Nishiyama, Yoshihiko; Robinson, Peter M.

Working Paper

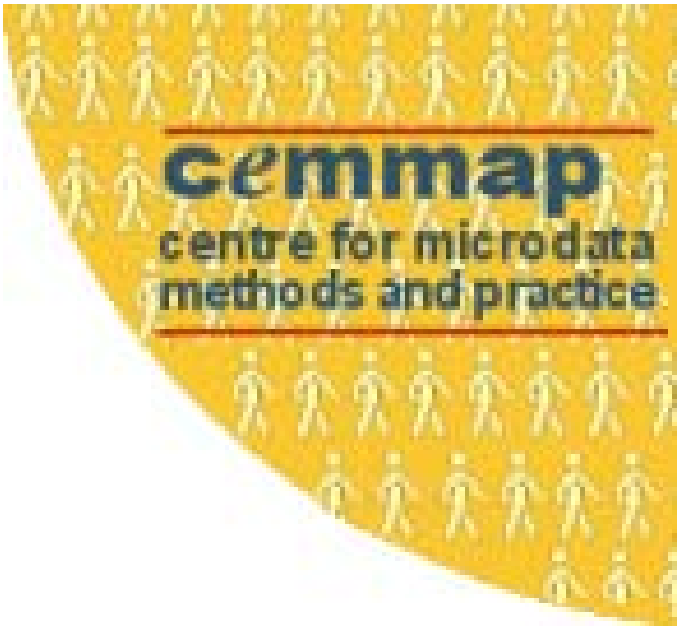
The bootstrap and the edgeworth correction for semiparametric averaged derivatives

cemmap working paper, No. CWP12/04

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Nishiyama, Yoshihiko; Robinson, Peter M. (2004) : The bootstrap and the edgeworth correction for semiparametric averaged derivatives, cemmap working paper, No. CWP12/04, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2004.1204

This Version is available at:
https://hdl.handle.net/10419/79262



THE BOOTSTRAP AND THE EDGEWORTH CORRECTION FOR SEMIPARAMETRIC AVERAGED DERIVATIVES

Y. Nishiyama
P.M. Robinson

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP12/04

The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives[1]

Y. Nishiyama
Institute of Economic Research,
Kyoto University
Sakyo, Kyoto, 606-8501, Japan

P.M. Robinson
Department of Economics,
London School of Economics
Houghton Street, London WC2A 2AE, UK

Abstract

In a number of semiparametric models, smoothing seems necessary in order to obtain estimates of the parametric component which are asymptotically normal and converge at parametric rate. However, smoothing can inflate the error in the normal approximation, so that refined approximations are of interest, especially in sample sizes that are not enormous. We show that a bootstrap distribution achieves a valid Edgeworth correction in case of density-weighted averaged derivative estimates of semiparametric index models. Approaches to bias-reduction are discussed. We also develop a higher order expansion, to show that the bootstrap achieves a further reduction in size distortion in case of two-sided testing. The finite sample performance of the methods is investigated by means of Monte Carlo simulations from a Tobit model.

Keywords: Bootstrap; Edgeworth correction; Semiparametric averaged derivatives
JEL Classification: C23

---

[1] We are grateful to Joel Horowitz and three referees for comments that have led to considerable improvements. The first author's research was supported by Japan Society for the Promotion of Science Grant 12730021. The second author's research was supported by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000238212.

1. Introduction.

Semiparametric estimates involving smoothing have become established as an important tool in the analysis of econometric data. On the one hand, these estimates avoid some stringent assumptions, largely unsupported by economic theory, that are entailed in a fully parametric analysis. On the other, they typically enjoy the same convergence rate as estimates in parametric models, being $n^{1/2}$-consistent, where n is sample size, and satisfying a central limit theorem which is conveniently usable in statistical inference. However, the semiparametric parameter estimates may have finite-sample properties that are inferior to those of estimates (such as maximum likelihood ones) of a fully specified parametric model.

Detailed evidence concerning this issue has emerged in case of density-weighted averaged derivative estimates (ADE) of semiparametric single-index models. These had in the first place been shown to possess the basic $n^{1/2}$-consistency and asymptotic normality properties referred to above by Powell, Stock and Stoker (1989) in case of independent observations, and by Robinson (1989) in case of weakly dependent observations. Robinson (1995) (hereafter R) established that a refinement that is classic in the case of fully parametric statistics (such as the sample mean) may also obtain, namely a Berry-Esseen bound of order $n^{-1/2}$, meaning that the difference between the actual distribution function of the ADE, after centering at its probability limit and norming by $n^{1/2}$, and its normal limit, can be of order $n^{-1/2}$ uniformly. This bound can achieve the $n^{-1/2}$ rate typical for parametric statistics, but only if the bandwidth, h, employed in the averaged derivative estimation is chosen to lie within a restricted window, and then subject to the use of a kernel of suitably high order; otherwise, the normal approximation error order depends on h and exceeds $n^{-1/2}$. In particular, R found its error dominated by three additive components, one of order $n^{-1/2}$, one that decreases as $h \to 0$ (a 'bias' term) and one that increases as $h \to 0$ (a 'variance' term). His result was sharp, in that Nishiyama and Robinson (2000) (hereafter NR0) and Nishiyama and Robinson (2001) (hereafter NR1) found that a valid Edgeworth approximation to the distribution of studentized and unstudentized ADE has three, leading, correction terms of precisely these orders of magnitude. NR0, NR1 evaluated these terms and showed how

they can be consistently estimated, in order to provide feasible, improved inference, including more accurate confidence intervals. They also employed their Edgeworth approximation to derive a choice of h that is optimal in the sense of minimizing the error in the normal approximation, and thus seems more relevant in the context of statistical inference than choices based on minimizing mean squared error.

The Edgeworth correction terms of NR0, NR1 are, however, somewhat complicated to estimate and there is always concern that imprecision in the estimation of such terms (which essentially involve higher moments) might frustrate the goal of improved finite-sample properties. Moreover, Edgeworth approximations to the distribution function need not be monotone.

The present paper shows that a bootstrap achieves the same improvement. In general, given independent observations Z_i, $i = 1,\ldots,n$, from a population with distribution function Ψ, suppose one is interested in conducting inference on the unknown (scalar) parameter $\mu = \mu(\Psi)$, using the (pivotal) statistic $T_n = T(Z_1, \cdots, Z_n; \Psi)$. Let Ψ_n be the empirical distribution function of $Z_1, \cdots, Z_n$, and $Z_1^*, \cdots, Z_n^*$ a random sample (with replacement) of size n from Ψ_n. The bootstrap approach (see Efron, 1979) is based on a belief that the distribution of $T_n^* = T(Z_1^*, \cdots, Z_n^*; \Psi_n)$, across replications $Z_1^*, \cdots, Z_n^*$ from Ψ_n, approximates that of T_n. In particular this has been shown for the case that T_n is asymptotically standard normal in many settings, see e.g. Bickel and Freedman (1981), Freedman (1981, 1984) and Arcones and Giné (1992, 1994). For many statistics, the construction of a bootstrap distribution by resampling is unnecessary for the sake of conducting first-order asymptotic inference, because the ordinary central limit theorem for T_n itself is convenient to use. However, the bootstrap is still widely used in such circumstances, because it has the reputation for providing greater finite-sample accuracy. Singh (1981) provided theoretical explanation for this numerical experience in case T_n is the centred and normalized sample mean: the Edgeworth expansions of T_n and T_n^* have the same leading term, of order $n^{-1/2}$, so their distributions differ by $o(n^{-1/2})$. Equivalently, the distribution of T_n^* achieves the Edgeworth correction of T_n. For

extensions to other models, see e.g. Babu and Singh (1989), Hall (1986, 1991, 1992), Qumsiyeh (1986), Bhattacharya and Qumsiyeh (1989), Strawderman and Wells (1997). Technically closest to our present work is Helmers (1991) (see also Maesono, 1997), who considered standard U-statistics. Our (2.1) below is a U-statistic but with a "kernel" (in the U-statistic sense) that, unlike Helmers', not only depends on n but whose second moment can diverge as $n \to \infty$. Thus, while we borrow some of Helmers' techniques, our situation is substantially more complicated than his and requires a substantially different treatment.

The possibility of Edgeworth correction by bootstrapping for semiparametric estimates involving smoothing was pointed out by R and by Horowitz (1997). Horowitz (1998, 2002) was first to rigorously demonstrate an Edgeworth correction property here, in connection with symmetric t and χ^2 tests. Horowitz (1998) considered smoothed least absolute deviations (LAD) estimates of a median-regression model. The original LAD estimates require no smoothing in order to be $n^{-1/2}$- consistent, but their higher-order properties are difficult to analyze. The smoothed estimates have the same first-order limit distribution, and Horowitz's (1998) motivation for introducing smoothing was to achieve more tractable higher-order asymptotics (though smoothing is in any case needed to form t-ratios of unsmoothed LAD estimates). Horowitz (2002) established similar properties for smoothed maximum score (MS) estimates, but the background to this work was rather different. Horowitz (1992) had introduced a smoothed MS estimate on the basis of having a simpler (normal) limit distribution and faster convergence rate than the original MS estimate of Manski (1985).

The motivation for smoothing in ADE is more fundamental: it is essential for consistency in a semiparametric setting and influences the first-order asymptotic distribution, even though the rate of convergence here is parametric. In the following section we establish a valid Edgeworth expansion of the bootstrap distribution for the ADE, indicating, by comparison with NR0, NR1, that an Edgeworth correction property is achieved. Section 3 discusses rival methods of bias-correction. Our results show that in one-sided testing the bootstrap approximation incurs less size distortion than the normal one. To demonstrate a

further improved outcome in two-sided testing we extend the Edgeworth expansion so as to include extra terms. This is carried out in Section 4, but under an additional condition that is unprimitive, in view of the greater complexity of detail. An optimal method of bandwidth selection is also introduced there. Section 5 examines finite sample performance of the methods via Monte Carlo simulation. Proof details are left to appendices.

2. Valid first order Edgeworth expansion

For a $d \times 1$ variate X with probability density function $f(x)$, and a scalar variate Y, suppose the regression function $g(X) = E(Y|X)$ is known to have single index form

$$g(X) = G(\beta^\tau X) \quad ,$$

for some differentiable function G: $\Re \to \Re$ and some unknown column vector β, τ denoting transposition. Suppose f and G are nonparametric functions. Given a sample (X_i^τ, Y_i), $i = 1, \cdots, n$, of independent observations on (X^τ, Y), the ADE is

$$U = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U_{ij} \quad , \tag{2.1}$$

where

$$U_{ij} = K_{ij}' (Y_i - Y_j) h^{-d-1} \quad , \quad K_{ij}' = K'\left(\frac{X_i - X_j}{h}\right) \quad ,$$

such that $K'(u) = (\partial / \partial u) K(u)$ where K: $\Re^d \to \Re$ is a differentiable kernel function that integrates to 1, and $h = h_n$ is a positive bandwidth sequence.

Consider a bootstrap sample $(X_i^{*\tau}, Y_i^*)$, $i = 1, \cdots, n$, drawn with replacement from (X_i^τ, Y_i), $i = 1, \cdots, n$. Our bootstrapped ADE is

$$U^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U_{ij}^* \quad ,$$

where

$$U_{ij}^* = K_{ij}^{*\prime} (Y_i^* - Y_j^*) h^{-d-1} \quad , \quad K_{ij}^{*\prime} = K'\left(\frac{X_i^* - X_j^*}{h}\right) \quad .$$

As in R, NR0 and NR1, we focus, for simplicity, on inference on a single linear combination of $\bar{\theta} \equiv -E[g'(X) f(X)] = -\beta E[G'(\beta^\tau X) f(X)]$, namely $\bar{\mu} = \nu^\tau \bar{\theta}$, where ν is a given $d \times 1$ vector. Bearing in mind (2.1), we thus cover the familiar problem of testing whether a single element of β is zero. We estimate $\bar{\mu}$ by $u = \nu^\tau U$, and write $u^* = \nu^\tau U^*$. (Note here and elsewhere changes in notation from NR0, NR1.)

The estimated variance of u is $\hat{\sigma}^2 = \nu^\tau \hat{\Sigma} \nu$, where

$$\hat{\Sigma} = \frac{4}{(n-1)(n-2)^2} \sum_{i=1}^{n} \left\{ \sum_{j \neq i}^{n} (U_{ij} - U) \right\} \left\{ \sum_{j \neq i}^{n} (U_{ij} - U)^\tau \right\} .$$

Its bootstrapped version is $\hat{\sigma}^{*2} = \nu^\tau \hat{\Sigma}^* \nu$, where

$$\hat{\Sigma}^* = \frac{4}{(n-1)(n-2)^2} \sum_{i=1}^{n} \left\{ \sum_{j \neq i}^{n} (U_{ij}^* - U^*) \right\} \left\{ \sum_{j \neq i}^{n} (U_{ij}^* - U^*)^\tau \right\} .$$

We center u^* at

$$\bar{\mu}^* = E^*(u^*) - \bar{b}^* , \tag{2.2}$$

where

$$\bar{b}^* = h^L \sigma^* \tilde{\kappa}_1$$

for $\sigma^{*2} = E^*[2\nu^\tau (U_1^* - E^* U^*)]^2$, $E^*(\cdot) = E(\cdot \mid (X_i^\tau, Y_i), i = 1, \cdots, n)$,

$$U_i^* = E^*(U_{ij}^* \mid X_i^*, Y_i^*) = \frac{1}{n} \sum_{j=1}^{n} h^{-d-1} K^{\prime}\left(\frac{X_i^* - X_j}{h}\right)(Y_i^* - Y_j) ,$$

$$\tilde{\kappa}_1 = \frac{2(-1)^L}{\hat{\sigma} L!} \sum_{\substack{0 \le l_1, \cdots, l_d \le L \\ l_1 + \cdots + l_d = L}} \cdots \sum \left\{ \int \prod_{i=1}^{d} x_i^{l_i} K(x)\, dx \right\} \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta^{(l_1, \cdots, l_d)} \nu^\tau \tilde{f}_i^{\prime}(X_i) \right\} Y_i ,$$

in which

$$\Delta^{(l_1, \cdots, l_d)} = \frac{\partial^{(l_1 + \cdots + l_d)}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} , \quad \tilde{f}_i(x) = \frac{1}{(n-1) c^d} \sum_{j \neq i} H\left(\frac{x - X_j}{c}\right) ,$$

L is the order of the kernel K (see assumption (viii) below), $H: \Re^d \to \Re$ is a differentiable kernel function and c is a positive bandwidth. The definition of $\tilde{\kappa}_1$ was discussed in NR0. Note that in (2.2)

$$E^*(U^*) = \frac{n-1}{n} U . \tag{2.3}$$

Denoting $P^*(\cdot) = P(\cdot \mid (X_i^\tau, Y_i), i = 1, \cdots, n)$, we compare the distribution functions

$$\hat{F}(z) = P(\hat{Z} \le z)$$

and

$$\hat{F}^*(z) = P^*(\hat{Z}^* \le z)$$

of, respectively, the centered and studentized ADE

$$\hat{Z} = n^{1/2} \hat{\sigma}^{-1} (u - \bar{\mu}) \tag{2.4}$$

and the centered and bias-corrected bootstrapped studentized ADE

$$\hat{Z}^* = n^{1/2} \hat{\sigma}^{*-1} (u^* - \bar{\mu}^*) . \tag{2.5}$$

Denote, for a function $k: \Re^d \to \Re$, $k = k(x)$, $k' = k'(x) = (\partial/\partial x)\, k(x)$, $k'' = k''(x) = (\partial^2/\partial x\, \partial x^\tau)\, k(x)$ and $k''' = k'''(x) = (\partial/\partial x^\tau)\, vec\,(k''(x))$. Define $\Sigma = 4E(\theta - \bar{\theta})(\theta - \bar{\theta})^\tau$, where $\theta = \theta(X, Y) = Yf' - e'$, $e = fg$ and $q = E(Y^2|X)$. The following assumptions are nearly identical to those of NR0, the only difference worth mentioning being the stronger moment conditions on Y in our present (i) and (vi).

(i) $E(Y^{16}) < \infty$.

(ii) Σ is finite and positive definite.

(iii) The distribution of X is absolutely continuous with respect to a σ-finite measure of the form μ_X, where μ_X is Lebesgue measure on $\Re^d$. The measure ν of (X^τ, Y) can be written $\nu = \mu_X \times \mu_Y$. (X_i^τ, Y_i) are iid observations on (X^τ, Y).

(iv) f is $(L+2)$-times differentiable, and f and its first $(L+2)$ derivatives are bounded for $2L > d+3$.

(v) g is $(L+2)$-times differentiable, and e and its first $(L+2)$ derivatives are bounded.

(vi) q is twice differentiable and q', q'', g', g'', g''', $E(Y^4|X)f$ and qf' are bounded.

(vii) f, gf, $g'f$ and qf vanish on the boundaries of their convex (possibly infinite) supports.

(viii) $K(x)$ is even, differentiable,

$$\int_{\Re^d} \{(1 + \|x\|^L)\,|K(x)| + \|K'(x)\|\}\, dx + \sup_{x \in \Re^d} \|K'(x)\| < \infty,$$

and

$$\int_{\Re^d} x_1^{l_1} \cdots x_d^{l_d} K(x)\, dx \begin{cases} = 1, & \text{if } l_1 + \cdots + l_d = 0 \\ = 0, & \text{if } 0 < l_1 + \cdots + l_d < L \\ \neq 0, & \text{if } l_1 + \cdots + l_d = L, \text{ some } l_1, \cdots, l_d. \end{cases}$$

(ix) $\dfrac{(\log n)^9}{nh^{d+3}} + nh^{2L} \to 0$ as $n \to \infty$.

(x) $\limsup_{|t| \to \infty} |E \exp[\{i\, t\, 2\nu^\tau(\theta - \bar{\theta})\}]| < 1$ for any vector ν satisfying $\nu^\tau \nu = 1$.

(xi) $H(x)$ is even and $(L+1)$-times differentiable,

$$\int_{\Re^d} H(x)\, dx = 1,$$

$$\int_{\Re^d} \|\Delta^{(l_1, \cdots, l_d)} H'(x)\|\, dx + \sup_{x \in \Re^d} \|\Delta^{(l_1, \cdots, l_d)} H'(x)\| < \infty$$

for any non-negative integers $l_1 \ldots l_d$ satisfying $0 \le l_1 + \cdots + l_d \le L$.

(xii) $c \to 0$ and $\dfrac{(\log n)^2}{nc^{d+2+2L}} = O(1)$ as $n \to \infty$.

Theorem 1 *Under assumptions* (i)-(xii), *as* $n \to \infty$,

$$\sup_{v:\, v^\tau v = 1} \sup_{z} |\hat{F}(z) - \hat{F}^*(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) \quad \text{a.s.} \tag{2.6}$$

In NR1 the valid Edgeworth expansion

$$\hat{F}(z) = F^+(z) + o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L)$$

is established, where

$$F^+(z) = \Phi(z) - \phi(z)\left[n^{1/2}h^L\kappa_1 - \frac{\kappa_2}{nh^{d+2}}z - \frac{4}{3n^{1/2}}\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}\right], \tag{2.7}$$

for

$$\kappa_1 = \frac{2(-1)^L}{\sigma L!} \sum_{\substack{0 \le l_1, \cdots, l_d \le L \\ l_1 + \cdots + l_d = L}} \cdots \sum \left\{ \int \prod_{i=1}^{d} x_i^{l_i} K(x)\, dx \right\} E[(\Delta^{(l_1, \cdots, l_d)} v^\tau f')g] ,$$

$$\kappa_2 = 2\sigma^{-2} \int \{v^\tau K'(x)\}^2 dx \; E(sf) ,$$

$$\kappa_3 = \sigma^{-3}E\left[(r - 3sg - g^3)(v^\tau f')^3 - 3s(v^\tau f')^2(v^\tau a) - (v^\tau a)^3 \right] ,$$

$$\kappa_4 = -\sigma^{-3}E[fs(v^\tau f')(v^\tau a' v) - f(v^\tau f')(v^\tau s')(v^\tau a)$$

$$- fs(v^\tau a)(v^\tau f'' v) + f(v^\tau g')(v^\tau a)^2] ,$$

in which $\sigma^2 = v^\tau \Sigma v$, $a = a(X) = g'f + \bar{\theta}$, $r = r(X) = E(Y^3|X)$ and $s = s(X) = Var(Y|X) = q - g^2$. It follows that the bootstrap distribution function $\hat{F}^*(z)$ approximates $\hat{F}(z)$ asymptotically as well as $F^+(z)$.

3. Discussion of Bias-Correction

The centering of u^* at $\bar{\mu}^*$, rather than at the more obvious u (see (2.2) and (2.3)) is explained as follows. From Lemma 11 of NR0, $Eu - \bar{\mu} = h^L\kappa_1(1 + o(1))$. The

bootstrap distribution of $n^{1/2}(u^* - u)$ thus will approximate only the distribution of the first term on the right of

$$n^{1/2}(u - \bar{\mu}) = n^{1/2}(u - Eu) + n^{1/2}h^L\kappa_1 + o(n^{1/2}h^L) \quad ,$$

since though $n^{1/2}h^L$ tends to zero it may do so too slowly to prevent the second term on the right from affecting higher-order properties. Thus we cannot rely fully on the bootstrap to achieve the Edgeworth correction property, but bias correct "manually". Of course this would not be necessary were h regarded as tending to zero sufficiently fast, or if a large enough kernel order L were chosen, indeed NR0 showed how the Edgeworth approximation simplifies for various choices of h and L.

The first possibility, "undersmoothing", was also employed by Horowitz (1998, 2002) and Hall (1991) in a higher-order context, in semiparametric and nonparametric estimation respectively. In our problem, undersmoothing seems acceptable only when d is small, since the "second order variance" is of order $n^{-1}h^{-d-2}$. But even with $d=2$, the Monte Carlo study of NR0 found undersmoothing to significantly inflate variance, indeed for $h=0.4$ it was twice that for $h=1.0$, independent of L. On the other hand the same study found that studentization can reduce variance inflation, in which case undersmoothing would be more acceptable in hypothesis testing than in point estimation.

The other approach to bias reduction, of choosing L larger than required in assumption (iv), also inflates the variance-related term κ_2 , in view of the factor $\int\{K'(x)\}^2dx$. To get some idea of the actual effect, we performed some numerical calculations, taking $K(x)$ to have the product form

$$K(x) = \prod_{i=1}^{d} k(x_i), \quad x=(x_1, \ldots, x_d)^\tau, \tag{3.1}$$

where $k: \Re \to \Re$ is given by

$$k(x) = p_{\frac{L}{2}-1}(x)\,\phi(x) \tag{3.2}$$

for even L, with no loss of generality, with $p_r(x)$ the polynomial

$$p_r(x) = \sum_{i=0}^{r} \zeta_i x^{2i} \quad , \tag{3.3}$$

and the ζ_i chosen to satisfy assumption (vii). Such kernels were considered in Robinson (1988), who proposed use of higher-order kernels to achieve $n^{1/2}$- consistency of semiparametric estimates in the context of a different model. For $L=2,4,6,8,10,12$ we found that $\pi\int\{K'(x)\}^2dx$ takes the values 0.443, 1.523, 3.075, 5.004, 7.304, 9.899, respectively, illustrating the relative increase in imprecision with L. Earlier, NR0 found in Monte Carlo that the variance of U doubles when going from $L=4$ to $L=10$, independent of h, when $n=100$. Therefore, as R pointed out, empirical researchers may not wish to use a very high order kernel unless n is very large indeed. Indeed even when $d=3$, the condition $L>2(d+2)$ discussed in R and NR0 already requires $L \geq 12$.

This discussion rather brings us back to the original bias-correction $\bar{b}^*$ in (2.2), which serves its purpose for all h, L covered by the Theorem, indeed $\bar{b}^*$ also arose in the empirical Edgeworth expansion justified in NR0. However, its cumbersome form and dependence on an additional kernel and bandwidth lead us to seek an alternative, more "direct", method of bias correction. Consider in place of the statistic $\hat{Z}^*$ (2.5) featured in the Theorem,

$$\hat{Z}^+ = n^{1/2}\hat{\sigma}^{*-1}(u^* - \bar{\mu}^+) \quad , \tag{3.4}$$

where

$$\bar{\mu}^+ = E^*(u^*) - \bar{b}^+ \quad , \tag{3.5}$$

for

$$\bar{b}^+ = v^\tau B^+, \qquad B^+ = \frac{1}{2}\{\frac{h^{-d-1}}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_i - Y_j)\, M(\frac{X_i - X_j}{h}) - U\}$$

in which the function M: $\Re^d \to \Re^d$, is

$$M(t) = \int\int K'(t+w+x)\, K(w)\, K(x)\, dwdx \ .$$

In Appendix C the following is proved.

Theorem 2 *Under assumptions (i)-(iv), (vii), (viii) and* $n^2 h^{2L+d+2} \to \infty$,

$$B^+ = EU - \bar{\theta} + o_p(h^L), \quad as \quad n \to \infty \ .$$

Comparison with Lemma 11 of NR0 indicates that therefore $\bar{b}^+ = h^L \kappa_1 + o_p(h^L)$. Thus, defining

$$\hat{F}^+(z) = P^*(\hat{Z}^+ \le z) \tag{3.6}$$

we deduce from Theorem 2 and the proof of Theorem 1:

Theorem 3 *Under assumptions (i)-(xii) and* $n^2 h^{2L+d+2} \to \infty$

$$\sup_{v:\, v^\tau v = 1} \sup_{z} |\hat{F}(z) - \hat{F}^+(z)| = o_p(n^{-1/2} + n^{-1} h^{-d-2} + n^{1/2} h^L) \ , \quad as \quad n \to \infty \ .$$

Thus, for a narrower range of bandwidths, and a weaker mode of convergence, bias correction by $\bar{b}^+$ can also achieve an Edgeworth correction. It remains to discuss computation of $M(t)$. At first sight the $2d$-dimensional integral is forbidding, but if a product form of K (3.1) is used we reduce to computing 2-dimensional integrals. Then we can even obtain a closed form expression for $M(t)$.

Theorem 4 *Let $K(x)$ be given by (3.1)-(3.3). Then if t_i denotes the i-th element of the $d \times 1$ vector t, the j-th element of $M(t)$ is*

$$m_1(t_j) \prod_{\substack{i=1 \\ i \ne j}}^{d} m_2(t_i) \ , \tag{3.7}$$

where the functions $m_i : \Re \to \Re$, $i = 1, 2$, are given in Appendix D.

4. Hypothesis Testing and Higher-Order Expansion

An important application of Theorem 1 is reduction in size distortion of tests. Consider the null hypothesis

$$H_0: \ \bar{\mu} = \mu_0, \tag{4.1}$$

for some given μ_0, and the one-sided alternative

$$H_1: \ \bar{\mu} > \mu_0 \tag{4.2}$$

(with no need to refer also to the alternative $\bar{\mu} < \mu_0$, due to arbitrariness of ν). The nominal size-α test based on the normal approximation, which we call the *N*-test, rejects H_0 against H_1 if $\hat{Z}_0 > z_\alpha$, where $z_\alpha = \Phi^{-1}(1-\alpha)$ and $\hat{Z}_0$ is $\hat{Z}$ with $\mu = \mu_0$. It has size distortion

$$-\phi(z_\alpha)\left[n^{1/2}h^L\kappa_1 - \frac{\kappa_2}{nh^{d+2}}z_\alpha - \frac{4}{3n^{1/2}}\{(2z_\alpha^2+1)\kappa_3 + 3(z_\alpha^2+1)\kappa_4\}\right] + o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L), \tag{4.3}$$

from Theorem 3 of NR0. Now consider a test based on the bootstrap distribution. This, called the *B*-test, rejects H_0 (4.1) against H_1 (4.2) if $\hat{Z}_0 > z_\alpha^*$, where $z_\alpha^* = \hat{F}^{*-1}(1-\alpha)$. In view of Theorem 1, this has size distortion of order only $o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$ a.s. A test based on the Edgeworth expansion attains the same size distortion, where the critical value z_α^* above is replacing by that from a Cornish-Fisher expansion (see NR0, pp.958); we call this the *E*-test.

Next consider two-sided testing. The alternative to H_0 (4.1) is now

$$H_1: \ \bar{\mu} \neq \mu_0. \tag{4.4}$$

The (two-sided) *N*-test rejects H_0 (4.1) against H_1 (4.4) if $|\hat{Z}_0| > z_{\alpha/2}$. Theorem 3 of NR0, and symmetry, implies the reduced size distortion

$$\kappa_2\phi(z_{\alpha/2})z_{\alpha/2}n^{-1}h^{-d-2} + o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L).$$

The (two-sided) *B*-test rejects H_0 (4.1) against H_1 (4.4) if $|\hat{Z}_0| > z_{\alpha/2}^*$. Theorem 1 implies that this has size distortion $o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$, as in the one-sided case.

The distortion (4.3) of the one-sided *N*-test can be minimized with respect to h, by

choosing h of order $n^{-3/\{2(L+d+2)\}}$ (the rate of the optimal bandwidth proposed by NR0), implying (4.3) has minimal order $n^{-(2L-d-2)/\{2(L+d+2)\}}$. The size distortion reported above for the two-sided N-test, however, is not susceptible to optimization, since the leading term is decreasing in h and we have insufficient information about components, absorbed in the remainder term, which increase in h. So far as the B-test is concerned, the information provided by Theorem 1 is too vague both to determine an optimal h and to investigate whether any size distortion is achieved by double-siding.

Both difficulties can be avoided by developing further the Edgeworth expansions of $\hat{Z}$ and $\hat{Z}^*$, to provide information on the remainder terms in Theorem 3 of NR0 and Theorom 2. Unfortunately, as expected, the details become even more complicated, and the following theorem relies on a highly unprimitive condition for validity. Introduce the additional assumptions

(i)' $E(|Y|^{25}) < \infty$.

(xiii) $\int_{p\le|t|\le N_0} \left|\frac{Ee^{itB}}{t}\right| dt = o(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-3})$,

$\int_{p\le|t|\le N_0} \left|\frac{E^*e^{itB^*}}{t}\right| dt = o(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-3})$ a.s.,

where B, p, N_0 are defined within the proof of Theorem 5 below, which is given in Appendix E, and B^* is the bootstrap analogue of B.

Define

$$F^{++}(z) = \Phi(z) - \phi(z)\left[n^{1/2}h^L\delta_1 - \frac{\delta_2}{nh^{d+2}}z + \frac{\delta_1^2}{2}nh^{2L}z \right.$$
$$\left. \frac{p_1(z)}{n^{1/2}} + \frac{p_2(z)}{n^{3/2}h^{d+3}} + \frac{p_3(z)}{n^{1/2}h^{d+2-L}} + \frac{q_1(z)}{n} + \frac{q_2(z)}{n^2h^{2d+4}} + q_3(z)h^L\right] ,$$

$$F^{++*}(z) = \Phi(z) - \phi(z)\left[n^{1/2}h^L\delta_1^* - \frac{\delta_2^*}{nh^{d+2}}z + \frac{\delta_1^{*2}}{2}nh^{2L}z \right.$$
$$\left. \frac{p_1^*(z)}{n^{1/2}} + \frac{p_2^*(z)}{n^{3/2}h^{d+3}} + \frac{p_3^*(z)}{n^{1/2}h^{d+2-L}} + \frac{q_1^*(z)}{n} + \frac{q_2^*(z)}{n^2h^{2d+4}} + q_3^*(z)h^L\right] ,$$

where $\delta_1 = h^{-L}\sigma^{-1}(Eu - \bar{\mu})$, $\delta_2 = h^{d+2}E(W_{12}^2)$, and the $p_i(z)$ are even polynomials

and the $q_i(z)$ are odd polynomials, defined in Appendix E; δ_1, δ_2 and the coefficients of the $p_i(z)$, $q_i(z)$ are $O(1)$ as $n\to\infty$. δ_i^*, $p_i^*(z)$ and $q_i^*(z)$ are respectively the bootstrap analogues of δ_i, $p_i(z)$ and $q_i(z)$ satisfying

$$\delta_i^* - \delta_i = o(1),\ p_i^*(z) - p_i(z) = o(1),\ q_i^*(z) - q_i(z) = o(1),\ \text{a.s.}$$

Theorem 5: *Under assumptions (i)', (ii)-(x) and (xiii), as* $n\to\infty$,

$$\sup_{\nu:\,\nu^\tau\nu=1}\sup_{z}\left|\hat{F}(z) - F^{++}(z)\right| = o(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-3}), \tag{4.5}$$

$$\sup_{\nu:\,\nu^\tau\nu=1}\sup_{z}\left|\hat{F}^*(z) - F^{++*}(z)\right| = o(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-3}) \text{ a.s.} \tag{4.6}$$

We point out that $\phi(z)q_1(z)$ is of similar form to the correction term in $P_n(x)$ in Theorem 5 of Maesono (1997).

We can deduce from Theorem 5, symmetry and Lemmas 11 and 12 of NR0, that the size distortion of the two-sided *N*-test is approximately

$$2\phi(z_{\frac{\alpha}{2}})\left\{\frac{\kappa_2}{nh^{d+2}}z_{\frac{\alpha}{2}} - \frac{\kappa_1^2 nh^{2L}}{2}z_{\frac{\alpha}{2}} + q_3(-z_{\frac{\alpha}{2}})h^L\right\}, \tag{4.7}$$

and that of the two-sided *B*-test is $o(n^{-1}h^{-d-2}+nh^{2L})$, noting that $h^L = O(n^{-1}h^{-d-2}+nh^{2L})$. The latter bound is slightly better than both the $o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$ a.s. one for the one-sided *B*-test (if h is chosen suitably large), and (4.7) for the two-sided *N*-test. On the other hand, (4.7) indicates that if we choose

$$h = \left(\frac{2\kappa_2}{\kappa_1^2 n^2}\right)^{\frac{1}{2L+d+2}} \tag{4.8}$$

the size-distortion of the two-sided test is only $o(n^{-1}h^{-d-2}+nh^{2L})$. Note that (4.8) is larger than the optimal h of NR0 (which minimizes the largest absolute error in the normal approximation) and has rate $n^{-3/(2L+d+2)}$, but of identical order to the minimum mean squared error bandwidth of Härdle, et al. (1992) and Powell and Stoker (1996). Desirably,

(4.8) does not depend on α. As in NR0, we can estimate the scale factor in (4.8) by inserting consistent estimates of κ_1 and κ_2.

5. Monte Carlo study

We compare normal, (empirical) Edgeworth and bootstrap approximations for the Tobit model used in the Monte Carlo study of NR0,

$$Y_i = (\beta^\tau X_i + \epsilon_i)\, I(\beta^\tau X_i + \epsilon_i \geq 0) \quad , \tag{5.1}$$

where $d=2$, $X_i=(X_{1i}, X_{2i})^\tau$, $(X_i^\tau, \epsilon_i) \sim N(0, I_3)$, $\beta=(1, 1)^\tau$. Given the normal assumption on ϵ_i, the maximum likelihood estimate of β comes to mind, but whereas Amemiya (1973) provided first order asymptotic theory for it, we know of no higher-order theory, let alone higher-order justification of the bootstrap (though doubtless these can be achieved).

From (5.1), $g(x)=\beta^\tau x\{1-\Phi(-\beta^\tau x)\}+\phi(-\beta^\tau x)$, $\bar{\theta}=-\beta/(8\pi)$. As in NR0, we employed three values of L, $L=4$, 8 and 10, which respectively correspond to the cases when $(d+2)/2<L<2(d+2)$, $L=2(d+2)$ and $L>2(d+2)$ (discussed in Section 4 of NR0), with K as in (3.1)-(3.3). We considered both elements of $\bar{\theta}$ individually, but since the results are very similar we report those for the first only, which we thus take as $\bar{\mu}$.

Bootstrap distributions (BDs) are estimated as follows.

1. We generate a Tobit data set of size n based on (5.1).
2. From this, we draw a random sample of size n, with replacement, from which we calculate U for each (h,L).
3. We repeat step 2 600 times.
4. We obtain empirical distributions, for each (h,L), of the two bias-corrected statistics $\hat{Z}^*$ (2.5), $\hat{Z}^+$ (3.4) and the bias-uncorrected statistic $\hat{Z}^- = n^{1/2}\hat{\sigma}^{*-1}(u^*-u)$.

5. We repeat steps 1-4 100 times, and average over the 100 empirical distributions to estimate the three BDs $\hat{F}^*(z)$, $\hat{F}^+(z)$ and $\hat{F}^-(z) = P^*(\hat{Z}^- \leq z)$.

Figures 1-12 compare the estimated BDs with the empirical distribution (ED), the Edgeworth expansion (EE) (or rather the empirical version of (2.7), with all three correction terms, discussed by NR0, NR1), and the normal distribution (N). We exclude the theoretical Edgeworth expansion, because the Monte Carlo study of NR0 found it inferior to EE. We used $c=1.2h$ for $\tilde{\kappa}_1$ as in NR0. ED and EE were produced from 600 replications for each combination of (n,h,L), as in NR0. In the figures, ED is represented as a solid line, N as a dotted line, BD as a dotted-and-broken line, and EE as a broken line. We employed for all combinations of $h=1.0, 0.8, 0.6, 0.4, 0.2$, $n=100, 400$ and $L=4, 8, 10$, but only report the results for $h= 0.8, 0.6, 0.4, 0.2$, $n=100$ and $L=4, 8, 10$. For $h=1.0$, none of N, EE or BD approximates ED well. The figures for $n=400$ are qualitatively similar to those for $n=100$, and the approximation errors are not only smaller but too small to read in some cases.

(Figures 1-12 about here)

Figures 1-4, 5-8, 9-12 correspond to $L=4, 8, 10$. We show results for $\hat{F}^+(z)$ as the BD when $L=4$, and for $\hat{F}^-(z)$ when $L=8,10$, in view of the discussion of Section 3. Generally BD and EE outperform N across h, and for smaller h, EE is actually better than BD (see Figures 4, 8, 12). The bias correction in (3.5) works rather well when $L=4$. With respect to bandwidth, $h=0.8$ (Figure 1) is too large, none of EE, BD, N working well. For the suppressed $n=400$ case, in general BD approximates ED as well as EE and better than N.

(Figures 13-16 about here)

Figures 13-16 focus more on effects of bias correction, comparing the BD

approximations of F^*, F^+, F^-, ED and N. The dashed and dotted line indicates F^*, the shorter dashed line is F^+, while the longer dashed line is F^-. Figure 13 considers $(n,h,L)=(100,0.6,4)$, where the bias in $\hat{Z}^-$ is expected to be relatively large, and indeed $\hat{Z}^*$ and $\hat{Z}^+$ both work much better. On the other hand in Figures 14, 15 and 16 $h=0.2$, so small bias is expected due to undersmoothing. Clearly $\hat{Z}^*$ performs worst, whereas $\hat{Z}^-$ is best for $L=8,10$ and it is hard to choose between $\hat{Z}^*$ and $\hat{Z}^-$. For the larger L bias corrections increase variance, and even when $L=4$ $\hat{Z}^*$ does very poorly; in general $\tilde{\kappa}_1$ is quite sensitive to choice of c, and overall $\hat{Z}^+$ appears to slightly outperform $\hat{Z}^*$.

We next compare hypothesis tests on $\bar{\mu}$, based on the normal, Edgeworth and bootstrap approximations to $\hat{F}(z)$, namely the *N*-test, *E*-test and *B*-test introduced in the previous section.

(Table 1 about here)

Table 1 (a) reports results for two-sided tests, with nominal size 5%, of the null hypothesis H_0: $\bar{\mu} = \mu_0$ against the alternative H_1: $\bar{\mu} \neq \mu_0$, for $L=4,8,10$, $n=100,400$ and $h=0.2,0.4,0.6,0.8$ (with $\mu_0 = -1/(8\pi)$, to correspond to $\beta=(1,1)^\tau$, $\nu=(1,0)^\tau$, under the normality). The empirical size of the *N*-test increases with h, consistent with the fact that the first term of the size distortion in (4.5), or the only first-order correction term remaining after double-siding, increases with h, since $\kappa_2 > 0$. The empirical size of the *B*-test is closer to 5% than that of the *N*-test in 21 cases out of 24, with 2 ties, the *N*-test dominating, slightly, in only the case $(n,h,L)=(400,0.8,8)$. The poor performance of the *N*-test can be explained by (2.6). Since $\kappa_2 > 0$, small h yields larger dispersion, especially around $z = \pm 1$, so the critical values of the N-test tend to be too large in absolute value, resulting in smaller empirical size. For larger h, size distortion is due to bias. Double-siding

should theoretically cancel the bias, but our Monte Carlo study shows that the distortion is not negligible. The *E*-test tends to outperform the *B*-test for small h, and vice versa, but overall they behave similarly.

Table 1(b),(c) reports corresponding results for one-sided tests against H_1: $\bar{\mu} < \mu_0$ and H_1: $\bar{\mu} > \mu_0$. The *N*-test performs very poorly in all cases, and almost always worse than the *B*-test and *E*-test.

We also studied the ability of the bootstrap to cope with the curse of dimensionality, considering larger values of d. We tried d=3,4, but report results only for d=4 because those for d=3 are somehow intermediate between the ones for d=2 and d=4. The model here is still (4.1), with $(X_i^{\tau}, \epsilon_i) \sim N(0, I_5)$, $\beta=(1, 1, 1, 1)^{\tau}$, $\nu=(1, 0, 0, 0)^{\tau}$. We obtained results for all combinations of h=1.0,0.8,0.6,0.4,0.2, n=100 and L=4,8,10, but only report results for h=1.0,0.8,0.6,0.4, bearing in mind that h should increase with d. Table 2 (a) and Table 2 (b), (c) respectively show empirical size of two-sided and one-sided tests for 5% nominal size.

(Table 2 about here)

Comparing Tables 1 (a) and 2 (a), and Tables 1 (b), (c) and 2 (b), (c), we find in general that increasing d clearly increases size distortion in magnitude for the *N*-test, but has relatively little impact on the *B*-test. For example, taking L=4, the *N*-test has empirical size 0.3%, 2.1%, 6.3% and 13.0% for h=0.2,0.4,0.6,0.8 respectively when d=2, compared with 0.8%,3.1%,13.2% and 38.0% for h=0.4,0.6,0.8,1.0, respectively, when d=4, whereas the *B*-test has sizes 1.3%,3.0%,5.1%,6.8% when d=2, and 1.2%,4.8%,5.7%,6.0% when d=4. Overall the *B*-test manifests greater resistance to the curse of dimensionality, especially in the one-sided tests of Table 2.

6. Conclusions

With the aim of improving normal-based inference, we have established that a bootstrap achieves a valid Edgeworth correction for the density-weighted ADE. However, the bootstrap does not always produce the necessary bias reduction, and methods for incorporating it have been presented. This provides a refinement in one-sided testing. To discuss two-sided testing we developed an Edgeworth expansion of still higher order, and found that double-siding cancels some terms, to suggest that the bias-corrected bootstrap here incurs even less size distortion. We have proposed a choice of bandwidth, which is optimal in the sense of minimizing two-sided size distortion while being independent of nominal size. In Monte Carlo study of finite sample performance, the bootstrap was found to usually improve on the normal approximation, though there is variability depending on the method of bias-correction, if any. We also presented limited Monte Carlo evidence that the bootstrap can suffer less from the curse of dimensionality than the normal approximation.

Our focus on the ADE of single index models was prompted by its relative simplicity, but is extremely narrow. In more general statistics, and other semiparametric estimates, details corresponding to those in the paper, both with respect to formal expansions and proofs of validity, will be even more complicated, possibly prohibitively so. Actual Edgeworth correction terms will depend on the particular model studied. However, it may be relatively obvious to see how to construct a sensible bootstrap procedure, and reasonable to suppose that our findings, building on those of Horowitz (1998, 2002), point to the potential for bootstrap-based inference to again provide improvements in moderate-sized samples. Of course, the choice of bandwidth, and bias correction, should be more or less dependent on the underlying model and the basic statistic. Undersmoothing informally "solves" both problems, but with respect to bias correction by $\bar{b}^*$ and $\bar{b}^+$, the latter, more "direct"

approach, may be the more convenient to extend.

APPENDIX A: Proof of Theorem 1

The qualification "almost surely" will be omitted hereafter. Let C denote a generic, finite, positive constant. We will make use of the following result for an unstudentized statistic. The proof is available from the authors upon request.

Proposition *Assuming* $E|Y|^{8+\delta} < \infty$ *for some* $\delta > 0$ *and* (ii)-(xii),

$$\sup_{\nu:\, \nu^\tau \nu = 1} \sup_{z} |F(z) - F^*(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad \text{a.s.}$$

where

$$F(z) = P(Z \le z) \quad , \quad Z = n^{1/2}\sigma^{-1}(u - \bar{\mu}) \quad ,$$

and

$$F^*(z) = P^*(Z^* \le z) \quad , \quad Z^* = n^{1/2}\sigma^{*-1}(u^* - \bar{\mu}^*) \quad .$$

Proof of Theorem 1:

In view of Theorem A of NR1, it suffices to show

$$\sup |\hat{F}^*(z) - F^+(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad , \qquad \text{(A.1)}$$

where we write *sup* for $\sup_{\nu:\, \nu^\tau \nu = 1} \sup_{z \in \Re}$. Note that $\hat{Z}^* = \sigma^* \hat{\sigma}^{*-1} Z^*$. Taylor's theorem gives $\sigma^* \hat{\sigma}^{*-1} = 1 + \tilde{R}^* + \tilde{\tilde{R}}^*$, where

$$\tilde{R}^* = \frac{1}{2}(1 - \sigma^{*-2}\hat{\sigma}^{*2}) \, , \quad \tilde{\tilde{R}}^* = \frac{3}{8}\{1 + \rho(\sigma^{*-2}\hat{\sigma}^{*2} - 1)\}^{-5/2}(\sigma^{*-2}\hat{\sigma}^{*2} - 1)^2 \, ,$$

for some $\rho \in [0, 1]$. Define $U_i = Y_i \int K(x) f'(X_i - hx)\, dx - \int K(x) e'(X_i - hx)\, dx$,

$u_i = \nu^\tau U_i$, $V_i = \sigma^{-1}(u_i - Eu)$, $u_{ij} = \nu^\tau U_{ij}$, $W_{ij} = \sigma^{-1}(u_{ij} - Eu) - V_i - V_j$. Introduce the

abbreviations $S_{ij} = V_i W_{ij}$, $T_{ij} = W_{ij} + V_j$, $N_{ij} = W_{ij} + V_i + V_j$. Also define the bootstrap-related variates $V_i^* = \sigma^{*-1}(u_i^* - E^* u^*)$, $u_{ij}^* = \nu^\tau U_{ij}^*$, $W_{ij}^* = \sigma^{*-1}(u_{ij}^* - E^* u^*) - V_i^* - V_j^*$, $S_{ij}^* = V_i^* W_{ij}^*$,

$$\bar{V}^* = \frac{2}{n^{1/2}} \sum_{i=1}^{n} V_i^*, \quad \bar{W}^* = n^{1/2} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} W_{ij}^*, \quad \Delta^* = n^{1/2} h^L \tilde{\kappa}_1, \qquad (A.2)$$

$\tilde{V}_i^* = E^*(V_j^* W_{ij}^* | i^*)$, $\tilde{W}_{jk}^* = E^*(W_{ij}^* W_{ik}^* | j^*, k^*)$, such that $E^*(\cdot | i_1^*, \ldots, i_r^*) = E^*(\cdot | (X_{i_j}^{*\tau}, Y_{i_j}^*), j = 1, \ldots, r)$, we have (cf. NR1) $\tilde{R}^* = T^* + Q^* + R^*$,

$T^* = T_1^* + T_2^* + T_3^*$, $R^* = R_1^* + R_2^* + R_3^* + R_4^* + R_5^* + R_6^*$, where

$$T_1^* = \frac{4n}{(n-2)^2} E^*(W_{12}^{*2}), \quad T_2^* = \frac{1}{n} \sum_{i=1}^{n} \left\{ (4V_i^{*2} - 1) + 8\tilde{V}_i^* \right\}, \quad T_3^* = 4\binom{n-1}{2}^{-1} \sum_{i<j}^{n} \tilde{W}_{ij}^*,$$

$$Q^* = 4\binom{n}{2}^{-1} \sum_{i<j}^{n} \left\{ (V_i^* + V_j^*) W_{ij}^* - \tilde{V}_i^* - \tilde{V}_j^* \right\},$$

$$R_1^* = -4\binom{n}{2}^{-1} \sum_{i<j}^{n} V_i^* V_j^*, \quad R_2^* = \frac{4}{n-2}\binom{n-1}{2}^{-1} \sum_{i=1}^{n} \sum_{k<m}^{n}{}^{(i)} (W_{ik}^* W_{im}^* - \tilde{W}_{km}^*),$$

$$R_3^* = \frac{4n}{(n-2)^2}\binom{n}{2}^{-1} \sum_{i<j}^{n} \{ W_{ij}^{*2} - \tilde{W}_{ii}^* - \tilde{W}_{jj}^* + E^*(W_{12}^{*2}) \},$$

$$R_4^* = \frac{8}{(n-2)^2} \sum_{i=1}^{n} \{ \tilde{W}_{ii}^* - E^*(W_{12}^{*2}) \}, \quad R_5^* = -\frac{4n(n-1)}{(n-2)^2} \left\{ \binom{n-1}{2}^{-1} \sum_{i<j}^{n} W_{ij}^* \right\}^2,$$

$$R_6^* = -\frac{8}{n}\binom{n-1}{2}^{-1} \sum_{i=1}^{n} \sum_{k<m}^{n}{}^{(i)} V_i^* W_{km}^*.$$

where $\sum_{k<m}^{n}{}^{(i)}$ denotes summation with respect to k and m for $1 \le k < m \le n$ but excluding $k=i$ and $m=i$. Because $Z^* = \bar{V}^* + \bar{W}^* + \Delta^*$, we have $\hat{Z}^* = (1 + \tilde{R}^* + \tilde{\tilde{R}}^*)(\bar{V}^* + \bar{W}^* + \Delta^*)$, so by a standard inequality the left side of (A.1) is bounded by

$$\sup_z |P^*((1 + T^* + Q^*)(\bar{V}^* + \bar{W}^*) + \Delta^* \le z) - F^+(z)|$$

$$+ P^*\left(|(R^* + \tilde{\tilde{R}}^*)(\bar{V}^* + \bar{W}^* + \Delta^*) + (T^* + Q^*)\Delta^*| \ge a_n \right) + O(a_n) \qquad (A.3)$$

for $a_n > 0$. Taking $a_n = \max(n^{-1/2}, n^{-1} h^{-d-2}, n^{1/2} h^L)/\log n$, we bound the second term on the right of (A.3) by

$$P^*\left(|(R^* + \tilde{\tilde{R}}^*)(\bar{V}^* + \bar{W}^* + \Delta^*)| \ge \frac{a_n}{2} \right) + P^*\left(|(T^* + Q^*)\Delta^*| \ge \frac{n^{1/2} h^L}{2 \log n} \right)$$

$$\le P^*\left(|R^* + \tilde{\tilde{R}}^*| \ge \frac{a_n}{2 \log n} \right) + P^*(|\bar{V}^* + \bar{W}^* + \Delta^*| \ge \log n)$$

$$+ P^*\left(|(T^*+Q^*)\Delta^*| \geq \frac{n^{1/2}h^L}{2\log n} \right) . \qquad \text{(A.4)}$$

The second term on the right of (A.4) is $o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$ similarly to the derivation of (3.14) of NR1 using Theorem 1 of NR0 and the Proposition. The first term can be handled as in (3.5)-(3.8) of NR1, appealing to Lemmas B.15-24, and shown to be $o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$. The last term on the right of (A.4) is $o(n^{-1/2}+n^{-1}h^{-d-2})$, similarly to (3.15) of NR1, by Markov's inequality, Lemmas B.15-18 and

$$\Delta^* = n^{1/2}h^L\{\kappa_1 + o(1)\} , \qquad \text{(A.5)}$$

which is straightforward because $\tilde{\kappa}_1 \to \kappa_1$ (see the proof of Theorem 2 of NR0).

Write $b_2^* = \bar{V}^*$, $b_3^* = \bar{W}^*$, $\mathcal{B}_2^* = (T^*+Q^*)\bar{V}^*$, $\mathcal{B}_3^* = (T^*+Q^*)\bar{W}^*$, $b_1^* = b_2^* + b_3^*$, $\mathcal{B}_1^* = \mathcal{B}_2^* + \mathcal{B}_3^*$, $B^* = b_1^* + \mathcal{B}_1^*$, and define

$$\chi^+(t) = \int e^{itz} dF^+(z)$$

$$= e^{-\frac{t^2}{2}}\left[1 + \left\{ n^{1/2}h^L\kappa_1 - \frac{4(\kappa_3+2\kappa_4)}{n^{1/2}} \right\}(it) - \frac{\kappa_2}{nh^{d+2}}(it)^2 - \frac{4(2\kappa_3+3\kappa_4)}{3n^{1/2}}(it)^3 \right] ,$$

using (2.7). For $N_0 = \log n \min(\epsilon^* n^{1/2}, nh^{d+2})$, where $\epsilon^* = (E^*|2V_1^*|^3)^{-1} \leq \{E^*(2V_1^*)^2\}^{-3/2} = 1$, Esseen's smoothing lemma bounds the first term on the right of (A.3) by

$$\int_{-N_0}^{N_0} \left| \frac{E^* e^{it(B^*+\Delta^*)} - \chi^+(t)}{t} \right| dt + O(N_0^{-1})$$

$$= \text{(I)} + \text{(II)} + \text{(III)} + o(n^{-1/2}+n^{-1}h^{-d-2}) , \qquad \text{(A.6)}$$

where

$$\text{(I)} = \int_{-p'}^{p'} \left| \frac{E^* e^{it(B^*+\Delta^*)} - \chi^+(t)}{t} \right| dt \quad , \quad \text{(II)} = \int_{p' \leq |t| \leq N_0} \left| \frac{E^* e^{it(B^*+\Delta^*)}}{t} \right| dt \quad ,$$

$$\text{(III)} = \int_{|t|\geq p} \left| \frac{\chi^{+}(t)}{t} \right| dt \quad ,$$

for $p = \min(\epsilon n^{1/2}, \log n)$, $p' = \min(\log n, \varepsilon n^{1/2}, nh^{d+2})$. It remains to show that

$$\text{(I), (II), (III)} = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) . \tag{A.7}$$

(III) is handled in (3.47) of NR1. We estimate (I). Writing $b_2^* = b_2^{*\prime} + b_2^{*\prime\prime}$ where $b_2^{*\prime} = T^* \bar{V}^*$ and $b_2^{*\prime\prime} = Q^* \bar{V}^*$, we obtain, similarly to (3.18)-(3.28) of NR1,

$$E^*\{e^{it(B^*+\Delta^*)}\} = e^{-\frac{t^2}{2}}\left[1 + \left\{ \Delta^* - \frac{4E^*(V_1^{*3}) + 8E^*(W_{12}^* V_1^* V_2^*)}{n^{1/2}} \right\}(it) - \frac{E^*(W_{12}^{*2})}{n}(it)^2 \right.$$
$$\left. - \frac{4\{2E^*(V_1^{*3}) + 3E^*(W_{12}^* V_1^* V_2^*)\}}{3n^{1/2}}(it)^3 \right] + D_n \quad , \tag{A.8}$$

where

$$D_n = O\left(\left\{ e^{-\frac{t^2}{2}} \frac{|t|^3}{n^{1/2}} + o(n^{-1/2}(t^6 + |t|^3) e^{-\frac{t^2}{4}}) \right\} \left\{ \frac{t^2}{n^2 h^{d+2}} + \frac{|t|^3 + |t|}{n^{1/2}} + A_n' \right\} \right.$$
$$+ e^{-\frac{t^2}{2}}(|t| n^{1/2} h^L + t^2 h^{2L} + |t| n^{1/2} h^{2L}) \left(\frac{t^2}{n^2 h^{d+2}} + \frac{|t|^3 + |t|}{n^{1/2}} + A_n' \right)$$
$$+ (|t| n^{1/2} h^L + t^2 n h^{2L}) \left\{ e^{-\frac{t^2}{2}} \frac{|t|^3}{n^{1/2}} + o(n^{-1/2}(t^6 + |t|^3) e^{-\frac{t^2}{4}}) \right\}$$
$$\left. \times \left(1 + \frac{t^2}{n^2 h^{d+2}} + \frac{|t|^3 + |t|}{n^{1/2}} + A_n' \right) \right)$$
$$+ (|t| + t^2 n^{1/2} h^L + |t|^3 n h^{2L}) \{ E^* |b_2^{*\prime\prime} + b_3^*| + |t| (E^*(b_2^{*2}) + E^* |b_2^{*\prime} b_3^*|) \} \quad , \tag{A.9}$$

$$A_n' = o\left(\frac{t^2 + t^{10}}{n h^{d+2}} + \frac{t^2 + t^6}{n^{1/2}} \right) . \tag{A.10}$$

By Hölder's inequality and Lemmas B.5 and B.14-B.18,

$$E^* |b_2^{*\prime\prime}| \leq [E^*(Q^{*2}) E^*(\bar{V}^{*2})]^{1/2} = O(n^{-1} h^{-\frac{d+2}{2}}) \quad , \tag{A.11}$$

$$E^* |b_3^*| \leq [E^* |T^* + Q^*|^2 E^*(\bar{W}^{*2})]^{1/2} = O((n^{-1/2} + n^{-1}h^{-d-2})(n^{-1}h^{-d-2})^{1/2}) . \tag{A.12}$$

Lemmas B.14-B.17 and Hölder's inequality give

$$E^*(b_2^{*2}) \leq C\{ T_1^{*2} E^*(\bar{V}^{*2}) + E^* |T_2^* \bar{V}^*|^2 + E^* |T_3^* \bar{V}^*|^2 \} \quad ,$$
$$\leq C\{ T_1^{*2} E^*(\bar{V}^{*2}) + [E^*(T_2^{*4}) E^*(\bar{V}^{*4})]^{1/2} + [E^*(T_3^{*4}) E^*(\bar{V}^{*4})]^{1/2} \}$$
$$= O(n^{-1} + n^{-2}h^{-2d-4}) . \tag{A.13}$$

Hölder's inequality, (A.13) and Lemma B.5 yield

$$E^*|b_2^{*\prime} b_3^*| = [E^*(b_2^{*2})E^*(b_3^{*2})]^{1/2} = O((n^{-1/2}+n^{-1}h^{-d-2})(n^{-1}h^{-d-2})^{1/2}) \quad . \quad (A.14)$$

We have, putting $v_1 = \sigma^{-1} v^\tau \{ Y_i f'(X_i) - e'(X_i) \}$,

$$E^*(V_1^*)^3 = \frac{1}{n\sigma^{*3}} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} u_{ij} - \frac{n-1}{n} u \right)^3 \to E(v_1^3) \quad , \quad (A.15)$$

due to (B.2) of NR0 and Lemma B.1-(d), and

$$E^*(W_{12}^{*2}) = h^{-d-2}\{\kappa_2 + o(1)\}, \quad (A.16)$$

due to Lemma 12 of NR0 and Lemma B.4-(a). By (A.5), (A.8), (A.16)

$$(I) \leq \int_{-\log n}^{\log n} \left| \frac{E^* e^{it(B^*+\Delta^*)} - \chi^+(t)}{t} \right| dt = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) \quad . \quad (A.17)$$

We evaluate (II) by partitioning the range of integration into two parts, $p \leq |t| \leq N_1$ and $N_1 \leq |t| \leq N_0$, for $N_1 = \min(\epsilon^* n^{1/2}, nh^{d+2})$. Because (3.33), (3.35)-(3.43) of NR1 hold for bootstrapped statistics and bootstrap moments due to Lemmas B.1-B.4, B.6-B.13, we obtain, using Lemma B.25-(a), (b),

$$\int_{p \leq |t| \leq N_1} \left| \frac{E^* e^{it(B^*+\Delta^*)}}{t} \right| dt = \int_{p \leq |t| \leq N_1} \left| \frac{E^* e^{itB^*}}{t} \right| dt = o(n^{-1/2} + n^{-1}h^{-d-2}) \quad , (A.18)$$

$$\int_{N_1 \leq |t| \leq N_0} \left| \frac{E e^{itB^*}}{t} \right| dt = o(n^{-1/2} + n^{-1}h^{-d-2}) \quad , \quad (A.19)$$

similarly to (3.45) and (3.46) of NR1. Therefore, by (A.18) and (A.19),

$$(II) = o(n^{-1/2} + n^{-1}h^{-d-2}) \quad , \quad (A.20)$$

to complete the proof. □

APPENDIX B: Technical Lemmas for the Proof of Theorem 1

The proofs of Lemmas B.1-6 and B.19-25 below are contained in a separate document which may be obtained from the authors on request. The proofs of Lemmas B.7-13 involve identical techniques and the orders of bootstrapped moments are the same as those of the corresponding moments, given in Lemmas 1-8 of NR1.

LEMMA B.1:

(a) $E^*(V_1^*) = 0, \quad E^*(2V_1^*)^2 = 1, \quad E^*(W_{12}^*) = 0.$

(b) $E^*(W_{12}^* \mid 1^*) = E^*(W_{12}^* \mid 2^*) = 0.$

Under (i), (iii), (iv), (v), (vii), (viii) *and* (ix),

(c) $E^*u^* - Eu = o((\log n)^{-3})$.

Under (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

(d) $\sigma^{*2} = \sigma^2 + o((\log n)^{-3})$, $\sigma^{*-r} = \sigma^{-r} + o((\log n)^{-3})$, $r \geq 1$.

Under $E(Y^8) < \infty$, (iii), (iv), (v), (vi), (vii), (viii) *and* (ix),

(e) $E^*(V_1^{*4}) = O(1)$.

Under (i), (iii), (iv), (v), (vi), (vii), (viii) *and* (ix),

(f) $E^*(V_1^{*8}) = O(1)$.

LEMMA B.2:

$\bar{V}^* - \bar{V}^*(m)$ *and* $\bar{W}^* - \bar{W}^*(m)$ *are independent of* $(X_1^{*\tau}, Y_1^*), \cdots,$ $(X_m^{*\tau}, Y_m^*)$ *conditional on* $(X_1^\tau, Y_1), \cdots, (X_n^\tau, Y_n)$.

LEMMA B.3: *Under* (i), (iii), (iv), (v), (vi), (vii), *and* (viii),

$$|E^*(W_{12}^* V_1^* V_2^*)| = |E(W_{12} V_1 V_2)| + o(1) = |E(W_{12} \nu_1 \nu_2)| + o(1) = \kappa_4 + o(1) = O(1) \quad .$$

LEMMA B.4: *Under* (i), (iii), (iv), (v), (vi), (vii), *and* (viii),

(a) $E^*|W_{12}^*|^r = E|W_{12}|^r + o(h^{-(r-1)d-r})$ *for* $r = 1, 2, 3, 4$.

(b) $E^*(|W_{12}^* W_{13}^*|^r) = O(h^{-2(r-1)d-2r})$ *for* $r = 1, 2$.

LEMMA B.5: *Under* (i), (iii), (iv), (v), (vi), (vii), (viii) *and* (ix),

$$E^*|\bar{W}^*(m)|^3 = O\left(\left(\frac{m}{n^2 h^{d+2}} \right)^{\frac{3}{2}} \right) ,$$

in particular $E^*|W^*|^3 = E^*|\bar{W}^*(n-1)|^3 = O\left((n^{-1}h^{-d-2})^{\frac{3}{2}} \right)$.

LEMMA B.6: *Under* (i)', (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*|V_1^* W_{12}^*|^r = E^*|V_2^* W_{12}^*|^r = O(h^{-(r-1)d-r}) \quad \text{for} \quad r = 2, 3, 4 .$$

LEMMA B.7: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*|\tilde{V}_1^*|^r = O(1) \quad \text{for} \quad 1 \le r \le 6 .$$

LEMMA B.8: *Under* (i), (iii), (iv), (v), (vi) *and* (viii),

(a) $E^*|\tilde{W}_{11}^*|^r = O(h^{-r(d+2)})$ for $1 \le r \le 3$,

(b) $E^*|\tilde{W}_{12}^*|^r = O(h^{-(r-1)d-2r})$ for $1 \le r \le 6$.

LEMMA B.9: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii), *for*

$$d_i^* = (4V_i^{*2} - 1) + 8\tilde{V}_i^* ,$$

(a) $E^*|d_1^* V_2^*|^r = O(1)$ for $1 \le r \le 3$,

(b) $E^*|d_1^* V_1^*|^r = O(1)$ for $1 \le r \le 2$.

LEMMA B.10: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E^*|W_{12}^* V_i^* V_3^*|^r = O(h^{-(r-1)d-r})$ for $1 \le r \le 3$, $i = 1, 2$,

(b) $E^*|W_{12}^* V_1^{*2}|^r = E^*|W_{12}^* V_2^{*2}|^r = O(h^{-(r-1)d-r})$ for $1 \le r \le 2$.

LEMMA B.11: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii), *for*

$$e_{ij}^* = 4\left\{ (V_i^* + V_j^*) W_{ij}^* - \tilde{V}_i^* - \tilde{V}_j^* + \frac{n}{(n-2)} \tilde{W}_{ij}^* \right\} ,$$

(a) $E^*|e^*_{12}V^*_3|^r = O(h^{-(r-1)d-2r})$ *for* $1 \le r \le 3$,

(b) $E^*|e^*_{12}V^*_1|^r = E^*|e^*_{12}V^*_2|^r = O(h^{-(r-1)d-2r})$ *for* $1 \le r \le 2$.

LEMMA B.12: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E^*|d^*_1 W^*_{23}|^r = O(h^{-(r-1)d-r})$ for $1 \le r \le 3$,

(b) $E^*|d^*_1 W^*_{12}|^r = E^*|d^*_2 W^*_{12}|^r = O(h^{-(r-1)d-r})$ for $1 \le r \le 2$.

LEMMA B.13: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E^*|\tilde{W}^*_{12}W^*_{12}|^r = O(h^{-(2r-1)d-3r})$ for $1 \le r \le 3$,

(b) $E^*|\tilde{W}^*_{12}W^*_{i3}|^r = O(h^{-2(r-1)d-3r})$ for $1 \le r \le 3$, $i = 1, 2$,

(c) $E^*|\tilde{W}^*_{12}W^*_{34}|^r = O(h^{-2(r-1)d-3r})$ for $1 \le r \le 6$.

LEMMA B.14: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$E^*|\bar{V}^*|^r = O(1)$ for $2 \le r \le 4$.

LEMMA B.15: *Under* (i), (v), (vi), (vii) *and* (viii)

$|T^*_1|^r = O(n^{-r}h^{-r(d+2)})$ for $r \le 4$.

LEMMA B.16: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$E^*|T^*_2|^r = O(n^{-\frac{r}{2}})$ for $r = 2, 3, 4$.

LEMMA B.17: *Under* (i), (iii), (iv), (v), (vi) *and* (viii),

$E^*(T^{*4}_3) = O(n^{-r}h^{-(r-1)d-2r})$ for $r = 2, 3, 4$.

LEMMA B.18: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*(Q^{*2}) = O(n^{-2}h^{-d-2}) \quad .$$

LEMMA B.19: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*(R_1^{*2}) = O(n^{-2}) \quad .$$

LEMMA B.20: *Under* (i), (iii), (iv), (v), (vi) *and* (viii),

$$E^*(R_2^{*2}) = O(n^{-3}h^{-2d-4}) \quad .$$

LEMMA B.21: *Under* (i), (iii), (iv), (v), (vi) *and* (viii),

$$E^*(R_3^{*2}) = O(n^{-4}h^{-3d-4}) \quad .$$

LEMMA B.22: *Under* (i), (iv), (v), (vi) *and* (viii),

$$E^*(R_4^{*2}) = O(n^{-3}h^{-2(d+2)}) \quad .$$

LEMMA B.23: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*(R_5^{*2}) = O(n^{-4}h^{-2(d+2)}) \quad .$$

LEMMA B.24: *Under* (i), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E^*(R_6^{*2}) = O(n^{-3}h^{-d-2}) \quad .$$

LEMMA B.25: *Under* (i), (ii), (iii), (iv), (v), (vii), (viii), (ix) *and* (x),

(a) *for* $p \le |t| \le \epsilon^* n^{1/2}$, $|\gamma^*(t)| \le \exp\left(-\frac{t^2}{3}n\right)$,

(b) *for* $\epsilon^* n^{1/2} \le |t| \le n^{1/2}\log n$, *there exists* $\eta \in (0, 1)$ *such that, for large enough*

n, $|\gamma^*(t)| \le 1-\eta$, where $\gamma^*(t) = E^*[w_1^*(t)]$, $w_j^*(t) = e^{2it\,n^{-1/2}V_j^*}$.

APPENDIX C: Proof of Theorem 2

Because $K(-u)=K(u)$ and $K'(-u)=-K'(u)$, $\bar{b}^+$ has a U statistic form with kernel

$$P_{ij} = \frac{(Y_i - Y_j)}{2h^{d+1}} \int\int \{K'(\frac{X_i - X_j}{h} + u - v) - K'(\frac{X_i - X_j}{h})\} K(u) K(v)\, du\, dv \ .$$

Putting $P_i = E(P_{ij} \mid X_i, Y_i) - E(P_{12})$ and $Q_{ij} = P_{ij} - P_i - P_j - E(P_{12})$,

$$\bar{b}^+ = \frac{2}{n}\sum_{i=1}^{n} P_i + \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q_{ij} + E(P_{12}) \ .$$

It suffices to show the following:

$$EP_{12} - (EU - \bar{\theta}) = o(h^L) \ , \tag{C.1}$$

$$\frac{1}{n}\sum_{i=1}^{n} P_i = o_p(h^L) \ , \tag{C.2}$$

$$\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q_{ij} = o_p(h^L) \ . \tag{C.3}$$

We first show (C.1). Change of variables and integration by parts give

$$E(P_{12} \mid X_1, Y_1) = \frac{Y_1}{2} \int\int\int K(z) K(u) K(v) \{f'(X_1 - h(z-u+v)) - f'(X_1 - hz)\} du\, dv\, dz$$
$$- \frac{1}{2} \int\int\int K(z) K(u) K(v) \{e'(X_1 - h(z-u+v)) - e'(X_1 - hz)\} du\, dv\, dz \ . \tag{C.4}$$

Using (v) and (viii) and assuming without loss of generality that L is even, Taylor expansion around $h=0$ gives

$$\int\int\int K(z) K(u) K(v)\, e'(X_1 - hz)\, dz\, du\, dv = \int K(z)\, e'(X_1 - hz)\, dz$$
$$= e'(X_1) + \frac{h^L}{L!} \sum_{\substack{0 \le l_1, \ldots, l_d \le L \\ l_1 + \ldots + l_d = L}} \cdots \sum \bar{K}_l e_l(X_1) + o(h^L) \ ,$$

where $\bar{K}_l = \int \prod_{i=1}^{d} x_i^{l_i} K(x)\, dx$, $g_l(x) = \Delta^{(l_1, \ldots, l_d)} g'(x)$ for a generic function g. Similar calculation gives

$$\int\int\int K(z)\, e'(X_1 - h(z-u+v))\, dz\, K(u) K(v)\, du\, dv$$
$$= e'(X_1) + \frac{3h^L}{L!} \sum_{\substack{0 \le l_1, \ldots, l_d \le L \\ l_1 + \ldots + l_d = L}} \cdots \sum \bar{K}_l e_l(X_1) + o(h^L) \ .$$

Therefore the last integral on the right of (C.4) is

$$\frac{2h^L}{L!}\sum_{\substack{0\le l_1,\ldots,l_d\le L\\ l_1+\ldots+l_d=L}}\cdots\sum \bar{K}_l e_l(X_1) + o(h^L) \quad . \tag{C.5}$$

Likewise the first integral on the right of (C.4) is

$$\frac{2h^L}{L!}\sum_{\substack{0\le l_1,\ldots,l_d\le L\\ l_1+\ldots+l_d=L}}\cdots\sum \bar{K}_l f_l(X_1) + o(h^L) \quad . \tag{C.6}$$

Because

$$EU-\bar{\theta} = E\Big[Y_1\int K(z)\{f'(X_1-hz)-f'(X_1)\}dz - \int K(z)\{e'(X_1-hz)-e'(X_1)\}dz \Big]$$
$$= E[\frac{Y_1h^L}{L!}\sum_{\substack{0\le l_1,\ldots,l_d\le L\\ l_1+\ldots+l_d=L}}\cdots\sum \bar{K}_l f_l(X_1) - \frac{h^L}{L!}\sum_{\substack{0\le l_1,\ldots,l_d\le L\\ l_1+\ldots+l_d=L}}\cdots\sum \bar{K}_l e_l(X_1)\ ] + o(h^L) \ ,$$

we obtain $E(P_{12}) - (EU-\bar{\theta}) = E[\,E(P_{12}\,|X_1, Y_1)\,] - (EU-\bar{\theta}) = o(h^L)$ by (i).

Manipulation similar to that in the proof of Lemmas 2 and 7 of NR0 yields $Var\,(\sum_{i=1}^{n} P_i) = O(nh^{2L})$ using (C.4)- (C.6), (i), (iv) and (v), which justifies (C.2). To prove (C.3), noting we can show $E(Q_{12}^2) = O(h^{-d-2})$ similarly to Lemma 4 of R, we have

$$Var\,\{h^{-L}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Q_{ij}\} = \frac{E(Q_{12}^2)}{h^L n(n-1)} = O(\frac{1}{h^{2L}n^2h^{d+2}}) = o(1) \quad . \qquad \square$$

APPENDIX D: Formulae for m_1, m_2 and proof of Theorem 3

Writing $l=L/2-1$, we have

$$m_1(t) = \frac{\phi(t/\sqrt{3})}{\sqrt{3}}\sum_{h=0}^{l}\sum_{i=0}^{l}\sum_{j=0}^{l}\xi_h\zeta_i\zeta_j\tau_{hij}\,(2h+1)!\,(2i)!\,(2j)!\ ,$$
$$m_2(t) = \frac{\phi(t/\sqrt{3})}{\sqrt{3}}\sum_{h=0}^{l}\sum_{i=0}^{l}\sum_{j=0}^{l}\zeta_h\zeta_i\zeta_j\omega_{hij}\,(2h)!\,(2i)!\,(2j)!\ ,$$

where

$$\xi_i = 2\zeta_{i+1}-\zeta_i\ ,\quad i=0,\ldots,l-1,$$

$$\xi_l = -\zeta_l\ ,$$

$$\tau_{hij}(t) = {\sum_{h_1+h_2+h_3=2h+1}}' \frac{(1/\sqrt{2})^{h_1}(1/\sqrt{6})^{h_2}(1/3)^{h_3}}{h_1!\,h_2!\,h_3!}$$
$$\times {\sum_{i_1+i_2=2i}}' \frac{(\sqrt{2/3})^{i_1}(-1/3)^{i_2}}{i_1!\,i_2!}\ {\sum_{j_1+j_2+j_3=2j}}' \frac{(1/\sqrt{2})^{j_1}(-1/\sqrt{6})^{j_2}(-1/3)^{j_3}}{j_1!\,j_2!\,j_3!}$$
$$\times\chi_{h_1+j_1}\chi_{h_2+i_1+j_2}t^{h_3+i_2+j_3}\ ,$$

where each primed sum is over all non-negative values of the subscript indices subject to the stated equality, and

$$\begin{aligned}\chi_j &= 1, \quad j = 0, \\ &= 0, \quad j \text{ odd}, \\ &= \prod_{i=1}^{j/2} (j - 2i + 1), \quad j \text{ even},\end{aligned}$$

Thus, many of the summands in $\tau_{hij}(t)$, $\omega_{hij}(t)$ will be zero. Note that, for $i=0,\ldots,l$, ζ_i is the $(i+1)$-th element of the vector $V^{-1}\mathrm{v}$, where v is the first column of the $(l+1)\times(l+1)$ identity matrix and V is the $(l+1)\times(l+1)$ Hankel matrix with (i,j)-th element $\chi_{2i+2j-4}$.

To establish (3.7), with these m_i, note from (3.1) that $K'(x)$ has j-th element $k'(x_j)\prod_{i=1, i\neq j}^{d} k(x_i)$, where $k'(x) = (d/dx)k(x)$. Thus $M(t)$ has j-th element $\tilde{m}_1(t_j)\prod_{i=1, i\neq j}^{d}\tilde{m}_2(t_i)$, where

$$\tilde{m}_1(t) = \int k'(t+w+x)\,k(w)\,k(x)\,dwdx, \quad \tilde{m}_2(t) = \int k(t+w+x)\,k(w)\,k(x)\,dwdx.$$

It remains to show that $\tilde{m}_i = m_i$, $i = 1, 2$, under (3.2), (3.3). We do this only for $i=1$, the proof for $i=2$ being essentially identical. Change of variables gives

$$\tilde{m}_1(t) = \int q_l(w)\,p_l(x)\,p_l(w-x-t)\,\phi(w)\,\phi(x)\,\phi(w-x-t)\,dwdx,$$

where

$$q_l(x) = k'(x)/\phi(x) = \sum_{i=0}^{l} \xi_i x^{2i+1}.$$

Using the identity

$$w^2 + x^2 + (w-x-t)^2 = (\sqrt{2}w - x/\sqrt{2} - j/\sqrt{2})^2 + (\mathrm{v}\sqrt{3/2} + z/\sqrt{6})^2 + z^2/3,$$

a change of variables gives

$$\tilde{m}_1(t) = \frac{\phi(t/\sqrt{3})}{\sqrt{3}} \int q_l\left(\frac{w}{\sqrt{2}} + \frac{x}{\sqrt{6}} + \frac{t}{3}\right) p_l\left(\sqrt{\frac{2}{3}}x - \frac{t}{3}\right) p_l\left(\frac{w}{\sqrt{2}} - \frac{x}{\sqrt{6}} - \frac{t}{3}\right) \times \phi(w)\,\phi(x)\,dwdx.$$

The proof is completed by applying the multinomial theorem to the summands of p_l and q_l and noting that χ_j is the j-th moment of a standard normal variate.

APPENDIX E: Details and proof of Theorem 5

We have

$$p_1(z) = -\frac{4}{3}\{(2z^2+1)\delta_3 + 3(z^2+1)\delta_4\},\ p_2(z) = \delta_6(z^2-13)/3,$$

$$p_3(z) = -\delta_1\delta_2(z^2+1),\ q_1(z) = \xi_1 z + \xi_2(z^3-3z) + \xi_3(z^5-10z^3+15z),$$

$$q_2(z) = \delta_2^2(19z^3/6 - 3z/2),\ q_3(z) = -\delta_1\{4(\delta_3+2\delta_4)z + (8\delta_3/3+4\delta_4)(z^3-3z)\},$$

where

$$\delta_3 = E(V_1^3),\ \delta_4 = E(W_{12}V_1V_2),\ \delta_5 = h^{d+2}E\{W_{12}^2(V_1+V_2)\},\ \delta_6 = h^{d+3}E(W_{12}W_{13}W_{23}),$$

$$\delta_7 = E\{W_{12}V_1V_2(V_1+V_2)\},\ \delta_8 = E(W_{12}W_{13}V_2V_3),\ \delta_9 = E(V_1^4),$$

$$\xi_1 = 64(\delta_3+2\delta_4)^2 + 1 - 16\delta_7 - 16\delta_8,$$

$$\xi_2 = \frac{64}{3}(\delta_3+2\delta_4)(2\delta_3+3\delta_4) + \frac{1}{2} - 8\delta_7 - 8\delta_8 - \frac{4}{3}\delta_9,\ \xi_3 = \frac{8}{9}(2\delta_3+3\delta_4)^2.$$

Proof of Theorem 5

Consider (4.5). Write

$$n^{1/2}\sigma^{-1}(u-\bar{\mu}) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} V_i + n^{1/2}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij} + n^{1/2}\sigma^{-1}(Eu-\bar{\mu})$$
$$= \bar{V} + \bar{W} + \Delta.$$

Letting $S = 4Var(U_i)$, $s^2 = \sigma^{-2}v^{\tau}Sv$, Taylor's theorem gives

$$\sigma\hat{\sigma}^{-1} = s^{-1} - \frac{s^{-3}}{2}(\sigma^{-2}\hat{\sigma}^2 - s^2) + \frac{3s^{-5}}{8}(\sigma^{-2}\hat{\sigma}^2 - s^2)^2$$
$$-\frac{5}{16}\{s^2+\theta(\sigma^{-2}\hat{\sigma}^2-s^2)\}^{-7/2}(\sigma^{-2}\hat{\sigma}^2-s^2)^3$$
$$= s^{-1} + \tilde{R} + \tilde{\tilde{R}} + R'$$

for some $\theta \in [0, 1]$. Put $\tilde{V}_i = E(V_j W_{ij} | i)$, $\tilde{W}_{jk} = E(W_{ij} W_{ik} | j, k)$, and let

$$\tilde{R} = T + Q + R,\quad T = T_1 + T_2 + T_3,\quad Q = Q_1 + Q_2,\quad R = R_1 + R_2 + R_3 + R_4 + R_5,$$

where the T_i, Q_i, R_i are defined as in NR1 (noting that S^2 should be s^2 in their T_2).

Writing $\hat{Z} = (s^{-1} + \tilde{R} + \tilde{\tilde{R}} + R')(\bar{V} + \bar{W} + \Delta)$, a standard inequality gives

$$\sup|\hat{F}(z) - F^{++}(z)| \le \sup|P(B \le z) - F^{++}(z)| + P\left(|\hat{Z} - B| \ge a_n\right) + O(a_n), \quad (E.1)$$

where $B = b_2 + b_3 + \tilde{b}_2 + \tilde{b}_3 + b_4 + b_5 + s^{-1}(\Delta + T_1\Delta)$, for

$$b_2 = s^{-1}\bar{V} \ , \ b_3 = s^{-1}\bar{W} \ , \ \tilde{b}_2 = (T+Q+R)\bar{V} \ , \ \tilde{b}_3 = (T+Q)\bar{W} \ ,$$

$$b_4 = (T_2+T_3+Q)\Delta \ , \ b_5 = \frac{3}{2}\{(T+Q)^2\bar{V} + T(T+2Q)\bar{W} + (T+Q)R_4\bar{V}\} \ ,$$

and $a_n > 0$. Taking $a_n = \max(n^{-1}, n^{-2}h^{-2d-4}, nh^{2L}, n^{-3/2}h^{-d-3})/\log n$, we can show that the second term on the right of (E.1) is $o(n^{-1} + n^{-2}h^{-2d-4} + nh^{2L} + n^{-3/2}h^{-d-3})$, similarly to (3.4)-(3.16) and Lemmas 9-19 of NR1 and Lemma 6 of R.

Putting $\chi^{++}(t) = \int e^{itz} dF^{++}(z)$, Esseen's smoothing lemma gives, for $N_0 = \log n \min(\eta n, n^2 h^{2d+4}, n^{3/2}h^{d+3})$, $\eta > 0$,

$$\sup_z |P(B \le z) - F^{++}(z)| \le \int_{-N_0}^{N_0} \left| \frac{Ee^{itB} - \chi^{++}(t)}{t} \right| dt + O(N_0^{-1})$$

$$\le \int_{-p}^{p} \left| \frac{Ee^{itB} - \chi^{+}(t)}{t} \right| dt + \int_{p \le |t| \le N_0} \left| \frac{Ee^{itB}}{t} \right| dt$$

$$+ \int_{|t| \ge p} \left| \frac{\chi^{++}(t)}{t} \right| dt + o(n^{-1} + n^{-2}h^{-2d-4})$$

$$= (I) + (II) + (III) + o(n^{-1} + n^{-2}h^{-2d-4}) .$$

for $p = \min(\log n, \epsilon n^{1/2})$, some $\epsilon > 0$. (III) is easily shown to be $o(n^{-1}, n^{-2}h^{-2d-4}, nh^{2L}, n^{-3/2}h^{-d-3})$, while (II) has the same order by (xiii).

Since $s^{-1}\Delta$ and T_1 are nonstochastic, putting $B_2 = B - b_2 - s^{-1}\Delta(1 + T_1)$,

$$E(e^{itB}) = e^{its^{-1}(\Delta + \Delta T_1)} E(e^{it(b_2 + B_2)}) \ , \qquad \text{(E.2)}$$

where, due to $s^{-1} = 1 + O(h^L)$,

$$e^{its^{-1}(\Delta + \Delta T_1)} = 1 + it\Delta(1 + T_1) + \frac{(it)^2\Delta^2}{2} + O(|t|^3\Delta^3 + |t|(\Delta + T_1)h^L) \ . \qquad \text{(E.3)}$$

Writing

$$E(e^{it(b_2+B_2)}) = E(e^{itb_2}) + itE(e^{itb_2}B_2) + \frac{(it)^2}{2}E(e^{itb_2}B_2^2)$$

$$+ \frac{(it)^3}{6}E(e^{itb_2}B_2^3) + \frac{(it)^4}{24}E(e^{itb_2}B_2^4) + O(|t|^5 E|B_2|^5) \ , \qquad \text{(E.4)}$$

we examine each of the right hand side terms. Note that the largest order term in B_2 is b_3 , and $E|b_3|^5 = O((n^{-1}h^{-d-2})^{5/2})$, by similar manipulation to equation (14) of R.

In the sequel, we say $A \approx B$ when

$$A = B + O(\{ \gamma(t) \}^{n-k} |t|^{m} (n^{-1} + n^{-2} h^{-2d-4} + n^{-3/2} h^{-d-3}))$$

for some integers k and m free of n, and $\gamma(t) = E w_1(t/s)$. We have

$$E(e^{itb_2 B_2}) = E\{ e^{itb_2} (b_3 + \tilde{b}_2 + \tilde{b}_3 + b_4 + b_5) \} \quad ,$$

with

$$E(e^{itb_2} b_3) \approx \{ \gamma(t) \}^{n-2} [\frac{4(it)^2}{\sqrt{n}} E(W_{12} V_1 V_2) + \frac{4(it)^3}{n} E(W_{12} V_1 V_2 (V_1 + V_2))] \quad ,$$

$$E(e^{it b_2} \tilde{b}_2) \approx \{ \gamma(t) \}^{n-3} \Big[-\frac{2it}{n} E(W_{12}^2) - \frac{8(it)^2}{n^{3/2}} E(V_1^3) E(W_{12}^2)$$

$$- \frac{4}{\sqrt{n}} \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \} - \frac{2it}{n} \{ 4E(V_1^4) - \frac{1}{4} + 4E(W_{12} V_1 V_2 (V_1 + V_2)) \}$$

$$- \frac{4(it)^2}{\sqrt{n}} \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \}$$

$$- \frac{(it)^3}{n} [16 \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \} E(V_1^3) + 4E(V_1^4) - \frac{1}{4} + 4E(W_{12} V_1 V_2 (V_1 + V_2))]$$

$$- \frac{16it}{n} E(W_{12} W_{13} V_2 V_3) - \frac{8(it)^3}{n} E(W_{12} W_{13} V_2 V_3)$$

$$- \frac{16it}{n} E\{ W_{12} V_1 V_2 (V_1 + V_2) \} - \frac{8(it)^3}{n} E\{ W_{12} V_1 V_2 (V_1 + V_2) \}$$

$$+ \frac{it}{n} + \frac{(it)^3}{2n} - \frac{4 + 8(it)^2}{n^{3/2}} E\{ W_{12}^2 (V_1 + V_2) \} \Big] \quad ,$$

$$E(e^{it b_2} \tilde{b}_3) \approx \{ \gamma(t) \}^{n-3} \Big[-\frac{8(it)^2}{n^{3/2}} E(W_{12}^2) E(W_{12} V_1 V_2)$$

$$- \frac{16(it)^3}{n} \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \} E(W_{12} V_1 V_2)$$

$$- \frac{4it}{n} [E\{ W_{12} V_1 V_2 (V_1 + V_2) \} + 4E(W_{12} W_{13} V_2 V_3)] - \frac{4}{n^{3/2}} E(W_{12} W_{13} W_{23})$$

$$- \frac{4}{n^{3/2}} E\{ W_{12}^2 (V_1 + V_2) \} + \frac{4}{n^{3/2}} E\{ W_{12}^2 (V_1 + V_2) \} \Big] \quad ,$$

$$E(e^{it b_2} b_4) \approx -\{ \gamma(t) \}^{n-2} 4(it) \frac{\Delta}{\sqrt{n}} \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \} \quad ,$$

$$E(e^{it b_2} b_5) \approx \frac{3}{2} \{ \gamma(t) \}^{n-4} \Big[\frac{4(it)}{n^2} \{ E(W_{12}^2) \}^2 + \frac{it}{4n} E(4V_1^2 - s^2 + 8\tilde{V}_1)^2$$

$$+ \frac{16 \{ it + (it)^3 \}}{n} \{ E(V_1^3) + 2E(W_{12} V_1 V_2) \}^2$$

$$+\frac{16E(W_{12}^2)}{n}\frac{1+(it)^2}{\sqrt{n}}\{E(V_1^3)+2E(W_{12}V_1V_2)\}\Bigg] .$$

The third expectation on the right of (E.4) is

$$E(e^{itb_2}B_2^2) = E\{e^{itb_2}(b_3+\bar{b}_2+\bar{b}_3+b_4+b_5)^2\} \approx E\{e^{itb_2}(b_3^2+\bar{b}_2^2+2b_3\bar{b}_2)\} ,$$

with

$$E(e^{itb_2}b_3^2) \approx \{\gamma(t)\}^{n-3}\Bigg[\frac{2}{n}E(W_{12}^2)+\frac{4it}{n^{3/2}}E\{W_{12}^2(V_1+V_2)\}$$
$$+\frac{16(it)^2}{n}E(W_{12}W_{13}V_2V_3)+\frac{16(it)^4}{n}\{E(W_{12}V_1V_2)\}^2\Bigg] ,$$

$$E(e^{itb_2}\bar{b}_2^2) \approx \{\gamma(t)\}^{n-4}\Bigg[\frac{4\{1+2(it)^2\}}{n^2}\{E(W_{12}^2)\}^2$$
$$+\frac{16\{2+5(it)^2+(it)^4\}}{n}\{E(V_1^3)+2E(W_{12}V_1V_2)\}^2+\frac{1+(it)^2}{4n}E(4V_1^2-s^2+8\tilde{V}_1)^2$$
$$+\frac{8\{3(it)+(it)^3\}}{n^{3/2}}E(W_{12}^2)\{E(V_1^3)+2E(W_{12}V_1V_2)\}\Bigg] ,$$

$$E(e^{itb_2}b_3\bar{b}_2) \approx \{\gamma(t)\}^{n-3}\Bigg[-\frac{4\{2(it)+(it)^3\}}{n^{3/2}}E(W_{12}^2)E(W_{12}V_1V_2)$$
$$-\frac{4\{1+(it)^2\}}{n}\{E(W_{12}V_1V_2(V_1+V_2)+4E(W_{12}W_{13}V_2V_3)\}$$
$$-\frac{16\{3(it)^2+(it)^4\}}{n}E(W_{12}V_1V_2)\{E(V_1^3)+2E(W_{12}V_1V_2)\}$$
$$-\frac{it}{n^{3/2}}E(W_{12}W_{13}W_{23})\Bigg] .$$

The fourth expectation on the right of (E.4) is

$$E(e^{itb_2}B_2^3) = E\{e^{itb_2}(b_3+\bar{b}_2+\bar{b}_3+b_4+b_5)^3\} \approx E\{e^{itb_2}(b_3^3+3b_3^2\bar{b}_2)\} ,$$

with

$$E(e^{itb_2}b_3^3) \approx \{\gamma(t)\}^{n-3}\Bigg[\frac{8}{n^{3/2}}E(W_{12}W_{13}W_{23})+\frac{8(it)^2}{n^{3/2}}E(W_{12}^2)E(W_{12}V_1V_2)\Bigg] ,$$

$$E(e^{itb_2}b_3^2\bar{b}_3) \approx \{\gamma(t)\}^{n-4}\Bigg[-\frac{2(it)}{n^2}\{E(W_{12}^2)\}^2$$
$$-\frac{8+8(it)^2}{n^{3/2}}E(W_{12}^2)\{E(V_1^3)+2E(W_{12}V_1V_2)\}\Bigg] .$$

The fifth expectation on the right of (E.4) is

$$E(e^{itb_2}B_2^4) = E\{e^{itb_2}(b_3+b_2+b_3+b_4+b_5)^4\} \approx E(e^{itb_2}b_3^4)$$

$$\approx \{\gamma(t)\}^{n-4}\frac{4}{n^2}\{E(W_{12}^2)\}^2 .$$

As in, e.g. Bickel et.al. (1986)

$$\{\gamma(t)\}^{n-k} = e^{-\frac{t^2}{2}}\Big[1+\frac{4(it)^3}{3\sqrt{n}}E(V_1^3) + \frac{1}{72n}[3(it)^4\{16E(V_1^4)-3\}+64(it)^6\{E(V_1^3)\}^2]\Big] + o\Big(\frac{|t|}{n}e^{-t^2/4}\Big) \qquad \text{(E.5)}$$

for $|t| \le \epsilon n^{1/2}$. Combining the above results gives eventually

$$E(e^{itB}) = \chi^{++}(t) + o(e^{-t^2/2}|t|^m(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-2})) + o(e^{-t^2/4}|t|n^{-1}) + O(|t|^5(n^{-1}h^{-d-2})^{5/2}) ,$$

where

$$\chi^{++}(t) = e^{-\frac{t^2}{2}}\Big[1+\frac{r_1(t)}{\sqrt{n}}-\frac{(it)^2\delta_2}{nh^{d+2}}+it\delta_1 n^{1/2}h^L + \frac{s_1(t)}{n}+\frac{\{8(it)^2+\frac{19}{6}(it)^4\}\delta_2^2}{n^2h^{2d+4}}+\frac{(it)^2\delta_1^2}{2}nh^{2L} + s_2(t)\delta_1 h^L+\frac{r_2(t)}{n^{3/2}h^{d+3}}-\frac{\{2(it)+(it)^3\}\delta_1\delta_2}{n^{1/2}}h^{L-d-2}\Big] .$$

$$r_1(t) = -4(it)(\delta_3+2\delta_4)-(it)^3(\frac{8}{3}\delta_3+4\delta_4), \quad r_2(t) = \{-4(it)+\frac{(it)^3}{3}\}\delta_6 ,$$

$$s_1(t) = (it)^2\xi_1+(it)^4\xi_2+(it)^6\xi_3 , \quad s_2(t) = (it)r_1(t) ,$$

and thus

$$\text{(I)} = \int_{-p}^{p}\Big|\frac{E(e^{itB})-\chi^{++}(t)}{t}\Big|dt = o(n^{-1}+n^{-2}h^{-2d-4}+nh^{2L}+n^{-3/2}h^{-d-3}) .$$

This proves (4.5). The proof of (4.6) proceeds by developing an analogous higher-order expansion for $\hat{F}^*(z)$, conditional on the sample, the algebraic calculations being identical to those above; see also Hall (1992, pp.88, 89, Section 5.2). □

REFERENCES

AMEMIYA, T. (1973): "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41, 997-1016.

ARCONES, M.A., AND E. GINÉ (1992): "On the Bootstrap of U and V Statistics," Annals of Statistics, 20, 655-674.

ARCONES, M.A., AND E. GINÉ (1994): "U-Process Indexed by Bapnik-Červonenkis Classes of Functions with Applications to Asymptotics and Bootstrap of U-Statistics with Estimated Parameters," Stochastic Processes and their Applications, 52, 17-38.

BABU, G.J., AND K. SINGH (1984): "On One Term Edgeworth Correction by Efron's Bootstrap", Sankhya, Series A, 46, 219-232.

BHATTACHARYA, R.N., AND M. QUMSIYEH (1989): "Second Order and L_p - Comparisons Between the Bootstrap and Empirical Edgeworth Expansion Methodologies,"Annals of Statistics, 17, 160-169.

BICKEL, P.J., AND D.A. FREEDMAN (1981): "Some Asymptotic Theory for the Bootstrap," Annals of Statistics, 9, 1196-1217.

BICKEL, P.J., F. GÖTZE, AND W.R. VAN ZWET (1986): "The Edgeworth Expansion for U-statistics of Degree two," Annals of Statistics, 14, 1463-1484.

DHARMARDHIKARI, S.W., V. FABIAN, AND K. JOGDEO (1968): "Bounds on the Moments of Martingales", Annals of Mathematical Statistics, 39, 1719-1723.

EFRON, B. (1979): "Bootstrap Methods: Another look at the Jackknife," Annals of Statistics," 7, 1-26.

FEUERVERGER, A., AND R.A.MUREIKA (1977): "The Empirical Characteristic Function and Its Applications", Annals of Statistics, 5, 88-97.

FREEDMAN, D.A. (1981): "Bootstrapping Regression Models," Annals of Statistics, 9, 1218-1228.

FREEDMAN, D.A. (1984): "Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models," Annals of Statistics, 12, 827-842.

HALL, P. (1986): "On the Bootstrap of Confidence Intervals," Annals of Statistics, 14, 1431-1452.

HALL, P. (1991): "Edgeworth Expansions for Nonparametric Density Estimators, with Applications," Statistics, 22, 215-232.

HALL, P. (1992): The Bootstrap and Edgeworth Expansion. New York: Springer-Verlag.

HÄRDLE, W., J. HART, J.S. MARRON, AND A.B. Tsybakov (1992): "Bandwidth Choice for Average Derivative Estimation", Journal of the American Statistical Association, 84, 218-226.

HELMERS, R. (1991): "On the Edgeworth Expansion and the Bootstrap Approximation for a Studentized *U*-Statistic," Annals of Statistics, 19, 470-484.

HOROWITZ, J.L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, 60, 3, 505-531.

HOROWITZ, J.L. (1997): "Bootstrap Methods in Econometrics: Theory and Numerical Performance," in Advances in Economics and Econometrics (D.M.Kreps and K.F. Wallis, eds.), Cambridge University Press, 188-222.

HOROWITZ, J.L. (1998): "Bootstrap Methods for Median Regression Models," Econometrica, 66, 6, 1327-1351.

HOROWITZ, J.L. (2002): "Bootstrap Critical Values for Tests Based on the Smoothed Maximum Score Estimator," Journal of Econometrics, 111, 141-167.

MAESONO, C.F. (1997): "Edgeworth Expansion of a Studentized U-Statistic and a Jackknife Estimator of Variance," Journal of Statistical Planning and Inference, 61, 61-84.

MANSKI, C.F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," Journal of Econometrics, 32, 65-108.

NISHIYAMA, Y., AND P.M.ROBINSON (2000): "Edgeworth Expansions for Semiparametric Averaged Derivatives", Econometrica, 68, 4, 931-980.

NISHIYAMA, Y., AND P.M.ROBINSON (2001): "Studentization in Edgeworth Expansions for Estimates of Semiparametric Single Index Models", Nonlinear Statistical Modeling (Festschrift for Takeshi Amemiya) (C.Hsiao, K.Morimune and J.Powell, eds.), Cambridge University Press, 197-240.

POWELL, J.L., J.H. STOCK, AND T.M. STOKER (1989): "Semiparametric Estimation of Index Coefficients", Econometrica, 57, 1403-1430.

POWELL, J.L., AND T.M. STOKER (1996): "Optimal Bandwidth Choice for Density-Weighted Averages", Journal of Econometrics, 75, 291-316.

QUMSIYEH, M. (1986): "Edgeworth Expansions in Regression and Comparison of Edgeworth Expansion and Bootstrap methodologies," PhD Dissertation, Indiana University.

ROBINSON, P.M. (1988): "Root-N-Consistent Semiparametric Regression", Econometrica, 56, 931-954.

ROBINSON, P.M. (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series", Review of Economic Studies, 56, 511-534.

ROBINSON, P.M. (1995): "The Normal Approximation for Semiparametric Averaged Derivatives", Econometrica, 63, 667-680.

SINGH, K. (1981): "On the Asymptotic Accuracy of Efron's Bootstrap," Annals of Statistics, 9, 1187-1195.

STRAWDERMAN, R.L., AND M.T.WELLS (1997): "Accurate Bootstrap Confidence Limits for the Cumulative Hazard and Survivor Functions Under Random Censoring", Journal of the American Statistical Association, 92, 1356-1374.

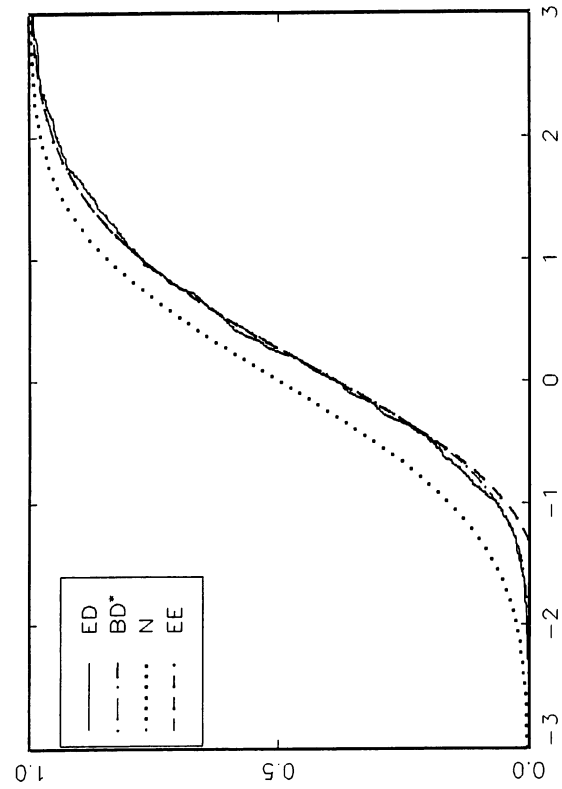


FIGURE 2. n=100 , h=0.6 , L=4

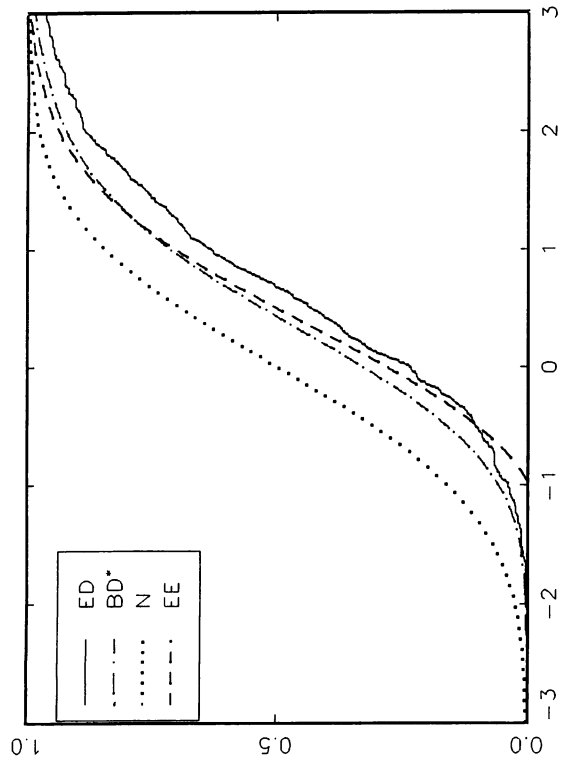


FIGURE 1. n=100 , h=0.8 , L=4

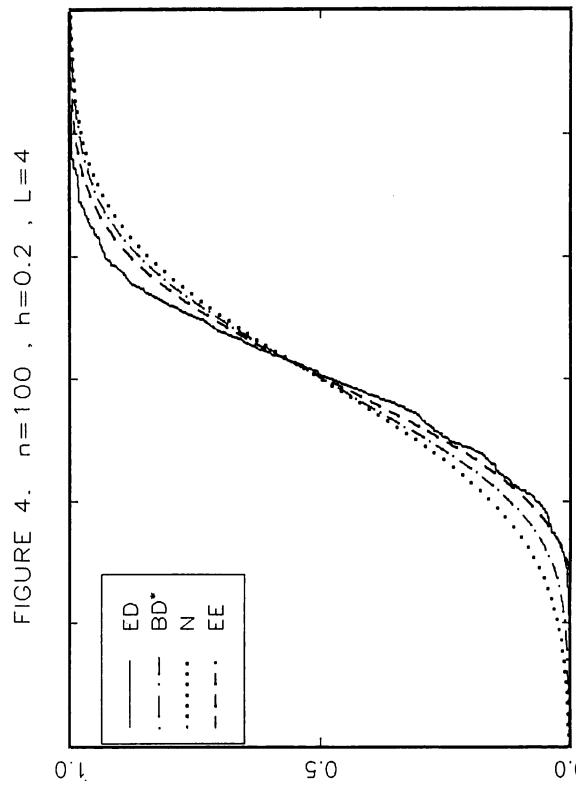


FIGURE 4. n=100 , h=0.2 , L=4

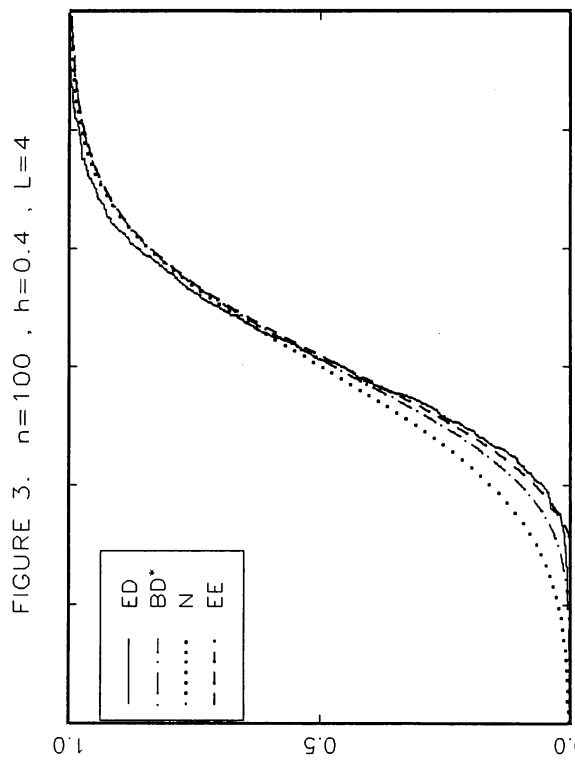


FIGURE 3. n=100 , h=0.4 , L=4

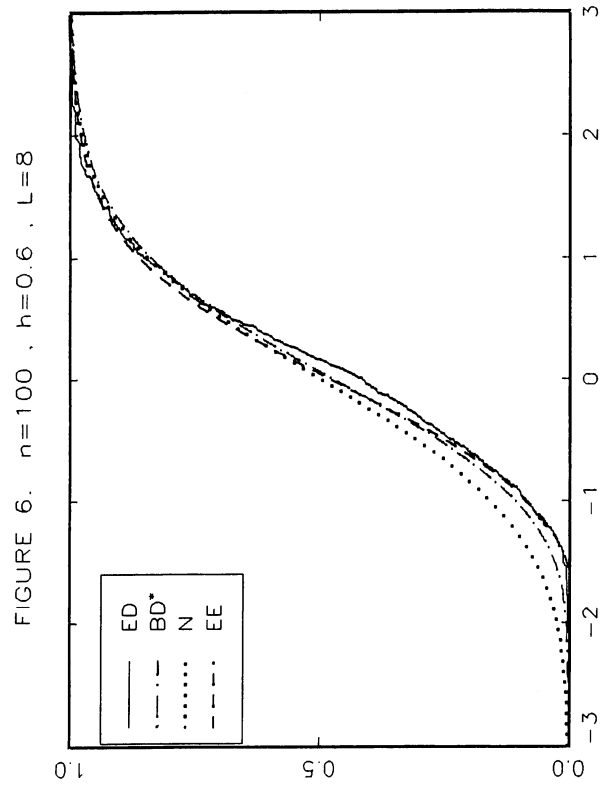


FIGURE 6. n=100 , h=0.6 , L=8

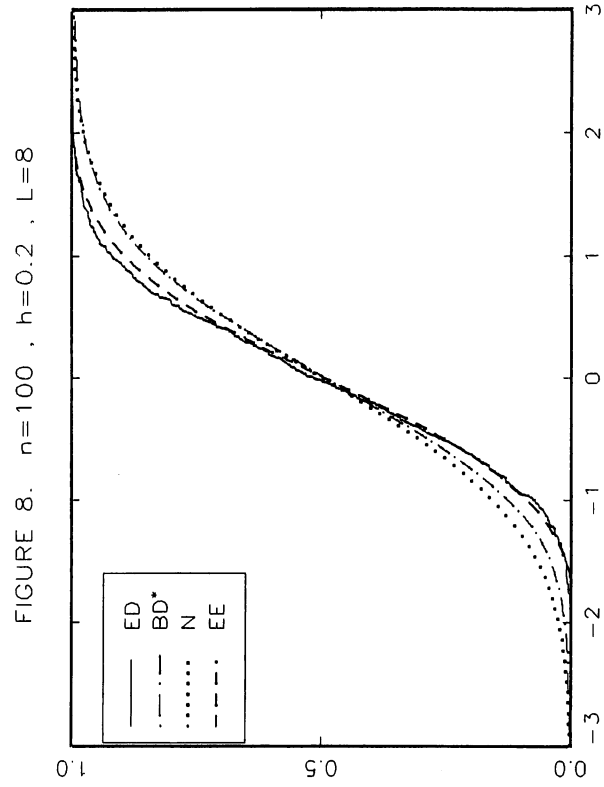


FIGURE 8. n=100 , h=0.2 , L=8

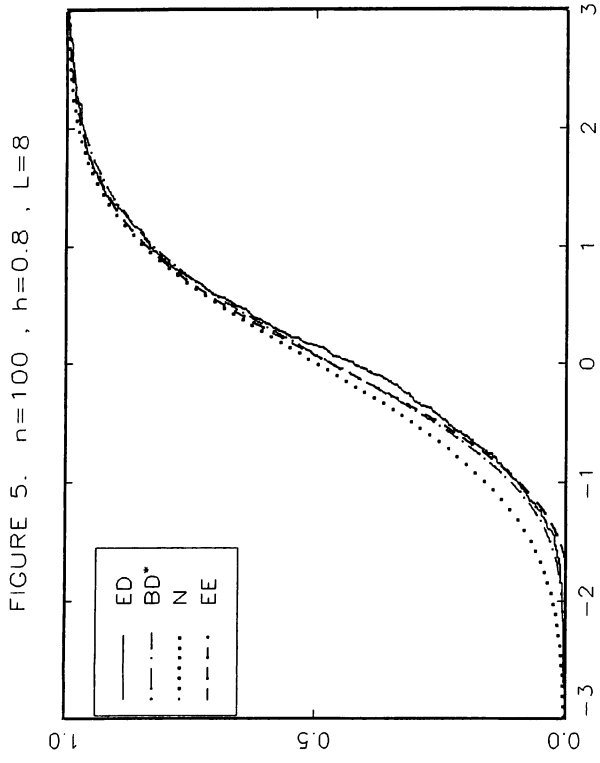


FIGURE 5. n=100 , h=0.8 , L=8

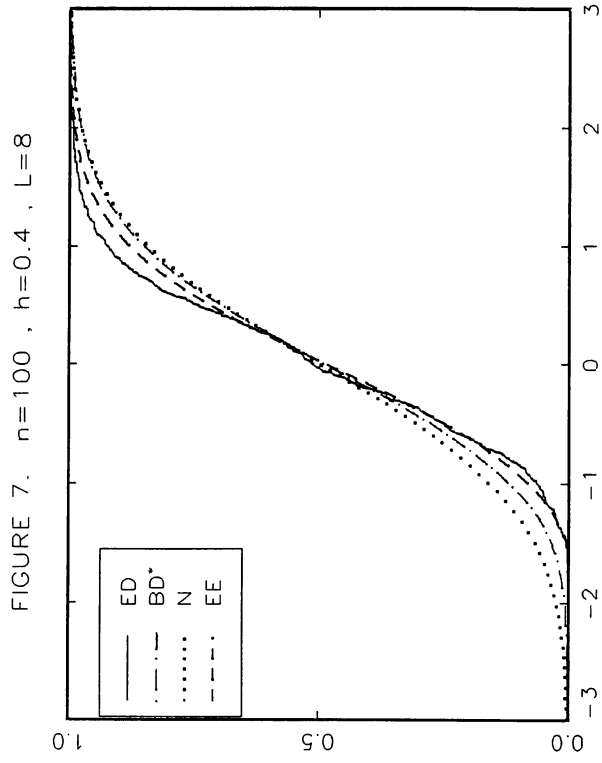


FIGURE 7. n=100 , h=0.4 , L=8

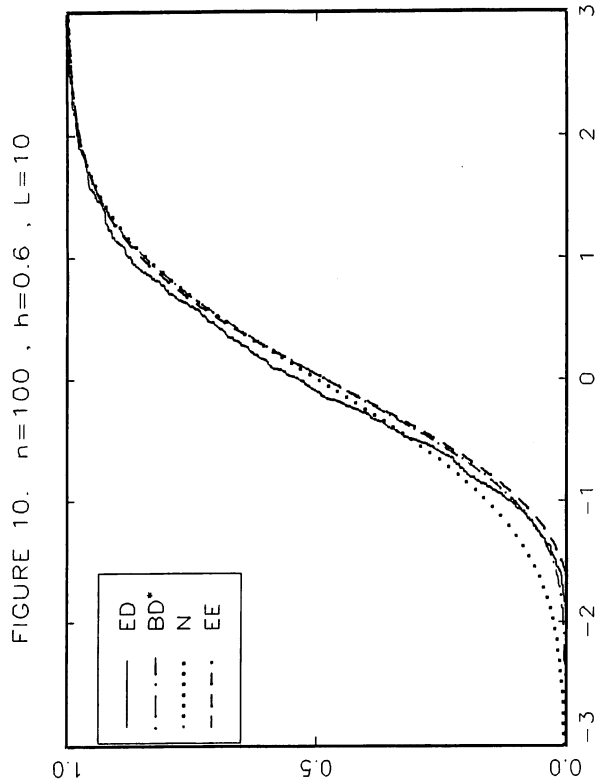


FIGURE 10. n=100 , h=0.6 , L=10

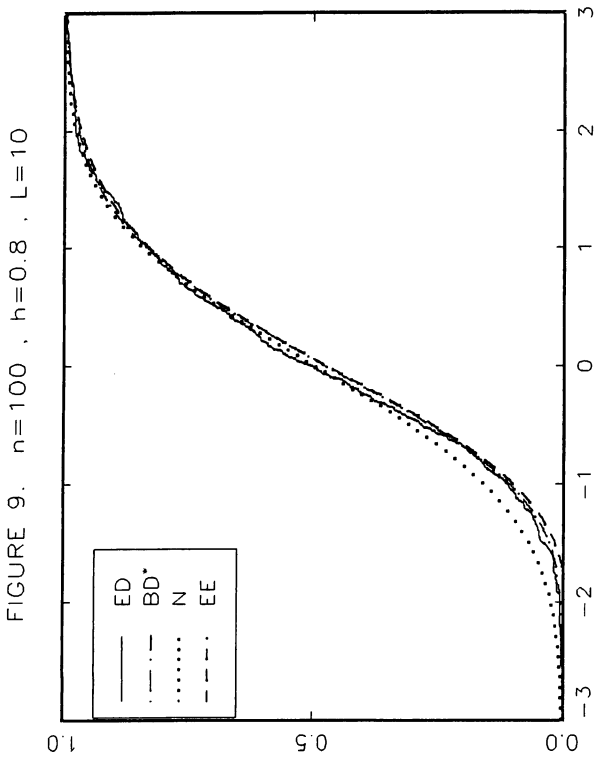


FIGURE 9. n=100 , h=0.8 , L=10

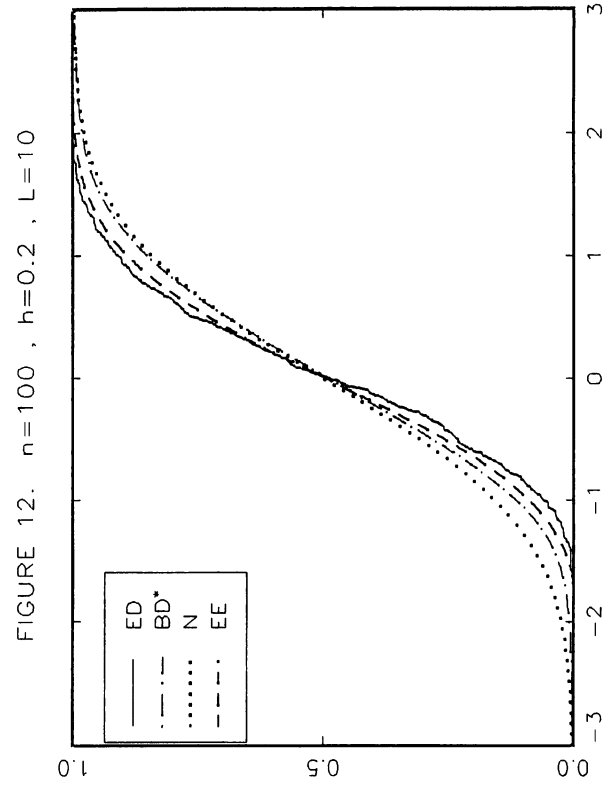


FIGURE 12. n=100 , h=0.2 , L=10

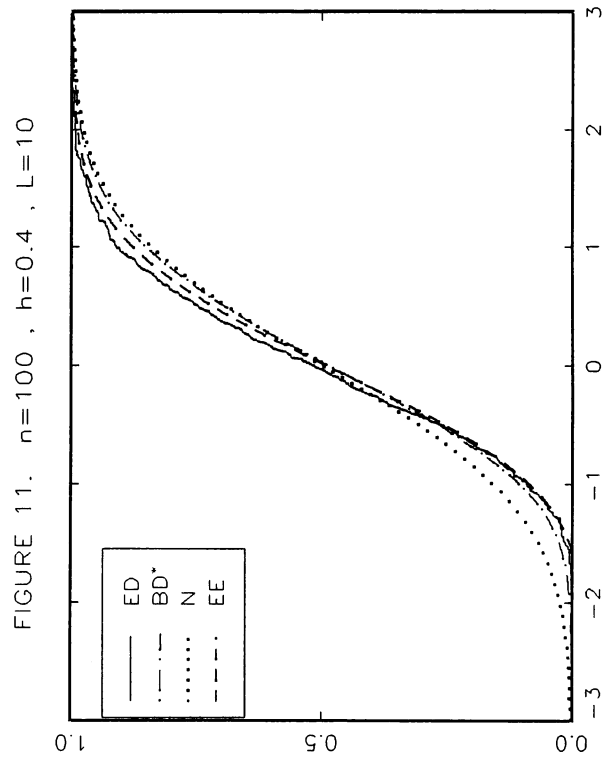


FIGURE 11. n=100 , h=0.4 , L=10

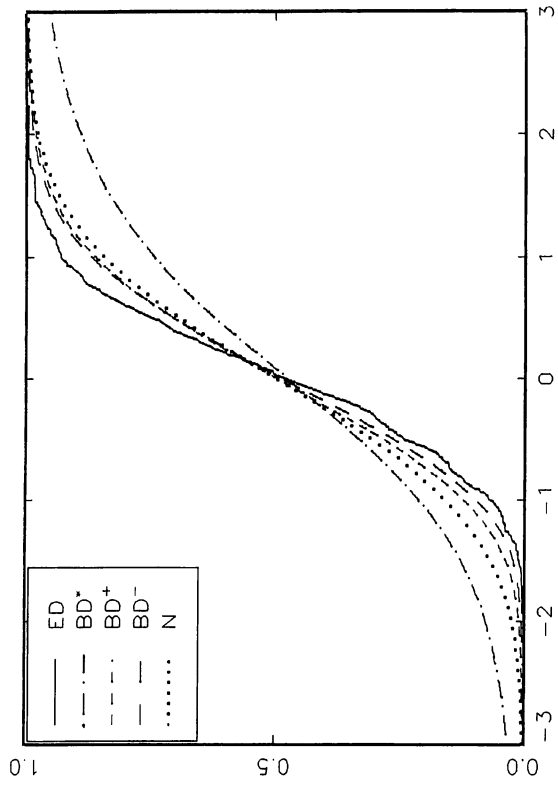


FIGURE 14. n=100 , h=0.2 , L=4

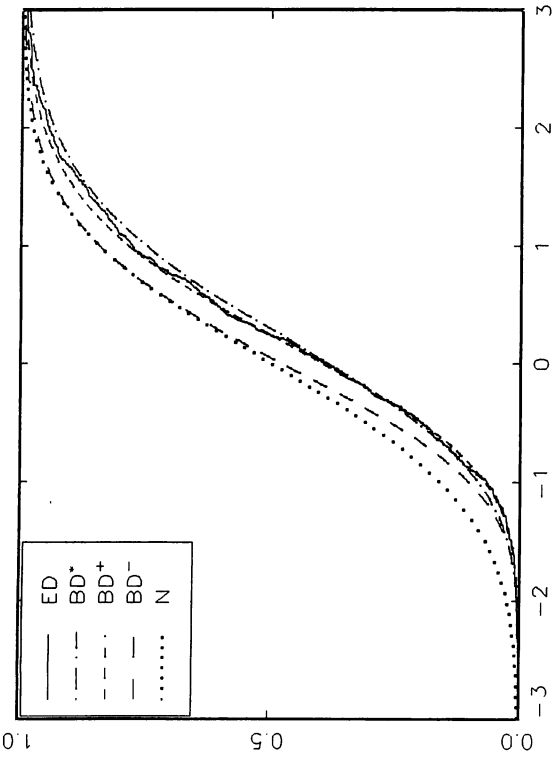


FIGURE 13. n=100 , h=0.6 , L=4

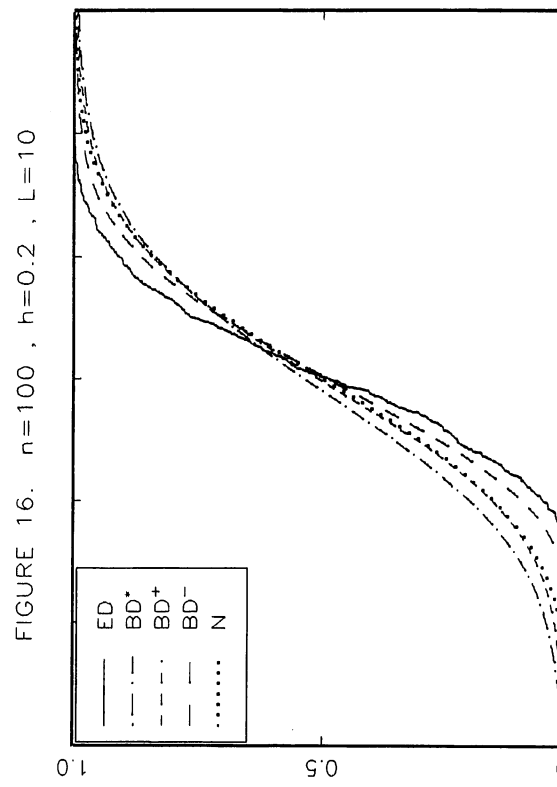


FIGURE 16. n=100 , h=0.2 , L=10

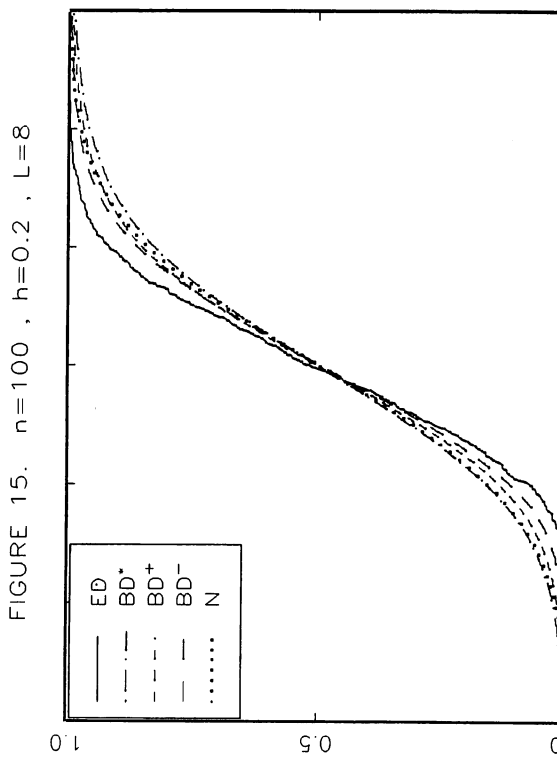


FIGURE 15. n=100 , h=0.2 , L=8

Table 1. Empirical size (%) of tests for $H_0: \bar{\mu} = \mu_0$ with nominal size 5%, $d=2$, $n=100$

L	n	h	(a) $H_1: \bar{\mu} \neq \mu_0$			(b) $H_1: \bar{\mu} < \mu_0$			(c) $H_1: \bar{\mu} > \mu_0$		
			B-test	*E*-test	*N*-test	*B*-test	*E*-test	*N*-test	*B*-test	*E*-test	*N*-test
4	100	0.2	1.3	3.3	0.3	2.5	4.5	0.0	1.7	2.0	0.3
		0.4	3.0	5.3	2.1	3.7	5.2	0.3	3.8	3.0	1.8
		0.6	5.1	6.6	6.3	3.8	7.5	0.3	5.7	5.2	6.0
		0.8	6.8	8.8	13.0	3.3	6.3	0.2	7.3	8.1	12.8
	400	0.2	1.6	3.3	0.3	3.8	4.5	0.0	2.3	3.5	0.3
		0.4	5.1	6.3	3.5	5.3	8.0	0.3	4.7	4.8	3.2
		0.6	4.8	6.5	7.8	4.1	6.5	0.0	5.6	5.0	7.8
		0.8	8.8	9.1	33.3	1.5	4.1	0.0	12.8	10.1	33.3
8	100	0.2	1.3	3.5	0.3	2.3	4.6	0.0	1.2	3.3	0.3
		0.4	2.0	3.5	0.7	2.5	5.0	0.0	1.9	2.5	0.7
		0.6	3.2	4.6	1.6	2.7	6.5	0.2	3.3	4.7	0.1
		0.8	4.7	6.5	4.0	4.3	7.0	0.5	4.1	4.8	3.5
	400	0.2	1.0	3.1	0.3	2.0	3.6	0.0	1.2	3.5	0.3
		0.4	2.3	5.0	1.7	3.3	5.3	0.3	3.2	4.0	1.3
		0.6	4.8	5.7	3.5	4.0	5.2	0.7	4.3	4.8	2.8
		0.8	6.0	6.3	5.7	4.2	4.6	1.2	5.7	5.6	4.5
10	100	0.2	0.0	2.0	0.0	0.0	3.2	0.0	0.5	3.5	0.0
		0.4	2.3	4.5	0.7	3.5	6.3	0.0	3.2	3.5	0.7
		0.6	3.5	6.8	2.0	4.2	7.7	0.0	4.0	4.0	2.0
		0.8	4.8	6.8	2.8	5.3	7.5	0.5	3.7	4.0	2.3
	400	0.2	3.0	4.5	0.2	1.7	7.0	0.0	2.8	3.8	0.2
		0.4	3.6	4.2	1.5	4.2	6.0	1.0	2.7	3.7	0.5
		0.6	5.8	5.5	4.0	6.1	5.3	1.5	3.1	4.3	2.5
		0.8	4.5	5.2	4.5	5.3	5.6	1.0	5.2	5.0	3.5

Table 2. Empirical size (%) of tests of $H_0: \bar{\mu} = \mu_0$ with nominal size 5%, $d=4$, $n=100$

L	h	(a) $H_1: \bar{\mu} \neq \mu_0$		(b) $H_1: \bar{\mu} < \mu_0$		(c) $H_1: \bar{\mu} > \mu_0$	
		B-test	*N*-test	*B*-test	*N*-test	*B*-test	*N*-test
4	0.4	1.2	0.8	2.4	0.0	1.6	0.2
	0.6	4.8	3.1	5.0	0.0	4.2	3.1
	0.8	5.7	13.2	3.7	0.1	6.2	12.8
	1.0	6.0	38.0	2.1	0.0	7.0	38.8
8	0.4	0.0	0.0	0.1	0.0	0.2	0.3
	0.6	1.3	0.3	3.0	0.1	2.0	0.8
	0.8	3.0	1.4	3.4	0.0	2.5	1.3
	1.0	4.3	3.1	3.7	0.1	4.5	3.0
10	0.4	0.1	0.6	1.7	0.0	1.5	0.6
	0.6	2.3	0.8	3.8	0.0	2.3	0.8
	0.8	3.1	2.1	4.5	0.1	5.0	2.0
	1.0	4.8	5.0	5.1	0.1	6.1	4.8