

2011s-06

A test of singularity for distribution functions

Victoria Zinde-Walsh, John W. Galbraith

Série Scientifique
Scientific Series

Montréal
Janvier 2011

© 2011 Victoria Zinde-Walsh, John W. Galbraith. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaire majeur

Ministère du Développement économique, de l'Innovation et de l'Exportation

Partenaires corporatifs

Banque de développement du Canada
Banque du Canada
Banque Laurentienne du Canada
Banque Nationale du Canada
Banque Royale du Canada
Banque Scotia
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Fédération des caisses Desjardins du Québec
Financière Sun Life, Québec
Gaz Métro
Hydro-Québec
Industrie Canada
Investissements PSP
Ministère des Finances du Québec
Power Corporation du Canada
Raymond Chabot Grant Thornton
Rio Tinto
State Street Global Advisors
Transat A.T.
Ville de Montréal

Partenaires universitaires

École Polytechnique de Montréal
HEC Montréal
McGill University
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 1198-8177

Partenaire financier

Développement
économique, Innovation
et Exportation
Québec 

A test of singularity for distribution functions^{*}

Victoria Zinde-Walsh[†], John W. Galbraith[‡]

Abstract

Many non- and semi- parametric estimators have asymptotic properties that have been established under conditions that exclude the possibility of singular parts in the distribution. It is thus important to be able to test for absence of singularities. Methods of testing that focus on specific singularities do exist, but there are few generally applicable approaches. A general test based on kernel density estimation was proposed by Frigyesi and Hössjer (1998), but this statistic can diverge for some absolutely continuous distributions. Here we use a result in Zinde-Walsh (2008) to characterize distributions with varying degrees of smoothness, via functionals that reveal the behavior of the bias of the kernel density estimator. The statistics proposed here have well defined asymptotic distributions that are asymptotically pivotal in some class of distributions (e.g. for continuous density) and diverge for distributions in an alternative class, at a rate that can be explicitly evaluated and controlled.

Keywords: generalized function, kernel density estimator, singularity

Codes JEL : C14

^{*} Both authors gratefully acknowledge the support of this research by the Social Sciences and Humanities Research Council of Canada (SSHRC) and the *Fonds québécois de la recherche sur la société et la culture* (FQRSC).

[†] McGill University and CIREQ

[‡] McGill University, CIREQ, and CIRANO. Email: john.galbraith@mcgill.ca

1 Introduction

Many non- or semi-parametric estimators use nonparametric kernel estimators of density. The literature on the kernel estimator of density itself, as well as on many other estimators into which the kernel estimator enters in some form or another, typically assumes continuity of the underlying density function and thus excludes the possibility of singular parts in the distribution; often more stringent smoothness assumptions are made.

There are many situations in which it is useful to know whether the distribution has singularities or is sufficiently smooth; smoothness of the distribution is required to establish convergence rates for many widely used semi- and nonparametric statistics such as density estimators, various estimators of conditional mean and their derivatives.² Insufficient smoothness or singularities may lead to non-standard convergence rates. Testing methods have typically focused on specific singularities, following Müller (1992), but there are few generally applicable approaches. A general test of singularity based on the kernel density estimator was proposed by Frigyesi and Hössjer (1998) (hereafter ‘FH’); however, that test is not well suited for verifying the validity of assumptions in nonparametric estimation.

If there is singularity an ordinary density function cannot be defined. However, as discussed in Zinde-Walsh (2008) (hereafter ‘VZW’), density can always be defined as a generalized function, specifically as the generalized derivative, f , of the distribution function, F , which is an ordinary function.³ Consider a space D_l of “test functions”, ψ , that are l times continuously differentiable and have bounded supports. Then a generalized function is defined as a linear continuous functional on D_l and belongs to the space denoted D_l' . For example, a generalized density function f defines a functional on the space of continuously differentiable functions, D_1 as a generalized derivative of the corresponding distribution function, F . For any $\psi \in D_1$ define the value of the functional (f, ψ) as

$$(f, \psi) = - \int F(x)\psi'(x)dx.$$

²Generally a distribution function can be represented as a mixture of absolutely continuous, discrete and singular components. The known discrete part of the distribution should be treated separately both for estimation and testing (as discussed in e.g. Li and Racine, 2007); the approach here deals with the distribution over a continuum domain and from now on we do not consider discrete components. Of course, it is possible that the discrete part is mistakenly treated as continuous in kernel estimation (e.g., Zinde-Walsh, 2008, Example 3); in such a case the discussion in this paper also applies.

³See e.g. VZW, Appendix B, for the definitions of a generalized function and related concepts.

Depending on the context it may be useful to think of a generalized function as an element in one of a variety of spaces, D , where differentiability of the test function substitutes for lack of required smoothness in the distribution function. For example, if the interest is in the density itself we may view it as a functional on the space of continuously differentiable functions, D_1 . If the density function exists and is continuous it can represent a functional on the space D_0 of continuous functions: for $\psi \in D_0$ the value (f, ψ) is equal to $\int f(x)\psi(x)dx$.

In VZW asymptotic results for the kernel estimator of density were obtained under general conditions that did not require existence of density; in this general setup the asymptotic process for the kernel estimator was described as a generalized Gaussian process in the sense of Gel'fand and Vilenkin (1964). Here we explore the relation between the properties of the density and the corresponding rate of convergence on the space D_l . In the class of distributions that possess a continuous density one can select a bandwidth $h(l)$ and kernel K of order no less than l , such that the kernel density estimator $\widehat{f(l)}$ based on $h(l)$ and K converges on the space D_l ; the rate of convergence of the (generalized) bias functional determines the convergence rate for $\widehat{f(l)}$. From VZW the limit distribution is Gaussian, and so for an appropriate bandwidth a statistic that has the standard normal as its asymptotic distribution can be constructed. In the case of singularity there will exist $\psi \in D_0$ such that the estimator diverges as a result of the behavior of the asymptotic bias functional. The statistic then diverges and the divergence rate can be established. Of course due to the nonparametric nature of the estimator the rates are non-parametric and divergence may be slow.

The performance of the test will depend on the function ψ . We propose such a function constructed based on the distribution itself, to magnify the possible violations of properties of continuity and smoothness, and in this our test bears a similarity to that of FH; however, our test has established asymptotic distributional properties in a class of interest and a known asymptotic rate of divergence for an alternative class. The statistic uses the discrepancy between a kernel density estimator and the density (generalized derivative of the distribution function) weighted by the weighting function ψ . The statistic behaves differently in the classes where density exists relative to cases in which the distribution is singular. A feasible version of the statistic is shown to have the same asymptotic properties as the infeasible statistic.

In section two the new statistic is proposed and its asymptotic properties

established. In section three the properties of the test (asymptotic size and power and uniform control of power over some non-parametric subclasses of distributions) are established; results of a Monte Carlo experiment show reasonable conformity with the asymptotic properties.

2 The test statistic and its properties

We need to distinguish several classes of distribution functions. Consider the following classes, H :

H_o : $F(x)$ is absolutely continuous; density exists;

$H_c \subset H_o$: the density $f(x)$ is a continuous function;

H_s : $F(x)$ is not absolutely continuous and has singularity points (where density does not exist as an ordinary locally summable function).

$H_s(U) \subset H_s$: the set of singularities of $F(x)$ has a non-empty intersection with an open bounded set (interval) $U \subset R$.

Many of the assumptions in the nonparametric literature place the density in one or another of these classes, and our aim is to test such assumptions. However it is known that in a fully nonparametric context, even for some parametric hypotheses, no valid tests exist (Bahadur and Savage, 1956); Lalley and Nobel (2003) showed that there is no consistent test that can distinguish the H_o from the H_s class. This problem stems from the fact that near any distribution of interest (e.g. from H_o) there are distributions that are empirically (at any given sample size, for any decision rule) indistinguishable from H_s . Nonetheless, although we cannot distinguish between H_o and H_s , we may be able to distinguish some subclasses. Note as well that in practice we have a limited degree of precision in data, say up to k decimal points; for every sample size, n , of observations from a (possibly absolutely continuous) distribution there is therefore a limit to the nature of the deviation from continuity which we can detect.

Frigyesi and Hössjer (1998) proposed a test that would distinguish a subclass of H_o from H_s ; they constructed a functional $\int H(f(x))dx$ such that for the kernel density estimator, \hat{f} , the functional $\int H(\hat{f}(x))dx$ diverges for distributions in H_s ; the subclass of H_o that can be then isolated is such that the value of the functional $\int H(f(x))dx$ is bounded by a constant, C . However, having the test statistic itself define the null class means that to establish absolute continuity of a distribution, different statistics need to be constructed. The statistic proposed here is applicable to any absolutely continuous distribution.

2.1 Weighting function

We first introduce a weighting function with order of differentiability that is controlled by the properties of the distribution. If $f(x)$ is continuous, $F(x)$ is continuously differentiable and its integral, $IF(x) = \int_{-\infty}^x F(w)dw$, is twice continuously differentiable; if $f(x)$ exists but is not continuous $F(x)$ can be at most continuous but not continuously differentiable, while $IF(x) = \int_{-\infty}^x F(w)dw$, is at most once but not twice continuously differentiable; $IF(x)$ for a singular distribution is at least Lipschitz continuous with

$$|IF(x + \Delta) - IF(x)| < |\Delta|;$$

$IF(x)$ grows at infinity no faster than linearly in $|x|$.

We construct the weight function ψ_{IF} (which we now subscript to indicate the dependence on the integral of the CDF) to inherit the non-differentiability properties of IF and to dampen its growth at infinity. For some (either bounded or unbounded) open set $U \in R$, denote by Ω the set $(\text{supp } F) \cap U$, and by $\bar{\Omega}$ the closure of this set.

Assumption 1 (weighting function).

(a) The weighting function is $\psi_{IF}(x) \equiv IF(x)\phi(x)$, where $\phi(x) > 0$ on Ω ; $\phi(x)$ is at least twice continuously differentiable on $\bar{\Omega}$; $\phi(x)$ is bounded on $\bar{\Omega}$; $\int x^4 \phi(x) dx < \infty$.

(b) $E\psi_{IF}^2 > \chi > 0$; $E\psi_{IF} > A > 0$.

For example, if $\Omega = (a, b)$, an open bounded interval, $\phi(x)$ could be defined as $\left[1 - \left(\frac{2x-a-b}{b-a}\right)^2\right]^2$ for $x \in (a, b)$, 0 otherwise. For a bounded Ω and a given ϕ there is a uniform upper bound on ψ_{IF} . Alternatively for $\Omega = R$, $\phi(x)$ could be $\phi_N(x)$, the standard normal density function; since $IF(x) = O(|x|)$ as $x \rightarrow \infty$ it follows that $\psi_{IF}(x)$ is bounded in that case as well.

The condition (b) is satisfied if for some set B we have $\mu_F(B) > \pi$ and $\inf_B \phi > \pi$, for some $\pi > 0$, where $\mu_F(B) > \pi$ denotes the measure associated with the distribution F .

2.2 The infeasible statistic

The next subsections give the standard kernel density estimator used in the statistics and then introduce the infeasible statistic and examine its limit properties.

2.2.1 Density estimator

The kernel estimator of density uses a kernel that satisfies standard assumptions.

Assumption 2 (kernel).

- (a). $K(w)$ is an ordinary bounded function on R ; $\int K(w)dw = 1$;
- (b). The support of $K(w)$ belongs to $[-1, 1]$;
- (c). $K(w)$ is an l -th order kernel: the integral

$$\int w^v K(w)dw \begin{cases} = 0 & \text{if } v < l; \\ \neq 0, < \infty & \text{if } v = l. \end{cases}$$

- (d) $K(w)$ is continuously differentiable.

Typically a second-order kernel, $l = 2$, is considered: a (bounded) example is the Epanechnikov kernel. Alternatively, the finite support assumption that somewhat simplifies the derivation here could be relaxed and a Gaussian kernel could then be used.

Define the kernel density estimator

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$$

and denote the expectation of $\widehat{f}(x)$, given the kernel function and bandwidth, by $E(\widehat{f})$.

2.2.2 The infeasible statistic and its limit properties

The generalized derivative of the distribution function will be denoted by $f(x)$ (even though it may not be defined point-wise); it appears only in integrals where it is multiplied by a function with a summable derivative, such as $\psi_{IF}(x)$.

Define the infeasible statistic:

$$\begin{aligned} z &= \frac{(nh)^{\frac{1}{2}} \int \left(\widehat{f}(x) - f(x)\right) \psi_{IF}(x) dx}{(\sigma_n^2)^{\frac{1}{2}}}; \\ \sigma_n^2 &\equiv \text{var} \left[(nh)^{\frac{1}{2}} \int \left(\widehat{f}(x) - f(x)\right) \psi_{IF}(x) dx \right]. \end{aligned} \tag{1}$$

Note that $\int f(x)\psi_{IF}(x)dx$ is not random, so that σ_n^2 can be simplified to $\text{var} \left[(nh)^{\frac{1}{2}} \int \widehat{f(x)}\psi_{IF}(x)dx \right]$.

Then we can decompose the statistic (recalling that $f_{hK} = E\hat{f}$) as:

$$\begin{aligned} z &= (nh)^{\frac{1}{2}} [z_1 + z_2]; \text{ with} \\ z_1 &= \frac{\int \left(\widehat{f(x)} - E(\hat{f}) \right) \psi_{IF}(x)dx}{(\sigma_n^2)^{\frac{1}{2}}}; \\ z_2 &= \frac{\int \left(E(\hat{f}) - f(x) \right) \psi_{IF}(x)dx}{(\sigma_n^2)^{\frac{1}{2}}} \text{ (bias part)}. \end{aligned}$$

First consider the functional in the numerator of z ,

$$(nh)^{\frac{1}{2}} \int \left(\widehat{f(x)} - f(x) \right) \psi_{IF}(x)dx,$$

and its expected value, m_n , and variance, σ_n^2 . It follows that

$$m_n = (nh)^{\frac{1}{2}} \int \left(E(\hat{f}) - f(x) \right) \psi_{IF}(x)dx.$$

Lemma 1. *Under Assumptions 1 and 2 as $n \rightarrow \infty$, $h \rightarrow 0$, and $nh \rightarrow \infty$, the variance $\sigma_n^2 =$*

$$\begin{cases} \int K(w)^2 dw E(\psi_{IF}^2) + & \text{for } H_o, H_c \\ h \left[\int K(w)^2 w dw E(\psi_{IF}^2)' - (E\psi_{IF}(x))^2 \right] + o(h), & \\ \int K(w)^2 dw E(\psi_{IF}^2) + O(h), & \text{for } H_s. \end{cases} \quad (2)$$

Proof. See the Appendix.

The theorem below establishes convergence of the statistic z to standard normal (as long as the bandwidth converges appropriately) in classes H_o and H_c , and shows divergence to infinity for singular distributions (as long as the set U captures at least one singularity point).

Theorem 1. *For any given distribution under Assumptions 1 and 2 for $n \rightarrow \infty$ with bandwidth $h \rightarrow 0$, $nh \rightarrow \infty$, (i) $(nh)^{\frac{1}{2}} z_1 \xrightarrow{D} N(0, 1)$; (ii) If the distribution is in H_c , then $(nh)^{\frac{1}{2}} z_2$ is $O(nh^5)^{\frac{1}{2}}$ and in H_o is $(nh)^{\frac{1}{2}} z_2 =$*

$O(nh^3)^{\frac{1}{2}}$; (iii) If the distribution is in $H_s(U)$ the function $(nh)^{\frac{1}{2}} z_2$ diverges to (negative) infinity at the rate $O((nh)^{\frac{1}{2}})$.

Proof. See the Appendix.

Corollary 1.

(i) Suppose that for some constant c_h we have (undersmoothing in H_c) $h = c_h([n \ln n]^{-\frac{1}{5}}) = o(n^{-\frac{1}{5}})$. For any distribution in H_c the test statistic converges in distribution to $N(0, 1)$; for any distribution in H_s it diverges at rate $c'_h(n^{\frac{2}{5}}(\ln n)^{-\frac{1}{10}})$, where c'_h is a constant dependent on c_h .

(ii) Suppose that (undersmoothing in H_o) $h = c_h([n \ln n]^{-\frac{1}{3}}) = o(n^{-\frac{1}{3}})$. Then for a distribution in H_o we have convergence in distribution to $N(0, 1)$ and in H_s divergence at rate $c'_h(n^{\frac{1}{3}}(\ln n)^{-\frac{1}{6}})$.

(iii) If the optimal bandwidth rate is used, then $z = O_p(1)$ in H_o or in H_c , but diverges for any distribution in H_s at the corresponding rate.

2.3 The feasible statistic

To obtain a feasible statistic we replace the integrals and expectations in the expression for the statistic z involving the unknown function $\psi_{IF}(x)$ by estimators. This feasible test statistic can be defined as

$$\hat{z} = (nh)^{\frac{1}{2}} \frac{\int(\widehat{\psi_{IF}f}) - \widehat{E\psi_{IF}}}{(\hat{\sigma}_n^2)^{\frac{1}{2}}}, \tag{3}$$

where $\hat{\sigma}_n^2$ is an estimator of $var(nh)^{\frac{1}{2}} [\int(\widehat{\psi_{IF}f}) - \widehat{E\psi_{IF}}]$.

The next subsection proposes estimators for the functionals in the numerator of z , subsection 2.3.2 provides limit properties of those functionals and the expansion for the variance of the numerator, subsection 2.3.3 examines the bootstrap estimator of the variance and 2.3.4 shows that the feasible statistic has the same asymptotic properties as the infeasible, z .

2.3.1 Estimators for the unknown functionals in the numerator

Define $\hat{F}(x) = \frac{1}{n} \sum_{j=1}^n I(x_j \leq x)$ and define the estimator

$$\widehat{IF}(x) = \int_{-\infty}^x \hat{F}(y) dy = \frac{1}{n} \sum_{j=1}^n (x - x_j) I(x_j \leq x),$$

alternatively

$$\widehat{IF}_N(x) = \frac{1}{N} \sum_{i=1}^N \{I(x'_i \leq x)\} \widehat{F}(x'_i), \quad (4)$$

where x'_i denotes one of the N points in an equally spaced grid over the relevant range. The two estimators differ by $O(N^{-1})$ for every sample and there is no need to distinguish between them here. Then define

$$\widehat{\psi}_{IF}(x) = \phi(x) \widehat{IF}(x).$$

Using the kernel density estimator, the integral $\int \widehat{f}(x) \widehat{\psi}_{IF}(x) dx$ is estimated by

$$\begin{aligned} \int \widehat{(\widehat{f})\widehat{\psi}_{IF}} &= \int \widehat{f}(x) \widehat{\psi}_{IF}(x) dx \\ &= \int \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x_i - x}{h}\right) \right) \widehat{IF}(x) \phi(x) dx. \end{aligned} \quad (5)$$

This expression can be rewritten as

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{x_i - x}{h}\right) \phi(x) (x - x_i) I(x - x_i > 0) dx \\ &+ \frac{1}{n^2} \sum_{i \neq j}^n \int \frac{1}{h} K\left(\frac{x_i - x}{h}\right) \phi(x) (x - x_j) I(x - x_j > 0) dx = \end{aligned}$$

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^n \int_0^1 K(w) \phi(x_i - hw) h w dw \\ &+ \frac{1}{n^2} \sum_{i \neq j}^n \int_{(x_i - x_j)/h} K(w) \phi(x_i - hw) (x_i - x_j - hw) dw. \end{aligned} \quad (6)$$

Of course, the integral can be replaced by a sum for computation, e.g.,

$$\widehat{(\widehat{f})\widehat{\psi}_{IF}} = \frac{1}{N} \sum_{i=1}^N \widehat{\psi}_{IF}(x'_i) \widehat{f}(x'_i). \quad (7)$$

Here again the difference between the integral sum in (7) and the integral in (6) is of $O(N^{-1})$. The expectation is estimated by a sample average:

$$\widehat{E\psi}_{IF} = \frac{1}{n} \sum \widehat{\psi}_{IF}(x_i). \quad (8)$$

2.3.2 Limit properties and moments of the estimators

Assume that conditions of Theorem 1 hold.

We can establish convergence for each of the estimators. By the law of iterated logarithms for the continuous distribution function F (e.g. Kiefer 1961, Theorem 2) we have that in H_o and H_c

$$P \left(\limsup_{n \rightarrow \infty} \sup_x \frac{n^{\frac{1}{2}}}{(2^{-1} \log \log n)^{\frac{1}{2}}} \left| \hat{F}_n(x) - F(x) \right| = 1 \right) = 1.$$

Thus $\sup \left| \hat{F}(x) - F(x) \right| = O_p(n^{-\frac{1}{2} + \nu})$ in H_o for arbitrary $\nu > 0$; in H_s we have $\sup \left| \hat{F}(x) - F(x) \right| < 1$.

Then using, e.g. the expression in (4),

$$\begin{aligned} & \sup \left| \widehat{\psi}_{IF}(x) - \psi_{IF}(x) \right| \\ & \leq \sup \left| \left\{ \frac{1}{N} \sum_{i=1}^N \{I(x'_i \leq x)\} \hat{F}(x'_i) - \int_{-\infty}^x F(w) dw \right\} \phi(x) \right| \\ & \leq \sup \left| \hat{F}(x) - F(x) \right| \sup |2\phi(x)x| \\ & = \begin{cases} O_p(n^{-\frac{1}{2} + \nu}) & \text{in } H_o, \\ O_p(1) & \text{in } H_s \end{cases} ; \end{aligned} \quad (9)$$

and the same rate applies to $\widehat{\psi}_{IF}(x) - \psi_{IF}(x)$. Then for N design points

$$\begin{aligned} & \left| \frac{1}{N} \sum_{i=1}^N \widehat{\psi}_{IF}(x'_i) \hat{f}(x'_i) - \int \hat{f}(x) \psi_{IF}(x) dx \right| = \\ & \left| \frac{1}{N} \sum_{i=1}^N \left[\widehat{\psi}_{IF}(x'_i) - \psi_{IF}(x'_i) \right] \hat{f}(x'_i) + \frac{1}{N} \sum_{i=1}^N \psi_{IF}(x'_i) \hat{f}(x'_i) - \int \hat{f}(x) \psi_{IF}(x) dx \right| \\ & = \begin{cases} O_p(n^{-\frac{1}{2} + \nu}) & \text{in } H_o, \\ O_p(1) & \text{in } H_s \end{cases} \end{aligned} \quad (10)$$

as long as N is large enough, $Nn^{-1} \rightarrow \infty$.

We now obtain the variance of the numerator of \hat{z} in (3).

Lemma 2. *Under the conditions of Theorem 1,*

$$\text{var} \left(\int (\widehat{\psi}_{IF} \hat{f}) - E \widehat{\psi}_{IF} \right) = \frac{1}{nh} \int K(w)^2 dw E(\psi_{IF}^2) + \frac{1}{n} A + o(n^{-1}), \quad (11)$$

where the expression for A is provided in the proof of the Lemma.

Proof. See the Appendix.

2.3.3 The bootstrap estimator for the variance of the numerator

Consider next the bootstrap variance estimator, $v\hat{a}r_B = \hat{\sigma}_{n,B}^2$. We show that this estimator gives a good approximation to $\text{var}(nh)^{\frac{1}{2}} [\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}}]$.

Lemma 3. *Under the conditions of Theorem 1*

$$\left| v\hat{a}r_B \left\{ (nh)^{\frac{1}{2}} \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right) \right\} - \text{var}(nh)^{\frac{1}{2}} \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right) \right| = O_p((nh)^{-1});$$

and if additionally $h^3n \rightarrow \infty$, then

$$v\hat{a}r_B \left\{ (nh)^{\frac{1}{2}} \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right) \right\} = \int K(w)^2 dw E(\psi_{IF}^2) + hA + o_p(h).$$

Proof. See the Appendix.

Thus the bootstrap estimator of variance is consistent with the rate $O_p((nh)^{-1})$, and preserves the variance expansion in Lemma 2 if a bandwidth satisfying $h^3n \rightarrow \infty$ is used for the bootstrap.

2.3.4 The limit properties of the feasible statistic

Theorem 2. *Under the conditions of Theorem 1 the asymptotic properties of \hat{z} with $\hat{\sigma}_{n,B}^2 = v\hat{a}r_B \left\{ (nh)^{\frac{1}{2}} \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right) \right\}$ are the same as those established for z in Theorem 1 and*

$$\hat{z} = \begin{cases} z + O_p(n^\nu h^{\frac{1}{2}}) & \text{in } H_o \text{ for } \nu > 0; \\ z + O_p(1) & \text{in } H_s. \end{cases}$$

Proof. See the Appendix.

Corollary 1 above, concerning behavior of the statistic for different classes of distribution, then applies to the feasible statistic as well.

3 The size and power of the test

In this section the size and power of the proposed test based on \hat{z} are examined; asymptotic properties are derived in section 3.1 and the finite sample performance is reported in 3.2.

3.1 Asymptotic properties of the test

Denote the size and power on a sample of size n at the quantile α by $\beta_n(\alpha) = \Pr(|\hat{z}| > \Phi^{-1}(\alpha))$, and asymptotic size/power by $\beta(\alpha) = \lim \beta_n(\alpha)$.

Theorem 2 and the properties of the infeasible statistic lead to the following conclusions regarding asymptotic power and size.

Corollary 2. Under the conditions of Theorem 2,

$$\beta(\alpha) = \begin{cases} \alpha & \text{for } h = o(n^{-\frac{1}{5}}) \text{ if } F \in H_c; \\ \alpha & \text{for } h = o(n^{-\frac{1}{3}}) \text{ if } F \in H_o; \\ 1 & \text{for } h = o(n^{-\frac{1}{5}}) \text{ if } F \in H_s. \end{cases}$$

This corollary provides appropriate asymptotic size in H_o for the test and shows that the power goes to 1 for any distribution from H_s .

It is known that a fully nonparametric test that has nontrivial power uniformly is not possible. However, by restricting testing to a subclass in $H_c(H_o)$ vs a subclass in H_s it is possible to control uniformly the power of the test for a given size. FH provide subclasses of distribution in H_o and H_s for their test and show that β_n converges uniformly to 0 over the subclass of H_o and to 1 over the subset of H_s . The problem for their classes is that different statistics are required for different subclasses in H_o . For example, the statistic $\int \hat{f}^2$ is not suitable for a class that includes χ_1^2 since the functional $\int f^2$ diverges; one could use $\int \hat{f}^{(1+\frac{1}{2})}$. One therefore requires *a priori* knowledge of the class containing the null in order to select the statistic.

Define subclasses $H_o(C, \varepsilon) = \{F \in H_o : \|f - f_v\|_{L_1} < \varepsilon \text{ for } f_v \text{ such that } |f_v| < C\}$; for any ε , any distribution with a density will eventually fall into some such class; essentially any density can be arbitrarily close to some bounded density. These subclasses do not contain any singular distributions and are (partially) ordered as $H_o(C_1, \varepsilon_1) \sqsubseteq H_o(C_2, \varepsilon_2)$ for $C_1 \leq C_2$ and $\varepsilon_1 \leq \varepsilon_2$; the intersection of these classes for a fixed C is the class of all distributions with density bounded by C . The subclasses for singular distributions, $H_s^{\delta, \zeta}$ are defined in FH (denoted there $F_s^{\delta, \alpha}$); they control the “importance” of the singular part of the measure. For a set B , denote by

B^γ the set $\{x : \inf_{y \in B} \|x - y\| \leq \gamma\}$, and by μ_F the measure associated with the distribution F , by λ the Lebesgue measure. Then

$$H_s^{\delta, \zeta} = \{F; \text{ such that for some } B \text{ with } \mu_F(B) \geq \delta; \lambda(B^\gamma) \leq \zeta \text{ for some } \gamma\}.$$

Note that any $H_s^{\delta, \zeta}$ contains absolutely continuous as well as singular distributions. A partial ordering is $H_s^{\delta_1, \zeta_1} \subseteq H_s^{\delta_2, \zeta_2}$ for $\delta_1 \geq \delta_2$ and $\zeta_1 \leq \zeta_2$. A singular distribution assigns a positive measure to some set of zero Lebesgue measure so that as $\gamma \rightarrow 0$, $\zeta \rightarrow 0$ if $F \in H_s$ and B belongs to the support of the singularity set. The parameters C, ε and δ, ζ can be selected so that the classes $H_o(C, \varepsilon)$ and $H_s^{\delta, \zeta}$ do not intersect. These properties are proved in the following lemma.

Lemma 4. Under the conditions of Theorem 1

(a) For any distribution $F \in H_o$ and any ε there exists C such that $F \in H_o(C, \varepsilon)$.

(b) For any distribution $F \in H_s$ and any ζ there exist B, δ (and γ) such that $F \in H_s^{\delta, \zeta}$.

(c) For any C, ε there exists ζ such that $H_o(C, \varepsilon) \cap H_s^{\delta, \zeta} = \emptyset$.

Proof. See the Appendix.

Denote by \bar{H} the class of distributions over which, for a given ϕ , Assumption 1(b) is satisfied uniformly. This excludes, for example, distributions with support that is outside an arbitrary bounded set.

Lemma 5. Suppose that $F \in \bar{H}$. Then for a given kernel K and for h that satisfy conditions of Theorem 1, as $n \rightarrow \infty$,

$$(a) \sup_{F \in \bar{H}} \left| \Pr \left[(nh)^{\frac{1}{2}} z_1 > \Phi^{-1}(\alpha) \right] - \alpha \right| \rightarrow 0;$$

$$(b) \lim_{\varepsilon \rightarrow \infty} \sup_{F \in \bar{H}} \Pr [|\hat{z} - z| > \varepsilon] = 0;$$

$$(c) \sup_{F \in \bar{H} \cap H_0} \Pr [|\hat{z} - z| > \varepsilon] \rightarrow 0.$$

Proof. See the Appendix.

For any of the subclasses defined here denote the intersection with \bar{H} by overbar on the notation for the class: $\bar{H}^{\cdot\cdot} = H^{\cdot\cdot} \cap \bar{H}$. The next theorem establishes uniform consistency and limit unbiasedness of the test.

Theorem 3. Under the assumptions of Theorem 1

$$(a) \text{ Power: } \inf_{F \in \bar{H}_s^{\delta, \zeta}} \beta_n(\alpha) \geq \bar{\beta}_n(\alpha) \nearrow 1;$$

$$(b) \text{ Size: } \lim_{h = o(n^{-\frac{1}{5}})} \sup_{F \in \bar{H}_c(C, \varepsilon)} \beta_n(\alpha) \leq \alpha$$

3.2 Finite sample evaluation

The simulations that we now report all use the bootstrap variance estimator $\hat{v}ar_B$, which other experiments (not recorded here) showed to provide better finite-sample conformity with the asymptotic distribution than was available from the use of the asymptotic variance.

We consider three null and three corresponding alternative cases.

In the first, the data are generated from the $U[0,1]$ under the null (case A) and under the alternative (case B) are generated from the singular distribution considered by FH (the discrete uniform distribution on 100 equally spaced discrete points in $[0,1]$). The second pairing consists of data generated under the null by a χ_2^2 (case C) and under the alternative (case D) by a 70% mixture of the a χ_2^2 and 30% of a point distribution on the discrete values $\{4, 5, 6, 7\}$. The third pairing uses a mixture of normal distributions under the null, with mean and variance parameters $(-0.767,1)$, $(0.33,100)$ and $(4.33,100)$ in proportions $\frac{1}{11}$, $\frac{6}{11}$ and $\frac{4}{11}$ (case E). Note that the large difference in variances makes this a potentially difficult case, in which the low-variance component of the mixture begins to take on some of the character of a singularity. The alternative in this case is based on the same continuous mixture, with again a 30% mixture of a point distribution on the discrete values $\{1, 2, 3, 4\}$.

Bandwidth h is selected by a simple rule in all cases:

$$h = (5 \times 10^{-4}) \left(1 + \frac{2000}{n} \right) n^{-\frac{1}{5}}.$$

Bandwidth selection of course plays an important role in tests of this type; this simple rule converges to the asymptotic rate of $n^{-\frac{1}{5}}$ but allows for faster decline at smaller sample sizes.

In the uniform-null case A, conformity of the estimated density of the statistic with the asymptotic normal distribution is very good at all sample sizes. The corresponding case B alternative, used by FH, shows distributions concentrated at large negative values; unsurprisingly, power is very high against this alternative, which embodies 100 discrete steps.

Note that the alternative here can be thought of as a uniform distribution on $[0,1]$, reported to two digits of measurement precision; the case therefore illustrates the importance of precision. In applications, finite precision of measurement implies that our empirical observations never literally come from a continuous density. In typical empirical samples from continuous densities, precision is sufficiently high that repeated observations will not occur, and the distance between observable values is negligible. This

need not always be the case, however, as in this alternative where only 100 values are possible, and so with the sample sizes used here, many repeated observations will occur. The test statistic will therefore reject either because the true distribution is non-smooth, because measurement precision is inadequate, or both: we implicitly test a joint hypothesis of a continuous density and arbitrarily precise measurement. Since measurement precision is known, it should in practice be straightforward to evaluate whether inadequate precision contributes non-negligibly to any observed rejection.

Case C also shows fairly good conformity with the asymptotic normal, although rare large values do occur. In comparing with Case D, which embodies a mixture with masses on discrete points, note the difference in scales: under the alternative much less probability mass is concentrated near zero, and much more in regions that would correspond with rejection at conventional levels. There is substantial power, but we do not observe in these examples much increase with sample size. The additional probability mass concentrated at a few points is much more difficult to detect than the discrete alternative of Case B.

Case E is a relatively challenging one of a normal mixture with differing variances, but the null distributions again appear to be converging on the asymptotic standard normal, albeit relatively slowly and with lower than unit variance, indicating an under-sized test. The corresponding alternative distributions again show relatively little change with sample size, and only moderate power.

4 Concluding remarks

The test of singularity developed here has a number of desirable properties. It is appropriate for any pre-specified domain, not necessarily a single point; by testing in various intervals the support of the singular (or non-smooth) part of the distribution can be narrowed down. Feasible test statistics can be computed straightforwardly, and have a pivotal asymptotic distribution so that bootstrap testing can be implemented. Finite-sample conformity with the asymptotic normal distribution is reasonable, and there is little evidence of over-rejection. The divergence rate under the alternative is nonparametric, but is known.

The test can be extended to handle multivariate densities. If the distribution function $F(x_1, \dots, x_k)$ is absolutely continuous then the density can be

defined as an ordinary function

$$f(x) = f(x_1, \dots, x_k) = \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k},$$

where $f(x)$ integrates to $F(x)$. However, whether $f(x)$ exists as an ordinary function or not, it can be defined as a generalized function, and the methods used above can be extended to test singularity jointly and in specified directions.

5 Appendix

Proof of Lemma 1.

The proof uses the derivations in the proof of Theorem 3 in VZW; here as in that paper integrals involving generalized functions $f(x)$, $f'(x)$ (with some abuse of notation since they may not be defined pointwise) are understood as the appropriate functionals applied to differentiable functions. Then using the substitutions $w = \frac{x_i - x}{h}$ and $y = x + hw$ leads to

$$\begin{aligned}
& \text{var} \left[(nh)^{\frac{1}{2}} \int (\widehat{f}(x) - f(x)) \psi_{IF}(x) dx \right] \\
&= \text{var} \left[(nh)^{\frac{1}{2}} \int (\widehat{f}(x) - E(\widehat{f})) \psi_{IF}(x) dx \right] \\
&= \int \int K(w)^2 f(x + hw) dw (\psi_{IF}(x))^2 dx \\
&\quad - h \left(\int \int K(w) f(x + hw) dw \psi_{IF}(x) dx \right)^2 \\
&= \int \int K(w)^2 f(y) (\psi_{IF}(y - hw))^2 dw dy \\
&\quad - h \left(\int \int K(w) f(y) \psi_{IF}(y - hw) dw dy \right)
\end{aligned}$$

Expanding in H_o, H_c where the function ψ_{IF} is continuously differentiable provides

$$\int K(w)^2 dw E \psi_{IF}^2 + h \int K(w)^2 dw \int f(x) (\psi_{IF}(x)^2)' dx - h (E \psi_{IF}(x))^2 + o(h) \tag{12}$$

Applying the Lipschitz condition for H_s concludes. ■

Proof of Theorem 1.

(i) Similarly to the proof of Theorem 3 (VZW) for $\eta_{hn}(x)$ defined there, but with the function $\psi_{IF}(x)$ substituted as a test function, the random variable

$$(\eta_{hn}(x), \psi_{IF}) = (nh)^{\frac{1}{2}} \int (\widehat{f}(x) - E(\widehat{f})) \psi_{IF}(x) dx;$$

converges to the limit $N(0, \sigma_n^2)$ with σ_n^2 derived in Lemma 1.

(ii) The bias of the estimator as a generalized function was provided in VZW (Theorem 2) for $\psi \in D = D_{l+1}$ and a kernel of order $l + 1$, as

$$\begin{aligned} & \left(E(\hat{f}), \psi \right) - (f(x), \psi) \\ &= (-1)^l h^l \int F(\tilde{x}) \frac{1}{l!} \frac{\partial^{l+1} \psi}{\partial x^{l+1}}(\tilde{x}) d\tilde{x} \int K(w) w^l dw + R(h) \end{aligned}$$

with $R(h) = o(h^l)$ and is thus $O(h^l)$. When density exists integration by parts provides

$$(-1)^{l-1} h^l \int f(\tilde{x}) \frac{1}{l!} \frac{\partial^l \psi}{\partial x^l}(\tilde{x}) d\tilde{x} \int K(w) w^l dw + R(h)$$

and so gives the same rate on the space D_l rather than D_{l+1} . When density exists ψ_{IF} is twice differentiable, once continuously, but unless H_c holds $\psi_{IF} \notin D_2$. So for H_o , $\psi_{IF} \in D_1$ and the rate $O(h)$ holds for the bias, thus $(nh)^{\frac{1}{2}} O(h) = O(n^{\frac{1}{2}} h^{\frac{3}{2}})$. Similarly for H_c , the rate $O(h^2)$ for bias provides $O(nh^5)^{\frac{1}{2}}$ for z_2 .

(iii) Consider $\left(E(\hat{f}), \psi_{IF} \right) - (f(x), \psi_{IF})$. Denote by μ_F the measure corresponding to F . If F is singular (by e.g. Lemma 2.1 of FH) there is a set $B \subset R$ such that $\mu_F(B) = \mu > 0$, while for $\delta \rightarrow 0$ the Lebesgue measure of a δ -neighbourhood of B , $\lambda(B^\delta) \rightarrow 0$. Then $\left[\left(E(\hat{f}), \psi_{IF} \right) - (f(x), \psi_{IF}) \right] =$

$$\begin{aligned} & \int E(\hat{f}) \psi_{IF}(x) dx - \int f(x) \psi_{IF}(x) dx \\ & \leq \sup_{x \in B} \left\{ E(\hat{f}) \psi_{IF}(x) \right\} \lambda(B^\gamma) - E \psi_{IF}(x). \end{aligned}$$

Note that $\sup |E(\hat{f})| \leq \sup |K(w)|$; $\sup \psi_{IF}(x) \leq \sup |x \phi(x)|$ and denote the bound on f_{hK} by $C_{f_{hK}}$ and the bound on ψ_{IF} by $C_{\psi_{IF}}$. Then for any $\varepsilon > 0$ there is $\gamma(\varepsilon)$ such that for any $\gamma < \gamma(\varepsilon)$,

$$\sup_{x \in \bar{B}(\varepsilon)} \left\{ E(\hat{f}) \psi_{IF}(x) \right\} \lambda(B^\gamma) \leq C_{f_{hK}} C_{\psi_{IF}} \varepsilon.$$

By Assumption 1(b), $E \psi_{IF}(x) > A > 0$. Then for $\varepsilon < \frac{A}{2C_{f_{hK}} C_{\psi_{IF}}}$,

$$\left[\left(E(\hat{f}), \psi_{IF} \right) - (f(x), \psi_{IF}) \right] \leq -\frac{A}{2}. \quad (13)$$

From Lemma 1, Assumption 1(b) and (13)

$$(nh)^{\frac{1}{2}} z_2 \leq - (nh)^{\frac{1}{2}} \frac{A}{2\chi \int K(w)^2 dw} (1 + o(h))$$

and the result follows. ■

Proof of Lemma 2.

To find the variance, $E \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right)^2 - \left[E \left(\widehat{\int(\psi_{IF}\hat{f})} - \widehat{E\psi_{IF}} \right) \right]^2$ consider

$$\begin{aligned} E \left(\widehat{\int(\psi_{IF}\hat{f})} \right)^2 = & \\ E \left(\int \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left(\frac{x_i - x}{h} \right) \right) \frac{1}{n} \sum_{j=1}^n (x - x_j) I(x_j \leq x) \phi(x) dx \right) & (14) \\ \cdot \left(\int \left(\frac{1}{n} \sum_{l=1}^n \frac{1}{h} K \left(\frac{x_l - y}{h} \right) \right) \frac{1}{n} \sum_{k=1}^n (y - x_k) I(x_k \leq y) \phi(y) dy \right); & \end{aligned}$$

$$\begin{aligned} E \left(\widehat{\int(\psi_{IF}\hat{f})} \times \widehat{E\psi_{IF}} \right) = & \\ E \left(\int \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left(\frac{x_i - x}{h} \right) \right) \frac{1}{n} \sum_{j=1}^n (x - x_j) I(x_j \leq x) \phi(x) dx \right) & (15) \\ \cdot \left(\frac{1}{n^2} \sum_{l \neq k}^n (x_l - x_k) I(x_k \leq x_l) \phi(x_l) \right); & \end{aligned}$$

$$\begin{aligned} E \left(\widehat{E\psi_{IF}} \right)^2 = & \\ E \left[\left(\frac{1}{n^2} \sum_{i \neq j}^n (x_i - x_j) I(x_j \leq x_i) \phi(x_i) \right) \left(\frac{1}{n^2} \sum_{l \neq k}^n (x_l - x_k) I(x_k \leq x_l) \phi(x_l) \right) \right]. & (16) \end{aligned}$$

Let $\alpha_1(x_i, x) = \frac{1}{h} K \left(\frac{x_i - x}{h} \right)$; $\alpha_2(x_i, x) = (x - x_j) I(x_j \leq x)$ and $\alpha_3(x_i, x_j) = (x_i - x_j) I(x_j \leq x_i) \phi(x_i)$. Then

$$\begin{aligned} E \left(\widehat{\int(\psi_{IF}\hat{f})} \right)^2 &= \frac{1}{n^4} \Sigma E \left(\int \int \alpha_1(x_i, x) \alpha_2(x_j, x) \alpha_1(x_l, y) \alpha_2(x_k, y) \phi(x) \phi(y) dx dy \right); \\ E \left(\widehat{\int(\psi_{IF}\hat{f})} \times \widehat{E\psi_{IF}} \right) &= \frac{1}{n^4} \Sigma E \left(\int \alpha_1(x_i, x) \alpha_2(x_j, x) \alpha_3(x_l, x_k) \phi(x) dx \right); \\ E \left(\widehat{E\psi_{IF}} \right)^2 &= \frac{1}{n^4} \Sigma E (\alpha_3(x_i, x_j) \alpha_3(x_l, x_k)). \end{aligned} \quad (17)$$

Note that

$$\begin{aligned}
& \left(E \widehat{\int (\psi_{IF} f)} \right)^2 = \\
& \frac{1}{n^4} \Sigma E \left(\int \int \alpha_1(x_i, x) \alpha_2(x_j, x) \alpha_1(x_l, y) \alpha_2(x_k, y) I(\{i, j\} \cap \{k, l\} = \emptyset) \phi(x) \phi(y) dx dy \right); \\
& E \left(\widehat{\int (\psi_{IF} f)} \right) \times E \left(\widehat{E \psi_{IF}} \right) = \\
& \frac{1}{n^4} \Sigma E \left(\int \alpha_1(x_i, x) \alpha_2(x_j, x) \alpha_3(x_l, x_k) I(\{i, j\} \cap \{k, l\} = \emptyset) \phi(x) dx \right); \\
& \left(E \widehat{E \psi_{IF}} \right)^2 = \frac{1}{n^4} \Sigma E \left(\alpha_3(x_i, x_j) \alpha_3(x_l, x_k) I(\{i, j\} \cap \{k, l\} = \emptyset) \right).
\end{aligned}$$

Thus the terms in the sums in (14-16) that would contribute to the variance are those where the sets $\{i, j\}$ and $\{k, l\}$ have exactly one or two elements in common. The sums involving subsets with any two common elements among $\{i, j, l, k\}$ will contribute at most $O(\frac{1}{n^2 h})$ to the variance: this follows from the fact that the number of such terms is proportionate to n^2 up to $O(n)$ and the expectation of each such term is at most $O(h^{-1})$, as will follow from the computations of moments below.

Consider now the cases when $\{i, j\}$ and $\{k, l\}$ have exactly one element in common and $i \neq j, l \neq k$; there are $4n(n-1)(n-2)$ terms of this sort that fit into four cases:

- (a) $i = l$; (b) $i = k$; (c) $j = l$ and (d) $j = k$.

Note that $\int (y-z) I(y-z > 0) f(z) dz = IF(y)$ and grows as $|y|$ at infinity (as $y \rightarrow +\infty$). Define the following functions given by integrals and note the order of growth at infinity:

$$\begin{aligned}
I_1(y) &= \int (y-z) I(y-z > 0) f(z) \phi(z) dz; \text{ bounded;} \\
I_2(y) &= \int (z-y) I(z-y > 0) f(z) \phi(z) dz; \text{ bounded;} \\
I_3(y) &= \int (y-z) I(y-z > 0) I_2(z) f(z) dz; \text{ grows as } |y|; \\
I_4(y) &= \int (y-z) I(y-z > 0) IF(z) f(z) \phi(z) dz; \text{ bounded.}
\end{aligned}$$

The evaluation of growth uses Assumption 1.

The contribution to the variance resulting from case (a) is the term $E \left(\widehat{\int (\psi_{IF} f)} \right)^2$ for

$i = l \neq j \neq k$, noting that $\int (y - x_k)I(x_k \leq y)f(x_k)dx_k = IF(y)$ and

$$\begin{aligned} & \int \int \frac{1}{nh} \int \frac{1}{h} K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_i - y}{h}\right) f(x_i) dx_i \phi(x) IF(x) \phi(y) IF(y) dx dy \\ &= \frac{1}{nh} \int K(w)^2 dw E \psi_{IF}^2 + \frac{1}{n} \int K(w)^2 dw E (\psi_{IF}(x)^2)' dx - \frac{1}{n} (E \psi_{IF}(x))^2 + o\left(\frac{1}{n}\right); \end{aligned}$$

similarly to the derivation in Lemma 1. From the term $-2E\left(\widehat{\int(\psi_{IF} f)} \times \widehat{E\psi_{IF}}\right)$,

$$\begin{aligned} & -2 \int \int \frac{1}{n} \int \frac{1}{h} K\left(\frac{x_i - x}{h}\right) (x - x_j) I(x_j < x) \phi(x) dx (x_i - x_k) I(x_k \leq x_i) \phi(x_i) f(x_i) f(x_j) f(x_k) dx_i dx_j dx_k \\ &= -2 \frac{1}{n} \int \int \left(\int K(w) (x_i - x_j - hw) I(x_i - x_j - hw > 0) \phi(x_i - hw) dw \right) IF(x_i) \phi(x_i) f(x_i) f(x_j) dx \\ &= -2 \frac{1}{n} E(\psi_{IF}^2) + O\left(\frac{h}{n}\right); \end{aligned}$$

from the term $E(\widehat{E\psi_{IF}})^2$ the contribution is, by a similar integration, $\frac{1}{n} E(\psi_{IF}^2)$.

For (b) $i = k$ from the first part the contribution is the same as for case (c): $j = l$ and is

$$\begin{aligned} & \int \int \left[\int K(w) f(x + hw) dw \right] \frac{1}{n} \int (x - x_j) I(x_j < x) \\ & \quad \frac{1}{h} K\left(\frac{x_j - y}{h}\right) f(x_j) dx_j \int (y - x_k) I(x_k < y) f(x_k) dx_k \phi(x) \phi(y) dx dy = \\ & \frac{1}{n} \int \int \phi(x - hw) K(w) dw f(x) \left(\int (x - y - hv) I(y + hv < x) K(v) f(y + hv) dv IF(y) \phi(x) \phi(y) dx dy \right) \\ & \quad \frac{1}{n} \int \int f(x) \phi(x) (x - z) I(z < x) \phi(x) dx \int K(v) f(z) \phi(z - hv) IF(z - hv) \\ & \quad \frac{1}{n} \int I_2(z) \phi(z) IF(z) f(z) dz + O\left(\frac{h}{n}\right) = \frac{1}{n} E(I_2 \psi_{IF}) + O\left(\frac{h}{n}\right). \end{aligned}$$

From the second term the contribution to variance is

$$-\frac{2}{n} E(I_4 \psi_{IF}) + O\left(\frac{h}{n}\right);$$

and from the third the contribution is

$$\frac{1}{n} E(\psi_{IF} I_4).$$

For (c) the contributions to variance from the first through third terms are $\frac{1}{n}E(I_2\psi_{IF}) + O(\frac{h}{n})$, $-\frac{2}{n}E(I_2\psi_{IF}) + O(\frac{h}{n})$, and $\frac{1}{n}E(\psi_{IF}I_2)$.

Finally for (d), $j = k$, from the first term $\frac{1}{n}E(I_2^2) + O(\frac{h}{n})$; from the second $-\frac{2}{n}E(I_3\phi)$; from the third $\frac{1}{n}E(I_2^2)$.

Summing up, we find that the variance equals

$$\begin{aligned}
& \frac{1}{nh} \int K(w)^2 dw E\psi_{IF}^2 + \frac{1}{n} \int K(w)^2 dw E(\psi_{IF}(x)^2)' dx - \frac{1}{n} (E\psi_{IF}(x))^2 \\
& - 2\frac{1}{n}E(\psi_{IF}^2) + \frac{1}{n}E(\psi_{IF}^2) \\
& + \frac{1}{n}E(I_2\psi_{IF}) - \frac{2}{n}E(I_4\psi_{IF}) + \frac{1}{n}E(\psi_{IF}I_4) \\
& + \frac{1}{n}E(I_2\psi_{IF}) - \frac{2}{n}E(I_2\psi_{IF}) + \frac{1}{n}E(\psi_{IF}I_2) \\
& + \frac{1}{n}E(I_2^2) - \frac{2}{n}E(I_3\phi) + \frac{1}{n}E(I_2^2) + o(\frac{1}{n}) \\
= & \frac{1}{nh} \int K(w)^2 dw E\psi_{IF}^2 + \frac{1}{n}A + o(\frac{1}{n}). \quad \blacksquare
\end{aligned}$$

Proof of Lemma 3.

For a sufficiently large number of bootstraps the bootstrap estimator of a moment replaces the expectation operator E by expectation E_n with respect to the empirical distribution, F_n . In the computation of the terms in E_n the density f (here treated as a generalized function) is replaced by the generalized density for the empirical distribution which is the average of δ -functions: $f_n(x) = \frac{1}{n}\sum\delta(x - x_i)$, so that for any continuous function g one has $E_n g = \int f_n(x)g(x)dx = \frac{1}{n}\sum g(x_i)$. For pairwise distinct indices $f_n(x_i, x_j)I(x_i \neq x_j) = \frac{1}{n(n-1)}\sum_{l \neq k}\delta(x_i - x_l)\delta(x_j - x_k)$, etc.

Consider the terms in the sums (17); any such term, $\omega_n(i, j, l, k)$ is given by an integral involving some products of the α 's; consider these terms for the cases (a)-(d) in Lemma 2. As was shown in the proof of that lemma the contribution from the sum of such terms, $E\frac{1}{n^4}\sum_{case(\cdot)}\omega_n(i, j, l, k)$ is either $O(\frac{1}{nh})$, or $O(\frac{1}{n})$; denote $E\frac{1}{n^4}\sum_{case(\cdot)}\omega_n(i, j, l, k)$ by $E(\omega_{case(\cdot)})$. From the Chebyshev inequality, denoting $E(nh\omega_n)$ by ω ,

$$\Pr(|E_n(\omega_{case(\cdot)}) - E(\omega_{case(\cdot)})| > \varepsilon) \leq \frac{\text{var}(E_n(\omega_{case(\cdot)}))}{\varepsilon^2}$$

and

$$\Pr(|nhE_n(w) - nhE(w)| > \varepsilon) \leq \frac{\text{var}(nhE_n(w))}{\varepsilon^2} = \frac{(nh)^2 \text{var}E_n(w)}{\varepsilon^2}.$$

Similarly to the argument in the proof of Lemma 2 the terms that will determine the variance $\text{var}(E_n(\omega_{\text{case}(\cdot)}))$ now involve two sets of four indices, (i_1, j_1, l_1, k_1) and (i_2, j_2, l_2, k_2) . Considerations of the number of terms with different numbers of coinciding indices provides that the terms that will determine the variance have exactly two indices in common between the two sets, in addition to the fact that in each set of indices there are already exactly two indices in common between $\{i., j.\}$ and $\{l., k.\}$. There are thus $O(n^5)$ terms of that type. The largest expectation comes from the terms involving products of four factors of the type $\alpha_1(x_m, x) = \frac{1}{h}K(\frac{x_m-x}{h})$ where all the indices are the same: $i_1 = l_1 = i_2 = l_2$. The expectation then will be $O(\frac{1}{h^3})$. Thus $\text{var}(E_n(\omega_{\text{case}(\cdot)})) = O(n^5 \frac{1}{n^3 h^3}) = O(n^{-3} h^{-3})$. Multiplying $E_n(\omega_{\text{case}(\cdot)})$ by nh provides the rate of consistency for the bootstrap variance estimator as $O((nh)^{-1})$.

To preserve the smallest order term of $O(n^{-1})$ in the expansion of the variance in Lemma 2 the largest order of term in $\text{var}(E_n(\omega_{\text{case}(\cdot)}))$ needs to be no larger than $o(n^{-2})$. If h is selected to be such that $nh^3 \rightarrow \infty$ implying $n^{-1}h^{-3} \rightarrow 0$ we have the result. ■

Proof of Theorem 2.

We show that in H_o the difference $\hat{z} - z$ is $o_p(1)$ and that in H_s the feasible statistic \hat{z} diverges at the same rate as the infeasible statistic z , $O_p((nh)^{\frac{1}{2}})$. Consider the expression in the numerator of the statistic \hat{z} . From (9,10) it follows that in H_o the numerators of z (denoted $\text{num}(z)$) and of \hat{z} ($\text{num}(\hat{z})$) differ by $\hat{r} = O_p(n^{-\frac{1}{2}+\nu} \cdot n^{\frac{1}{2}}h^{\frac{1}{2}}) = O_p(n^\nu h^{\frac{1}{2}}) = o_p(1)$ since ν is arbitrarily small; in H_s the difference is $O_p(1)$. By Lemma 3 the denominator of \hat{z} differs from the positive (bounded from zero and infinity) quantity $d = \int K(w)^2 dw E\psi_{IF}^2$ by $O_p(h)$ both in H_o and H_s . Then

$$\begin{aligned} \hat{z} &= (\text{num}(z) + \hat{r}) \left[d \left(1 + \frac{O_p((nh)^{-1})}{d} \right) \right]^{-1} \\ &= \begin{cases} (d^{-1})\text{num}(z)(1 + O_p(n^\nu h^{\frac{1}{2}})) & \text{in } H_o \\ (d^{-1})\text{num}(z)(1 + O_p(1)) & \text{in } H_s. \end{cases} \end{aligned}$$

The theorem follows. ■

Proof of Lemma 4.

(a) For $F \in H_o$ the density function, f , is in L_1 ; consider for any constant C the set

$$B(f, C) = \{x : f(x) \geq C\};$$

then $\int_{B(f,C)} f(x)dx \geq C\lambda(B(f,C))$. If $C \nearrow \infty$, then $\lambda(B(f,C)) \searrow 0$ and since f is absolutely integrable $\int_{B(f,C)} f(x)dx \searrow 0$; thus $C\lambda(B(f,C)) \searrow 0$. Then select a large enough C such that $\int_{B(f,C)} f(x)dx < \frac{\varepsilon}{4}$. Consider the function $\sup\{f(x), C\}$; this is in L_1 and can be approximated by a step-function: for $\frac{\varepsilon}{2}$ there exists a step function f_{step} such that $|f_{step}| \leq C$ and $\|\sup\{f(x), C\} - f_{step}\|_{L_1} < \frac{\varepsilon}{2}$. Then

$$\begin{aligned} \|f(x) - f_{step}\|_{L_1} &\leq \int_{B(f,C)} f(x)dx + C\lambda(B(f,C)) + \|\sup\{f(x), C\} - f_{step}\|_{L_1} \\ &\leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

(b) This follows from Lemma 2.1 in FH.

(c) Suppose that $F \in H_o(C, \varepsilon)$; then for any set B with $\lambda(B^\gamma) < \zeta$ we have $\mu_F(B) \leq \varepsilon + \zeta C$. If δ, ζ are such that $\delta > \varepsilon + \zeta C$, then $F \notin H_s^{\delta, \zeta}$. ■

Proof of Lemma 5.

(a) Since for $A_i \equiv \frac{1}{h} \int K(\frac{x-x_i}{h})\psi_{IF}(x)dx$,

$$z_1 = \frac{\frac{1}{n}\Sigma(A_i - EA_i)}{var A_i},$$

the conditions of the Lindeberg-Feller central limit theorem are fulfilled via moments that are uniformly bounded on \bar{H} , by the boundedness of $K(w)$ and Assumption 1.

(b) Write

$$\begin{aligned} |\hat{z} - z| &= \left| \frac{num(\hat{z})}{den(\hat{z})} - \frac{num(z)}{den(z)} \right| \\ &= \left| \frac{num(\hat{z}) - num(z) + num(\hat{z}) \frac{den(z)-d}{d} - num(z) \frac{den(\hat{z})-d}{d}}{d \left(1 + \frac{den(\hat{z})-d}{d}\right) \left(1 + \frac{den(z)-d}{d}\right)} \right|. \end{aligned}$$

This expression is bounded in probability uniformly, since we can apply uniform bounds on the moments for $|num(\hat{z}) - num(z)|$ and use Assumption 1 (b) for the bound on d ; for $F \in \bar{H}$ note that $|den(z) - d|$ goes to zero in probability uniformly.

(c) In H_o the quantity $|num(\hat{z}) - num(z)|$ goes to zero by Theorem 2 and since the moments are bounded uniformly the convergence is uniform. ■

Proof of Theorem 3.

(a) Write the statistic as $\hat{z} = (nh)^{\frac{1}{2}}z_2 + (nh)^{\frac{1}{2}}z_1 + (z - \hat{z})$. Consider $\tilde{c}_n > \Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu}$ (with arbitrarily small ν). By Lemma 5(a,b),

$$\sup_{F \in \bar{H}} \Pr \left(\left| (nh)^{\frac{1}{2}}z_1 + (z - \hat{z}) \right| > \Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu} \right) \rightarrow 0.$$

Then by the uniform condition (13) in the proof of Theorem 1 (iii),

$$\begin{aligned} \inf_{H_s^{\delta, \zeta}} \Pr(|\hat{z}| > \Phi^{-1}(\alpha)) & \\ & \geq \Pr \left(\left| (nh)^{\frac{1}{2}}z_2 \right| > 2\Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu} \cap \left| (nh)^{\frac{1}{2}}z_1 + (z - \hat{z}) \right| < \Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu} \right) \\ & > \Pr \left(\left| (nh)^{\frac{1}{2}}z_2 \right| > 2\Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu} \right) - \Pr \left(\left| (nh)^{\frac{1}{2}}z_1 + (z - \hat{z}) \right| > \Phi^{-1}(\alpha) (nh)^{\frac{1}{2}-\nu} \right) \\ & \rightarrow 1. \end{aligned}$$

Therefore $\inf_{\bar{H}_s^{\delta, \zeta}} \beta_n \rightarrow 1$.

(b) The result follows from (a) and (c) of Lemma 5 and (ii) of Theorem 1.. ■

References

- [1] Bahadur, R.R. and Savage, L.J. (1956) The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics* 27, 1115-1122.
- [2] Donoho, D.L. (1988) One-sided inference about functionals of a density. *Annals of Statistics* 16, 1390-1420.
- [3] Frigyesi, A. and O. Hössjer (1998) A Test for Singularity. *Statistics and Probability Letters* 40, 215-226.
- [4] Härdle, W. , G. Kerkycharian, D. Picard and A. Tsybakov (1998) *Wavelets, Approximations and Statistical Applications*. Springer-Verlag.
- [5] J. Kiefer (1961) On large deviations of the empiric D.F. of vector chance variables and a law of the iterated logarithm. *Pacific Journal of Mathematics* 11, 649-660.
- [6] Lalley, S.P. and A.Nobel (2003) Indistinguishability of Absolutely Continuous and Singular Distributions. *Statistics and Probability Letters* 62, 145-154.
- [7] Li, Q. and J. Racine (2007) *Nonparametric Econometrics*. Princeton University Press.
- [8] Müller, H-G. (1992) Change-points in Nonparametric Regression Analysis. *Annals of Statistics* 20, 737-761.
- [9] Schwartz, L. (1950) *Théorie des Distributions*, v.1,2. Hermann, Paris.
- [10] Sobolev, S. (1992) *Cubature formulas and Modern Analysis*. Gordon and Breach Science Publishers, S.A.
- [11] Zinde-Walsh, V. (2008) Kernel Estimation when Density May not Exist. *Econometric Theory* 24, 696-725.
- [12] Zinde-Walsh, V. and P. C. B. Phillips (2003) Fractional Brownian motion as a differentiable generalized Gaussian process. In K. Athreya, M. Majumdar, M. Puri and W. Waymire (eds.) *Probability, Statistics and their Applications: Papers in Honor of Rabi Bhattacharya*. Institute of Mathematical Statistics Lecture Notes-Monograph Series, Beachwood, Ohio, v. 41, 285-292.

FIGURE 1
Empirical densities of test statistics
5000 replications; 100 bootstrap samples

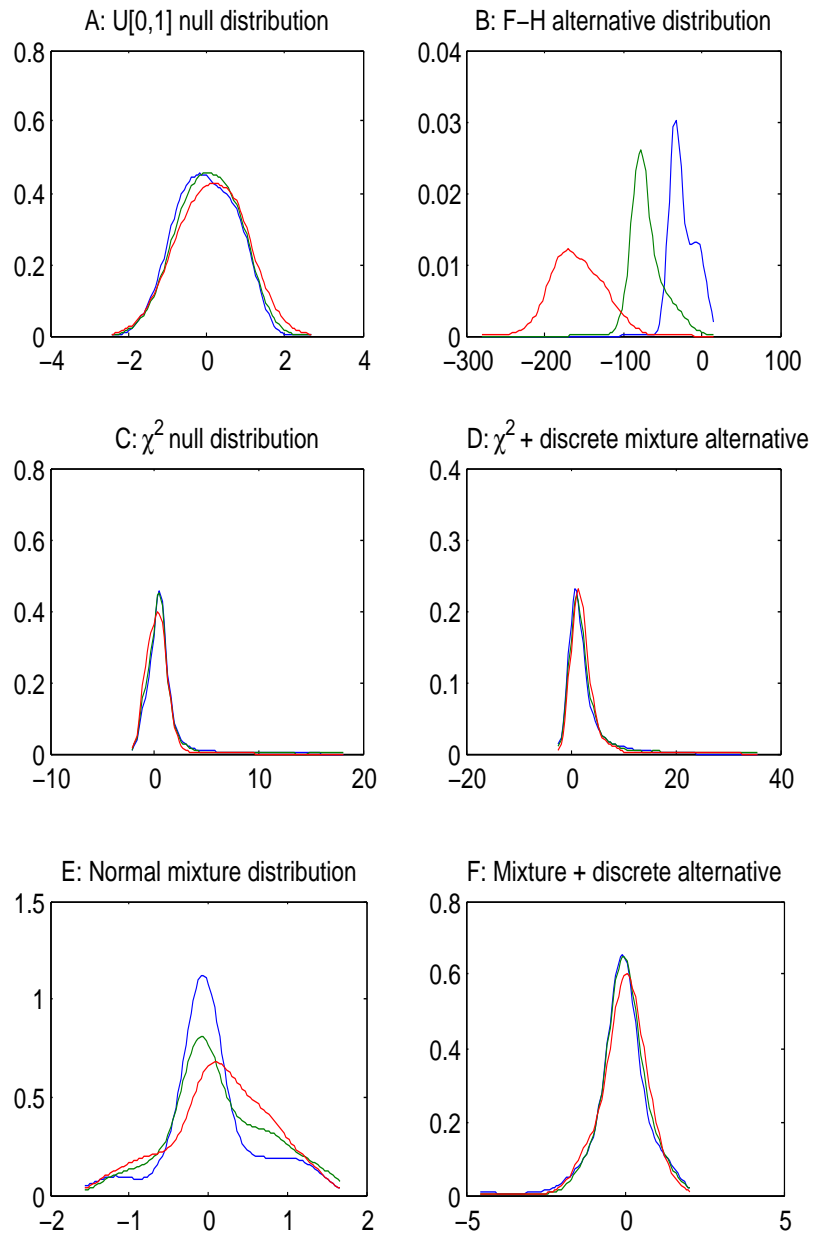


FIGURE 2
Empirical tail probabilities at nominal values from $N(0,1)$
5000 replications; 100 bootstrap samples

