h}) - K_h^{\partial(\omega)}(\cdot, \hat{h}) * f(x)\big\}\hat{\delta}_n(x)dx, \end{aligned}$$

where from Lemma 1 we get

$$J_{n,3}^1 = O_p\big(n^{-1}h^{-1/r} + n^{-1/2}\xi_n^{\omega}\big). \tag{34}$$

On the other hand, from (23) we have

$$\begin{aligned} J_{n,3}^2 = &\int \{K_h * f(x) - f(x)\}\hat{\delta}_n(x)dx \\ &+ \sum_{\ell=1}^{\omega-1} \xi_n^{\ell} \int K_h^{\partial(\ell)} * f(x)\hat{\delta}_n(x)dx + \xi_n^{\omega} \int K_h^{\partial(\omega)}(\cdot, \hat{h}) * f(x)\hat{\delta}_n(x)dx, \end{aligned} \tag{35}$$

where for all $\epsilon \in]0, \eta[$ and for $|\hat{h}/h - 1| \leq \epsilon$ we have

$$\left| \int K_h^{\partial(\omega)}(\cdot, \hat{h}) * f(x)\hat{\delta}_n(x)dx \right| \leq ||K^{\partial(\omega),\eta}||_1 ||f||_2 ||\hat{\delta}_n||_2.$$

Moreover, as $\int K^{\partial(\ell)}(u)du = \int uK^{\partial(\ell)}(u)du = 0$ for $\ell \geq 1$, a Taylor expansion of second order leads to

$$K_h^{\partial(\ell)} * f(x) = h^2 \iint_0^1 (1-t)u^2 K^{\partial(\ell)}(u) f''(x - tuh) dt du.$$

Therefore, for $\ell \geq 1$ we have

$$\left| \int K_h^{\partial(\ell)} * f(x) \hat{\delta}_n(x) dx \right| \leq h^2 \int u^2 |K^{\partial(\ell)}(u)| du ||f''||_2 ||\hat{\delta}_n||_2.$$

Taking into account (27) and the fact that $||\hat{\delta}_n||_2 = O_p(n^{-1/2})$ under the null hypothesis, from (35) we get

$$J_{n,3}^2 = \int \{K_h * f(x) - f(x)\} \hat{\delta}_n(x) dx + O_p\big(n^{-1/2}(h^2 \xi_n + \xi_n^\omega)\big).$$

Finally, from (29) and (32), and assumptions (E) and (G'), we have

$$\begin{aligned}
&\int \{K_h * f(x) - f(x)\} \hat{\delta}_n(x) dx \\
&\qquad = h^2 \iiint_0^1 (1-t) u^2 K(u) f''(x - tuh) \hat{\delta}_n(x) dt du dx \\
&\qquad = -\frac{1}{\sqrt{n} h^{-2}} \big(W_n + o_p(1)\big) + O_p\big(n^{-1} h^2\big),
\end{aligned}$$

where

$$W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_k \psi_k(X_i; \theta_1(f), \theta_2(f)) D_k(f), \tag{36}$$

with

$$D_k(f) = \frac{1}{2} \mu_2(K) \int \bar{f}''(x) \frac{\partial g}{\partial \theta_k}(x; \theta_1(f)\theta_2(f)) dx,$$

as

$$\iiint_0^1 (1-t) u^2 K(u) f''(x - tuh) \frac{\partial g}{\partial \theta_k}(x; \theta_1(f)\theta_2(f)) dt du dx = D_k(f) + o(1),$$

and

$$\left| \iiint_0^1 (1-t) u^2 K(u) f''(x - tuh) v_n(x) dt du dx \right| \leq \mu_2(K) ||f''||_s ||v_n||_r,$$

with $1/r + 1/s = 1$. Thus

$$J_{n,3}^2 = -\frac{1}{\sqrt{n} h^{-2}} \big(W_n + o_p(1)\big) + O_p\big(n^{-1} h^2 + n^{-1/2}(h^2 \xi_n + \xi_n^\omega)\big) \tag{37}$$

The proposition follows from (33), (34) and (37). ∎

We can now conclude the proof of Theorem 2. From Proposition 4 and assumption (B') we have

$$J_{n,1} = O_p\big((nh)^{-1} + \xi_n^\omega\big).$$

Therefore, from expansion (30) and Propositions 5 and 6, we get

$$(nh)^{-1}\big(J_n(\hat{h}) - R(K) - c_n(f;K)\big) = R\big(f - g(\cdot;\theta_1(f),\theta_2(f)\big) + O_p\big((nh)^{-1/2} + \xi_n^\omega\big),$$

which completes the proof of part *b*). Moreover, from Propositions 4, 5 and 6, under the null hypothesis we also have

$$\begin{aligned} h^{-1/2}&\big(J_n(\hat{h}) - R(K) - c_n(f;K)\big) \\ &= U_n + \big(nh^5\big)^{1/2}(V_n - 2W_n) \\ &\quad + O_p\Big(h^{-1/2}\xi_n + h^{1/2-1/r} + nh^{1/2}\xi_n^\omega\Big) + o_p(1). \end{aligned}$$

Taking into account hypothesis (14), this completes the proof of part *a*) as $r > 2$ and, from the central limit theorem for degenerate U-statistics with variable kernels established in Tenreiro (1997, Theorem 1, p. 190), the sum $U_n + \big(nh^5\big)^{1/2}(V_n - 2W_n)$ is asymptotically normal with zero mean and variance equal to $\sigma_f^2 = 2R(K*K)R(f) + \lambda_f\mu_2(K)^2\mathrm{Var}_f(\varphi_f(X))\big)$. ∎

6.3. Proof of Theorem 3. We consider only the case of the test based on the critical region $\mathscr{C}(J_n(\hat{h}),\alpha) = \{J_n(\hat{h}),\alpha) > q(J_n(\hat{h}),\alpha)\}$ given in (15), where $q(J_n(\hat{h}),\alpha)$ is the quantile of order $1-\alpha$ of the null distribution of $J_n(\hat{h})$, but similar arguments can be used to establish the consistency of the test based on $\mathscr{C}(I_n(\hat{h}),\alpha)$. From Theorem 2.*a*) and for $f \in \mathscr{F}_0$ we have $\upsilon_f^{-1}h(f)^{-1/2}\big(q(J_n(\hat{h}),\alpha) - R(K) - c_n(f;K)\big) \to \Phi^{-1}(1-\alpha)$. Therefore,

$$q(J_n(\hat{h}),\alpha) \to R(K) + c(f_0;K), \tag{38}$$

because $h(f)$ tends to zero, as $n \to \infty$, and $c_n(f;K) = c_n(f_0;K) = c(f_0;K)$ $(1+o(1))$, with $c(f;K) = \frac{1}{4}\lambda_f\mu_2(K)^2R(f'')(1+o(1))$ (see Wand and Jones, 1995, pp. 19–23, and Bosq and Lecoutre, 1987, pp. 80–81). On the other hand, from Theorem 2.*b*) and for $f \in \mathscr{F}\backslash\mathscr{F}_0$ we have

$$(nh(f))^{-1}\big(J_n(\hat{h}) - R(K) - c_n(f;K)\big) \xrightarrow{p} R\big(f - g(\cdot;\theta_1(f),\theta_2(f))\big) \neq 0,$$

which enables us to conclude that

$$J_n(\hat{h}) \xrightarrow{p} +\infty, \quad \text{for all} \;\; f \in \mathscr{F} \backslash \mathscr{F}_0, \tag{39}$$

as $nh(f) \to \infty$, and $c_n(f;K) = c(f;K)(1 + o(1))$. The consistency of the test based on $\mathscr{C}(J_n(\hat{h}), \alpha)$ follows now from (38) and (39). ■

Funding

Research partially supported by the Centre for Mathematics of the University of Coimbra – UID/MAT/00324/2019, funded by the Portuguese Government through FCT/MEC and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020.

References

Anderson, T.W., Darling, D.A. (1954). A test of goodness of fit. *J. Amer. Statist. Assoc.* 49, 765–769.

Arcones, M.A., Wang, Y. (2006). Some new tests for normality based on U-processes. *Statis. Probab. Lett.* 76, 69–82.

Baringhaus, L., Danschke, R., Henze, N. (1989). Recent and classical tests for normality. A comparative study. *Comm. Statist. Simulation Comput.* 18, 363–379.

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

Bosq, D., Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle.* Paris: Economica.

Bowman, A.W. (1992). Density based tests for goodness-of-fit normality. *J. Stat. Comput. Simul.* 40, 1–13.

Bowman, A.W., Foster, P.J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.* 88, 529–537.

Cao, R., Lugosi, G. (2005). Goodness-of-fit tests based on the kernel density estimator. *Scand. J. Statist.* 32, 599–616.

Cao, R., Van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *Canad. J. Statist.* 34, 61–77.

Chacón, J.E., Montanero, J., Nogales, A.G., Pérez, P. (2007). On the existence and limit behavior of the optimal bandwidth in kernel density estimation. *Statist. Sinica* 17, 289–300.

Chacón, J.E., Tenreiro, C. (2013). Data-based choice of the number of pilot stages for plug-in bandwidth selection. *Comm. Statist. Theory Methods* 42, 2200–2214.

Devroye, L., Györfi, L. (1985). *Nonparametric density estimation: the L_1 view.* New York: Wiley.

Ebner, B., Henze, N. (2020). Tests for multivariate normality - a critical review with emphasis on weighted L^2-statistics. *TEST* 29, 845–892.

Epps, T.W. (2005). Tests for location-scale families based on the empirical characteristic function. *Metrika* 62, 99–114.

Epps, T.W., Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* 70, 723–726.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.

Fan, Y. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test. *Econometric Rev.* 14, 367–382.

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* 14, 604–621.

Ghosh, B. K., Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Statist.* 19, 999–1009.

Gouriéroux, C., Tenreiro, C. (2001). Local power properties of kernel based goodness of fit tests. *J. Multivariate Anal.* 78, 161–190.

Gürtler, N. (2000). Asymptotic theorems for the class of BHEP-tests for multivariate normality with fixed and variable smoothing parameter (in German). Doctoral dissertation, University of Karlsruhe, Germany.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* 14, 1–16.

Hall, P., Marron, J.S. (1987). Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* 74, 567–581.

Hall, P., Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation *Probab. Theory Related Fields* 90, 149–173.

Hall, P., Marron, J.S., Park, B.U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* 92, 1–20.

Hall, P., Sheather, S.J., Jones, M.C., Marron, J.S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78, 263–269.

Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statist. Papers* 43, 467–506.

Henze, N., Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Comm. Statist. Theory Methods* 19, 3595–3617.

Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). *Continuous univariate distributions 2.* New York: John Wiley.

Krzanowski, W.J., Hand, D.J. (2009). *ROC curves for continuous data.* Boca Raton: CRC Press.

Marron, J.S., Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.* 20, 712–736.

Martínez-Camblor, P., Uña-Álvarez, J., Corral, N. (2008). k-Sample test based on the common area of kernel density estimator. *J. Statist. Plann. Inference* 138, 4006–4020.

Martínez-Camblor, P., Uña-Álvarez, J. (2009). Nonparametric k-sample tests: density function vs. distribution function. *Comput. Statist. Data Anal.* 53, 3344–3357.

Martínez-Camblor, P., Uña-Álvarez, J. (2013). Density comparison for independent and right censored samples via kernel smoothing. *Comput. Statist.* 28, 269–288.

Meintanis, S.G. (2004). Goodness-of-fit tests for the logistic distribution based on empirical transforms. *Sankhya Ser. A* 66, 306–326

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.

R Development Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org

Romão, X., Delgado, R., Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *J. Stat. Comput. Simul.* 80, 545–591.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* 27, 832–837.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* 3, 1–14.

Scott, D.W., Terrel, G.R. (1987). Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* 82, 1131–1146.

Shapiro, S.S, Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

Simonoff, J.S. (1996). *Smoothing methods in statistics.* New York: Springer.

Stephens, M.A. (1986). Tests based on EDF statistics, In: Goodness-of-Fit Techniques, D'Agostino, R.B. and Stephens, M.A. (eds), New York: Marcel Dekker, 97–193.

Tenreiro, C. (1997). Loi asymptotique des erreurs quadratiques intégrées des estimateurs à noyau de la densité et de la régression sous des conditions de dépendance. *Portugal. Math.* 54, 187–213.

Tenreiro, C. (2001). On the asymptotic behaviour of the integrated square error of kernel density estimators with data-dependent bandwidth. *Statis. Probab. Lett.* 53, 283–292.

Tenreiro, C. (2003). On the asymptotic normality of multistage integrated density derivatives kernel estimators. *Statist. Probab. Lett.* 64, 311–322.

Tenreiro, C. (2007). On the asymptotic behaviour of location-scale invariant Bickel-Rosenblatt tests. *J. Statist. Plann. Inference* 137, 103–116. Erratum: 139, 2115, 2009.

Tenreiro, C. (2017). A weighted least-squares cross-validation bandwidth selector for kernel density estimation. *Comm. Statist. Theory Methods* 46, 3438–3458.

Tsybakov, A.B. (2009). *Introduction to nonparametric estimation.* London: Springer.

Wand, M.P., Jones, M.C. (1995). *Kernel smoothing.* New York: Chapman & Hall.

Carlos Tenreiro

CMUC, Department of Mathematics, University of Coimbra, 3001–501 Coimbra, Portugal

E-mail address: `tenreiro@mat.uc.pt`

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
			$\alpha=0.01$			
$\hat{h}_{H_0}$	0.009	0.008	0.009	0.010	0.009	0.009
$\hat{\lambda}_1\hat{h}_{H_0}$	0.009	0.009	0.009	0.009	0.010	0.009
$\hat{\lambda}_2\hat{h}_{H_0}$	0.010	0.010	0.010	0.010	0.011	0.010
$\hat{\lambda}_3\hat{h}_{H_0}$	0.010	0.010	0.010	0.010	0.010	0.010
$\hat{h}_{\mathrm{CT}}$	0.009	0.009	0.009	0.010	0.010	0.009
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.009	0.010	0.009	0.008	0.009	0.008
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.009	0.010	0.010	0.010	0.010	0.010
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.010	0.011	0.011	0.010	0.010	0.010
			$\alpha=0.05$			
$\hat{h}_{H_0}$	0.049	0.048	0.051	0.049	0.048	0.050
$\hat{\lambda}_1\hat{h}_{H_0}$	0.049	0.047	0.047	0.052	0.051	0.048
$\hat{\lambda}_2\hat{h}_{H_0}$	0.049	0.050	0.049	0.052	0.050	0.052
$\hat{\lambda}_3\hat{h}_{H_0}$	0.050	0.051	0.048	0.051	0.050	0.052
$\hat{h}_{\mathrm{CT}}$	0.050	0.050	0.051	0.049	0.047	0.051
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.050	0.050	0.049	0.050	0.049	0.050
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.050	0.053	0.051	0.052	0.048	0.050
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.051	0.052	0.050	0.050	0.051	0.051
			$\alpha=0.10$			
$\hat{h}_{H_0}$	0.098	0.098	0.103	0.101	0.096	0.100
$\hat{\lambda}_1\hat{h}_{H_0}$	0.099	0.094	0.099	0.100	0.098	0.101
$\hat{\lambda}_2\hat{h}_{H_0}$	0.100	0.099	0.095	0.102	0.101	0.100
$\hat{\lambda}_3\hat{h}_{H_0}$	0.100	0.099	0.098	0.105	0.102	0.103
$\hat{h}_{\mathrm{CT}}$	0.099	0.101	0.101	0.100	0.098	0.101
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.100	0.100	0.097	0.100	0.098	0.102
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.102	0.101	0.100	0.101	0.099	0.099
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.101	0.100	0.097	0.099	0.100	0.101

TABLE 9. *Estimates of the nominal levels of significance, for a preassigned level α, for the goodness-of-fit tests for the* logistic distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors. The number of replications for each case is $20,000$.*

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n = 20$	$n = 50$	$n = 80$	$n = 20$	$n = 50$	$n = 80$
			$\alpha = 0.01$			
$\hat{h}_{H_0}$	0.009	0.009	0.009	0.009	0.009	0.010
$\hat{\lambda}_1\hat{h}_{H_0}$	0.010	0.010	0.010	0.010	0.010	0.010
$\hat{\lambda}_2\hat{h}_{H_0}$	0.010	0.008	0.009	0.011	0.009	0.011
$\hat{\lambda}_3\hat{h}_{H_0}$	0.012	0.009	0.011	0.010	0.010	0.011
$\hat{h}_{\mathrm{CT}}$	0.010	0.009	0.009	0.010	0.009	0.009
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.010	0.009	0.009	0.010	0.009	0.011
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.011	0.009	0.009	0.009	0.010	0.010
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.010	0.010	0.010	0.010	0.011	0.011
			$\alpha = 0.05$			
$\hat{h}_{H_0}$	0.047	0.049	0.051	0.050	0.049	0.049
$\hat{\lambda}_1\hat{h}_{H_0}$	0.050	0.044	0.048	0.049	0.051	0.052
$\hat{\lambda}_2\hat{h}_{H_0}$	0.050	0.048	0.048	0.051	0.048	0.048
$\hat{\lambda}_3\hat{h}_{H_0}$	0.053	0.050	0.050	0.051	0.052	0.050
$\hat{h}_{\mathrm{CT}}$	0.049	0.049	0.049	0.049	0.049	0.050
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.050	0.046	0.047	0.047	0.046	0.050
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.053	0.048	0.051	0.051	0.051	0.050
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.054	0.049	0.049	0.051	0.049	0.050
			$\alpha = 0.10$			
$\hat{h}_{H_0}$	0.097	0.097	0.097	0.098	0.099	0.098
$\hat{\lambda}_1\hat{h}_{H_0}$	0.099	0.097	0.098	0.098	0.098	0.098
$\hat{\lambda}_2\hat{h}_{H_0}$	0.098	0.098	0.099	0.101	0.098	0.098
$\hat{\lambda}_3\hat{h}_{H_0}$	0.104	0.101	0.099	0.102	0.101	0.102
$\hat{h}_{\mathrm{CT}}$	0.098	0.099	0.100	0.099	0.100	0.097
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.101	0.097	0.098	0.096	0.096	0.098
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.104	0.098	0.099	0.099	0.101	0.101
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.102	0.096	0.099	0.101	0.099	0.099

TABLE 10. *Estimates of the nominal levels of significance, for a preassigned level α, for the goodness-of-fit tests for the* Gumbel distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors. The number of replications for each case is $20,000$.*

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
	Uniform					
$\hat{h}_{H_0}$	0.33	0.86	0.99	0.54	0.95	1.00
$\hat{\lambda}_1\hat{h}_{H_0}$	0.32	0.84	0.98	0.48	0.92	0.99
$\hat{\lambda}_2\hat{h}_{H_0}$	0.19	0.76	0.93	0.13	0.14	0.42
$\hat{\lambda}_3\hat{h}_{H_0}$	0.22	0.54	0.89	0.37	0.93	0.96
$\hat{h}_{\text{CT}}$	0.20	0.69	0.95	0.43	0.94	0.99
$\hat{\lambda}_1\hat{h}_{\text{CT}}$	0.21	0.70	0.95	0.32	0.90	0.98
$\hat{\lambda}_2\hat{h}_{\text{CT}}$	0.27	0.75	0.94	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\text{CT}}$	0.29	0.77	0.94	0.06	0.13	0.61
A^2	0.30	0.83	0.99	0.30	0.83	0.99
	Exponential					
$\hat{h}_{H_0}$	0.68	0.99	1.00	0.52	0.97	1.00
$\hat{\lambda}_1\hat{h}_{H_0}$	0.64	0.98	1.00	0.46	0.95	1.00
$\hat{\lambda}_2\hat{h}_{H_0}$	0.52	0.96	1.00	0.02	0.15	0.61
$\hat{\lambda}_3\hat{h}_{H_0}$	0.26	0.81	0.98	0.06	0.26	0.60
$\hat{h}_{\text{CT}}$	0.60	0.98	1.00	0.54	0.94	1.00
$\hat{\lambda}_1\hat{h}_{\text{CT}}$	0.60	0.98	1.00	0.48	0.92	1.00
$\hat{\lambda}_2\hat{h}_{\text{CT}}$	0.58	0.96	1.00	0.00	0.01	0.03
$\hat{\lambda}_3\hat{h}_{\text{CT}}$	0.53	0.94	1.00	0.08	0.17	0.67
A^2	0.72	0.99	1.00	0.72	0.99	1.00
	Laplace					
$\hat{h}_{H_0}$	0.11	0.20	0.31	0.03	0.03	0.04
$\hat{\lambda}_1\hat{h}_{H_0}$	0.11	0.20	0.30	0.04	0.05	0.07
$\hat{\lambda}_2\hat{h}_{H_0}$	0.10	0.20	0.29	0.02	0.01	0.00
$\hat{\lambda}_3\hat{h}_{H_0}$	0.11	0.21	0.31	0.02	0.00	0.00
$\hat{h}_{\text{CT}}$	0.13	0.25	0.36	0.08	0.08	0.12
$\hat{\lambda}_1\hat{h}_{\text{CT}}$	0.13	0.24	0.35	0.07	0.06	0.09
$\hat{\lambda}_2\hat{h}_{\text{CT}}$	0.11	0.17	0.26	0.03	0.01	0.00
$\hat{\lambda}_3\hat{h}_{\text{CT}}$	0.09	0.16	0.25	0.02	0.01	0.01
A^2	0.12	0.21	0.31	0.12	0.21	0.31
	Cauchy					
$\hat{h}_{H_0}$	0.78	0.99	1.00	0.52	0.92	0.99
$\hat{\lambda}_1\hat{h}_{H_0}$	0.77	0.98	1.00	0.57	0.93	0.99
$\hat{\lambda}_2\hat{h}_{H_0}$	0.74	0.98	1.00	0.29	0.70	0.94
$\hat{\lambda}_3\hat{h}_{H_0}$	0.65	0.97	1.00	0.45	0.86	0.96
$\hat{h}_{\text{CT}}$	0.80	0.99	1.00	0.73	0.96	1.00
$\hat{\lambda}_1\hat{h}_{\text{CT}}$	0.80	0.99	1.00	0.71	0.96	0.99
$\hat{\lambda}_2\hat{h}_{\text{CT}}$	0.76	0.98	1.00	0.22	0.63	0.86
$\hat{\lambda}_3\hat{h}_{\text{CT}}$	0.73	0.98	1.00	0.41	0.74	0.95
A^2	0.80	0.99	1.00	0.80	0.99	1.00

TABLE 11. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* logistic distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*

$\hat{h}$	$I_n(\hat{h})$ $n=20$	$n=50$	$n=80$	$J_n(\hat{h})$ $n=20$	$n=50$	$n=80$
			Normal			
$\hat{h}_{H_0}$	0.05	0.07	0.11	0.09	0.16	0.21
$\hat{\lambda}_1\hat{h}_{H_0}$	0.05	0.07	0.10	0.09	0.14	0.18
$\hat{\lambda}_2\hat{h}_{H_0}$	0.05	0.07	0.09	0.08	0.15	0.19
$\hat{\lambda}_3\hat{h}_{H_0}$	0.06	0.05	0.06	0.10	0.19	0.27
$\hat{h}_{\mathrm{CT}}$	0.04	0.05	0.07	0.07	0.16	0.24
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.04	0.05	0.07	0.07	0.14	0.21
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.05	0.07	0.08	0.06	0.10	0.13
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.06	0.07	0.08	0.07	0.11	0.16
A^2	0.05	0.06	0.09	0.05	0.06	0.09
			Gumbel			
$\hat{h}_{H_0}$	0.19	0.46	0.68	0.15	0.38	0.60
$\hat{\lambda}_1\hat{h}_{H_0}$	0.18	0.42	0.63	0.12	0.29	0.46
$\hat{\lambda}_2\hat{h}_{H_0}$	0.12	0.30	0.38	0.04	0.04	0.02
$\hat{\lambda}_3\hat{h}_{H_0}$	0.07	0.11	0.24	0.06	0.07	0.03
$\hat{h}_{\mathrm{CT}}$	0.14	0.36	0.58	0.12	0.25	0.45
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.14	0.35	0.57	0.10	0.20	0.35
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.13	0.28	0.45	0.03	0.02	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.12	0.24	0.40	0.04	0.02	0.02
A^2	0.20	0.49	0.74	0.20	0.49	0.74
			Student(6)			
$\hat{h}_{H_0}$	0.07	0.07	0.08	0.05	0.05	0.04
$\hat{\lambda}_1\hat{h}_{H_0}$	0.07	0.07	0.07	0.05	0.05	0.04
$\hat{\lambda}_2\hat{h}_{H_0}$	0.06	0.06	0.06	0.05	0.04	0.03
$\hat{\lambda}_3\hat{h}_{H_0}$	0.05	0.06	0.06	0.04	0.04	0.03
$\hat{h}_{\mathrm{CT}}$	0.07	0.07	0.08	0.06	0.05	0.04
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.07	0.07	0.08	0.06	0.05	0.04
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.06	0.06	0.07	0.04	0.04	0.04
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.06	0.06	0.06	0.05	0.04	0.04
A^2	0.07	0.08	0.09	0.07	0.08	0.09
			Lognormal(0.5)			
$\hat{h}_{H_0}$	0.37	0.79	0.95	0.24	0.67	0.90
$\hat{\lambda}_1\hat{h}_{H_0}$	0.34	0.74	0.93	0.20	0.57	0.82
$\hat{\lambda}_2\hat{h}_{H_0}$	0.24	0.60	0.77	0.03	0.02	0.06
$\hat{\lambda}_3\hat{h}_{H_0}$	0.11	0.30	0.62	0.05	0.05	0.06
$\hat{h}_{\mathrm{CT}}$	0.29	0.70	0.92	0.22	0.53	0.81
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.28	0.69	0.91	0.19	0.45	0.73
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.25	0.59	0.84	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.22	0.53	0.79	0.03	0.02	0.14
A^2	0.38	0.82	0.97	0.38	0.82	0.97

TABLE 12. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* logistic distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n = 20$	$n = 50$	$n = 80$	$n = 20$	$n = 50$	$n = 80$
			Tukey(5)			
$\hat{h}_{H_0}$	0.05	0.09	0.18	0.04	0.05	0.06
$\hat{\lambda}_1\hat{h}_{H_0}$	0.07	0.12	0.22	0.06	0.08	0.12
$\hat{\lambda}_2\hat{h}_{H_0}$	0.11	0.23	0.40	0.06	0.06	0.09
$\hat{\lambda}_3\hat{h}_{H_0}$	0.14	0.31	0.39	0.04	0.05	0.06
$\hat{h}_{\mathrm{CT}}$	0.10	0.22	0.36	0.07	0.09	0.14
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.11	0.22	0.36	0.07	0.07	0.10
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.10	0.23	0.34	0.03	0.02	0.009
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.11	0.24	0.37	0.04	0.01	0.01
A^2	0.06	0.09	0.16	0.06	0.09	0.16
			Gamma(4,1)			
$\hat{h}_{H_0}$	0.19	0.49	0.73	0.17	0.45	0.70
$\hat{\lambda}_1\hat{h}_{H_0}$	0.18	0.44	0.69	0.14	0.35	0.57
$\hat{\lambda}_2\hat{h}_{H_0}$	0.12	0.32	0.42	0.05	0.06	0.02
$\hat{\lambda}_3\hat{h}_{H_0}$	0.07	0.13	0.26	0.07	0.10	0.05
$\hat{h}_{\mathrm{CT}}$	0.14	0.37	0.63	0.13	0.31	0.56
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.14	0.36	0.62	0.11	0.24	0.45
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.13	0.31	0.49	0.03	0.02	0.008
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.12	0.26	0.45	0.04	0.02	0.03
A^2	0.19	0.51	0.78	0.19	0.51	0.78
			Weibul(2,1)			
$\hat{h}_{H_0}$	0.12	0.33	0.55	0.16	0.41	0.63
$\hat{\lambda}_1\hat{h}_{H_0}$	0.11	0.30	0.50	0.14	0.32	0.50
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.22	0.30	0.08	0.12	0.06
$\hat{\lambda}_3\hat{h}_{H_0}$	0.07	0.10	0.20	0.12	0.25	0.20
$\hat{h}_{\mathrm{CT}}$	0.08	0.21	0.41	0.12	0.32	0.56
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.09	0.21	0.40	0.10	0.25	0.45
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.09	0.21	0.33	0.04	0.04	0.04
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.10	0.20	0.32	0.06	0.05	0.05
A^2	0.12	0.31	0.56	0.12	0.31	0.56
			Beta(2,3)			
$\hat{h}_{H_0}$	0.12	0.38	0.63	0.23	0.57	0.80
$\hat{\lambda}_1\hat{h}_{H_0}$	0.11	0.36	0.60	0.19	0.48	0.70
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.27	0.42	0.12	0.23	0.14
$\hat{\lambda}_3\hat{h}_{H_0}$	0.08	0.14	0.30	0.20	0.59	0.68
$\hat{h}_{\mathrm{CT}}$	0.07	0.23	0.45	0.16	0.57	0.81
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.08	0.23	0.45	0.12	0.48	0.73
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.10	0.26	0.42	0.04	0.05	0.04
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.11	0.27	0.42	0.08	0.15	0.14
A^2	0.11	0.33	0.60	0.11	0.33	0.60

TABLE 13. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* logistic distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*

$\hat{h}$	$I_n(\hat{h})$ $n=20$	$I_n(\hat{h})$ $n=50$	$I_n(\hat{h})$ $n=80$	$J_n(\hat{h})$ $n=20$	$J_n(\hat{h})$ $n=50$	$J_n(\hat{h})$ $n=80$
			Uniform			
$\hat{h}_{H_0}$	0.34	0.79	0.96	0.50	0.90	0.99
$\hat{\lambda}_1\hat{h}_{H_0}$	0.32	0.76	0.95	0.41	0.84	0.97
$\hat{\lambda}_2\hat{h}_{H_0}$	0.27	0.58	0.94	0.33	0.75	0.93
$\hat{\lambda}_3\hat{h}_{H_0}$	0.19	0.46	0.73	0.05	0.14	0.30
$\hat{h}_{\mathrm{CT}}$	0.22	0.66	0.92	0.39	0.89	0.99
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.22	0.65	0.90	0.24	0.80	0.96
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.23	0.57	0.82	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.20	0.52	0.69	0.11	0.20	0.91
A^2	0.26	0.76	0.96	0.26	0.76	0.96
			Exponential			
$\hat{h}_{H_0}$	0.27	0.63	0.86	0.12	0.37	0.65
$\hat{\lambda}_1\hat{h}_{H_0}$	0.25	0.60	0.84	0.10	0.29	0.56
$\hat{\lambda}_2\hat{h}_{H_0}$	0.23	0.48	0.82	0.21	0.55	0.81
$\hat{\lambda}_3\hat{h}_{H_0}$	0.17	0.32	0.58	0.33	0.44	0.33
$\hat{h}_{\mathrm{CT}}$	0.24	0.61	0.86	0.16	0.40	0.67
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.23	0.60	0.84	0.12	0.32	0.58
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.18	0.48	0.76	0.01	0.00	0.00
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.15	0.48	0.73	0.13	0.07	0.52
A^2	0.38	0.82	0.97	0.38	0.82	0.97
			Laplace			
$\hat{h}_{H_0}$	0.43	0.83	0.96	0.35	0.77	0.93
$\hat{\lambda}_1\hat{h}_{H_0}$	0.43	0.83	0.95	0.27	0.69	0.90
$\hat{\lambda}_2\hat{h}_{H_0}$	0.40	0.76	0.95	0.42	0.79	0.94
$\hat{\lambda}_3\hat{h}_{H_0}$	0.35	0.70	0.89	0.18	0.35	0.55
$\hat{h}_{\mathrm{CT}}$	0.43	0.82	0.95	0.40	0.76	0.92
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.43	0.82	0.95	0.34	0.70	0.89
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.32	0.73	0.92	0.02	0.05	0.12
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.34	0.75	0.92	0.24	0.40	0.85
A^2	0.46	0.86	0.97	0.46	0.86	0.97
			Cauchy			
$\hat{h}_{H_0}$	0.87	1.00	1.00	0.75	0.99	1.00
$\hat{\lambda}_1\hat{h}_{H_0}$	0.86	1.00	1.00	0.68	0.98	1.00
$\hat{\lambda}_2\hat{h}_{H_0}$	0.85	0.99	1.00	0.85	1.00	1.00
$\hat{\lambda}_3\hat{h}_{H_0}$	0.80	0.99	1.00	0.72	0.97	1.00
$\hat{h}_{\mathrm{CT}}$	0.88	1.00	1.00	0.85	0.99	1.00
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.87	1.00	1.00	0.81	0.99	1.00
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.81	0.99	1.00	0.37	0.86	0.97
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.81	0.99	1.00	0.76	0.97	1.00
A^2	0.89	1.00	1.00	0.89	1.00	1.00

TABLE 14. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* Gumbel distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*

	$I_n(\hat{h})$			$J_n(\hat{h})$		
$\hat{h}$	$n=20$	$n=50$	$n=80$	$n=20$	$n=50$	$n=80$
			Normal			
$\hat{h}_{H_0}$	0.20	0.47	0.67	0.23	0.50	0.68
$\hat{\lambda}_1\hat{h}_{H_0}$	0.20	0.44	0.65	0.17	0.41	0.60
$\hat{\lambda}_2\hat{h}_{H_0}$	0.17	0.31	0.63	0.18	0.37	0.53
$\hat{\lambda}_3\hat{h}_{H_0}$	0.12	0.18	0.28	0.03	0.02	0.03
$\hat{h}_{\mathrm{CT}}$	0.16	0.38	0.59	0.19	0.45	0.66
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.16	0.37	0.58	0.13	0.33	0.55
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.12	0.24	0.45	0.04	0.01	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.11	0.24	0.36	0.06	0.04	0.40
A^2	0.21	0.55	0.78	0.21	0.55	0.78
			Logistic			
$\hat{h}_{H_0}$	0.28	0.63	0.82	0.26	0.58	0.78
$\hat{\lambda}_1\hat{h}_{H_0}$	0.28	0.61	0.79	0.20	0.50	0.71
$\hat{\lambda}_2\hat{h}_{H_0}$	0.25	0.49	0.78	0.25	0.52	0.71
$\hat{\lambda}_3\hat{h}_{H_0}$	0.18	0.37	0.56	0.07	0.09	0.15
$\hat{h}_{\mathrm{CT}}$	0.25	0.57	0.77	0.24	0.53	0.74
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.25	0.56	0.76	0.18	0.43	0.65
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.17	0.42	0.68	0.03	0.01	0.02
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.18	0.45	0.64	0.11	0.14	0.54
A^2	0.31	0.69	0.88	0.31	0.69	0.88
			Student(6)			
$\hat{h}_{H_0}$	0.31	0.65	0.84	0.28	0.60	0.80
$\hat{\lambda}_1\hat{h}_{H_0}$	0.30	0.63	0.83	0.21	0.52	0.74
$\hat{\lambda}_2\hat{h}_{H_0}$	0.27	0.53	0.82	0.27	0.55	0.74
$\hat{\lambda}_3\hat{h}_{H_0}$	0.19	0.42	0.63	0.09	0.14	0.22
$\hat{h}_{\mathrm{CT}}$	0.28	0.60	0.81	0.28	0.56	0.76
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.27	0.59	0.79	0.20	0.46	0.69
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.19	0.47	0.72	0.03	0.03	0.05
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.19	0.50	0.69	0.13	0.19	0.60
A^2	0.33	0.71	0.89	0.33	0.71	0.89
			Lognormal(0.5)			
$\hat{h}_{H_0}$	0.08	0.11	0.15	0.04	0.05	0.06
$\hat{\lambda}_1\hat{h}_{H_0}$	0.07	0.10	0.13	0.05	0.05	0.06
$\hat{\lambda}_2\hat{h}_{H_0}$	0.07	0.07	0.12	0.06	0.07	0.09
$\hat{\lambda}_3\hat{h}_{H_0}$	0.06	0.06	0.09	0.12	0.15	0.18
$\hat{h}_{\mathrm{CT}}$	0.06	0.09	0.13	0.05	0.04	0.05
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.06	0.09	0.12	0.04	0.04	0.04
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.06	0.07	0.09	0.04	0.03	0.02
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.05	0.07	0.08	0.06	0.07	0.07
A^2	0.09	0.15	0.22	0.09	0.15	0.22

TABLE 15. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* Gumbel distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*

$\hat{h}$	$I_n(\hat{h})$ $n=20$	$I_n(\hat{h})$ $n=50$	$I_n(\hat{h})$ $n=80$	$J_n(\hat{h})$ $n=20$	$J_n(\hat{h})$ $n=50$	$J_n(\hat{h})$ $n=80$
			Tukey(5)			
$\hat{h}_{H_0}$	0.27	0.72	0.93	0.26	0.71	0.93
$\hat{\lambda}_1\hat{h}_{H_0}$	0.28	0.70	0.92	0.20	0.61	0.88
$\hat{\lambda}_2\hat{h}_{H_0}$	0.25	0.62	0.90	0.33	0.75	0.92
$\hat{\lambda}_3\hat{h}_{H_0}$	0.29	0.46	0.51	0.08	0.15	0.29
$\hat{h}_{\mathrm{CT}}$	0.32	0.73	0.92	0.32	0.69	0.91
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.33	0.71	0.91	0.25	0.59	0.85
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.26	0.62	0.85	0.01	0.00	0.01
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.27	0.55	0.75	0.15	0.21	0.78
A^2	0.34	0.81	0.97	0.34	0.81	0.97
			Gamma(4,1)			
$\hat{h}_{H_0}$	0.05	0.05	0.06	0.06	0.07	0.08
$\hat{\lambda}_1\hat{h}_{H_0}$	0.05	0.05	0.06	0.06	0.07	0.08
$\hat{\lambda}_2\hat{h}_{H_0}$	0.05	0.05	0.06	0.05	0.06	0.06
$\hat{\lambda}_3\hat{h}_{H_0}$	0.05	0.05	0.05	0.06	0.07	0.09
$\hat{h}_{\mathrm{CT}}$	0.05	0.05	0.06	0.05	0.06	0.08
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.05	0.05	0.05	0.05	0.06	0.07
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.05	0.05	0.05	0.05	0.06	0.07
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.05	0.05	0.05	0.06	0.06	0.09
A^2	0.05	0.05	0.06	0.05	0.05	0.06
			Weibul(2,1)			
$\hat{h}_{H_0}$	0.06	0.09	0.11	0.10	0.16	0.21
$\hat{\lambda}_1\hat{h}_{H_0}$	0.06	0.09	0.11	0.09	0.14	0.18
$\hat{\lambda}_2\hat{h}_{H_0}$	0.05	0.07	0.10	0.07	0.10	0.12
$\hat{\lambda}_3\hat{h}_{H_0}$	0.06	0.06	0.06	0.03	0.03	0.03
$\hat{h}_{\mathrm{CT}}$	0.05	0.06	0.08	0.08	0.17	0.24
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.05	0.07	0.09	0.07	0.11	0.19
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.06	0.07	0.09	0.06	0.07	0.11
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.06	0.06	0.05	0.05	0.05	0.12
A^2	0.05	0.08	0.11	0.05	0.08	0.11
			Beta(2,3)			
$\hat{h}_{H_0}$	0.10	0.20	0.32	0.17	0.35	0.51
$\hat{\lambda}_1\hat{h}_{H_0}$	0.09	0.19	0.29	0.15	0.28	0.40
$\hat{\lambda}_2\hat{h}_{H_0}$	0.08	0.10	0.27	0.10	0.19	0.28
$\hat{\lambda}_3\hat{h}_{H_0}$	0.07	0.06	0.08	0.01	0.01	0.01
$\hat{h}_{\mathrm{CT}}$	0.07	0.13	0.22	0.12	0.40	0.59
$\hat{\lambda}_1\hat{h}_{\mathrm{CT}}$	0.07	0.13	0.21	0.09	0.27	0.47
$\hat{\lambda}_2\hat{h}_{\mathrm{CT}}$	0.08	0.11	0.17	0.05	0.03	0.05
$\hat{\lambda}_3\hat{h}_{\mathrm{CT}}$	0.08	0.07	0.07	0.05	0.03	0.29
A^2	0.08	0.19	0.33	0.08	0.19	0.33

TABLE 16. *Empirical power results, at level $\alpha = 0.05$, for the goodness-of-fit tests for the* Gumbel distribution *based on $I_n(\hat{h})$ and $J_n(\hat{h})$, for the indicated bandwidth selectors, and A^2. The power estimates are based on $10,000$ samples from the considered alternatives.*