Econometric Theory, **14**, 1998, 604–621. Printed in the United States of America.

GOODNESS-OF-FIT TESTS BASED ON KERNEL DENSITY ESTIMATORS WITH FIXED SMOOTHING PARAMETERS

YANQIN FAN
University of Windsor

In this paper, we study the bias-corrected test developed in Fan (1994). It is based on the integrated squared difference between a kernel estimator of the unknown density function of a random vector and a kernel smoothed estimator of the parametric density function to be tested under the null hypothesis. We provide an alternative asymptotic approximation of the finite-sample distribution of this test by fixing the smoothing parameter. In contrast to the normal approximation obtained in Fan (1994) in which the smoothing parameter shrinks to zero as the sample size grows to infinity, we obtain a non-normal asymptotic distribution for the bias-corrected test. A parametric bootstrap procedure is proposed to approximate the critical values of this test. We show both analytically and by simulation that the proposed bootstrap procedure works. Consistency and local power properties of the bias-corrected test with a fixed smoothing parameter are also discussed.

1. INTRODUCTION AND MOTIVATION

Let X be a $d \times 1$ random variable on the measurable space $(\mathcal{X}, \mathcal{A})$ with distribution function $F(x)$ and density function $f(x)$, $x \in \mathcal{X} \subset R^d$, where d is a positive integer. Let $f_0(x, \beta)$ be a parametric density function on $\mathcal{X}$ that is measurable in x for every $\beta \in \mathcal{B}$, a compact subset of R^p, and continuous in β for every $x \in \mathcal{X}$. Suppose n independent and identically distributed (i.i.d.) observations, $X_1, \ldots, X_n$, are available from the unknown distribution F, we are interested in testing

$$H_0: f(x) = f_0(x, \beta) \quad \text{for some } \beta \in \mathcal{B} \quad \text{against}$$

$$H_1: f(x) \neq f_0(x, \beta) \quad \text{for all } \beta \in \mathcal{B}.$$

Recently, a number of papers have employed nonparametric density estimators such as kernel and series estimators in constructing consistent tests for H_0 against H_1. The first paper in this direction is probably the one by Bickel and Rosenblatt (1973) in which they used kernel density estimators to develop con-

I thank the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for their support. I am grateful to Bruce Hansen, John Knight, Aman Ullah, a co-editor, and two referees for helpful comments. Address correspondence to: Yanqin Fan, Department of Economics, University of Windsor, Windsor, Ontario N9B 3P4, Canada; e-mail: Fan1@uwindsor.ca.

 © 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

sistent goodness-of-fit tests. Later, Rosenblatt (1975) and Fan (1994) extended the Bickel–Rosenblatt test in several directions. For details about the extensions and more recent references, see Fan (1994, 1997). It is well known that for a kernel density estimator to be consistent, the smoothing parameter must decrease to zero and must not decrease too fast as the sample size n approaches infinity. Similar conditions on the smoothing parameter are also adopted in Bickel and Rosenblatt (1973), Rosenblatt (1975), and Fan (1994). Under these conditions, they show that their test statistics are asymptotically normally distributed under H_0. Although currently there are well-developed ways of choosing an optimal smoothing parameter for kernel density estimation, such methods have not yet been developed for the corresponding tests. The conditions imposed on the smoothing parameter to validate these tests do not provide any guidance on how its value should be chosen for a given data set. Moreover, simulation studies in Fan (1994, 1995) show that these tests are very sensitive to the choice of the smoothing parameter: For a wide range of values of the smoothing parameter, the true size and the nominal size of these tests can differ greatly and the test statistics can have large skewness and kurtosis exhibiting behaviors more like χ^2 tests than normal tests.

In this paper, we will study the bias-corrected test developed in Fan (1994). The main purpose is to understand the role played by the smoothing parameter and to provide an alternative asymptotic approximation for the finite-sample distribution of this test.

The bias-corrected test is based on $J_n(h)$ (say) defined as the integrated squared difference between a kernel density estimator and a kernel-smoothed parametric density estimator. In Section 2, we first review the existing properties of this test and then show that $J_n(h)$ can be written as a weighted integrated squared difference between the empirical characteristic function of the observations and the parametric estimate of the characteristic function implied by the null model, where the weighting function depends on the kernel function and the smoothing parameter. The weighting function in the alternative expression for $J_n(h)$ determines the weight given to different directions of departure from the null model. For a given kernel function, a small value of the smoothing parameter increases the weight allocated to relatively high-frequency components (see Sections 2 and 4 for more detailed discussion on this point). Hence, we expect that tests based on kernel smoothing (with vanishing smoothing parameters) are more powerful against high-frequency alternatives than tests based on empirical distribution functions such as the Cramér–von Mises test, because tests based on empirical distribution functions effectively only use information contained in the low frequencies (see Fan, 1996). On the basis of the alternative expression for $J_n(h)$, we derive the asymptotic distribution of the bias-corrected test statistic under H_0 for a fixed smoothing parameter in Section 3. It is given by the distribution of an infinite sum of weighted χ^2 random variables. This provides a theoretical justification for the simulation results alluded to previously. In Section 4, we show that one can construct a consistent test using $J_n(h)$ without having to resort to a vanishing smooth-

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

ing parameter (with respect to n). The local power properties of the resulting test are also established in Section 4. Section 5 proposes a parametric bootstrap procedure for approximating the critical values of the test. We show both analytically and by simulation that the proposed bootstrap procedure works. The last section concludes.

2. THE BIAS-CORRECTED TEST

The class of kernel density estimators of $f(x)$ is given by

$$\tilde{f}(x) = \frac{1}{nh^d}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) = \frac{1}{h^d}\int K\left(\frac{x-u}{h}\right) dF_n(u), \tag{1}$$

where $K(\cdot)$ is a kernel function integrating to one, h is a smoothing parameter, and $F_n(\cdot)$ is the empirical distribution function of the observations $X_1, \ldots, X_n$ assigning mass $1/n$ to each of the X_j's.

Let $L(\beta) \equiv n^{-1}\sum_{j=1}^{n} \log f_0(X_j, \beta)$ be the quasi-log-likelihood function and $\hat{\beta}$ the quasi-maximum-likelihood estimator (QMLE). Further, let β_0 denote the true value of β when H_0 is true. Then under H_0, $\hat{\beta}$ is the maximum likelihood estimator (MLE) of β_0. Hence, under H_0, the maximum likelihood estimator of $f_0(x, \beta_0)$ is $\hat{f}(x) \equiv f_0(x, \hat{\beta})$, which is known to be asymptotically efficient.

Fan (1994) proposed a test for H_0 against H_1 using the integrated squared difference between $\tilde{f}(x)$ and $K_h * \hat{f}(x)$ defined as

$$\begin{aligned} J_n(h) &= \int [\,\tilde{f}(x) - K_h * \hat{f}(x)]^2\, dx \\ &= \frac{1}{h^{2d}}\int \left\{ \int K\left(\frac{x-u}{h}\right) d[F_n(u) - \hat{F}(u)] \right\}^2 dx, \end{aligned} \tag{2}$$

where

$$K_h * \hat{f}(x) = \frac{1}{h^d}\int K\left(\frac{x-u}{h}\right)\hat{f}(u)\, du \tag{3}$$

and $\hat{F}(u)$ is the distribution function corresponding to $\hat{f}(u)$.

Fan (1994) showed that under appropriate conditions, if H_0 holds, then

$$nh^{d/2}\left[J_n(h) - \frac{1}{nh^d}\int K^2(z)\, dz\right] \to 2\sigma N(0,1), \quad \text{in distribution,}$$

where $\sigma^2 = \int f^2(x)\, dx \{\int [\int K(u)K(u+v)\, du]^2\, dv\}$. The conditions imposed on h are $h \equiv h_n \to 0$ and $nh^d \to \infty$, as $n \to \infty$. On the basis of this distribution, Fan (1994) proposed the following bias-corrected test statistic:[1]

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

$$\tilde{T}_n = \frac{nh^{d/2}\left[J_n(h) - (nh^d)^{-1}\int K^2(z)\,dz\right]}{2^{1/2}\tilde{\sigma}}, \tag{4}$$

where $\tilde{\sigma}^2 = \int \tilde{f}^2(x)\,dx\{\int[\int K(u)K(u+v)\,du]^2\,dv\}$. Hence, the bias-corrected test will reject H_0 at significance level α if $\tilde{T}_n > z_\alpha$, where z_α is the upper α-percentile of the standard normal distribution.

The consistency of the preceding test is based on the fact that under appropriate conditions including $h \to 0$ and $nh^d \to \infty$, $J_n(h) = \int[\,f(x) - f_0(x,\beta_*)]^2\,dx + o_p(1)$, where $\beta_* = \beta_0$ under H_0 (see assumption (P) in Section 3 for the definition of β_*). This seems to suggest that the conditions on h should be necessary for the consistency of any test based on $J_n(h)$, because after all they are necessary conditions for the consistency of the kernel estimator $\tilde{f}(x)$. However, an alternative expression for $J_n(h)$ given in Lemma 2.1, which follows, shows that these conditions are not needed. Moreover, h does not even have to depend on n.

LEMMA 2.1. *Suppose $K(\cdot)$ is symmetric about zero. Let $\bar{K}(\cdot)$ be the Fourier transform of $K(\cdot)$, i.e., $\bar{K}(t) = (2\pi)^{-d/2}\int\exp(it'x)K(x)\,dx$, where $i^2 \equiv -1$. Then, $J_n(h) = \int|C_n(t) - \hat{C}(t)|^2\bar{K}^2(ht)\,dt$, where $C_n(t) \equiv \int\exp(it'x)\,d[F_n(x)]$ is the empirical characteristic function of the observations and $\hat{C}(t) \equiv C_0(t,\hat{\beta}) \equiv \int\exp(it'x)\,d[\hat{F}(x)]$ is the characteristic function of $\hat{F}$.*

Proof. From the definition of $\bar{K}(\cdot)$, we get $\bar{K}(ht) = (2\pi)^{-d/2}h^{-d}\int\exp(it'x)K(x/h)\,dx$. Thus, $h^d\bar{K}(ht)$ is the Fourier transform of $K(\cdot/h)$. Now, we consider the Fourier transform of $\tilde{f}(x) - K_h * \hat{f}(x)$: It is given by

$$\begin{aligned}
&\frac{1}{(2\pi)^{d/2}}\int\exp(it'x)[\,\tilde{f}(x) - K_h * \hat{f}(x)]\,dx \\
&\quad= \frac{1}{(2\pi)^{d/2}h^d}\int\exp(it'x)\left\{\int K\left(\frac{x-u}{h}\right)d[F_n(u) - \hat{F}(u)]\right\}dx \\
&\quad= \frac{1}{h^d}\int\left[\frac{1}{(2\pi)^{d/2}}\int\exp(it'x)K\left(\frac{x-u}{h}\right)dx\right]d[F_n(u) - \hat{F}(u)] \\
&\quad= \int\exp(it'u)\bar{K}(ht)\,d[F_n(u) - \hat{F}(u)],
\end{aligned} \tag{5}$$

where the last equality is obtained by using the result that $h^d\bar{K}(ht)$ is the Fourier transform of $K(\cdot/h)$. By (2), (5), and Parseval's identity (see Debnath, 1995), we have

$$\begin{aligned}
J_n(h) &= \int\left|\int\exp(it'u)\bar{K}(ht)\,d[F_n(u) - \hat{F}(u)]\right|^2 dt \\
&= \int\left|\int\exp(it'u)\,d[F_n(u) - \hat{F}(u)]\right|^2\bar{K}^2(ht)\,dt,
\end{aligned}$$

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

because $\bar{K}(\cdot)$ is real given that $K(\cdot)$ is symmetric. This completes the proof. ■

Lemma 2.1 shows that $J_n(h)$ equals to a weighted integrated squared difference between the empirical characteristic function of the observations and an estimator of the characteristic function corresponding to the null model. Let $C(t)$ denote the characteristic function of F. It is well known that $C_n(t)$ is a consistent[2] estimator of $C(t)$. Also, under some regularity conditions, $\hat{C}(t)$ is a consistent estimator of $C_0(t,\beta_*) \equiv \int \exp(it'x) f_0(x,\beta_*)\, dx$, where $\beta_* = \beta_0$ under H_0. Thus, for a fixed h, we would expect $J_n(h) \to J(h)$ in probability, where $J(h) \equiv \int |C(t) - C_0(t,\beta_*)|^2 \bar{K}^2(ht)\, dt$. Given the one-to-one correspondence between distribution functions and characteristic functions, we get $J(h) \geq 0$ and $J(h) = 0$ iff H_0 holds, provided that the set $\{t \in R^d : \bar{K}(t) = 0\}$ has Lebesgue measure zero. Consequently, tests based on $J_n(h)$ for any fixed h would be consistent for H_0 against H_1. One interesting implication of this is that the consistency of $\tilde{f}(x)$ is not necessary for developing a consistent test for H_0 based on $J_n(h)$. See Anderson, Hall, and Titterington (1994) and Feuerverger (1993) on similar observations on two sample goodness-of-fit tests and tests for bivariate dependence. In our case, it is important that the bias of $\tilde{f}(x)$ be removed from $J_n(h)$. One can easily demonstrate that the smoothing parameter must vanish for tests based on direct comparison of $\tilde{f}(x)$ and $\hat{f}(x)$ to be consistent.[3] Of course, the asymptotic distribution of $J_n(h)$ for a fixed h and that for a decreasing h in n are different, leading to different critical values for tests based on $J_n(h)$. One apparent advantage of fixing h is that there is a one-to-one correspondence between values of h and the asymptotic distributions of $J_n(h)$. Thus, one would expect to obtain a better approximation[4] of the finite-sample distribution of $J_n(h)$ by its asymptotic distribution with a fixed h than with a vanishing h.

The alternative expression stated in Lemma 2.1 also facilitates the application of the tests based on $J_n(h)$ to testing the goodness-of-fit of the Stable law family for which closed-form expressions exist for the characteristic function but do not exist for the distribution function. This possibility will be explored elsewhere.

For both notational compactness and generality, in the remainder of this paper, we will study asymptotic properties of tests for H_0 based on the general weighted integrated squared difference between $C_n(\cdot)$ and $\hat{C}(\cdot)$, i.e.,

$$J_n \equiv \int |C_n(t) - \hat{C}(t)|^2 W(t)\, dt, \tag{6}$$

where $W(\cdot)$ is a non-negative weighting function.

If we choose $W(t) = \bar{K}^2(ht)$ in (6), then $J_n = J_n(h)$. The Cramér–von Mises test corresponds to J_n with $W(t) = t^{-2}$, which severely down weights high-frequency components. In general, by choosing a different weighting function $W(\cdot)$, one can direct power of the test based on J_n toward different directions of departure from the null model.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

3. THE ASYMPTOTIC DISTRIBUTION

We first collect the assumptions to be used to derive the asymptotic distribution of J_n.

Assumption (W). The weighting function $W(\cdot)$ is integrable and satisfies the condition that the set $\{t \in R^d : W(t) = 0\}$ has Lebesgue measure zero.

Assumption (f_0). The parametric density function f_0 is differentiable with respect to β up to the second order, inclusive, such that the following conditions are satisfied: (i) for almost all $x \in \mathcal{X}$, $D^2 f_0(x, \cdot)$ is continuous in a neighborhood of β_0; (ii) there exists a neighborhood $\mathcal{N}(\beta_0)$ of β_0 such that for almost all $x \in \mathcal{X}$ and all $\beta \in \mathcal{N}(\beta_0)$, every element of $D^2 f_0(x, \beta)$ is bounded by an integrable function $M(x)$; (iii) each element of $Df_0(x, \beta_0)$ is absolutely integrable, where D and D^2 are the first- and second-order differential operators with respect to β.

Assumption (P). There exists $\beta_* \in \mathcal{B}$ such that $\hat{\beta} \to \beta_*$ almost surely and

$$\hat{\beta} - \beta_* = \frac{1}{n} A(\beta_*)^{-1} \sum_{j=1}^{n} D \log f_0(X_j, \beta_*) + o_p(n^{-1/2}),$$

where $A(\beta_*) = E[D^2 \log f_0(X_j, \beta_*)]$. Here, and in what follows, expectations are taken with respect to the true distribution function, unless otherwise stated.

The first assumption ensures that $J \equiv \int |C(t) - C_0(t, \beta_*)|^2 W(t)\, dt$ is a distance measure between $C(\cdot)$ and $C_0(\cdot, \beta_*)$. Consequently, it is also a distance measure between $f(\cdot)$ and $f_0(\cdot, \beta_*)$. For the bias-corrected test, this assumption is satisfied, provided that the kernel function $K(\cdot)$ is symmetric about zero, is square integrable, and integrates to one such that its Fourier transform $\bar{K}(\cdot)$ satisfies the condition that the set $\{t \in R^d : \bar{K}(t) = 0\}$ has Lebesgue measure zero. Most of the commonly used kernel functions such as the uniform kernel, the standard normal kernel, the triangular kernel, the double exponential kernel, and the Epanechninov kernel satisfy this condition. The last two assumptions are introduced here to examine the effect of estimating $f_0(x, \beta_*)$ by $\hat{f}(x)$ on the asymptotic distribution of J_n. The validity of the last assumption (Assumption (P)) was verified by White (1982) under certain regularity conditions, and β_* was shown there to be the parameter vector that minimizes the Kullback–Leibler information criterion (KLIC) between the true distribution of the data and the parametric distribution with density function $f_0(x, \beta)$ (for details, see White, 1982). If the parametric distribution $f_0(x, \beta)$ correctly specifies the true distribution of the data, i.e., H_0 holds, then $\beta_* = \beta_0$, the true value of β. If H_0 is not true, then β_* can be regarded as a "pseudo-true" value of β. Throughout the rest of this paper, we assume the corresponding conditions in White (1982) are satisfied, so that Assumption (P) holds.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

It follows from (6) that

$$J_n = \int \left| \int \cos(t'x)\, d[F_n(x) - \hat{F}(x)] + i \int \sin(t'x)\, d[F_n(x) - \hat{F}(x)] \right|^2 W(t)\ dt$$

$$= \int \left\{ \int \cos(t'x)\, d[F_n(x) - \hat{F}(x)] \right\}^2 W(t)\, dt$$

$$+ \int \left\{ \int \sin(t'x)\, d[F_n(x) - \hat{F}(x)] \right\}^2 W(t) dt. \tag{7}$$

Let

$$I_n = \int \left\{ \int \cos(t'x)\, d[F_n(x) - F_0(x,\beta_0)] - \mu(t)(\hat{\beta} - \beta_0) \right\}^2 W(t)\, dt$$

$$+ \int \left\{ \int \sin(t'x)\, d[F_n(x) - F_0(x,\beta_0)] - v(t)(\hat{\beta} - \beta_0) \right\}^2 W(t)\, dt, \tag{8}$$

where $\mu(t) = \int \cos(t'x) D' f_0(x,\beta_0)\, dx$ and $v(t) = \int \sin(t'x) D' f_0(x,\beta_0)\, dx$.

By Taylor series expansion, we get $\hat{f}(x) \equiv f_0(x,\hat{\beta}) = f_0(x,\beta_0) + D' f_0(x,\beta_0)(\hat{\beta} - \beta_0) + (\hat{\beta} - \beta_0)' D^2 f_0(x,\bar{\beta})(\hat{\beta} - \beta_0)/2$, where $\bar{\beta}$ is between $\hat{\beta}$ and β_0. Substituting this into (7) and collecting terms yield

$$n[J_n - I_n] = -n \int \left\{ \int \cos(t'x)\, d[F_n(x) - F_0(x,\beta_0)] - \mu(t)(\hat{\beta} - \beta_0) \right\}$$

$$\times \left[(\hat{\beta} - \beta_0)' \int \cos(t'x) D^2 f_0(x,\bar{\beta})\, dx (\hat{\beta} - \beta_0) \right] W(t)\, dt$$

$$+ \frac{n}{4} \int \left[(\hat{\beta} - \beta_0)' \int \cos(t'x) D^2 f_0(x,\bar{\beta})\, dx (\hat{\beta} - \beta_0) \right]^2 W(t)\, dt$$

$$- n \int \left\{ \int \sin(t'x)\, d[F_n(x) - F_0(x,\beta_0)] - v(t)(\hat{\beta} - \beta_0) \right\}$$

$$\times \left[(\hat{\beta} - \beta_0)' \int \sin(t'x) D^2 f_0(x,\bar{\beta})\, dx (\hat{\beta} - \beta_0) \right] W(t)\, dt$$

$$+ \frac{n}{4} \int \left[(\hat{\beta} - \beta_0)' \int \sin(t'x) D^2 f_0(x,\bar{\beta})\, dx (\hat{\beta} - \beta_0) \right]^2 W(t)\ dt$$

$$\equiv -I_1 + \tfrac{1}{4} I_2 - I_3 + \tfrac{1}{4} I_4. \tag{9}$$

We will show in the following lemma that $I_j = o_p(1)$ for $j = 1,2,3,4$.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

LEMMA 3.1. *Under* H_0, $n[J_n - I_n] = o_p(1)$, *provided that Assumptions* (W), (f_0) *hold and* $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$.

Proof. First we consider I_1. Let $a_n(t) = \int \cos(t'x)\, d[F_n(x) - F_0(x, \beta_0)]$ and $b_n(t) = \int \cos(t'x) D^2 f_0(x, \bar{\beta})\, dx$. Then, from (9), it follows that

$$I_1 = [\sqrt{n}(\hat{\beta} - \beta_0)'] \int [a_n(t) b_n(t)] W(t)\, dt [\sqrt{n}(\hat{\beta} - \beta_0)]$$

$$- (\hat{\beta} - \beta_0)' \int \mu'(t) [\sqrt{n}(\hat{\beta} - \beta_0)'] b_n(t) W(t)\, dt [\sqrt{n}(\hat{\beta} - \beta_0)]. \qquad \textbf{(10)}$$

Now, for any fixed t, $a_n(t)$ converges to zero almost surely under H_0 by the strong law of large numbers. Also, $|a_n(t)| \leq 2$. By Assumption (f_0) and the dominated convergence, we have $b_n(t)$ converges to $\int \cos(t'x) D^2 f_0(x, \beta_0)\, dx$ and the absolute value of each element of $b_n(t)$ is bounded by $\int M(x) dx < \infty$. Hence, $a_n(t) b_n(t) W(t)$ converges to zero and the absolute value of each of its elements is bounded by $MW(t)$, where M is a positive generic constant. Moreover, Assumption (W) implies that $\int W(t)\, dt < \infty$. Hence, by applying the dominated convergence again, we get $\int a_n(t) b_n(t) W(t)\, dt$ converges to zero. Thus, the first term on the right-hand side of (10) is $o_p(1)$, because $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$. Similar arguments lead to the conclusion that the second term is $O_p(n^{-1/2})$. Thus, $I_1 = o_p(1)$.

For I_2, note that $I_2 = n^{-1} \int [\sqrt{n}(\hat{\beta} - \beta_0) b_n(t) \sqrt{n}(\hat{\beta} - \beta_0)]^2 W(t)\, dt = O_p(n^{-1})$. Similarly, one can show that $I_3 = o_p(1)$ and $I_4 = O_p(n^{-1})$. ■

One implication of Lemma 3.1 is that nJ_n has the same asymptotic distribution as nI_n under the null hypothesis. The effect of estimating β_0 by $\hat{\beta}$ on the asymptotic distribution of nJ_n is captured by the two terms with $\mu(t)$ and $v(t)$ on the right-hand side of (8). In particular, the asymptotic distribution of the bias-corrected test with a fixed smoothing parameter is affected by the estimation of β_0. This is in sharp contrast to the case of a vanishing smoothing parameter analyzed in Fan (1994).

To derive the asymptotic distribution of nI_n, we note from (8) and Assumption (P) that

$$I_n = \int \left\{ \frac{1}{n} \sum_{j=1}^{n} a(t, X_j) + \mu(t) o_p(n^{-1/2}) \right\}^2 W(t)\, dt$$

$$+ \int \left\{ \frac{1}{n} \sum_{j=1}^{n} b(t, X_j) + v(t) o_p(n^{-1/2}) \right\}^2 W(t)\, dt, \qquad \textbf{(11)}$$

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

where

$$a(t,X_j) = \cos(t'X_j) - \int \cos(t'x) f_0(x,\beta_0)\,dx - \mu(t)A(\beta_0)^{-1}D\log f_0(X_j,\beta_0), \tag{12}$$

$$b(t,X_j) = \sin(t'X_j) - \int \sin(t'x) f_0(x,\beta_0)\,dx - v(t)A(\beta_0)^{-1}D\log f_0(X_j,\beta_0). \tag{13}$$

Let

$$I_n^0 = \int \left[\frac{1}{n}\sum_{j=1}^n a(t,X_j)\right]^2 W(t)\ dt + \int \left[\frac{1}{n}\sum_{j=1}^n b(t,X_j)\right]^2 W(t)\,dt. \tag{14}$$

We have the following result.

LEMMA 3.2. *Under* H_0, $n[I_n - I_n^0] = o_p(1)$, *if Assumptions* (W), (f_0), *and* (P) *hold*.

Proof. Let J_1 and J_2 be the first and second terms on the right-hand side of (11), respectively. Then,

$$nJ_1 = n\int \left[\frac{1}{n}\sum_{j=1}^n a(t,X_j)\right]^2 W(t)\,dt + 2o_p(n^{1/2})\int \left[\frac{1}{n}\sum_{j=1}^n a(t,X_j)\right]\mu(t)W(t)\,dt + o_p(1)\int \mu^2(t)W(t)\,dt. \tag{15}$$

Let $Y_n \equiv n^{-1/2}\sum_{j=1}^n \int a(t,X_j)\mu(t)W(t)\,dt$. It is easy to see that under H_0, $E[Y_n] = 0$ and $\mathrm{Var}[Y_n] = \mathrm{Var}[\int a(t,X_1)\mu(t)W(t)\,dt] < \infty$. Hence, $Y_n = O_p(1)$. Thus, from (15), it follows that

$$nJ_1 = n\int \left[\frac{1}{n}\sum_{j=1}^n a(t,X_j)\right]^2 W(t)\,dt + o_p(1),$$

because the second term on the right-hand side of (15) equals $o_p(1)Y_n = o_p(1)$. Similarly, one can show that

$$nJ_2 = n\int \left[\frac{1}{n}\sum_{j=1}^n b(t,X_j)\right]^2 W(t)\,dt + o_p(1). \qquad \blacksquare$$

In view of Lemmas 3.1 and 3.2, we only need to consider the asymptotic distribution of nI_n^0. To this end, note that I_n^0 can be decomposed into the sum of two terms: one a degenerate U-statistic that determines the form of the asymptotic distribution of I_n^0 and the other term that provides the appropriate center for the asymptotic distribution of I_n^0:

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

$$I_n^0 = \frac{1}{n^2}\sum\sum_{j\neq k}\int [a(t,X_j)a(t,X_k) + b(t,X_j)b(t,X_k)]W(t)\,dt$$
$$+ \frac{1}{n^2}\sum_j \int [a^2(t,X_j) + b^2(t,X_j)]W(t)\,dt$$
$$\equiv \frac{1}{n^2}\sum\sum_{j\neq k} P(X_j,X_k) + \frac{1}{n^2}\sum_j P(X_j,X_j)$$
$$\equiv U_n + S_n, \tag{16}$$

where

$$P(x,y) = \int [a(t,x)a(t,y) + b(t,x)b(t,y)]W(t)\,dt.$$

Observe that $P(x,y)$ is a symmetric function and $\int P(x,y)\,f_0(x,\beta_0)\,dx = 0$ under H_0. Hence, U_n is a degenerate U-statistic whose kernel function does not depend on n. Suppose $\int\int P^2(x,y)\,f_0(x,\beta_0)\,f_0(y,\beta_0)\,dx\,dy < \infty$; then Theorem 1 in Lee (1990, p. 79) implies that under H_0, nU_n converges in distribution to a random variable of the form $\sum_{v=1}^{\infty}\lambda_v(Z_v^2 - 1)$, where $Z_1, Z_2, \ldots$ are independent standard normal random variables and the λ_v's are the eigenvalues of the integral equation

$$\int P(x_1,x_2)\psi(x_2)\,dF_0(x_2,\beta_0) = \lambda\psi(x_1). \tag{17}$$

Now we analyze S_n. Note that nS_n is a sample average of n i.i.d. random variables: $\{P(X_j,X_j)\}_{j=1}^n$. By the strong law of large numbers, $nS_n - E_0[nS_n] \to 0$ almost surely, provided that $E_0[|P(X_1,X_1)|] < \infty$, where $E_0[\cdot]$ is the expectation of $\cdot$ with respect to $f_0(x,\beta_0)$. Thus, $nS_n = E_0[P(X_1,X_1)] + o_p(1)$.

In addition, because $\sum_{v=1}^{\infty}\lambda_v = E_0[P(X_1,X_1)]$ by Dunford and Schwartz (1963, p. 1087), we have our main result, provided we can verify the following two conditions: $\int\int P^2(x,y)\,f_0(x,\beta_0)\,f_0(y,\beta_0)\,dx\,dy < \infty$ and $\int |P(x,x)|\,f_0(x,\beta_0)\,dx < \infty$. Noting that

$$|a(t,x)| \le 2 + |\mu(t)A(\beta_0)^{-1}D\log f_0(x,\beta_0)|$$
$$|b(t,x)| \le 2 + |v(t)A(\beta_0)^{-1}D\log f_0(x,\beta_0)|,$$

one can easily show that the preceding two conditions are satisfied under the assumptions stated in Theorem 3.3, which follows.

THEOREM 3.3. *Under H_0 and Assumption* (W), $nI_n^0 - E_0[P(X_1,X_1)] \to \sum_{v=1}^{\infty}\lambda_v(Z_v^2 - 1)$. *Or equivalently*, $nI_n^0 \to \sum_{v=1}^{\infty}\lambda_v Z_v^2$. *Moreover, if Assumptions* (f_0) *and* (P) *hold, then* $nJ_n \to \sum_{v=1}^{\infty}\lambda_v Z_v^2$ *in distribution.*

For the bias-corrected test with $W(t) = \bar{K}^2(ht)$ in (6), the degenerate U-statistic U_n in (16) may depend on n through the dependence of the smoothing parameter

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

h on n in such a way that it has an asymptotic normal distribution (see Section 2). However, if h is fixed, then Theorem 3.3 implies that the asymptotic distribution in $nJ_n(h)$ is not normal. Moreover, it follows from (17) that in general, the eigenvalues λ_v will depend on h. Thus, the distribution of the random variable $\sum_{v=1}^{\infty} \lambda_v Z_v^2$ will also depend on h, which implies that for a given significance level α, the critical value of the asymptotic distribution of $nJ_n(h)$ will be different for different values of h. There is one critical value for one and only one value of h. In view of this and the fact that the asymptotic distribution of $nJ_n(h)$ for a fixed h is affected by the estimate of β_0 (as would be the case for finite samples), one would expect that the distribution of $\sum_{v=1}^{\infty} \lambda_v Z_v^2$ would provide a better approximation to the finite-sample distribution of $nJ_n(h)$ than the corresponding normal distribution (see also note 4).

In general, λ_v will depend on the unknown parameter β_0. Except in simple cases, it is extremely difficult to solve (17). However, the eigenvalues λ_v from (17) may be consistently estimated (see Silverman, 1996). Given the estimates of the eigenvalues, $\hat{\lambda}_v$ (say), one can then use one of the following two distributions to approximate the distribution of nJ_n: (i) the distribution of $\sum_{v=1}^{V} \hat{\lambda}_v Z_v^2$, where V is a finite positive integer (see Gregory, 1980); (ii) the distribution of $\sum_{v=1}^{V} \hat{\lambda}_v Z_v^2 + k\chi_{[l]}^2$, where $\chi_{[l]}^2$ denotes a χ^2 random variable with l degrees of freedom and k and l are chosen such that the first two moments of $\sum_{v=1}^{V} \hat{\lambda} v Z_v^2 + k\chi_{[l]}^2$ match those of $\sum_{v=1}^{\infty} \lambda_v Z_v^2$ (see Durbin, Knott, and Taylor, 1975). Other methods may also be used to approximate the distribution or critical values of nJ_n. These include the conditional Monte-Carlo approach of Hansen (1996) and the application of case-invariant upper bounds of the critical values given in Bierens and Ploberger (1997) (for details about both approaches, see Bierens and Ploberger, 1997). In Section 5, we propose a parametric bootstrap procedure and show that it works.

Let C_α denote an estimate of the α-percentile of the distribution of the random variable $\sum_{v=1}^{\infty} \lambda_v Z_v^2$ obtained by one of the methods we just mentioned. Theorem 3.3 suggests the following asymptotic test: For each given h, the test will reject H_0 at significance level α, if

$$T_n \equiv nJ_n > C_\alpha.$$

4. CONSISTENCY AND LOCAL POWER PROPERTIES

For a fixed alternative distribution $F(x) \neq F_0(x, \beta)$ for all $\beta \in \mathcal{B}$, it follows from Assumptions (W), (P), and the dominated convergence that $J_n \to J = \int |C(t) - C_0(t, \beta_*)|^2 W(t)\, dt > 0$. Thus, for a given significance level α,

$$P(T_n > C_\alpha) = P(J_n > n^{-1} C_\alpha) \to P(J > 0) = 1.$$

THEOREM 4.1. *Suppose that Assumptions* (W) *and* (P) *hold. Then the test* T_n *is consistent.*

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

One implication of Theorem 4.1 is that the bias-corrected test with a fixed smoothing parameter is consistent.

To investigate the local power property of test T_n, we make use of the concept of contiguity of one sequence of probability measures to the other. Let P_n and Q_n be two sequences of probability measures defined on measurable spaces $(\mathcal{X}_n, \mathcal{A}_n)$, $n \geq 1$, respectively. If for any sequence of events $\{A_n\}$, $A_n \in \mathcal{A}_n$, $[P_n(A_n) \to 0]$ implies $[Q_n(A_n) \to 0]$, we say that Q_n are contiguous to P_n. For a detailed discussion on this, see Hájek and Šidák (1967).

One of the most useful properties of contiguity is that any sequence of random variables converging to zero in P_n probability converges to zero in Q_n probability as $n \to \infty$.

For $A \in \mathcal{A}$, define $P_0(A) = \int_{x \in A} f_0(x, \beta_0)\, dx$ and $P_n(A) = \int_{x \in A} f(x)\, dx$. Consider a Pitman alternative distribution with density function

$$f(x) \equiv f_n(x) = f_0(x, \beta_0)[1 + n^{-1/2}\delta_n(x)], \tag{18}$$

for some converging sequence $\delta_n(\cdot)$ in $L_2(\mathcal{X}, \mathcal{A}, P_0)$ with $\delta_n(\cdot) \to \delta(\cdot)$.

From well-known criteria for contiguity (see, e.g., Roussas, 1972), it follows that the sequence of probability measures $P_H^n = P_n \times \cdots \times P_n$ (n-times), $n \geq 1$, is contiguous to $P_0^n = P_0 \times \cdots \times P_0$ (n-times), $n \geq 1$. Because Lemmas 3.1 and 3.2 imply that $n(J_n - I_n^0)$ converges to zero in P_0^n probability (or under H_0), it follows from the contiguity of P_H^n to P_0^n that $n(J_n - I_n^0)$ also converges to zero in P_H^n probability (or under (18)). Hence, nJ_n has the same asymptotic distribution as nI_n^0 under (18), where $nI_n^0 = nU_n + nS_n$ with U_n and S_n defined in (16). By Theorem 2.1 in Gregory (1977), we have nU_n converges in distribution to $\sum_{v=1}^{\infty} \lambda_v[(Z_v + a_v)^2 - 1]$ under (18), where $a_v = \int \delta\psi_v\, dP_0$. By contiguity again and the proof of Theorem 3.3, we get $nS_n = E_0[P_h(X_1, X_1)] + o_p(1)$. Hence, we obtain the following theorem.

THEOREM 4.2. *Suppose assumptions in Theorem* 3.3 *hold. For local alternatives satisfying* (18), $T_n = nJ_n$ *converges in distribution to* $\sum_{v=1}^{\infty} \lambda_v(Z_v + a_v)^2$, *where* $a_v = \int \delta\psi_v\, dP_0$, λ_v *and* ψ_v *are eigenvalues and eigenfunctions of the integral equation* (17).

From independence of Z_v, $v = 1, 2, \ldots$, and the fact that for a normal random variable Z, $(Z + a)^2$ is stochastically larger than Z^2 for $a \neq 0$, it follows that $\sum_{v=1}^{\infty} \lambda_v(Z_v + a_v)^2$ is stochastically larger than $\sum_{v=1}^{\infty} \lambda_v Z_v^2$, provided that $a_v \neq 0$ for at least one v. This and Theorem 4.2 imply the unbiasedness of our test for Pitman alternatives in (18). For the bias-corrected test, this shows that tests based on fixed smoothing parameters have comparable power as parametric tests in the sense that they are able to detect Pitman local alternatives that differ from the null by $O(n^{-1/2})$ and hence are asymptotically more powerful than those based on decreasing h against Pitman local alternatives. However, we note that the power advantage of the bias-corrected test with a fixed smoothing parameter is with respect to the class of Pitman local alternatives. Fan (1994) considered another class of local alternatives where $\delta_n(\cdot)$ in (18) does not settle down to a well-

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

behaved function; instead it has many spikes and changes rapidly with respect to n. It is shown there that the bias-corrected test with a vanishing smoothing parameter has nontrivial power against such alternatives that differ from the null by $o(n^{-1/2})$, as long as the smoothing parameter is chosen to vanish at an appropriate rate. Hence for high-frequency alternatives, the bias-corrected test with a vanishing smoothing parameter may be more powerful than the corresponding test with a fixed smoothing parameter.

The preceding result can easily be understood by the alternative expression for $J_n(h)$ given in Lemma 2.1. Recall from Lemma 2.1 that $J_n(h) = \int |C_n(t) - \hat{C}(t)|^2 \bar{K}^2(ht)\,dt$, i.e., $W(t) = \bar{K}^2(ht)$ in (6). For a given kernel function $K(\cdot)$ and a fixed smoothing parameter h, $W(t) \to 0$ as $|t| \to \infty$. Hence the weight given by $W(t)$ to high-frequency components (corresponding to large $|t|$) is small. For two values of the smoothing parameter h_1 and h_2 (say) with $h_1 < h_2$, we have $h_1|t| < h_2|t|$. Hence, a small value of h tends to increase the weight given to high-frequency components.[5]

5. THE PARAMETRIC BOOTSTRAP

Suppose Assumption (P) holds. Then $\hat{\beta} \to \beta_*$ almost surely whether or not H_0 is true. Hence $\hat{f}(x) = f_0(x, \hat{\beta})$ mimics $f_0(x, \beta_*)$ whether or not H_0 is true. Noting that H_0 is a composite hypothesis and $f_0(x, \beta_*)$ belongs to H_0, we propose the following parametric bootstrap procedure to approximate the critical values of $T_n = nJ_n$, where the bootstrap samples are drawn from the parametric distribution with density function $\hat{f}(x) = f_0(x, \hat{\beta})$.

Step 1. Draw a random sample of size n, $\{X_j^*\}_{j=1}^n$ (say), from the distribution with density function $\hat{f}(x)$. This is the bootstrap sample. Hence conditional on the random sample $\{X_j\}_{j=1}^n$, the bootstrap sample satisfies H_0 with $\beta = \hat{\beta}$.

Step 2. Use the bootstrap sample $\{X_j^*\}_{j=1}^n$ in place of the data $\{X_j\}_{j=1}^n$ to compute T_n and call it T_n^*, i.e.,

$$T_n^* = nJ_n^* = n \int |C_n^*(t) - \hat{C}^*(t)|^2 W(t)\,dt,$$

where $C_n^*(t) = n^{-1} \sum_{j=1}^n \exp(it'X_j^*)$ and $\hat{C}^*(t) = \int \exp(it'x) f_0(x, \hat{\beta}^*)\,dx$ with $\hat{\beta}^*$ the maximum likelihood estimator based on the bootstrap sample. The distribution of T_n^* conditional on the random sample $\{X_j\}_{j=1}^n$ is called the bootstrap distribution of T_n.

Step 3. To approximate the bootstrap distribution of T_n, we repeat Step 1 and Step 2 for a large number of times, say B times, and obtain the empirical distribution of $T_{n1}^*, \ldots, T_{nB}^*$ (say). This empirical distribution estimates the bootstrap distribution of T_n.

Let $C_\alpha(\hat{\beta})$ be the upper α-percentile of the preceding bootstrap distribution. We will reject H_0 at significance level α if $T_n > C_\alpha(\hat{\beta})$. We now show that the preceding bootstrap procedure works under the additional assumption (P*), which follows.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

Assumption (P*).

$$\hat{\beta}^* - \hat{\beta} = \frac{1}{n} A(\hat{\beta})^{-1} \sum_{j=1}^{n} D \log \hat{f}(X_j) + o_{p^*}(n^{-1/2}),$$

where $A(\hat{\beta}) = E^*[D^2 \log f_0(X_j^*, \hat{\beta})]$ and $E^*[\cdot]$ is the expectation of $\cdot$ conditional on the random sample $\{X_j\}_{j=1}^n$.

THEOREM 5.1. *Suppose Assumptions* (W), (f_0), (P), *and* (P*) *hold. Then* $\mathcal{L}^*(T_n^*)$ *converges to the same limit as* $\mathcal{L}(T_n | f_0(x, \beta_*))$ *almost surely, where* $\mathcal{L}^*(T_n^*)$ *is the distribution of* T_n^* *conditional on the random sample* $\{X_j\}_{j=1}^n$ *and* $\mathcal{L}(T_n | f_0(x, \beta_*))$ *is the distribution of* T_n *when the random sample is generated from the distribution with density function* $f_0(x, \beta_*)$.

Proof. Under the assumptions stated in the theorem, one can easily show by following the proofs of Lemmas 3.1 and 3.2 that $T_n^* - nI_n^{*0} = o_{p^*}(1)$, where

$$nI_n^{*0} = \int \left[\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{a}(t, X_j^*) \right]^2 W(t)\, dt + \int \left[\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{b}(t, X_j^*) \right]^2 W(t)\, dt, \quad \textbf{(19)}$$

$$\hat{a}(t, X_j^*) = \cos(t' X_j^*) - \int \cos(t'x) \hat{f}(x)\ dx - \hat{\mu}(t) A(\hat{\beta})^{-1} D \log f_0(X_j^*, \hat{\beta}), \quad \textbf{(20)}$$

$$\hat{b}(t, X_j^*) = \sin(t' X_j^*) - \int \sin(t'x) \hat{f}(x)\, dx - \hat{v}(t) A(\hat{\beta})^{-1} D \log f_0(X_j^*, \hat{\beta}), \quad \textbf{(21)}$$

with $\hat{\mu}(t) = \int \cos(t'x) D' f_0(x, \hat{\beta})\, dx$ and $\hat{v}(t) = \int \sin(t'x) D' f_0(x, \hat{\beta})\, dx$.

Hence, we only need to analyze $\mathcal{L}^*(nI_n^{*0})$. The limiting distribution of nI_n^{*0} conditional on the random sample $\{X_j\}_{j=1}^n$ can be derived in a similar way to that of nI_n^0 by using results on degenerate U-statistics. However, to show that $\mathcal{L}^*(nI_n^{*0})$ converges to the same limit as $\mathcal{L}(T_n | f_0(x, \beta_*))$ or equivalently $\mathcal{L}(nI_n^0 | f_0(x, \beta_*))$ almost surely, it is more convenient to analyze the weak convergence of Hilbert space–valued random sequences $n^{-1/2} \sum_{j=1}^n \hat{l}(\cdot, X_j^*)$, where $l = a, b$.

By Theorem 10.5 in Ledoux and Talagrand (1991), we get $n^{-1/2} \sum_{j=1}^n \hat{a}(\cdot, X_j^*)$ converges weakly to $\hat{Z}_a$, where $\hat{Z}_a = \hat{Z}_a(\cdot)$ is a $L_2(R^d)$-valued Gaussian random element with mean function 0 and covariance function $\hat{\Sigma}_a : R^d \times R^d \to R^+$ given by

$$\begin{aligned}
\hat{\Sigma}_a(s,t) &= E^*[\hat{a}(t, X_1^*) \hat{a}(s, X_1^*)] \\
&= \int \cos(t'x) \cos(s'x) \hat{f}(x)\, dx - \left[\int \cos(t'x) \hat{f}(x)\, dx \right] \\
&\quad \times \left[\int \cos(s'x) \hat{f}(x)\, dx \right] \\
&\quad - \hat{\mu}(s) A(\hat{\beta})^{-1} \int \cos(t'x) \hat{f}(x) D \log \hat{f}(x)\, dx \\
&\quad - \hat{\mu}(t) A(\hat{\beta})^{-1} \int \cos(s'x) \hat{f}(x) D \log \hat{f}(x)\, dx + \hat{\mu}(t) A(\hat{\beta})^{-1} \hat{\mu}'(s).
\end{aligned} \quad \textbf{(22)}$$

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

Similarly, one can show that $n^{-1/2}\sum_{j=1}^{n}\hat{b}(\cdot, X_j^*))$ converges weakly to $\hat{Z}_b$ with covariance function $\hat{\Sigma}_b$, where $\hat{\Sigma}_b$ is defined in (22) with a replaced by b and μ by v.

It follows from the continuous mapping theorem that $\mathcal{L}^*(nI_n^{*0})$ converges to $\int[\hat{Z}_a(t)]^2W(t)\,dt + \int[\hat{Z}_b(t)]^2W(t)\,dt$. In view of (22) and the almost sure convergence of $\hat{\beta}$ to β_*, one can easily obtain the almost sure convergence of $\int[\hat{Z}_a(t)]^2W(t)\,dt + \int[\hat{Z}_b(t)]^2W(t)\,dt$ to $\int[Z_a(t)]^2W(t)\,dt + \int[Z_b(t)]^2W(t)\,dt$, where $Z_a(\cdot)$ and $Z_b(\cdot)$ are, respectively, $L_2(R^d)$-valued Gaussian random elements with mean functions 0 and covariance functions Σ_a and Σ_b that are defined similarly to $\hat{\Sigma}_a$ and $\hat{\Sigma}_b$ with $\hat{\beta}$ replaced by β_*. Applying Theorem 10.5 in Ledoux and Talagrand (1991) and the continuous mapping theorem again to nI_n^0, it is easy to see that $\int[Z_a(t)]^2W(t)\,dt + \int[Z_b(t)]^2W(t)\,dt$ is indeed the limiting distribution of nI_n^0 when the random sample is generated from the distribution with density function $f_0(x, \beta_*)$. ■

It is important to note that Theorem 5.1 holds whether or not H_0 is true. If H_0 is indeed true, then $\beta_* = \beta_0$. Hence Theorem 5.1 implies that under H_0, the bootstrap distribution of T_n, $\mathcal{L}^*(T_n^*)$, converges to the asymptotic null distribution of T_n almost surely, which ensures that the asymptotic significance level of the test T_n based on the bootstrap critical value $C_\alpha(\hat{\beta})$ is indeed α. If H_0 is not true, then Theorem 5.1 implies that the bootstrap critical value $C_\alpha(\hat{\beta})$ converges to $C_\alpha(\beta_*)$, which[6] is the upper α-percentile of the asymptotic distribution of T_n when the random sample is generated from the distribution with density function $f_0(x, \beta_*)$. In view of the fact that H_0 is a composite hypothesis and $f_0(x, \beta_*)$ belongs to H_0, Theorem 5.1 justifies the parametric bootstrap procedure and ensures that the resulting test has power (is consistent). See Andrews (1997) for a similar discussion.

To illustrate the preceding bootstrap procedure, we have applied it to testing univariate normality by the bias-corrected test $T_n(h) = nJ_n(h)$ with the standard normal kernel function. The smoothing parameter is taken to be $h = h_0 X_{sd}$, where X_{sd} is the standard deviation of the random sample $\{X_j\}_{j=1}^n$ and h_0 equals 0.25, 0.5, 0.75, and 1.00. In the simulation experiment, the number of replications is 500 and the number of bootstrap replications is 1,000. The results for $n = 50$ and $n = 100$ are reported in Table 1. As is expected, the rejection frequencies are not far from the sizes of the test.

6. CONCLUSION

In this paper, we provide an alternative asymptotic distribution for the bias-corrected test in Fan (1994) by fixing the smoothing parameter. Unlike in Fan (1994), the asymptotic distribution obtained in this paper is non-normal, given by that of an infinite weighted sum of χ^2 random variables. This is in agreement with the existing simulation results indicating that kernel-based tests are more like χ^2 tests than normal tests. A parametric bootstrap procedure is proposed to obtain critical values for the test. We show both analytically and by simulation

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

TABLE 1. Rejection frequencies

	$n = 50$			$n = 100$		
h_0	1%	5%	10%	1%	5%	10%
0.25	0.010	0.054	0.120	0.012	0.054	0.094
0.50	0.010	0.052	0.118	0.014	0.052	0.092
0.75	0.010	0.056	0.118	0.014	0.054	0.094
1.00	0.010	0.058	0.110	0.014	0.046	0.096

that the proposed bootstrap works. We also analyze the consistency and local power properties of the bias-corrected test with a fixed smoothing parameter.

Several extensions of the current paper are straightforward: First, one can easily extend the results in this paper to near-epoch-dependent processes by using the central limit theorem for Hilbert space–valued near-epoch-dependent processes in Chen and White (1997). Second, if the random variable is symmetric about the origin, then one can test the functional form of its density function by using only the real part of the characteristic function. Specifically, one can use the test based on $T_{ns} = n\int\{\int \cos(t'x)d[F_n(x) - \hat{F}(x)]\}^2 W(t)\,dt$. The properties of this test are similar to those of T_n. Last, one can use T_n to construct tests of the goodness-of-fit of the Stable law family for which closed-form expressions exist for the characteristic function but do not exist for the distribution function. In this case, it is more convenient to estimate β_0 directly from the empirical characteristic function than from the density function. See Feuerverger and McDunnough (1981c) for details on estimating the Stable law family based on the empirical characteristic function.

NOTES

1. It is called the bias-corrected test statistic because under H_0, the expected value of the kernel estimate $\tilde{f}(x)$ is $K_h * f_0(x, \beta_0)$, a consistent estimator of which is $K_h * \hat{f}(x)$.

2. See Feuerverger and Mureika (1977) and Feuerverger and McDunnough (1981a, 1981b) for the investigation of the asymptotic properties of the empirical characteristic function including consistency and the asymptotic distribution.

3. For example, consider tests based on the integrated squared difference between $\tilde{f}(x)$ and $\hat{f}(x)$. One can easily show that

$$\int [\,\tilde{f}(x) - \hat{f}(x)]^2\,dx = (2\pi)^{-d}\int |C_n(t)[(2\pi)^{d/2}\bar{K}(ht)] - \hat{C}(t)|^2\,dt.$$

Hence to construct a consistent test using $\int[\,\tilde{f}(x) - \hat{f}(x)]^2\,dx$, the smoothing parameter h must approach zero so that $\bar{K}(ht) \to (2\pi)^{-d/2}$, as $n \to \infty$.

4. In fact, one can show that the rate of approximation of the finite-sample distribution of $J_n(h)$ under H_0 by its asymptotic distribution with a fixed h is of order $n^{-1/2}$, whereas that by its asymptotic distribution with a vanishing h is of order $h^{d/2}$. Obviously, $n^{-1/2} = o(h^{d/2})$ given $nh^d \to \infty$.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

5. The power difference between the bias-corrected tests with fixed versus vanishing smoothing parameters is asymptotic in nature. In practice, for a given data set with a fixed sample size, only one value of the smoothing parameter will be used. Once adjustment is made for size, the power difference between the two tests will disappear. However, one may infer from the preceding comparison that relatively speaking, a small value of the smoothing parameter may lead to large power for high-frequency alternatives and a large value of the smoothing parameter may lead to large power for low-frequency alternatives.

6. Note that the bootstrap critical value approximates the asymptotic critical value $C_\alpha(\beta_*)$, which depends on the value of β_*. This simply reflects the fact that the test based on T_n is not distribution free.

REFERENCES

Anderson, N.H., P. Hall, & D.M. Titterington (1994) Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis* 50, 41–54.

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65 (5), 1097–1128.

Bickel, P.J. & M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics* 1 (6), 1071–1095.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65 (5), 1129–1151.

Chen, X. & H. White (1997) Central and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications (forthcoming in *Econometric Theory*).

Debnath, L. (1995) *Integral Transforms and Their Applications.* New York: CRC Press.

Dunford, N. & J.T. Schwartz (1963) *Linear Operators.* New York: Wiley.

Durbin, J., M. Knott, & C.C. Taylor (1975) Components of Cramér-von Mises statistics. II. *Journal of Royal Statistical Society, Series B* 37, 216–237.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association* 91, 674–688.

Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* 10 (2), 316–356.

Fan, Y. (1995) Bootstrapping a consistent nonparametric goodness-of-fit test. *Econometric Reviews* 14 (3), 367–382.

Fan, Y. (1997) Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis* 62 (1), 36–63.

Feuerverger, A. (1993) A Consistent Test for Bivariate Dependence. Manuscript, University of Toronto.

Feuerverger, A. & R.A. Mureika (1977) The empirical characteristic function and its applications. *Annals of Statistics* 5 (1), 88–97.

Feuerverger, A. & P. McDunnough (1981a) On some Fourier methods for inference. *Journal of the American Statistical Association* 76, 379–386.

Feuerverger, A. & P. McDunnough (1981b) On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B* 43, 147–156.

Feuerverger, A. & P. McDunnough (1981c) On efficient inference in symmetric stable laws and processes. In M. Csörgö, D.A. Dawson, J.N.K. Rao, & A.K.Md.E. Saleh (eds.), *Statistics and Related Topics.* Amsterdam: North-Holland.

Gregory, G.G. (1977) Large sample theory for U-statistics and Tests of Fit. *Annals of Statistics* 5, 110–123.

Gregory, G.G. (1980) On efficiency and optimality of quadratic tests. *Annals of Statistics* 8, 116–131.

Hájek, J. & Z. Šidák (1967) *Theory of Rank Tests.* New York: Academic.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press

Ledoux, M. & M. Talagrand (1991) *Probability in Banach Spaces: Isoperimetry and Processes.* New York: Springer-Verlag.
Lee, A.J. (1990) *U-Statistics: Theory and Practice.* New York: Marcel Dekker.
Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3 (1), 1–14.
Roussas, G.G. (1972) *Contiguity of Probability Measures: Some Applications in Statistics.* New York: Cambridge University Press.
Silverman, B.W. (1996) Smoothed functional principal components analysis by choice of norm. *Annals of Statistics* 24, 1–24.
White, H.(1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

https://doi.org/10.1017/S0266466698145036 Published online by Cambridge University Press