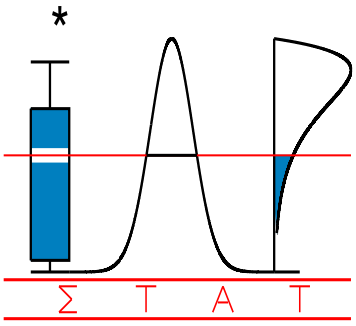
T E C H N I C A L

R E P O R T

0240

TESTING FOR MONOTONE INCREASING HAZARD RATE

P. HALL and I. VAN KEILEGOM



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

TESTING FOR MONOTONE INCREASING HAZARD RATE

Peter Hall[1] and Ingrid Van Keilegom[1,2]

ABSTRACT. A test of the null hypothesis that a hazard rate is monotone nondecreasing, versus the alternative that it is not, is proposed. Both the test statistic and the means of calibrating it are new. Unlike previous approaches, neither is based on the assumption that the null distribution is exponential. Instead, empirical information is used to effectively identify, and eliminate from further consideration, parts of the line where the hazard rate is clearly increasing; and to confine subsequent attention only to those parts that remain. This produces a test with greater apparent power, without the excessive conservatism of exponential-based tests. Our approach to calibration borrows from ideas used in certain tests for unimodality of a density, in that a bandwidth is increased until a distribution with the desired properties is obtained. However, the test statistic does not involve any smoothing, and is in fact based directly on an assessment of convexity of the distribution function, using the conventional empirical distribution. The test is shown to have optimal power properties in difficult cases, where it is called upon to detect a small departure, in the form of a bump, from monotonicity. More general theoretical properties of the test, and its numerical performance, are explored.

SHORT TITLE. Hazard rate test.

KEYWORDS. Bandwidth, bootstrap, convex function, cumulative hazard rate, kernel methods, local alternative, monotone function, power, survival analysis.

[1]Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

[2]Institut de Statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

The research of Van Keilegom was supported by the contract 'Projet d'Actions de Recherche Concertées' nr 98/03–217, and by the IAP research network nr P5/24 of the Belgian state.

1. INTRODUCTION

Estimation of hazard rate under the hypothesis that it is nondecreasing, and testing the validity of this assumption, are motivated by problems where failure rate of a machine part or a biological system can be expected to increase with lifetime. If, for some reason, a machine part becomes more reliable with time over at least part of its life cycle, then it can be particularly important to know that fact. The knowledge may lead to changes in the way the part is manufactured or finished, so as to remove the requirement for a running-in period where failure is relatively likely to occur. In this paper we suggest a new test statistic of the null hypothesis of monotone nondecreasing failure rate, and a new approach to calibrating the distribution of the statistic so as to determine a critical point for the test.

Our methods confer two advantages relative to existing approaches. First, our test statistic is focused on relatively "local" departures from the null hypothesis of nondecreasing hazard rate, and pays relatively little attention to those parts of the sample space where the hazard rate is indeed monotone nondecreasing. Nevertheless the method is easily localised still further, since it focuses on variation of the hazard rate over an interval which can be increased or decreased at the investigator's discretion, or indeed replaced by the union of two or more intervals.

Secondly, our new method of calibration makes the test statistic much more sensitive to relatively small departures from the null hypothesis. For a given nominal significance level, our calibration approach produces a test with greater apparent power than do standard methods based on calibration by comparison with the exponential distribution. Our technique for calibration also plays a role in removing much of the need for localising the test statistic, since it too places emphasis on locations in the sample space where the hazard rate of the sampled distribution appears (from empirical evidence) to depart from the hypothesis of monotonicity.

The result is that the test we propose has substantially greater apparent power in so-called "difficult cases" (compare Cheng and Hall, 1999) than do, for example, methods based on Proschan and Pyke's (1967) test statistic, calibrated using the exponential distribution (see Gijbels and Heckman, 2000). Indeed, we shall prove that our method has optimal power in this setting. That is, it is able to detect a very small perturbation of the empirical distribution, placed at a point where it produces a small non-monotone bump in the hazard rate, and so small that even a likelihood ratio test (requiring knowledge of the shape of the bump) is barely able to detect the bump.

Our calibration method is related to the "increasing bandwidth" approach first suggested by Silverman (1981) in the case of density estimation, and used in a range of other settings since; see Bowman, Jones and Gijbels (1998) for an application in the setting of monotone nonparametric regression. However, quite unlike those applications, we increase the bandwidth only for the purpose of calibrating the test. Our test statistic does not involve any smoothing at all, and is based directly on the standard empirical distribution function.

Contributions to the problem of testing for a *constant* hazard rate, against a monotone alternative, include those of Bickel and Doksum (1969), based, like the method of Proschan and Pyke (1967), on normalised spacings; Bickel (1969), on the existence of asymptotically most powerful tests; Barlow and Doksum (1972), on the more general problem of testing for convex orderings; and Ahmad (1975), Gail and Gastwirth (1978a,b) and Klefsjö (1983), who proposed tests of the hypothesis of an exponential distribution. However, these approaches share the drawbacks noted above for exponential-based methods. A related difficulty arises in the context of testing for unimodality of a probability density by calibrating against the most difficult case of a uniform density; see for example Hartigan and Hartigan (1985). Hall *et al.* (2001) have suggested methods for estimating hazard rate under the assumption of monotonicity, and surveyed earlier work on the topic.

Although our focus is on testing the null hypothesis of a monotone nondecreasing hazard rate, the case where the null asserts a monotone nonincreasing rate is related. In the former case the smoothed empirical hazard rate estimator is guaranteed to be monotone nondecreasing for all sufficiently large bandwidths, and this property is not available in the latter setting. The property makes it particularly easy to propose a bandwidth selection rule that ensures resampling from a distribution that satisfies the null; we may start with any conventional bandwidth selector, for example based on a plug-in rule, and steadily increase the bandwidth until the smoothed empirical distribution has a monotone nondecreasing hazard rate in the region where the test is to be conducted.

There is also a simple rule in the case where H_0 stipulates that the hazard rate is nonincreasing: starting with any conventional bandwidth selector, increase the bandwidth until a monotone nonincreasing hazard rate is obtained; or, if that does not occur no latter how large the bandwidth, reject the null hypothesis at this point, without passing to a further step. This rule is justified by the fact that, if the hazard rate is nonincreasing, then the probability that there exists a finite bandwidth (of

larger order than the conventional $n^{-1/5}$) such that the smoothed empirical hazard rate is nonincreasing, generally converges to 1 as sample size increases. Nevertheless, in the remainder of the paper we shall address only the more practically important case, where H_0 asserts a nondecreasing hazard rate.

2. METHODOLOGY

2.1. Test statistic. Suppose the random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ is drawn from a distribution with distribution function F. The standard empirical distribution function is $\widehat{F}(x) = n^{-1} \sum_i I(X_i \le x)$, where $I(\mathcal{E})$ denotes the indicator function of an event $\mathcal{E}$. The null hypothesis that F has monotone hazard rate on an interval $\mathcal{I}$ is equivalent to $H = -\log(1-F)$ being convex on $\mathcal{I}$, and hence, provided F is twice differentiable with a nonvanishing first derivative on $\mathcal{I}$, to H'' being nonnegative on $\mathcal{I}$. The function H is the cumulative hazard rate. Its derivative is the hazard rate.

The empirical form of H, $\widehat{H} = -\log(1 - \widehat{F})$, is not differentiable, however. Therefore it makes little sense to test the null hypothesis by checking for nonnegativity of the second derivative of $\widehat{H}$. We could investigate methods based directly on smoothed forms of $\widehat{H}$, but this would not necessarily lead to tests that have good power properties; see section 2.3. Instead we note that convexity of H on $\mathcal{I}$ is equivalent to nonnegativity of $H(x+y) + H(x-y) - 2\,H(x)$ for all x and y such that both $x+y$ and $x-y$ are elements of $\mathcal{I}$. It is not essential to take $\mathcal{I}$ to be an interval; it can be replaced by a disjoint union of intervals, for example.

Therefore, a test of the hypothesis of increasing hazard rate, or equivalently of

$$H_0\text{: } H \text{ is convex on } \mathcal{I}\,, \tag{2.1}$$

is to reject H_0 in favour of its complement if the value of

$$T = \int\!\!\int_{x,y\,:\,x+y,\,x-y\in\mathcal{I}} \max\left\{0, 2\,\widehat{H}(x) - \widehat{H}(x+y) - \widehat{H}(x-y)\right\}^r w(x,y)\,dx\,dy \tag{2.2}$$

is "too large". The exponent r is an arbitrary positive number, and w is a nonnegative weight function. By taking the maximum in the argument of the integral at (2.2) we have largely restricted attention to places where the sampled distribution has a decreasing hazard rate. (Here and below we use the words "increasing" and "decreasing" to mean "nondecreasing" and "nonincreasing," respectively.) Further restriction will be made through our method for calibration, which uses the data to determine where the hazard rate is more likely to be increasing or decreasing.

2.2. Calibration. Our approach to calibration will be based on bootstrap sampling from the distribution determined by a kernel density estimator,

$$\tilde{f}(x|h) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is a kernel and h a bandwidth. We shall choose K to be a smooth, symmetric density function, its graph being of conventional bell shape. Let $\widetilde{F}$ denote the distribution function corresponding to the density $\tilde{f}$, and let $\widetilde{H} = -\log(1 - \widetilde{F})$ be the associated cumulative hazard function. Then

$$\widetilde{H}''(x) = -(d/dx)^2 \log\{1 - \widetilde{F}(x)\} = \frac{\{1 - \widetilde{F}(x)\}\, \tilde{f}'(x) + \tilde{f}(x)^2}{\{1 - \widetilde{F}(x)\}^2}\,. \tag{2.3}$$

We shall write $\widetilde{H}''(x)$ as $\widetilde{H}''(x|h)$ when it is necessary to indicate dependence on bandwidth, and as at (2.3) we shall drop the notation h from quantities such as $\tilde{f}(\cdot|h)$ when it is not necessary for our argument. An empirical approach to bandwidth choice will be employed, as follows.

Let $\hat{h}$ denote a conventional empirical bandwidth, the asymptotic size of which is $n^{-1/5}$. We shall call $\hat{h}$ the "starting bandwidth". Examples include the bandwidths selected by the bootstrap, cross-validation or plug-in methods. Steadily increase the bandwidth, starting from $\hat{h}$, and stopping on the first occasion on which $\widetilde{H}''$ does not change sign on $\mathcal{I}$. Define

$$\hat{h}_{\rm crit} = \inf\left\{h \geq \hat{h} : \text{the equation } \widetilde{H}''(\cdot|h) = 0 \text{ has no solution on } \mathcal{I}\right\} \tag{2.4}$$

We claim that if $\mathcal{I}$ is a compact interval then for all sufficiently large h, $\widetilde{H}''(\cdot|h) > 0$ on $\mathcal{I}$, and so the set at (2.4) is not empty. Therefore, $\hat{h}_{\rm crit}$ is well defined.

To verify the claim, assume K has two continuous derivatives in a neighbourhood of the origin, $K(0) > 0$ and $K'(0) = 0$, and observe that as $h \to \infty$, $\tilde{f}(x|h) = h^{-1}\, K(0) + o(h^{-1})$ and $\tilde{f}'(x|h) = h^{-3}\, K''(0)\, n^{-1} \sum_i (x - X_i) + o(h^{-3})$, where both relations hold uniformly in $x \in \mathcal{I}$. It follows that for all sufficiently large h, $\tilde{f}(x)^2 > |\tilde{f}'(x)|$ for all $x \in \mathcal{I}$. The claim that $\widetilde{H}''(\cdot|h) > 0$ on $\mathcal{I}$ now follows from (2.3).

Having computed $\hat{h}_{\rm crit}$ we repeatedly create samples of size n by sampling randomly, with replacement, from the distribution with density $\tilde{f}(\cdot|\hat{h}_{\rm crit})$, and thereby repeatedly compute bootstrap values, T^* say, of the statistic T. Arguing thus, and given a nominal level α for the test, we may compute a critical point $\hat{c}(a)$ defined by

$$P\left\{T^* > \hat{c}(\alpha) \,\middle|\, \mathcal{X}\right\} = \alpha\,.$$

The test takes the form: reject the null hypothesis at level α if $T > \hat{c}(\alpha)$.

2.3. The road not taken: tests based on $\widetilde{H}''$. In the test described in sections 2.2 and 2.3 we have used smoothing methods only for calibration, not to construct the test statistic itself. An alternative approach would be to base a test directly on the property that, when H is twice continuously differentiable, the null hypothesis is satisfied if and only if $H'' \geq 0$ on $\mathcal{I}$. In particular we could construct a smoothed version, $\widetilde{H}$ say, of $\widehat{H}$ with the property that $\widetilde{H}''$ is a consistent estimator of H'', and reject H_0 if (for example) $S = \int_{\mathcal{I}} \{\max(0, -\widetilde{H}'')\}^2$ is "too large".

This approach has drawbacks, however. First, it requires a bandwidth to be chosen when constructing the test statistic S; a second bandwidth would be needed when calibrating the test, if calibration were to involve sampling from a smoothed distribution. Secondly, the power of the test will depend intimately on choice of the first bandwidth. Indeed, the minimum distance from the null hypothesis at which local alternative distributions can be detected by the test will generally be proportional to $n^{-1/2}h^{-c}$, where h is the bandwidth employed when constructing S, and $c > 0$ depends on the smoothing method used. Examples of this behaviour in more conventional testing problems may be found in work of Anderson, Hall and Titterington (1998), Lavergne and Vuong (2000) and Delecroix, Hall and Roget (2001).

3. THEORETICAL PROPERTIES

3.1. Summary of properties. Section 3.2 shows that, if H'' is bounded above zero on $\mathcal{I}$ (call this hypothesis H_{01}), the statistic T is of size n^{-1}, and asymptotically normally distributed. The bootstrap accurately captures this distribution. As the convexity of H becomes more marginal the stochastic fluctuations of T increase. Thus, if H'' vanishes at just a finite number of discrete points in $\mathcal{I}$ (call this H_{02}), then the size of T increases to $O(n^{-6/7})$, and its distribution becomes non-normal (see section 3.3). The size of T increases still further, to $O_p(n^{-1/2})$, if H'' vanishes on an interval, and in particular if F is an exponential distribution (section 3.7). Properties of our calibration method under H_{02} are treated in section 3.4, where it is shown that the asymptotic level is bounded away from zero. (Section 4 reports numerical properties in this case.) By way of contrast, if calibration is made ! against the exponential distribution then, under either H_{01} or H_{02}, the level converges to zero (section 3.7), implying that this approach gives ultra conservatism. Optimality of our approach, for identifying small, non-monotone "wiggles" in the hazard rate, is proved in section 3.5. The ability of our calibration method to identify any fixed departure from the null hypothesis is shown in section 3.6.

3.2. Strict monotonicity of hazard rate. Throughout section 3 we shall define the statistic T by taking $r = 1$ and $w \equiv 1$ in the definition at (2.2). Let H_{01}, a sub-hypothesis of H_0 at (2.1), be:

$$H_{01}\text{: } H'' \text{ has two continuous derivatives on } \mathcal{I} \text{ and } H'' > 0 \text{ on } \mathcal{I}\,.$$

Put $g = f^{1/2}/(1-F)$,

$$\begin{aligned}\mu &= -\int_{\mathcal{I}} dx \int_{-\infty}^{\infty} E\big[\min\big\{0\,,\, y^2\,H''(x) + g(x)\,(2|y|)^{1/2}\,N\big\}\big]\,dy > 0\,,\\ \sigma^2 &= \int \mathcal{I} dx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \operatorname{cov}\Big(\min\big\{0\,,\, y_1^2\,H''(x) + g(x)\,B(y_1)\big\}\,,\\ &\qquad\qquad \min\big[0\,,\, y_2^2\,H''(x) + g(x)\,\{B(y_2+y_3) - B(y_3)\}\big]\Big)\,dy_1\,dy_2\,dy_3\,,\end{aligned}$$

where the random variable N has the standard normal distribution and B denotes a standard Brownian motion. It is clear that μ is finite; our proof of Theorem 3.1 will show that σ^2 is also well defined and finite.

Theorem 3.1. *Assume the distribution function F has three continuous derivatives on an open interval $\mathcal{I}'$ which contains the compact bounded $\mathcal{I}$, and that the density $f = F' > 0$ on $\mathcal{I}$. If H_{01} holds then $T = n^{-1}\,\mu + n^{-7/6}\,\sigma\,N_n$, where N_n is asymptotically normally distributed with zero mean and unit variance.*

A version of the theorem continues to hold if the distribution function $F = F_n$ is allowed to depend on n. The main requirements in this case are that the regularity conditions hold in a contiguous way, and F_n converge sufficiently fast to a proper limiting distribution, G say. In particular, F_n and G (the former for all sufficiently large n) should satisfy the conditions of the theorem, and, for $j = 0$, 1 and 2, $F_n^{(j)} - G^{(j)}$ should converge to 0 at a faster rate than $n^{-1/6}$, uniformly on $\mathcal{I}'$. Under these assumptions the limiting distribution of T is that defined when, in the definitions of μ and σ, F is replaced by G. The proof requires only minor modifications.

This result may be used to prove that, under H_{01} and mild conditions on h and K, the bootstrap estimator of the distribution of T is strongly consistent for the limiting distribution of T. Our next theorem will state this result. To formulate it, put $\mathcal{H}(\xi_1, \xi_2) = [n^{-\xi_1}, n^{-\xi_2}]$ where

$$\tfrac{1}{12} < \xi_2 < \xi_1 < \tfrac{2}{9}\,. \tag{3.1}$$

Assume that

$$\begin{array}{l} K \text{ is a symmetric, compactly supported probability density with a Hölder-continuous derivative.} \end{array} \tag{3.2}$$

Note particularly that bandwidths of size $n^{-1/5}$ are in $\mathcal{H}(\xi_1, \xi_2)$ if (3.1) holds. Indeed, conventional bandwidth selectors, for example those based on bootstrap methods, cross-validation or plug-in rules, satisfy

$$P(C_1\, n^{-1/5} < \hat{h} < C_2\, n^{-1/5}) \to 1 \quad \text{as } n \to \infty, \text{ for some } 0 < C_1 < C_2 < \infty\,. \tag{3.3}$$

Let T^* denote the version of T, defined at (2.2) but with $r = 1$ and $w \equiv 1$, and computed from a sample drawn by sampling randomly from the distribution $\widetilde{F}$, conditional on $\mathcal{X}$. Let μ and σ be as in Theorem 3.1, and write Φ for the standard normal distribution function.

Theorem 3.2. *Assume that the (possibly random) bandwidth h lies in $\mathcal{H}(\xi_1, \xi_2)$, where ξ_1 and ξ_2 satisfy* (3.1)*, and that K satisfies* (3.2)*. Suppose too that F has four bounded derivatives on an open interval $\mathcal{I}'$ which contains the compact interval $\mathcal{I}$, that $f > 0$ on $\mathcal{I}$, and that H_{01} holds. Then, uniformly in x and with probability 1,*

$$P\big\{n^{7/6}\,\big(T^* - n^{-1}\,\mu\big)/\sigma \le x \,\big|\, \mathcal{X}\big\} \to \Phi(x) \tag{3.4}$$

as $n \to \infty$.

Since the conclusion of Theorem 3.1 may be stated equivalently as

$$P\big\{n^{7/6}\,\big(T - n^{-1}\,\mu\big)/\sigma \le x\big\} \to \Phi(x)\,,$$

then (3.4) may be interpreted as implying that the bootstrap distribution of T^* converges to the limiting distribution of T under H_{01}.

It should be mentioned too that if a starting bandwidth $\hat{h}$ is chosen using a standard method such as the bootstrap, cross-validation or plug-in, and if the method suggested in section 2.2 is employed to calculate the critical bandwidth $\hat{h}_{\rm crit}$, then, under the conditions imposed on F and K in Theorem 3.2, it is true with probability 1 that $\hat{h} = \hat{h}_{\rm crit}$ for all sufficiently large n. That is to say, the iterative process used to define $\hat{h}_{\rm crit}$ stops at the very first step. Furthermore, with probability 1, $\hat{h} \in \mathcal{H}(\xi_1, \xi_2)$ for all sufficiently large n. Therefore, when F satisfies H_{01} the calibration step in section 2.2 degenerates, in asymptotic terms, to simply using

the standard bandwidth selector, in which case its properties are covered by Theorem 3.2. In particular, using a standard bandwidth selector leads to consistent estimation of the limiting distribution of T under H_{01}.

3.3. Strict monotonicity at all but a finite number of points. Let H_{02}, a subhypothesis of H_0 at (2.1), be:

H_{02}: H'' has two continuous derivatives on $\mathcal{I}$, and $H'' > 0$ on $\mathcal{I}$, except for a finite number of distinct points $x_1, \ldots, x_m \in \mathcal{I}$, where H'' vanishes and $H^{(4)} > 0$.

We assume $m \geq 1$. Note that it is not possible for H'' to vanish at a point x, for $H^{(4)}$ to be strictly negative there, and at the same time for the hazard rate to be strictly increasing on sufficiently small intervals containing x.

The case of strict monotonicity at all but a finite number of points may fairly be interpreted as the boundary between cases where H_{01} holds and those where the hazard rate has decreasing parts in the vicinities of points $x_1, \ldots, x_n$. The assumption that $H''(x_i) = 0$ and $H^{(4)}(x_i) > 0$ implies that the hazard rate has a "shoulder" at x_i, and is on the verge of decreasing there. Therefore, testing in this context means attempting to identify alternative hypotheses in difficult cases; compare Cheng and Hall (1999). It offers the opportunity to assess performance against local alternative hypotheses, an opportunity we shall take up in section 3.5. The opportunity is virtually absent in the setting of section 3.2.

Let $Z_1, \ldots, Z_m$ be independent random variables, Z_i having the distribution of

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\left\{0\,,\, \left(\tfrac{1}{2}\,x^2\,y^2 + \tfrac{1}{12}\,y^4\right) H^{(4)}(x_i) + g(x_i)\,B(x+y)\right\} dx\,dy\,, \qquad (3.5)$$

where B denotes a standard Brownian motion. For simplicity we shall assume that

$$\text{no } x_i \text{ is an endpoint of } \mathcal{I}. \qquad (3.6)$$

Theorem 3.3 has an analogue in the contrary case; it involves altering the distribution of Z_i when x_i is an endpoint.

Theorem 3.3. *Assume F has four continuous derivatives on an open interval which contains the compact interval $\mathcal{I}$, and that $f = F' > 0$ on $\mathcal{I}$. Suppose too that F satisfies H_{02}, for points $x_1, \ldots, x_m$ in the formulation of that hypothesis, and that* (3.6) *holds. Then we may write $T = n^{-6/7} \sum_{1 \leq i \leq m} Z_{ni}$, where the joint distribution of $(Z_{n1}, \ldots, Z_{nm})$ converges to that of $(Z_1, \ldots, Z_m)$.*

Again a version of the theorem holds when $F = F_n$ varies with n. However, a direct analogue of Theorem 3.2 does not exist in this setting. Essentially, this

is because a bandwidth that is sufficiently large to ensure convergence of $\widetilde{H}^{(4)}$ to $H^{(4)}$, and so capture the role of $H^{(4)}(x_i)$ in the definition of the distribution of Z_i, is too large to allow sufficiently fast convergence for capturing other features of the limiting distribution. Thus, in the "boundary" case treated by Theorem 3.2 there is not a direct way, based on the estimator $\widetilde{F}$ and using a bandwidth that is asymptotic to a nonrandom quantity, of calibrating the test so as to capture the exact distribution of T.

Details behind this claim will be given in section 5.4. These difficulties persist even if $\widetilde{F}$ is computed using a high-order kernel.

One way of overcoming these difficulties would be to locally model the behaviour of F in the neighbourhood of points x where $\widetilde{H}''(x)$ was small, rather than leaving estimation there up to the generic estimator $\widetilde{F}$; and to use the model directly to estimate the distributions of $Z_1, \ldots, Z_n$. This approach is rather cumbersome, however, and so for simplicity we shall not consider it further. Moreover, the problems are largely overcome by the calibration method proposed in section 2.2, the theory of which we treat next.

3.4. Calibration based on $\hat{h}_{\rm crit}$. The calibration method suggested in section 2.2 produces a test whose level, under H_{02}, converges to a number that lies strictly between 0 and 1, and so suffers less from the difficulties noted above. First we describe limiting behaviour of the critical bandwidth, $\hat{h}_{\rm crit}$, under H_{02}. For simplicity we assume there is only a single point, x_1, at which H'' vanishes.

Define $c = \frac{1}{2} \int u^2 K(u)\, du$ and

$$\begin{aligned} S(q,x,y) = \bigl(c\,q^2 + \tfrac{1}{2}\,x^2 + \tfrac{1}{12}\,y^2\bigr)\, H^{(4)}(x_1) \\ + q^{-2}\, g(x_1) \int_{-\infty}^{\infty} K''(u)\, du \int_0^1 \{B(x+ty-qu) \\ + B(x-ty-qu)\}\,(1-t)\, dt\,, \end{aligned} \tag{3.7}$$

where $g = f^{1/2}/(1-F)$ and B is standard Brownian motion. Let $Q > 0$ denote the infimum of values $q > 0$ such that $S(q,x,y) \ge 0$ for all real x, y.

Theorem 3.4. *Assume the conditions of Theorem 3.3, but with $m = 1$. Suppose too that K is a symmetric, compactly supported probability density with two Hölder-continuous derivatives, and that the starting bandwidth $\hat{h}$, used to initiate the algorithm that produces $\hat{h}_{\rm crit}$, satisfies* (3.3). *Then $n^{1/7}\,\hat{h}_{\rm crit} \to Q$ in distribution as $n \to \infty$.*

Next we describe the asymptotic level of the test under H_{02}. For $0 < \alpha < 1$

define z_α to be the α-level quantile of the distribution defined at (3.5), in the case $i = 1$. Noting (3.5) we see that we may write z_α as a continuous function of $H^{(4)}(x_1)$ and $g(x_1)$, say $z_\alpha = \Gamma_\alpha\{H^{(4)}(x_1), g(x_1)\}$. Put

$$S(x) = S(Q, x, 0) = \left(c\,Q^2 + \tfrac{1}{2}\,x^2\right) H^{(4)}(x_1) + Q^{-2}\,g(x_1) \int_{-\infty}^{\infty} K''(u)\,B(x - Q\,u)\,du\,. \tag{3.8}$$

It follows from the definition of Q that with probability 1, (a) $S(x) \geq 0$ for $-\infty < x < \infty$, (b) there exists a unique (random) point $x = A$ at which $S(x) = 0$, and (c) $S'(A) = 0$ and $S''(A) > 0$. (To appreciate why, observe that S is asymptotically proportional to $\widetilde{H}''(x_1 + n^{-1/7}x)$, after taking the bandwidth to equal $n^{-1/7}Q$. Note that the second derivative of S is well defined and continuous as long as K has three continuous derivatives.)

Let Z_1 denote the random variable at (3.5) when $i = 1$, constructed using the same Brownian motion B as at (3.8). Therefore, Z_1 and $S''(A)$ are linked through B. In interpreting the theorem below, note that the probability that $Z_1 \leq \Gamma_\alpha\{H^{(4)}(x_1), g(x_1)\}$ equals α.

Theorem 3.5. *Assume the conditions of Theorem 3.4, but with the additional requirement that K have three continuous derivatives. Then the level of the bootstrap test converges, as $n \to \infty$, to the probability that $Z_1 \leq \Gamma_\alpha\{S''(A), g(x_1)\}$.*

3.5. Power against local alternatives, and optimality. Let F denote a four-times continuously-differentiable distribution satisfying H_{02}. Assume for simplicity that there is only one point at which, for this F, H'' vanishes on $\mathcal{I}$. Let this point be $x_1 = 0$, and take it to be an interior point of $\mathcal{I}$. Since H_{02} holds then $H^{(3)}(0) = 0$ and $H^{(4)}(0) > 0$.

We shall add a "wiggle" to F in the vicinity of the origin, such that the perturbed distribution violates the null hypothesis. The perturbation will be chosen so that it is only barely detectable using an optimal parametric method, i.e. the likelihood-ratio test. We shall then explore the performance of our nonparametric test, based on the statistic T, and show that it too is able to detect the wiggle.

The perturbation, $a\,\epsilon^4\,\Psi(x/\epsilon)$, is based on a four-times continuously-differentiable function Ψ, supported on $[-1, 1]$. The constant $a > 0$ represents the height of the wiggle, and $\epsilon = \epsilon(n) \to 0$ indicates the extent of the perturbation away from its centre, at the origin. We shall choose ϵ so small that the perturbation is only barely detectable by the likelihood-ratio test. Our construction of the perturbation

ensures that, like the distribution F to which it is added, it has four bounded derivatives near the origin.

The perturbed distribution is

$$F_n(x) = F(x) + a\,\epsilon^4\,\Psi(x/\epsilon)\,. \tag{3.9}$$

Let H_n denote the cumulative hazard rate corresponding to F_n. If we choose Ψ so that $\Psi(x) \equiv -x^4$ in a neighbourhood of the origin then, for each $a > 0$ and all sufficiently large n, H_n' is strictly monotone decreasing in a neighbourhood of 0. (This neighbourhood is of width $O(\epsilon)$.) Therefore, F_n fails to satisfy the null hypothesis of an increasing hazard rate.

The density $f_n = F_n'$ satisfies $f_n(x) = f(x) + a\,\epsilon^3\,\psi(x/\epsilon)$, where $\psi = \Psi'$. Since f_n must be a density then $\int \psi = 0$. Now,

$$\log\{f_n(x)/f(x)\} = \frac{a\,\epsilon^3\,\psi(x/\epsilon)}{f(x)} - \frac{a^2\,\epsilon^6\,\psi(x/\epsilon)^2}{2\,f(x)^2} + O\big(\epsilon^9\big)\,.$$

Therefore, putting $b(a) = \frac{1}{2}\,a^2\,f(0)^{-1}\,\int \psi^2$, $f_+ = f_n$ and $f_- = f$ we have, taking the $\pm$ signs respectively,

$$\int f_\pm(x)\,\log\{f_n(x)/f(x)\}\,dx = \pm\,b(a)\,\epsilon^7 + o\big(\epsilon^7\big)\,. \tag{3.10}$$

It follows from (3.10) that the expected log-likelihood ratio, for a sample of size n, is of size $n\epsilon^7$. Choosing ϵ such that this quantity is bounded away from zero and infinity, in particular $\epsilon = n^{-1/7}$, makes the perturbation only barely detectable. In that case a likelihood-ratio test for discriminating between f and f_n does not have asymptotically perfect accuracy. In particular, if $\pi_{1\alpha}(a)$ denotes the limiting probability that the α-level likelihood-ratio test correctly detects f_n when the null hypothesis is f, then $0 < \pi_{1\alpha}(a) < 1$ for each a, $\pi_{1\alpha}(a) \to \alpha$ as $a \to 0$, and $\pi_{1\alpha}(a) \to 1$ as $a \to \infty$. By the Neyman-Pearson lemma, if $\pi_{2\alpha}(a)$ denotes the analogous limiting probability in the case of any other test (for example, our own), then $\pi_{2\alpha}(a) \le \pi_{1\alpha}(a)$.

Our test is able to detect local alternatives, such as F_n, provided the function $\pi_{2\alpha}(a)$ for our test satisfies

$$\lim_{a\to\infty} \pi_{2\alpha}(a) = 1\,. \tag{3.11}$$

If (3.11) holds then our test shares the optimal performance of the likelihood-ratio test.

To establish (3.11), note first that for $j = 0, \ldots, 4$,

$$H_n^{(j)} = H^{(j)} + \frac{a\,\epsilon^{4-j}\,\Psi^{(j)}(x/\epsilon)}{1 - F(x)} + O\big\{\epsilon^{5-j}\,I(|x| \le \epsilon)\big\}\,,$$

uniformly in x. In particular, the second derivative of $H_n - H$ is of size $n^{-2/7}$, and the fourth derivative is asymptotic to $a\,\Psi^{(j)}(x/\epsilon)/\{1 - F(x)\}$. Using these properties, and the arguments leading to Theorem 3.5, the following result may be proved. It verifies (3.11) in the case where the test in question is the bootstrap-calibrated one proposed in section 2.

Theorem 3.6. *Assume the four-times continuously-differentiable distribution F satisfies H_{02} with $m = 1$ and $x_1 = 0$, and that F_n is given by* (3.9)*, where the four-times continuously-differentiable function Ψ is supported on $[-1, 1]$ and satisfies $\Psi(x) = -x^4$ in a neighbourhood of the origin. Suppose too that ϵ, in* (3.9)*, is $n^{-1/7}$. Let $p_\alpha(a, n)$ denote the probability that the α-level, bootstrap-calibrated test of the null hypothesis of monotone hazard rate rejects the null hypothesis when applied to data from F_n. Then (a) $p_\alpha(a, n)$ converges to a limit, $\pi_{2\alpha}(a)$ say, as $n \to \infty$, and (b) $\pi_{2\alpha}(a)$ satisfies* (3.11) *as $a \to \infty$.*

3.6. Level and power against fixed alternatives. The result below shows that the bootstrap-calibrated form of our test is asymptotically consistent in rejecting the null hypothesis whenever it is violated by a fixed alternative.

Theorem 3.7. *Assume F has two continuous derivatives on an open interval $\mathcal{I}'$ which contains the compact interval $\mathcal{I}$, that $f > 0$ on $\mathcal{I}$, but that the hazard rate for F is strictly decreasing in a subinterval of $\mathcal{I}$. Suppose too that K satisfies* (3.2)*, that $K(0) \ne 0$, that $E|X| < \infty$, and that the starting bandwidth $\hat{h}$, for the algorithm leading to $\hat{h}_{\rm crit}$ defined in section 2.2, satisfies* (3.3)*. Then $P\{T \ge \hat{c}(\alpha)\} \to 1$, as $n \to \infty$, for each $0 < \alpha < 1$, where $\hat{c}(\alpha)$ is the α-level bootstrap critical point introduced in section 2.2.*

3.7. Calibration against the exponential distribution. Put $A(x) = B_0\{F(x)\}/\{1 - F(x)\}$, where B_0 is a standard Brownian bridge, and define

$$T_0 = \int\!\!\int_{x,y\,:\,x+y,\,x-y \in \mathcal{I}} \max\big\{0, A(x+y) + A(x-y) - 2\,A(x)\big\}\,dx\,dy\,.$$

In this notation, and using standard Gaussian approximations to the empirical distribution $\widehat{F}$ (see e.g. Komlós, Major and Tusnády, 1975), it can be proved that

if F is taken to be exponential over $\mathcal{I}$ then $n^{1/2}\,T \to T_0$ in distribution. This result follows from that fact that, in the exponential case, the cumulative hazard rate is linear. In particular, neither H_{01} nor H_{02} holds in that setting.

Therefore, if we calibrate T by reference to an exponential distribution then the critical points for the test will be distant $n^{-1/2}$ from the origin. However, under H_{01} or H_{02}, this is much further from zero than the actual critical points of the distribution of T. Indeed, we know from Theorems 3.1 and 3.3 that those points are distant only $O(n^{-1})$ from zero when H_{01} holds, and only distant $O(n^{-6/7})$ when H_{02} is true. (The same is true of the bootstrap critical-point estimator suggested in section 2.2.) It follows that for each value of the nominal level of an exponentially calibrated test, the exact level (under H_{01} or H_{02}) will converge to 0 as $n \to \infty$.

Put another way, the exponentially calibrated test will become ultra-conservative as sample size increases. In particular it will fail, asymptotically, to detect the perturbation-type null hypothesis discussed in section 3.5. In order for detection to be even barely possible in that setting, the perturbation $\epsilon^4\,\Psi(x/\epsilon)$ (with $\epsilon = n^{-1/7}$) would have to be increased by the factor $n^{3/14}$.

4. SIMULATIONS

In order to illustrate the finite sample performance of the proposed test procedure, we carry out a small simulation study for the following hazard rate of the variable X :

$$H'(x) = a[(x-b)^3 + b^3] + c + dx^2\,,$$

where $x, a, b, c > 0$ and d is chosen such that $H'(x) > 0$ for all $x > 0$. The distribution corresponding to this hazard function is given by

$$F(x) = 1 - \exp\{-a[\frac{1}{4}(x-b)^4 + b^3 x] - cx - \frac{d}{3}x^3\}\,.$$

It is readily verified that hypothesis H_{01} is satisfied for $d > 0$, H_{02} for $d = 0$ and neither H_{01} nor H_{02} are satisfied when $d < 0$. Figure 1 shows the graph of $H'(x)$ for certain values of the parameters.

Figure 1 about here

The simulations are based on 2000 samples of size $n = 50$ and for each simulated sample 2000 resamples are generated. The interval $\mathcal{I}$ on which the test statistic T is based is given by $[0, F^{-1}(0.95)]$. The starting bandwidth $\hat{h}$ is determined from

the normal reference rule for plug-in estimation, i.e. $\hat{h} = 1.06 n^{-1/5} \hat{\sigma}$, where $\hat{\sigma}$ is the estimated standard error of X. The kernel function used is the normal kernel. The results for $a = 2.5, b = 0.75, c = 0.50$, for several values of d and for $\alpha = 0.10$ are presented in Figure 2. The power curve starts at -1.14, which is the smallest possible value of d for this choice of parameters. The results for other choices of the parameters and for $\alpha = 0.05$ are similar. For most choices slightly conservative alpha levels are observed. As a comparison, we also carried out the test proposed by Gijbels and Heckman (2000). The power/actual alpha level of their test was everywhere zero, even for cases where the new test has power close to 1. This confirms what was explained in Sections 1 and 3.7 about the lack of power of tests based on calibration with respect to the exponential distribution.

Figure 2 about here

5. TECHNICAL ARGUMENTS

5.1. Proof of Theorem 3.1. Define $\Delta_{0F} = \widehat{F} - F$, and observe that

$$\widehat{H} = H + \frac{\Delta_{0F}}{1 - F} + O_p\big(n^{-1}\big)\,, \quad \Delta_{0F} = n^{-1/2}\, W_0(F) + O_p\big(n^{-1} \log n\big)\,, \tag{5.1}$$

where the first result holds uniformly on $\mathcal{I}$, the second uniformly on the real line, and W_0 denotes a Brownian bridge the construction of which depends on the data. The first identity at (5.1) follows by simple Taylor expansion, while the second uses results of Komlós, Major and Tusnády (1975). Together the identities imply that

$$\widehat{H} = H + n^{-1/2}\, \frac{W_0(F)}{1 - F} + O_p\big(n^{-1} \log n\big)\,, \tag{5.2}$$

uniformly on $\mathcal{I}$.

Assume H_{01} obtains, and, given a function $\psi(x)$ defined for $x \in \mathcal{I}$, put $\psi(x, y) = \psi(x+y) + \psi(x-y) - 2\,\psi(x)$ whenever $x+y, x-y \in \mathcal{I}$. Now, $H(x, y) = y^2\, H''(x+\theta y)$, where $-1 \le \theta = \theta(x, y) \le 1$. Hence, under H_{01},

$$\inf_{x,y\,:\,x+y,\,x-y \in \mathcal{I}} y^{-2}\, H(x, y) > 0\,.$$

We may therefore deduce from (5.2) that for some constant $C_1 > 0$, $-\widehat{H}(x, y) > 0$ only if

$$C_1\, n^{1/2}\, y^2 \le \max\Big\{ \big|W_0\{F(x+y)\} - W_0\{F(x)\}\big|\,, \\ \big|W_0\{F(x-y)\} - W_0\{F(x)\}\big| \Big\} + n^{-1/2}\, (\log n)\, A_n\,, \tag{5.3}$$

where the random variable A_n does not depend on x or y and equals $O_p(1)$ as $n \to \infty$.

For each x let $Y(x)$ denote the supremum of values y such that $x+y$, $x-y \in \mathcal{I}$ and (5.3) holds. Then for each x, $Y(x) = O_p(n^{-1/3})$. Since

$$|W_0(t+u) - W_0(t)| = O_p\big(|u \log u|^{1/2}\big) \tag{5.4}$$

uniformly in t, u such that $0 < t, t+u < 1$, then

$$\sup \mathcal{I}\, Y(x) = O_p\big\{(n^{-1} \log n)^{1/3}\big\}\,. \tag{5.5}$$

Defining $\Delta_{1F} = W_0(F)$ and $\Delta_{2F} = W_0(F)/(1-F)$ we deduce by first Taylor expansion and then application of (5.4) that

$$\begin{aligned}
\Delta_{2F}(x,y) &= \frac{W_0\{F(x+y)\}}{1-F(x)} \left(1 + \frac{y\, f(x)}{1-F(x)}\right) \\
&\qquad + \frac{W_0\{F(x-y)\}}{1-F(x)} \left(1 - \frac{y\, f(x)}{1-F(x)}\right) - 2\, \frac{W_0\{F(x)\}}{1-F(x)} + O_p\big\{Y(x)^2\big\} \\
&= \frac{\Delta_{1F}(x,y)}{1-F(x)} + O_p\big\{Y(x)^{3/2}\, (\log n)^{1/2}\big\}\,,
\end{aligned} \tag{5.6}$$

uniformly in $x \in \mathcal{I}$ and $|y| \le Y(x)$. Therefore,

$$T = \int \mathcal{I} dx \int_{-Y(x)}^{Y(x)} \max\big\{0, -\widehat{H}(x,y)\big\}\, dy \tag{5.7}$$

$$\begin{aligned}
&= -\int \mathcal{I} dx \int_{-Y(x)}^{Y(x)} \min\big\{0\,,\, H(x,y) + n^{-1/2}\, \Delta_{2F}(x,y)\big\}\, dy \\
&\qquad\qquad + O_p\Big\{\big(n^{-1} \log n\big)^{4/3}\Big\} \\
&= -\int \mathcal{I} dx \int_{-Y(x)}^{Y(x)} \min\left\{0\,,\, H(x,y) + n^{-1/2}\, \frac{\Delta_{1F}(x,y)}{1-F(x)}\right\} dy \\
&\qquad\qquad + O_p\Big\{\big(n^{-1} \log n\big)^{4/3}\Big\}\,,
\end{aligned} \tag{5.8}$$

where the second identity follows from (5.2) and (5.5), and the third comes from (5.5) and (5.6).

Let W denote the standard Brownian motion through which W_0 may be expressed as $W_0(t) = W(t) - t\, W(1)$ for $0 \le t \le 1$. Put $\Delta_{3F} = W(F)$. Observe that $\Delta_{1F}(x,y) - \Delta_{3F}(x,y) = O_p\{Y(x)^2\}$ uniformly in $x \in \mathcal{I}$ and $|y| \le Y(x)$. Therefore, (5.5) and (5.8) imply that

$$\begin{aligned}
T = -\int \mathcal{I} dx \int_{-Y(x)}^{Y(x)} &\min\left\{0\,,\, H(x,y) + n^{-1/2}\, \frac{\Delta_{3F}(x,y)}{1-F(x)}\right\} dy \\
&+ O_p\Big\{\big(n^{-1} \log n\big)^{4/3}\Big\}\,.
\end{aligned} \tag{5.9}$$

Since H''' is bounded then $H(x,y) = y^2\,H''(x) + O(|y|^3)$ as $y \to 0$, uniformly in $x \in \mathcal{I}$. From this result, (5.5) and (5.9) we deduce that

$$T = T_1 + O_p\Big\{\big(n^{-1}\log n\big)^{4/3}\Big\}\,, \tag{5.10}$$

where

$$\begin{aligned} -T_1 &= \int \mathcal{I}dx \int_{-Y(x)}^{Y(x)} \min\left\{0\,,\, y^2\,H''(x) + n^{-1/2}\,\frac{\Delta_{3F}(x,y)}{1-F(x)}\right\} dy \\ &= n^{-1}\int \mathcal{I}dx \int_{\mathcal{I}_n(x)} \min\left\{0\,,\, z^2\,H''(x) + n^{1/6}\,\frac{\Delta_{3F}(x,n^{-1/3}z)}{1-F(x)}\right\} dz \end{aligned} \tag{5.11}$$

and $\mathcal{I}_n(x)$ denote the set of y such that both $x + n^{-1/3}y$ and $x - n^{-1/3}y$ lies in $\mathcal{I}$.

Put

$$W_x(y) = n^{1/6}\big[W\big\{F(x) + n^{-1/3}y\,f(x)\big\} - W\{F(x)\}\big]/f(x)^{1/2}\,,$$

which, like W, is standard Brownian motion. It may be proved from (5.11) that

$$\begin{aligned} -n\,E(T_1) = \int \mathcal{I}dx \int_{-\infty}^{\infty} E\Bigg\{ \min\Bigg(0\,,\, z^2\,H''(x) + \frac{f(x)^{1/2}}{1-F(x)} \\ \times\Big[W_x\big\{z + \tfrac{1}{2}\,n^{-1/3}z^2\,f'(x)\,f(x)^{-1}\big\} \\ + W_x\{-z + \tfrac{1}{2}\,n^{-1/3}z^2\,f'(x)\,f(x)^{-1}\}\Big]\Bigg)\Bigg\}\,dz + o\big(n^{-1/6}\big) \end{aligned}$$

¿From this result and the fact that, for $0 < |u| < |z|$, $W_x(z+u) + W_x(-z+u)$ has the normal $\mathrm{N}(0, 2|z|)$ distribution, we deduce that

$$n\,E(T_1) = \mu + o\big(n^{-1/6}\big)\,, \tag{5.12}$$

where μ is as defined in section 3.

To derive a central limit theorem for T_1 we first approximate T_1 by a sum of 3-dependent random variables, as follows. Define $\lambda_n = \log n$ and $\delta = \delta(n) = \lambda_n\,(n^{-1}\,\log n)^{1/3}$. Put

$$\begin{aligned} -T_2 &= \int \mathcal{I}dx \int_{-\delta}^{\delta} \min\left\{0\,,\, y^2\,H''(x) + n^{-1/2}\,\frac{\Delta_{3F}(x,y)}{1-F(x)}\right\} dy\,, \\ -T_2(i) &= \int_{\mathcal{I}\cap(i\delta,(i+1)\delta)} dx \int_{-\delta}^{\delta} \min\left\{0\,,\, y^2\,H''(x) + n^{-1/2}\,\frac{\Delta_{3F}(x,y)}{1-F(x)}\right\} dy\,; \end{aligned}$$

compare these definitions with the first identity at (5.11). Then $T_2 = \sum_i T_2(i)$. Note that, since Brownian motion has independent increments, $T_2(i)$ is stochastically independent of $T_2(j)$ for $|i-j| \geq 3$.

In view of (5.5), the probability that $\max \mathcal{I}|Y(x)| \le \delta$ converges to 1 as $n \to \infty$. Note too that $\max_{x \in \mathcal{I}} |Y(x)| \le \delta$ implies $T_1 = T_2$. Hence, if we prove that the following three results are true: (a) $\operatorname{var} T_2 \sim \operatorname{var} T_1 \sim \sigma^2 n^{-7/3}$, (b) $(T_2 - ET_2)/(\operatorname{var} T_2)^{1/2}$ has an asymptotic standard normal distribution, and (c) $n^{7/6}(ET_1 - ET_2) \to 0$; then it will follow that $n^{7/6}(T_1 - ET_1)/\sigma$ has an asymptotic standard normal distribution. Theorem 3.1 is a consequence of this property and (5.12).

Result (c) may be proved using the argument leading to (5.12), and the first asymptotic relation in (a) may be derived using the method giving the second. Therefore, it suffices to show that (b) holds and that (d) $\operatorname{var} T_1 \sim \sigma^2 n^{-7/3}$.

To prove (b), let $C > 0$ and define $T_3(i) = n^{7/6} \delta^{-1/2} T_2(i)$, $T_4(i) = T_3(i) \times I\{|T_3(i)| \le C\}$, $T_5(i) = T_3(i) - T_4(i)$ and $T_j = \sum_i T_j(i)$ for $j = 4, 5$. For all sufficiently large C the variance of T_4, and the number of nondegenerate summands $T_4(i)$, are both asymptotic to constant multiples of δ^{-1}; and the summands are uniformly bounded. Therefore, using a central limit theorem for uniformly bounded m-dependent random variables (see e.g. Theorem 7.3.1, p. 214 of Chung (1974)), we may prove that $(T_4 - ET_4)/(\operatorname{var} T_4)^{1/2}$ has an asymptotic standard normal distribution; call this result (e). The argument that we shall use to prove (d) may be employed to show that as $C \to \infty$, (f) $\lim_{n \to \infty} \delta \operatorname{var} T_4 \to \sigma^2$ and (g) $\lim_{n \to \infty} \delta \operatorname{var} T_5 \to 0$. Combining (e)–(g) we deduce that $(T_3 - ET_3)/(\operatorname{var} T_3)^{1/2}$ has an asymptotic normal distribution. This is equivalent to (b).

It remains to derive (d). Recall that $g = f^{1/2}/(1 - F)$, and define

$$
\begin{aligned}
U_x(y) &= n^{1/6} \big[W\big\{F(x + n^{-1/3} y)\big\} - W\{F(x)\} \big] / f(x)^{1/2} \,, \\
V_x(y) &= U_x(y) + U_x(-y) \,, \\
W_1(x_1, y_1) &= \min \Big\{ 0 \,,\, y_1^2 \, H''(x_1) + g(x_1) \, V_{x_1}(y_1) \Big\} \,, \\
W_2(x_1, x, y_2) &= \min \Big\{ 0 \,,\, y_2^2 \, H''(x_1 + n^{-1/3} x) + g(x_1 + n^{-1/3} x) \, V_{x_1 + n^{-1/3} x}(y_2) \Big\} \,,
\end{aligned}
$$

and $\mathcal{J}_n(x_1) = \big\{x : x_1 + n^{-1/3} x \in \mathcal{I}\big\}$. In this notation,

$$
\begin{aligned}
n^2 \operatorname{var} T_1 &= \int_{\mathcal{I}} dx_1 \int_{\mathcal{I}} dx_2 \int_{\mathcal{I}_n(x_1)} dy_1 \int_{\mathcal{I}_n(x_2)} \operatorname{cov} \Big[\min \big\{ 0 \,,\, y_1^2 \, H''(x_1) \\
&\qquad + g(x_1) \, V_{x_1}(y_1) \big\} \,,\, \min \big\{ 0 \,,\, y_2^2 \, H''(x_2) + g(x_2) \, V_{x_2}(y_2) \big\} \Big] \, dy_1 \, dy_2 \\
&= n^{-1/3} \int_{\mathcal{I}} dx_1 \int_{x \in \mathcal{J}_n(x_1)} dx_2 \int_{\mathcal{I}_n(x_1)} dy_1 \int_{\mathcal{I}_n(x_1 + n^{-1/3} x)} \\
&\qquad\qquad \operatorname{cov}\{W_1(x_1, y_1), W_2(x_1, x, y_2)\} \, dy_2 \,.
\end{aligned}
$$

In view of the independent increments of Brownian motion, the random variables $V_{x_1}(y_1)$ and $V_{x_1+n^{-1/3}x}(y_2)$ are independent if $|y_1| + |y_2| \le |x|$. In this case, the covariance in the second identity above vanishes. Therefore,

$$n^{7/3} \operatorname{var} T_1 = \int_{\mathcal{I}} dx_1 \int_{\mathcal{J}_n(x_1)} dx \iint\limits_{y_1,y_2\,:\,|y_1|+|y_2|>|x|\,;\,\mathcal{C}(x_1,x)} \operatorname{cov}\{W_1(x_1,y_1), W_2(x_1,x,y_2)\}\, dy_1\, dy_2\,, \tag{5.13}$$

where $\mathcal{C}(x_1, x)$ denotes the constraint that $y_1 \in \mathcal{I}_n(x_1)$ and $y_2 \in \mathcal{I}_n(x_1 + n^{-1/3}x)$.

The random variables $|W_1(x_1, y_1)|$ and $|W_2(x_1, x, y_2)|$ are respectively bounded by $C_1\,|N_1|\,I(|N_1| > C_2\,y_1^2)$ and $C_1\,|N_2|\,I(|N_2| > C_2\,y_2^2)$, where N_1 and N_2 are standard normal random variables, and C_1 and C_2 are positive constants not depending on x_1, x, y_1 or y_2, although the correlation between N_1 and N_2 does depend on these quantities. We may therefore deduce that, for constants $C_3, C_4 > 0$,

$$\begin{aligned}
&\bigl|\operatorname{cov}\{W_1(x_1,y_1), W_2(x_1,x,y_2)\}\bigr| \\
&\qquad \le C_1\, E\bigl\{|N_1|^2\, I\bigl(|N_1| > C_2\, y_1^2\bigr)\bigr\}^{1/2}\, E\bigl\{|N_2|^2\, I\bigl(|N_2| > C_2\, y_2^2\bigr)\bigr\}^{1/2} \\
&\qquad \le C_3\, \exp\bigl\{ - C_4\, \bigl(y_1^4 + y_2^4\bigr)\bigr\}\,.
\end{aligned} \tag{5.14}$$

Therefore, $|\operatorname{cov}\{W_1(x_1, y_1), W_2(x_1, x, y_2)\}|$ is bounded above by a function which does not depend on n and whose integral over $-\infty < x < \infty$ and over all real y_1, y_2 that satisfy $|y_1| + |y_2| > |x|$, is bounded uniformly in $x_1 \in \mathcal{I}$.

Furthermore, if V is a standard Brownian motion then

$$\begin{aligned}
&\operatorname{cov}\{W_1(x_1, y_1), W_2(x_1, x, y_2)\} \\
&\qquad \to \operatorname{cov}\Bigl(\min\bigl\{0\,,\, y_1^2\, H''(x_1) + g(x_1)\, V(y_1)\bigr\}\,, \\
&\qquad\qquad \min\bigl[0\,,\, y_2^2\, H''(x_1) + g(x_1)\,\{V(x + y_2) - V(x)\}\bigr]\Bigr)\,,
\end{aligned}$$

uniformly in $x_1 \in \mathcal{I}$ and x, y_1 and y_2 in any compact set. We may therefore deduce from (5.13) and the dominated convergence theorem that $\operatorname{var} T_1 \sim \sigma^2\, n^{-7/3}$, which is the desired result (d). Note that (5.14) also implies the finiteness of σ^2.

5.2. Proof of Theorem 3.2. Put $\nu_0 = 0$ and $\nu_j = 2j - 1$ for $j \ge 1$. Observe that with probability 1, and for $j = 0, 1, 2$ and each $\eta > 0$,

$$\widetilde{F}^{(j)} - F^{(j)} = O\bigl\{\bigl(nh^{\nu_j}\bigr)^{\eta-(1/2)} + h^2\bigr\}\,, \tag{5.15}$$

uniformly in $h \in \mathcal{H}(\xi_1, \xi_2)$ and $x \in \mathcal{I}'$. (The assumption that F has four bounded derivatives is needed to obtain the $O(h^2)$ remainder term in (5.15) when $j = 2$.)

It follows from this property and (2.3) that with probability 1, $\widetilde{H}''$ converges to H'' uniformly in $h \in \mathcal{H}(\xi_1, \xi_2)$ and $x \in \mathcal{I}'$. We may choose $\mathcal{I}'$ and $\epsilon > 0$ such that $H'' > \epsilon$ on $\mathcal{I}'$. In this case, and with probability 1, $\widetilde{H}'' > \frac{1}{2}\,\epsilon$ on $\mathcal{I}'$ for all sufficiently large n. In particular, for all sufficiently large n, $\widetilde{F}$ satisfies H_{01}.

The argument leading to Theorem 3.1 may now be used to prove that (3.4) holds when $\widetilde{F}$, rather than F, is the sampled distribution, provided μ and σ at (3.4) are replaced by the analogous functionals of $\widetilde{F}$. Let these be $\widetilde{\mu}$ and $\tilde{\sigma}$, respectively, and denote by (R) the corresponding version of (3.4). By (5.15),

$$|\widetilde{\mu} - \mu| + |\tilde{\sigma} - \sigma| = O_p\big\{\big(nh^3\big)^{\eta-(1/2)} + h^2\big\} = o_p\big(n^{-1/6}\big)\,, \tag{5.16}$$

the second identity holding uniformly in $h \in \mathcal{H}(\xi_1, \xi_2)$ and following from (3.1). Property (3.4) follows from (5.16) and (R).

We should mention that the assumption in Theorem 3.1 that F have three derivatives is imposed for simplicity, and is a little more stringent than necessary. At (5.10) we need only two derivatives and a Hölder condition of order $\frac{1}{2} + \epsilon$ on F'', in which case the O_p term at (5.10) becomes $O_p\{(n^{-1}\log n)^{(3+\epsilon)/3}\} = o_p(n^{-7/6})$ (as required), the identity holding provided $\epsilon > 0$. An empirical version of this argument can be developed provided $h \in \mathcal{H}(\xi_1, \xi_2)$ and ξ_1, ξ_2 satisfy (3.1).

5.3. Proof of Theorem 3.3. The assumption that the hazard rate is nondecreasing, and that $H''(x_i) = 0$, implies that $H'''(x_i) = 0$ for $1 \le i \le m$. To appreciate why, observe that

$$H(x, y) = y^2\, H''(x) + \tfrac{1}{12}\, y^4\, H^{(4)}(x + \theta y)\,,$$

where $-1 \le \theta = \theta(x, y) \le 1$. Taking $x = x_i + u$, where $|u|$ is small, and Taylor-expanding, we deduce that

$$H(x_i + u, y) = y^2\, u\, H'''(x_i) + \big(\tfrac{1}{2}\, u^2\, y^2 + \tfrac{1}{12}\, y^4\big)\, H^{(4)}\{x + \theta'\,(|u| + |y|)\}\,,$$

where $-1 \le \theta' \le 1$. If $H'''(x_i) \ne 0$ then, taking $|u| = |y|^{3/2}$ and choosing the sign of u such that $u\, H'''(x_i) < 0$, we find that as $y \to 0$, $H(x_i + u, y) = -|y|^{7/2}\, |H'''(x_i)| + o(|y|^{7/2})$. This implies that H is nonconvex near x_i, and so contradicts the assumption that the hazard rate is nondecreasing.

Result (5.2) continues to hold in the setting of Theorem 3.3, and so by (5.7),

$$T = T_2 + O_p\Big\{\big(n^{-1}\log n\big)^{8/7}\Big\}\,, \tag{5.17}$$

where

$$T_2 = -\int \mathcal{I} dx \int_{-Y(x)}^{Y(x)} \min\Big\{0\,,\, y^2\, H''(x) + \tfrac{1}{12}\, y^4\, H^{(4)}(x + \theta y) + n^{-1/2}\, \Delta_{2F}(x, y)\Big\}\, dy$$

and we redefine $Y(x)$ to equal the supremum of values y such that $x+y, x-y \in \mathcal{I}$ and

$$y^2\, H''(x) + \tfrac{1}{12}\, y^4\, H^{(4)}(x+\theta y) + n^{-1/2}\, \Delta_{2F}(x,y) + n^{-1}\, (\log n)\, A_n \le 0\,,$$

where the random variable $A_n = O_p(1)$ does not depend on x or y. In deriving (5.17) we have used the fact that, by employing arguments leading to (5.5), it may be proved that

$$\sup_{x \in \mathcal{I}} Y(x) = O_p\Big\{ \big(n^{-1} \log n\big)^{1/7} \Big\}\,.$$

More analogously to (5.5) it may be shown that if $\eta > 0$ and $\mathcal{J} = \mathcal{J}(\eta)$ is the subset of $\mathcal{I}$ all of whose points are distant at least η from each x_i then, using the new definition of $Y(x)$,

$$\sup_{x \in \mathcal{J}} Y(x) = O_p\big\{ (n^{-1} \log n)^{1/3} \big\}\,.$$

Using this result and the arguments leading to (5.9) and (5.10) we may show that, if $T_2(\mathcal{J})$ denotes the contribution to T_2 from the integral over $x \in \mathcal{J}$, rather than $x \in \mathcal{I}$, then $T_2(\mathcal{J}) = T_3(\mathcal{J}) + o_p(n^{-1})$, where

$$T_3(\mathcal{J}) = -\int_{\mathcal{J}} dx \int_{-Y(x)}^{Y(x)} \min\left\{ 0\,,\; y^2\, H''(x) + n^{-1/2}\, \frac{\Delta_{3F}(x,y)}{1-F(x)} \right\} dy\,.$$

The methods leading to (5.12) give that $E\{T_3(\mathcal{J})\} = O(n^{-1})$. Therefore,

$$T_2(\mathcal{J}) = O_p\big(n^{-1}\big)\,. \tag{5.18}$$

Let $\eta > 0$ be less than half the minimum of $x_{i+1} - x_i$ over $0 \le i \le m$, where x_0 denotes the lower limit of $\mathcal{I}$ and x_{m+1} is the upper limit. Write $T_2(x_i, \eta)$ for the contribution to T_2 from the integral over $x_i - \eta < x < x_i + \eta$. Then

$$T_2(x_i,\eta) = -\int_{-\eta}^{\eta} du \int_{-Y(x_i+u)}^{Y(x_i+u)} \min\Big[0\,, \big(\tfrac{1}{2}\, u^2\, y^2 + \tfrac{1}{12}\, y^4\big)\, H^{(4)}\{x_i + \theta_i\, (|u| + |y|)\} + n^{-1/2}\, \Delta_{2F}(x_i, y) \Big]\, dy\,,$$

where $-1 \le \theta_i \le 1$. Changing variable from (u, y) to (v, z), where $u = n^{-1/7} v$ and $y = n^{-1/7} z$, we deduce that

$$T_2(x_i,\eta) = -n^{-6/7} \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} \min\Big\{ 0\,, \big(\tfrac{1}{2}\, v^2\, z^2 + \tfrac{1}{12}\, z^4\big)\, H^{(4)}(x_i) + g(x_i)\, V_i(v+z) \Big\}\, dz + o_p\big(n^{-6/7}\big)\,, \tag{5.19}$$

where V_i is a standard Brownian motion. The processes V_i, $1 \le i \le m$, may be taken to be independent, without violating (5.19). Theorem 3.3 now follows on combining (5.18) and (5.19).

5.4. Reasons for failure of bootstrap version of Theorem 3.3. In order for $\widetilde{H}^{(4)}$ to consistently estimate $H^{(4)}$ it is necessary that the bandwidth h used to construct $\widetilde{F}$ be of larger order than $n^{-1/7}$. For simplicity we shall assume below that h is at least of size $n^{\xi-(1/7)}$ for some $\xi > 0$, although our argument may by pursued to an unaltered conclusion when the increase of h over $n^{-1/7}$ is by only a logarithmic factor.

Put $c = \frac{1}{2} \int u^2 K(u)\, du$. Observe that with probability 1 and for each $\eta > 0$, $\widetilde{F}'' = F'' + ch^2 F^{(4)} + O\{(nh^3)^{\eta-(1/2)}\} + o(h^2)$, uniformly in $x \in \mathcal{I}'$. (Here we have used the fact that $h \ge n^{\xi-(1/7)}$.) It follows that $\widetilde{H}'' = D^2 A(F + ch^2 F'') + O\{(nh^3)^{\eta-(1/2)}\} + o(h^2)$, uniformly in $x \in \mathcal{I}'$, where $A(u) = -\log(1-u)$ and D is the differential operator. Now, $D^2 A(F + ch^2 F'') = D^2 A(F) + ch^2 D^2\{F''\, A'(F)\} + o(h^2)$, and $D^2\{F''\, A'(F)\} = D^2\{D^2 A(F) - (F')^2\, A''(F)\} = D^2\{H'' - (H')^2\}$. Therefore, $D^2\{F''\, A'(F)\} = H^{(4)} - 2\,\{H' H''' + (H'')^2\}$. Hence, with probability 1,

$$\widetilde{H}'' = H'' + ch^2 \left[H^{(4)} - 2\,\{H' H''' + (H'')^2\}\right] + O\big\{(nh^3)^{\eta-(1/2)}\big\} + o\big(h^2\big)\,, \quad (5.20)$$

uniformly in $x \in \mathcal{I}'$.

The term of order $(nh^3)^{\eta-(1/2)}$ on the right-hand side of (5.20) is of course the result of stochastic error, and performance would only improve if it could be dropped. Let us assume this can be done. Then we can estimate $H''(x)$ with error equal to

$$ch^2 \left[H^{(4)}(x) - 2\,\{H'(x)\, H'''(x) + H''(x)^2\}\right] + o\big(h^2\big)\,. \quad (5.21)$$

Now, the limiting distribution of T under H_{02} is determined by properties of H in arbitrarily small neighbourhoods of the points x_i, and so it is there that we are most interested in properties of $\widetilde{H}''$. If x is in a decreasingly small neighbourhood of x_i, the expansion at (5.21) equals

$$ch^2 \left[H^{(4)}(x_i) - 2\,\{H'(x_i)\, H'''(x_i) + H''(x_i)^2\}\right] + o\big(h^2\big) = ch^2\, H^{(4)}(x_i) + o\big(h^2\big)\,,$$

the second identity holding since $H''(x_i) = H'''(x_i) = 0$. Therefore, if we ignore stochastic fluctuations (which are asymptotically equally likely to increase or decrease the value of $\widetilde{H}''$), $\widetilde{H}''(x)$ is distant at least order h^2 strictly above zero when x is in the neighbourhood of x_i. Since h is at least of order $n^{-1/7}$ then the distance

of $\widetilde{H}''$ above zero, in the neighbourhood of x_i, is (with probability at least $\frac{1}{2}+o(1)$) no less than a certain fixed constant multiple of $n^{-2/7}$; call this result (R).

Let $\widehat{H}^*$ denote the bootstrap version of $\widehat{H}$, and recall from the proof of Theorem 3.2 that that limit result derives entirely from fluctuations of $\widehat{H}(x,y)$ below zero when x is close to x_i and y is near zero. Under H_{02} these fluctuations occur with a probability that is bounded away from zero as n increases. The perturbations of $\widehat{H}^* - \widetilde{H}$ are of order only $n^{-1/2}$, and in particular are of strictly smaller order than $n^{-2/7}$. This property, and result (R), imply that the probability that the empirical fluctuations of $\widehat{H}^*$, near $x_1,\ldots,x_m$, ever protrude below zero converges to zero as $n\to\infty$. In consequence the limit results described by Theorem 3.2 do not apply in the bootstrap setting.

5.5. Proof of Theorem 3.4. Observe that

$$\widetilde{H}(x+y)+\widetilde{H}(x-y)-2\,\widetilde{H}(x)=y^2\int_0^1\big\{\widetilde{H}''(x+ty)+\widetilde{H}''(x-ty)\big\}\,(1-t)\,dt\,. \quad (5.21)$$

Let H_G denote the version of H that arises if F is replaced by a distribution G, and note that, by (2.3) and approximations based, for example, on the Komlós, Major and Tusnády (1975) embedding,

$$\widetilde{H}''=H''_{E(\widetilde{F})}+(1-F)^{-1}\,(\tilde f'-E\tilde f')+O_p\big\{(nh)^{\eta-(1/2)}\big\}\,, \quad (5.22)$$

uniformly in $x\in\mathcal{I}'$ and in $h\in\mathcal{H}$, for each $\eta>0$. The argument in section 5.4 (see particularly (5.20)) shows that

$$H''_{E(\widetilde{F})}=H''+ch^2\left[H^{(4)}-2\left\{H'H'''+(H'')^2\right\}\right]+o\big(h^2\big)$$

uniformly on $\mathcal{I}'$ and in $h\in\mathcal{H}$. Therefore,

$$\begin{aligned}\int_0^1&\big\{H''_{E(\widetilde{F})}(x+ty)+H''_{E(\widetilde{F})}(x-ty)\big\}\,(1-t)\,dt\\&=\int_0^1\big\{H''(x+ty)+H''(x-ty)\big\}\,(1-t)\,dt+c\,h^2\,H^{(4)}(x_1)+o\big(h^2\big)\\&=H''(x)+\big(\tfrac{1}{12}\,y^2+c\,h^2\big)\,H^{(4)}(x_1)+o\big(h^2+y^2\big)\,, \quad (5.23)\end{aligned}$$

uniformly in $h\in\mathcal{H}$, $|x-x_1|\le\delta(n)$ and $|y|\le\delta(n)$ for any sequence $\delta(n)\downarrow 0$.

Furthermore,

$$\begin{aligned}\tilde f'(x)-E\tilde f'(x)&=h^{-2}\int K''(u)\,\big\{\widehat{F}(x-hu)-F(x-hu)\big\}\,du\\&=h^{-2}\,n^{-1/2}\int K''(u)\,\big[W\{F(x-hu)\}-W\{F(x)\}\big]\,du\\&\qquad+O_p\big\{h^{-1}\,n^{-1/2}\,(\log n)^{1/2}\big\}\,,\end{aligned}$$

uniformly in $h \in \mathcal{H}$ and $x \in \mathcal{I}'$, where W is a standard Brownian motion. Put $h = n^{-1/7}q$, $x = x_1 + n^{-1/7}s + ty$ and $y = n^{-1/7}z$, and recall that $g = f^{1/2}/(1-F)$. Then there exists a standard Brownian motion V such that

$$\begin{aligned} h^{-2}\, n^{-1/2}\, \{1 - F(x)\}^{-1} \left[W\{F(x - hu)\} - W\{F(x)\} \right] \\ = n^{-2/7}\, q^{-2}\, g(x_1)\, \{V(s + tz - qu) - V(s + tz)\} \\ + O_p\big\{ n^{-5/14}\, (\log n)^{1/2}\, (|q| + |s| + |z|) \big\}\,, \end{aligned}$$

uniformly in $0 \le t \le 1$, $|u| \le C$ for any $C > 0$, and q, s, z such that $n^{-1/7}q \in \mathcal{H}$, $|s| \le n^{1/7}\, \delta(n)$ and $|z| \le n^{1/7}\, \delta(n)$. Therefore, defining $M = (1-F)^{-1}\, (\tilde{f}' - E\tilde{f}')$, we have:

$$\begin{aligned} &\int_0^1 \{M(x + ty) + M(x - ty)\}\, (1 - t)\, dt \\ &= n^{-2/7}\, q^{-2}\, g(x_1) \int K''(u)\, du \int_0^1 \{V(s + tz - qu) + V(s - tz - qu)\}\, (1 - t)\, dt \\ &\quad + O_p\big\{ n^{-5/14}\, (\log n)^{1/2}\, \big(|q| + |q|^{-1} + |s| + |z|\big) \big\} \end{aligned} \tag{5.24}$$

uniformly in $n^{-1/7}q \in \mathcal{H}$, $|s| \le n^{1/7}\, \delta(n)$ and $|z| \le n^{1/7}\, \delta(n)$.

Combining (5.21)–(5.24), and taking $x = x_1 + n^{-1/7}s$ and $y = n^{-1/7}z$, we deduce that

$$\begin{aligned} n^{4/7} \left\{ \widetilde{H}(x + y) + \widetilde{H}(x - y) - 2\, \widetilde{H}(x) \right\} \\ = z^2 \left(c\, q^2 + \tfrac{1}{2}\, s^2 + \tfrac{1}{12}\, z^2 \right) H^{(4)}(x_1) \\ + z^2\, q^{-2}\, g(x_1) \int K''(u)\, du \int_0^1 \{V(s + tz - qu) \\ + V(s - tz - qu)\}\, (1 - t)\, dt \\ + O_p\big\{ n^{-1/14}\, (\log n)^{1/2}\, z^2\, \big(|q| + |q|^{-1} + |s| + |z|\big) \big\} \\ + o_p\big\{ z^2 \left(q^2 + s^2 + z^2 \right) \big\}\,, \end{aligned} \tag{5.25}$$

uniformly in $n^{-1/7}q \in \mathcal{H}$, $|s| \le n^{1/7}\, \delta(n)$ and $|z| \le n^{1/7}\, \delta(n)$. The theorem follows from (5.25).

5.6. Proof of Theorem 3.5. Dividing both sides of (5.25) by z^2 and letting $z \to 0$ we deduce that when $h = \hat{h}_{\rm crit}$, $n^{2/7} \widetilde{H}''(x_1 + n^{-1/7}s) = S(s) + o_p(1)$, uniformly in $|s| \le n^{1/7}\, \delta(n)$, where $S(s)$ is defined as at (3.8). Thus, the bootstrap calibration step involves sampling from a distribution whose cumulative hazard rate $\bar{H}$ is convex on $\mathcal{I}$ and satisfies $\bar{H}''(x) > 0$ for all $x \in \mathcal{I}$, excepting a single point x which may be expressed as $x = x_1 + n^{-1/7}A + o_p(n^{-1/7})$, where A is uniquely defined by $S(A) = 0$.

At this point, $\bar{H}''$ vanishes. Reworking the proof of Theorem 3.3 we deduce that the α-level critical point $\hat{c}(\alpha)$ of T^*, defined conditional on the data $\mathcal{X}$, equals $n^{-6/7}\,\Gamma_\alpha\{S''(A), g(x_1)\} + o_p(n^{-6/7})$. (Here, T^* denotes the value of T computed from an n-sample drawn from the distribution $\widetilde{F}(\cdot|\hat{h}_{\rm crit})$.)

The distribution of T may be represented, in asymptotic form, as before, and in terms of the same Brownian motion B that was used to construct the representation for $\widetilde{H}$ at (5.25). In particular, the Brownian motion V_1 appearing at (5.19) (when $i = 1$) may be taken identical to the process V at (5.25). Letting B denote the common process, we see that the inequality $T \le \hat{c}(\alpha)$ may equivalently be written as

$$n^{-6/7}\,Z_1 + o_p\big(n^{-6/7}\big) \le n^{-6/7}\,\Gamma_\alpha\big\{S''(A), g(x_1)\big\} + o_p\big(n^{-6/7}\big)\,, \tag{5.26}$$

where Z_1 is defined by (3.5) with $i = 1$. Theorem 3.5 follows from (5.26).

5.7. Proof of Theorem 3.7. If the hazard rate H' is not increasing on $\mathcal{I}$ then for some $\epsilon > 0$ there exists a nondegenerate rectangle $\mathcal{R}$ such that for all $(x, y) \in \mathcal{R}$, both $x + y$ and $x - y$ lie in $\mathcal{I}$ and $H(x, y) \le -\epsilon$. Under the hypotheses of the theorem, $\widehat{H}(x, y) = H(x, y) + o_p(1)$ uniformly in $(x, y) \in \mathcal{R}$, and so $T \ge \epsilon\,|\mathcal{R}| + o_p(1)$, where $|\mathcal{R}|$ denotes the area of $\mathcal{R}$. Therefore, the theorem will follow if we prove that for each $\alpha \in (0, 1)$ the α-level critical point $\hat{c}(\alpha)$, derived using the bootstrap argument in section 2.2, satisfies

$$P\{\hat{c}(\alpha) > \eta\} \to 0 \quad \text{for each } \eta > 0\,. \tag{5.27}$$

As $h \to \infty$, $E\{\tilde{f}(x|h)\} = h^{-1}\,K(0) + o(h^{-1})$ and $E|\tilde{f}'(x|h)| = h^{-3}\,K''(0)\,E|x - X| + o(h^{-3})$, uniformly in $x \in \mathcal{I}$. Hence, there exists $h_0 > 0$ such that $\{E\tilde{f}(x|h_0)\}^2 \ge 2\,E|\tilde{f}'(x|h_0)|$ for all $x \in \mathcal{I}$. It may be proved from this property that with probability converging to 1, $\tilde{f}(x|h_0)^2 \ge |\tilde{f}'(x|h_0)|$ for all $x \in \mathcal{I}$. Therefore, by (2.3), the probability that $\widetilde{H}''(x|h_0) \ge 0$ for all $x \in \mathcal{I}$ converges to 1 as $n \to \infty$, and so

$$P(\hat{h}_{\rm crit} \le h_0) \to 1\,. \tag{5.28}$$

It can be proved that, under the conditions of the theorem, for each $h_1 > 0$ there exists $\epsilon(h_1) > 0$ such that for all sufficiently large n,

$$1 - E\{\widetilde{F}(x|h)\} \ge \epsilon(h_1) \text{ for all } x \in \mathcal{I} \text{ and all } h \in (0, h_1]\,,$$

and moreover that for each $h_1 > 0$,

$$\big|\widetilde{F}(x|h) - E\{\widetilde{F}(x|h)\}\big| = o_p(1) \text{ uniformly in } x \in \mathcal{I} \text{ and } h \in (0, h_1]\,.$$

Therefore,

$$P\Big\{\, 1 - \widetilde{F}(x|h) \ge \tfrac{1}{2}\,\epsilon(h_1) \text{ for all } x \in \mathcal{I} \text{ and all } h \in (0, h_1] \,\Big\} \to 1\,.$$

This result and (5.28) imply that

$$P\Big\{\, 1 - \widetilde{F}(x|\hat{h}_{\rm crit}) \ge \tfrac{1}{2}\,\epsilon(h_0) \text{ for all } x \in \mathcal{I} \,\Big\} \to 1\,. \tag{5.29}$$

If $\widehat{F}^*_h$ denotes the bootstrap version of $\widehat{F}$, computed from an n-sample drawn from the distribution $\widetilde{F}(\cdot|h)$ rather than from F, then for all $\lambda > 0$,

$$\sup_{x\in\mathcal{I}} \sup_{h\in(0,h_1]} E\big\{|\widehat{F}^*_h(x) - \widetilde{F}(x|h)|^\lambda\big\} = O\big(n^{-\lambda/2}\big)$$

for all $\lambda > 0$. We may deduce from this result, using a "continuity argument" (see Cheng and Hall, 2001), that for any $C > 0$, if $\tilde{h}$ is a random element of the interval $[Cn^{-1/5}, h_1]$,

$$E\Big\{\, \sup_{x\in\mathcal{I}} \big|\widehat{F}^*_{\tilde{h}}(x) - \widetilde{F}(x|\tilde{h})\big| \,\Big\} = o(1)\,. \tag{5.30}$$

Write simply $\widehat{F}^*$ for $\widehat{F}^*_{\hat{h}_{\rm crit}}$, and put $\widehat{H}^* = -\log(1-\widehat{F}^*)$ and $\widetilde{H} = -\log(1-\widetilde{F})$. Taking $h_1 = h_0$ and $\tilde{h} = \hat{h}_{\rm crit}$, which in view of (5.28) and the assumptions in the theorem satisfies $P(Cn^{-1/5} \le \hat{h}_{\rm crit} \le h_0) \to 1$ for some $C > 0$, we deduce from (5.30) that $|\widehat{F}^* - \widetilde{F}(\cdot|\hat{h}_{\rm crit})| = o_p(1)$ uniformly on $\mathcal{I}$. From this result and (5.29) we see that

$$\sup_{x\in\mathcal{I}} \big|\widehat{H}^*(x) - \widetilde{H}(x|\hat{h}_{\rm crit})\big| = R^*_1\,, \tag{5.31}$$

where, here and below, R^*_j denotes a random variable that is defined through Monte Carlo simulation conditional on $\mathcal{X}$, and satisfies $P(|R^*_j| > \eta) \to 0$ for each $\eta > 0$, where the probability is defined in the unconditional sense.

If T^* denotes the bootstrap version of T then

$$\begin{aligned} T^* &= \iint_{x,y\,:\,x+y,\,x-y\in\mathcal{I}} \max\big\{0, 2\,\widehat{H}^*(x) - \widehat{H}^*(x+y) - \widehat{H}^*(x-y)\big\}\,dx\,dy \\ &= \iint_{x,y\,:\,x+y,\,x-y\in\mathcal{I}} \max\big\{0, 2\,\widetilde{H}(x|\hat{h}_{\rm crit}) - \widetilde{H}(x+y|\hat{h}_{\rm crit}) \\ &\qquad\qquad - \widetilde{H}(x-y|\hat{h}_{\rm crit})\big\}\,dx\,dy + R^*_2 = R^*_2\,, \end{aligned}$$

where the second identity follows from (5.31) and the third from the fact that, by the definition of $\hat{h}_{\rm crit}$, $\widetilde{H}(\cdot|\hat{h}_{\rm crit})$ is convex on $\mathcal{I}$. Therefore, $P(T^* > \eta) \to 0$ for each $\eta > 0$. Hence, since $\hat{c}(\alpha)$ is defined by $P\{T^* > \hat{c}(\alpha)|\mathcal{X}\} = \alpha$, then (5.27) must hold.

REFERENCES

AHMAD, I.A. (1975). A nonparametric test for the monotonicity of a failure rate function. *Comm. Statist. A—Theory Meth.* **4**, 967–974.

ANDERSON, N.H., HALL, P. AND TITTERINGTON, D.M. (1998). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50**, 41–54.

BARLOW, R.E. AND DOKSUM, K.A. (1972). Isotonic tests for convex orderings. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* (Berkeley, Calif., 1970/1971), Vol. I: Theory of Statistics, pp. 293–323. Univ. California Press, Berkeley, Calif.

BICKEL, P.J. (1969). Tests for monotone failure rate. II. *Ann. Math. Statist.* **40** 1250–1260.

BICKEL, P.J. AND DOKSUM, K.A. (1969). Tests for monotone failure rate based on normalized spacings. *Ann. Math. Statist.* **40**, 1216–1235.

BOWMAN, A.W., JONES, M.C. AND GIJBELS, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7**, 489–500.

CHENG, M.-Y. AND HALL, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27**, 1294–1315.

CHENG, M.-Y. AND HALL, P. (2001). Methods for tracking support boundaries with corners. *Ann. Statist.*, under revision.

CHUNG, K.L. (1974). *A Course in Probability Theory*, 2nd edn. Academic Press, New York.

DELECROIX, M., HALL, P. AND ROGET, C. (2001). Powerful tests for single index models. Manuscript.

GAIL, M.H. AND GASTWIRTH, J.L. (1978a). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *J. Roy. Statist. Soc.* Ser. B **40**, 350–357.

GAIL, M.H. AND GASTWIRTH, J.L. (1978b). A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve. *J. Amer. Statist. Assoc.* **73**, 787–793.

GIJBELS, I. AND HECKMAN, N. (2000). Nonparametric testing for a monotone hazard function via normalized spacings. Manuscript.

HALL, P., HUANG, L.-S., GIFFORD, J.A. AND GIJBELS, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *J. Comput. Graph. Statist.* **10**, 592–614.

HARTIGAN, J.A. AND HARTIGAN, P.M. (1985). The dip test of unimodality. *Ann. Statist.* **13**, 70–84.

KLEFSJÖ, B. (1983). Some tests against aging based on the total time on test

transform. *Comm. Statist. A—Theory Meth.* **12**, 907–927.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's, and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 111–131.

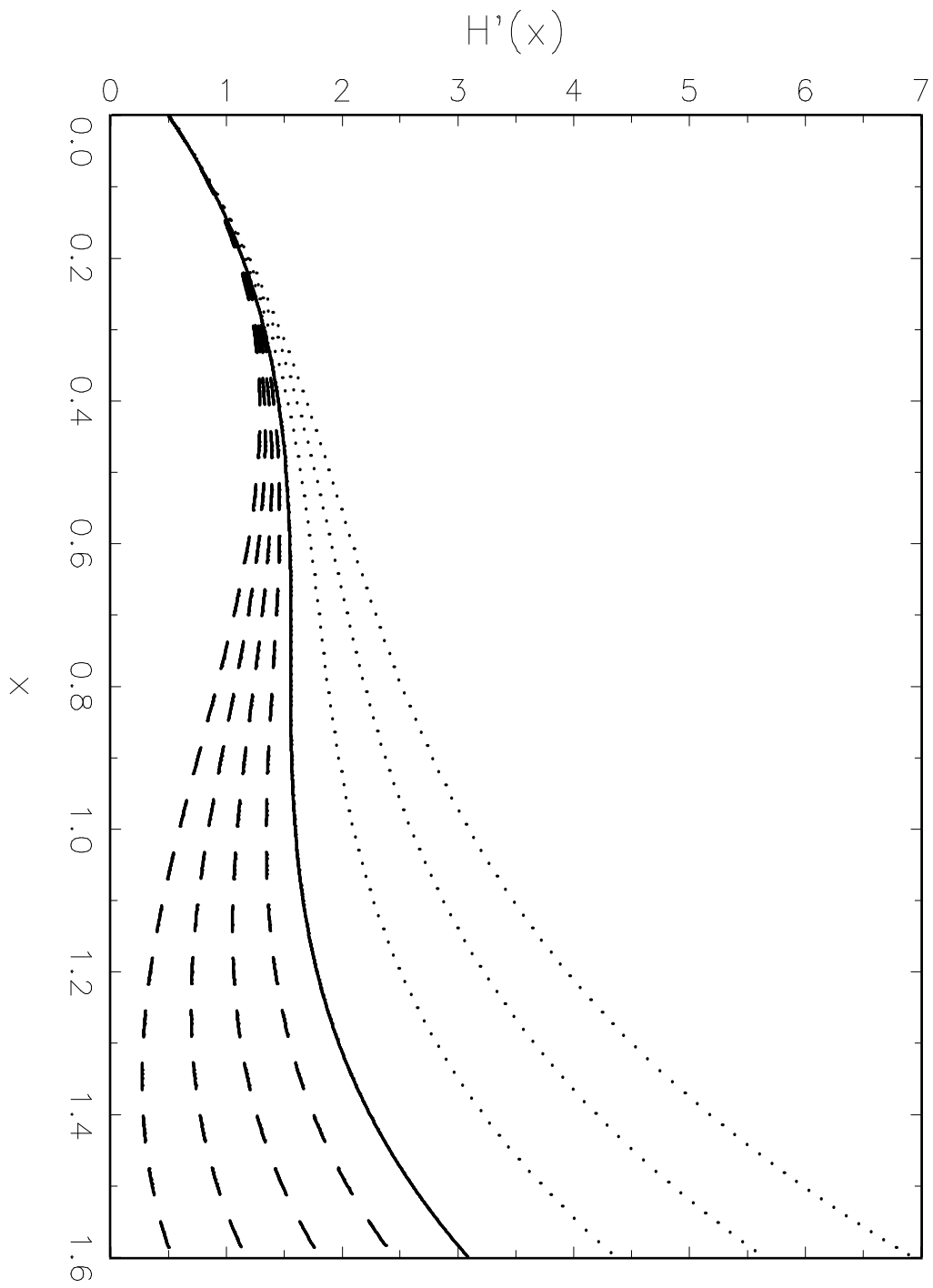
LAVERGNE, P. AND VUONG, Q. (2000). Nonparametric significance testing. *Econometric Theory* **16**, 576–601.

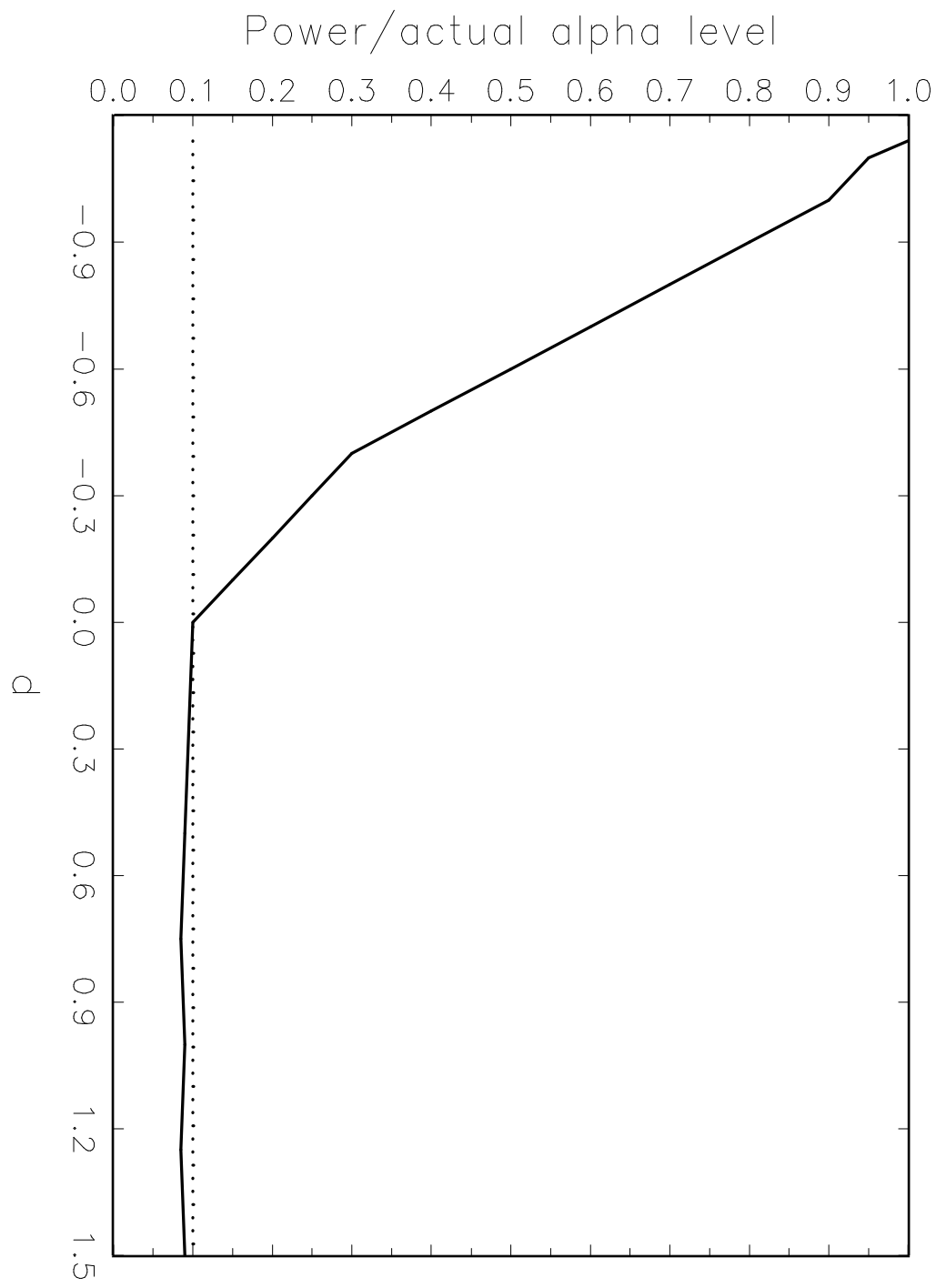
PROSCHAN, F. AND PYKE, R. (1967). Tests for monotone failure rate. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* (Berkeley, Calif., 1965/66), Vol. III: Physical Sciences, pp. 293–312 Univ. California Press, Berkeley, Calif.

SILVERMAN, B.W. (1981). Using kernel density estimates to investigate multimodality. *J. Royal Statist. Society, Series B*, **43**, 97–9.

Caption for Figure 1: *Graph of* $H'(x)$ *for* $a = 2.5, b = 0.75, c = 0.5$ *and* $d = -1, -0.75, -0.50$ *and* -0.25 *(dashed curve),* $d = 0$ *(full curve) and* $d = 0.5, 1$ *and* 1.5 *(dotted curve)*

Caption for Figure 2: *Power/actual alpha level for* $a = 2.5, b = 0.75, c = 0.5$*, for a range of values for* d *and for* $\alpha = 0.10$*.*


H'(x)
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
X


Power/actual alpha level
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ρ
−0.9 −0.6 −0.3 0.0 0.3 0.6 0.9 1.2 1.5