EDGEWORTH EXPANSIONS FOR SEMIPARAMETRIC AVERAGED DERIVATIVES*

by

Y Nishiyama and P M Robinson
Department of Economics, London School of Economics

Contents:

Discussion Paper
No.EM/99/373
October 1999

The Suntory Centre
Suntory and Toyota International Centres
for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London, WC2A 2AE
Tel.: 0171-405 7686

---

* Research supported by ESRC Grant R000235892. The second author's research was also supported by a Leverhulme Trust Personal Research Professorship. The comments of the referees and co-editor have led to significant improvements.

Abstract

A valid Edgeworth expansion is established for the limit distribution of density-weighted semiparametric averaged derivative estimates of single index models. The leading term that corrects the normal limit varies in magnitude, depending on the choice of bandwidth and kernel order. In general this term has order larger than the $n^{-½}$ that prevails in standard parametric problems, but we find circumstances in which it is $O(n^{-½})$, thereby extending the achievement of an $n^{-½}$ Berry-Essen bound in Robinson (1995). A valid empirical Edgeworth expansion is also established. We also provide theoretical and empirical Edgeworth expansions for a studentized statistic, where the correction terms are different from those for the unstudentized case. We report a Monte Carlo study of finite sample performance.

Keywords: Edgeworth expansion; semiparametric estimates; averaged derivatives.

JEL Nos.: C21, C24

1. INTRODUCTION

First-order large sample distribution theory of estimates of semiparametric econometric models has been extensively studied. A major recent focus has concerned inference on the parametric component, when the nonparametric curve is estimated by some method of smoothing, such as kernels or nearest neighbours (see e.g. Manski (1984), Robinson (1987), Powell, Stock and Stoker (1989), Newey (1990)). In some such cases, estimates actually achieve the first order efficiency of optimal ones based on a fully parametric model, and more generally they are asymptotically normal and achieve the same rate of convergence as parametric estimates, namely being $n^{\frac{1}{2}}$ -consistent, where n is sample size.

There is no reason to suppose that these correspondences even approximately occur in small or moderate sample sizes. Indeed the smoothed nonparametric estimates involved in the semiparametric estimation converge more slowly than $n^{\frac{1}{2}}$, which could reasonably be expected to affect finite sample performance, and indeed many Monte Carlo studies have demonstrated a sensitivity to the precise implementation of the nonparametric estimates. Analytic study of the finite-sample distribution theory for semiparametric estimates seems mathematically intractable, and indeed the precise distributional assumptions which such a theory would require are incompatible with the ethos of semiparametric inference.

On the other hand, higher-order asymptotic theory, which also has the potential to shed light on finite-sample performance, seems feasible for semiparametric estimates, under acceptably general conditions. Parametric estimates typically enjoy a Berry-Esseen bound of order $n^{-\frac{1}{2}}$ (see e.g. Pfanzagl (1971)), and Nagar (see e.g. Nagar (1959)) and Edgeworth (see e.g. Sargan and Mikhail (1971)) expansions in powers of $n^{-\frac{1}{2}}$. Due to the nonparametric smoothing, the semiparametric estimates might be expected to have a larger Berry-Esseen bound, and correction term of order greater than $n^{-\frac{1}{2}}$. If so, the semiparametric estimates are inferior to parametric ones in the sense that their distribution converges to the normal limit more slowly, while the Bartlett corrections advanced in the parametric literature will be unsuccessful and bootstrap replications (e.g. Hall (1992)) would not provide the usual second-order correctness.

Some recent papers have investigated higher-order properties of semiparametric estimates. The Berry-Esseen bound for averaged derivative estimates of semiparametric index models was derived by Robinson (1995). He found that while in general the bound is larger than $n^{-\frac{1}{2}}$, it is nevertheless possible to implement the estimate, by appropriate choice of smoothing or bandwidth number and of kernel order, to achieve the $n^{-\frac{1}{2}}$ bound, opening up the possibility that some semiparametric estimates can rival the higher-order and bootstrap properties of parametric estimates. Nagar expansions were developed by Linton (1995, 1996b), for estimates of the semiparametric partly linear model and of the linear regression model with disturbance heteroscedasticity of unknown form. Linton (1995, 1996b) found that the leading terms are of order greater than $n^{-\frac{1}{2}}$ and showed how their contribution might be minimized by appropriate choice of bandwidth. In another paper, Linton (1996a) established valid Nagar and Edgeworth expansions for a wide class of semiparametric estimates. Making assumptions of a high-level type, including that the nonparametric estimate converges suitably fast, Linton (1996a) showed that the nonparametric estimation has no effect on expansions to order n^{-1}, and indicated that his assumptions can be satisfied by a version of the partly linear model as well as in models where no smoothing is involved.

The present paper develops a valid Edgeworth expansion for semiparametric density-weighted averaged derivative estimates of semiparametric index models. Such estimates were shown to be $n^{\frac{1}{2}}$ - consistent and asymptotically normal for independent and identically distributed (iid) observations by Powell, Stock and Stoker (1989) and for weakly dependent observations by Robinson (1989), while Cheng and Robinson (1994) found that a non-normal limit could pertain in the event of some long range dependence. The single index model includes a number of practically important special cases, such as probit, Tobit and Box-Cox and other transformation models, and averaged derivative estimation has proved popular. However, as in the Berry-Esseen theory of Robinson (1995), density-weighted averaged derivatives are chosen for study in large part by virtue of their algebraic simplicity relative to the bulk of other semiparametric estimates; even in this case the details of higher-order theory are complicated, and they would be more so in others, such as in ones employing trimming to handle the effects of stochastic denominators, where we are unable to say whether similar qualitative conclusions to ours can be reached.

Our work differs significantly from the Edgeworth theory in the aforementioned Linton (1996a) reference. Averaged derivatives are not among the illustrations Linton employs, and in fact do not in general satisfy his orthogonality condition B4(2). Whereas Linton employs a fixed design (as in Linton (1995, 1996b)) we do not condition on our stochastic explanatory variables, as in the bulk of first-order theory for semiparametric econometric estimates, including that for averaged derivatives, such as in the Berry-Esseen theory of Robinson (1995). Unlike Linton (1996a) we do not achieve an expansion to order n^{-1}, but rather focus on the extent to which an $n^{-\frac{1}{2}}$ term may be dominated by other terms. These latter involve the bandwidth, such that the second term in the Edgeworth expansion varies with respect to the choice of bandwidth, which is suppressed in the treatment of Linton (1996a), due to his assumption of better - than - $n^{\frac{1}{4}}$ - consistency of the nonparametric estimates, which our conditions do not necessarily satisfy. We provide a valid empirical Edgeworth expansion for practical use. Linton's paper does not overlap with our detailed treatment of a different and more specialized problem, under primitive conditions.

Since our estimate is of U-statistic form, our work can also be compared with that on Edgeworth expansions of U-statistics in the mathematical statistics literature (see Callaert, Janssen and Veraverbeke (1980), Bickel, Götze and van Zwet (1986), and a recent treatment of more general symmetric statistics due to Bentkus, Götze and van Zwet (1997)). However the dependence of our U-statistic "kernel" on the bandwidth, and thence on sample size, prevents us from applying the results of these authors, and while our proofs sometimes employ similar techniques to those in the first two of these papers, our work can be seen more as an extension of the treatment of averaged derivatives in Robinson (1995), a number of whose intermediate results we use or extend. As in Robinson (1995), we overcome a serious bias problem by resorting to higher-order kernels (in the nonparametric estimation). Though our conclusions are substantially stronger than those of Robinson (1995), our conditions (for the theoretical Edgeworth expansion) do not seem to be, with the notable but predictable exception of the addition of a Cramér condition.

The following section describes the single index model, the averaged derivative estimate and theoretical and empirical Edgeworth expansions for the estimate normalized by its asymptotic variance matrix, with regularity conditions for validity. Section 3 proposes a jackknife estimate of the asymptotic variance matrix and provides valid Edgeworth expansions when the averaged derivative statistic is studentized by this variance estimate. Section 4 discusses special cases covered by the Edgeworth expansion in Section 3 and, based on this, derives an optimal bandwidth choice which minimizes the normal approximation error, and a data-dependent approximation to this for practical use. The proofs of Theorems in Section 2, along with a number of technical lemmas, are left to appendices, the substantial extra details needed to complete the proofs of Theorems 3 and 4 appearing in a companion paper, Nishiyama and Robinson (1998). Section 5 reports the results of a Monte Carlo study of finite sample performance based on a Tobit model.

2. EDGEWORTH EXPANSIONS : UNSTUDENTIZED CASE

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

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

for some $G: R\to R$ and some column vector β, τ denoting transposition. For example, letting V be a scalar variate independent of X and with distribution function F, and $1(\cdot)$ the indicator function, if $Y=1(\beta^{\tau}X+V>0)$ or $Y=(\beta^{\tau}X+V)1(\beta^{\tau}X+V>0)$ or, for some increasing function t, $t(Y)=\beta^{\tau}X+V$, we have (2.1) with respectively $G(u)=1-F(-u)$, $G(u)=u\{1-F(u)\}+\int vdF(v)$ and $G(u)=\int t^{-1}(u+v)dF(v)$.
In the first case we have respectively the probit or logit model when V is normal or logistic, in the second the Tobit and in the last, various transformation models arise on parameterising F and

t. Then β can be estimated $n^{\frac{1}{2}}-$ consistently and asymptotically normally and efficiently by maximum likelihood. If F has been misspecified, such parametric approaches in general lead to inconsistent estimates. Regarding F, and thus G, as nonparametric, β can be identified only up to scale, for example, by

(2.2) $$\bar{\mu} \overset{def}{=} -E\{g'(X)f(X)\} = c\beta \ ,$$

where $c = -E\{G'(\beta^{\tau}X)f(X)\}$, f is the density of X, the prime denotes differentiation and the final equality in (2.2) follows from the chain rule. On the other hand, under conditions imposed below, integration by parts gives

$$\bar{\mu} = 2E\{g(X)f'(X)\} = 2E\{Yf'(X)\}$$

which can be estimated by the density-weighted averaged derivative statistic

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

given the sample (X_i^{τ}, Y_i), $i = 1, \cdots, n$ from (X^{τ}, Y), where

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

$K : R^d \to R$ is even, differentiable and integrates to one and $h > 0$ converges to zero as $n \to \infty$.

For a function $k : R^d \to R$, write $k = k(X), k' = k'(X), k'' = k''(X), k''' = k'''(X)$ where

$$k'(x) = \partial k(x)/\partial x, k''(x) = \partial^2 k(x)/\partial x \partial x^{\tau}, k'''(x) = \partial\, vec(k''(x))/\partial x^{\tau},$$

and define $q = E(Y^2|X)$, $r = E(Y^3|X)$, $s = q - g^2$, $\mu = \mu(X,Y) = Yf' - e'$, $e = fg$, $a = g'f + \bar{\mu}$, $a' = g''f + g'f'^{\tau}$, $\Sigma = 4E(\mu - \bar{\mu})(\mu - \bar{\mu})^{\tau}$.

We introduce the following assumptions.

(i) $E|Y|^3 < \infty$.

(ii) Σ is finite and positive definite.

(iii) The underlying measure of (X^{τ}, Y) can be written as $\mu_X \times \mu_Y$, where μ_X and μ_Y are Lebesgue measure on R^d and R respectively. (X_i^{τ}, Y_i) are iid observations on (X^{τ}, Y) .

(iv) f is $(L+1)$ times differentiable, and f and its first $(L+1)$ derivatives are bounded, for $2L>d+2$.

(v) g is $(L+1)$ times differentiable, and e and its first $(L+1)$ derivatives are bounded.

(vi) q is twice differentiable and q', q'', g', g'', g''', $E(|Y|^3 | X)f$, and qf' are bounded.

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

(viii) $K(u)$ satisfies $K(u) = K(-u)$, is differentiable,

$$\int_{R^d}\{(1+\|u\|^L)|K(u)| + \|K'(u)\|\}du + \sup_{u \varepsilon R^d}\|K'(u)\| < \infty ,$$

and for the same L as in (iv) and (v),

$$\int_{R^d} u_1^{l_1} \cdots u_d^{l_d} K(u)du \begin{cases} = 1, & if \quad l_1+\cdots+l_d=0 \\ = 0, & if \; 0<l_1+\cdots+l_d<L \\ \neq 0, & if \quad l_1+\cdots+l_d=L \end{cases} .$$

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

(x) For a $d \times 1$ vector ν ,

$$\sup_{\nu:\nu^\tau\nu=1} \limsup_{|t|\to\infty} |E\exp[\{it2\sigma_\nu^{-1}\nu^\tau(\mu-\bar{\mu})\}]| < 1 ,$$

where $\sigma_\nu^2 = \nu^\tau \Sigma \nu$.

Assumptions (i)-(iv) and (viii) are identical to corresponding ones of Robinson (1995), which are discussed there, Assumption (viii) referring to a higher-order kernel *K*; such kernels have a long history in bias-reduction of nonparametric estimates, were used by Robinson (1988) and subsequent authors to achieve $\sqrt{n}$ - consistent semiparametric estimation, and by Robinson (1995) to control the Berry-Esseen bound of averaged derivative estimates. Assumptions (v)-(vii) and (ix) somewhat strengthen corresponding ones of Robinson (1995), and Assumption (x) is a Cramér condition (see e.g. Bhattacharya and Rao (1976)) (note that σ_ν^2 is bounded away from zero under (ii)). In their study of ordinary U-statistics and symmetric statistics, Callaert, Janssen and Veraverbeke (1986) and Bentkus, Götze and van Zwet (1997) employ more stringent conditions of Cramér type; however in the context of their Edgeworth expansion in powers of $n^{-\frac{1}{2}}$, the former authors establish an expansion to order n^{-1} (with remainder $o(n^{-1})$) while the latter authors expand to $n^{-\frac{1}{2}}$ with $O(n^{-1})$ remainder. Our expansion is to order g_n, say where g_n is a sum of $n^{-\frac{1}{2}}$ plus other terms, which may be of bigger or smaller order than $n^{-\frac{1}{2}}$ depending on the bandwidth h, with remainder term that is only shown to be $o(g_n)$. The ν^{τ} factor in Assumption (ix) is due to the fact that we consider only expansions for a single linear combination of the vector averaged derivative statistic U (and a studentized version of this). The development of full multivariate expansions would require further work (we cannot appeal to the Cramèr-Wold device). Our present set-up allows higher-order inference on individual elements of β (up to scale), which is of practical importance in itself, as well as on arbitrary single linear combinations of β.

Define further

$$Z = n^{1/2}\sigma_{\nu}^{-1}\nu^{\tau}(U-\bar{\mu}) \;,\; F(z) = P(Z \le z) \;,$$

(2.3)
$$\tilde{F}(z) = \Phi(z)-\varphi(z)\left\{n^{1/2}h^{L}\kappa_1 + \frac{\kappa_2}{nh^{d+2}}z + \frac{4(\kappa_3+3\kappa_4)}{3n^{1/2}}(z^2-1)\right\} \;,$$

where z is real-valued, Φ and φ are respectively the distribution and density function of a standard normal variate, and writing

$$\Delta^{(l_1,\cdots,l_d)} = \left.\frac{\partial^{(l_1+\cdots+l_d)}}{\partial x_1^{l_1}\cdots\partial x_d^{l_d}}\right|_{(x_1,\cdots,x_d)^{\tau}=X} \;,$$

(2.4)
$$\kappa_1 = \frac{2(-1)^{L}\sigma_{\nu}^{-1}}{L!}\sum_{\substack{0\le l_1,\cdots,l_d\le L\\ l_1+\cdots+l_d=L}}\cdots\sum\left\{\int\prod_{i=1}^{d}u_i^{l_i}K(u)du\right\}E\left[\left(\Delta^{(l_1,\cdots,l_d)}\nu^{\tau}f'\right)g\right] \;,$$

$$\kappa_2 = 2\sigma_{\nu}^{-2}\int\left\{\nu^{\tau}K'(u)\right\}^2 du\, E(sf) \;,$$

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

$$\kappa_4 = -\sigma_{\nu}^{-3}E\Big[fs(\nu^{\tau}f')(\nu^{\tau}a'\nu)-f(\nu^{\tau}f')(\nu^{\tau}s')(\nu^{\tau}a)$$
$$-fs(\nu^{\tau}a)(\nu^{\tau}f''\nu)+f(\nu^{\tau}g')(\nu^{\tau}a)^2\Big] \;,$$

where the κ_i are all finite under our assumptions. The nature and role of the κ_i are discussed in the following section.

THEOREM 1 : *Under Assumptions* (i)-(x), *as* $n\to\infty$,

$$\sup_{\nu:\nu^{\tau}\nu=1}\sup_{z\varepsilon R}\left|F(z)-\tilde{F}(z)\right| = o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^{L}) \;.$$

We call $\tilde{F}(z)$ a theoretical Edgeworth expansion of $F(z)$, Theorem 1 establishing its validity. We can derive an empirical Edgeworth expansion by replacing the population κ_i in

(2.3) by strongly consistent estimates

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

$$\tilde{\kappa}_2 = \hat{\sigma}_\nu^{-2} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} h^{d+2} \tilde{W}_{ij}^2 , \qquad \tilde{\kappa}_3 = \frac{\hat{\sigma}_\nu^{-3}}{n} \sum_{i=1}^{n} \tilde{V}_i^3 ,$$

$$\tilde{\kappa}_4 = \frac{\hat{\sigma}_\nu^{-3}}{n(n-1)} \sum_{i=1}^{n} \sum_{j\ne i}^{n} \nu^\tau U_{ij} \tilde{V}_i \tilde{V}_j ,$$

where $\hat{\sigma}_\nu^2 = \nu^\tau \hat{\Sigma} \nu$,

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

and for positive b and a function $H : R^d \to R$

$$\tilde{f}(X_i) = \frac{1}{(n-1)b^d} \sum_{j\ne i}^{n} H\left(\frac{X_i-X_j}{b}\right) ,$$

(2.5) $$\tilde{U}_i = \frac{1}{n-1} \sum_{j\ne i}^{n} U_{ij} ,$$

$$\tilde{V}_i = \nu^\tau(\tilde{U}_i - U) , \quad \tilde{W}_{ij} = \nu^\tau(U_{ij} - \tilde{U}_i - \tilde{U}_j + U) .$$

$\hat{\Sigma}$ is a jackknife estimate of Σ . It may be observed that, notwithstanding the form of κ_3 and κ_4, $\tilde{\kappa}_3$ and $\tilde{\kappa}_4$ do not entail explicit estimation of derivatives.

To establish validity of our empirical Edgeworth expansion we require some strengthening of some of Assumptions (i)-(x), and additional assumptions.

(i)' $E(Y^6) < \infty$.

(iv)' f is (L+2) times differentiable, and f and its first (L+2) derivatives are bounded, where 2L>d+2.

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

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

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

$$\int_{R^d} H(u)du = 1$$

and

$$\int_{R^d} \| \Delta^{(l_1,\cdots,l_d)} H'(u) \| du + \sup_{u \varepsilon R^d} \| \Delta^{(l_1,\cdots,l_d)} H'(u) \| < \infty$$

for all integers $l_1, \cdots, l_d$ satisfying $0 \le l_1 + \cdots + l_d \le L$.

(xii) $b \to 0$ and $\frac{(\log n)^2}{nb^{d+2+2L}} = O(1)$ as $n \to \infty$.

Notice that H need only be a second-order kernel, whereas K has to be a higher-order one unless $d=1$. It is possible to choose $H(u)=K(u)$ with Assumptions (viii) and (xi) simultaneously satisfied. However, in comparing (xii) with (ix) it seems that b might in general be chosen larger than h, while there is a case for avoiding the use of higher-order kernels when possible.

Define

(2.6) $$\tilde{\tilde{F}}(z) = \Phi(z) - \varphi(z)\left\{n^{1/2}h^L\tilde{\kappa}_1 + \frac{\tilde{\kappa}_2}{nh^{d+2}}z + \frac{4(\tilde{\kappa}_3 + 3\tilde{\kappa}_4)}{3n^{1/2}}(z^2-1)\right\} .$$

THEOREM 2 : *Under Assumptions* (i)', (ii), (iii), (iv)', (v)', (vi)-(viii), (ix)*and* (x)-(xii),

$$\sup_{\nu:\nu^\tau\nu=1}\sup_{z\varepsilon R}\left|F(z) - \tilde{\tilde{F}}(z)\right| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) \text{ almost surely.}$$

3. EDGEWORTH EXPANSIONS : STUDENTIZED CASE

Theorems 1 and 2 concern *Z* which involves unknown Σ through σ_ν^2 so that they fall short of being fully operational. The same criticism can be levelled against much of the econometric and statistical literature on Edgeworth expansions, but we nevertheless wish to develop the previous discussion by considering the studentized statistic $\hat{Z} = n^{1/2}\hat{\sigma}_\nu^{-1}\nu^\tau(U-\bar{\mu})$. We first validly approximate

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

by the theoretical Edgeworth expansion

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

THEOREM 3 : *Under Assumptions* (i)', (ii)-(x), *as* $n \to \infty$,

$$\sup_{\nu:\nu^\tau\nu=1}\sup_{z\varepsilon R}\left|\hat{F}(z) - F^+(z)\right| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) .$$

The correction terms in $F^{+}(z)$ are of the same orders as those in the unstudentized case (see Theorem 1), though their coefficients differ.

The κ_i are unknown, but a feasible, empirical Edgeworth expansion is

$$\hat{F}^{+}(z) = \Phi(z) - \varphi(z)\left[n^{1/2}h^{L}\tilde{\kappa}_1 - \frac{\tilde{\kappa}_2}{nh^{d+2}}z - \frac{4}{3n^{1/2}}\left\{(2z^2+1)\tilde{\kappa}_3 + 3(z^2+1)\tilde{\kappa}_4\right\}\right] .$$

THEOREM 4 : *Under Assumptions* (i)', (ii), (iii), (iv)', (v)', (vi)-(viii), (ix)' *and* (x)-(xii),

$$\sup_{\nu:\nu^{\tau}\nu=1}\sup_{z\varepsilon R}\left|\hat{F}(z) - \hat{F}^{+}(z)\right| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \text{ almost surely.}$$

The conditions in Theorem 3 strengthen those in Theorem 1 only with respect to the moment condition on Y, while Theorem 2's conditions are identical to Theorem 4's. The proofs of Theorems 3 and 4 entail considerable additional work beyond that in the proofs of Theorems 1 and 2 which are already lengthy and technical, so we have instead reported the former proofs in Nishiyama and Robinson (1998). However in the following section we analyze special cases of Theorem 3 and thereby deduce a novel form of optimal *h*, which can be approximated for practical use, and we include studentized statistics along with unstudentized ones, as well as our bandwidth proposal, in the Monte Carlo study of Section 5.

4. SPECIAL CASES AND BANDWIDTH CHOICE

Theorem 3 covers a number of situations, depending on the choice of kernel order *L*, relative to dimension *d*, and on the rate of decay of the bandwidth *h*. We classify these according to *L* and then *h*. Let $C_i,\ i=1,2,3,4,$ be finite positive constants.

I. $\frac{d+2}{2} < L < 2(d+2)$.

(a) If $n^3h^{2(L+d+2)} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z\varphi(z)}{nh^{d+2}}\{1 + o(1)\} .$$

(b) If $h \sim C_1 n^{-\frac{3}{2(L+d+2)}}$,

$$\hat{F}(z) = \Phi(z) - (C_1^{L}\kappa_1 - \frac{\kappa_2 z}{C_1^{d+2}})\varphi(z)n^{\frac{-2L+d+2}{2(L+d+2)}}\{1 + o(1)\} .$$

(c) If $n^3 h^{2(L+d+2)} \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1 \varphi(z) n^{1/2} h^L \{1 + o(1)\} .$$

II. $L = 2(d+2)$.

(a) If $n^{1/2} h^{d+2} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z \varphi(z)}{n h^{d+2}} \{1 + o(1)\} .$$

(b) If $h \sim C_2 n^{-\frac{1}{2(d+2)}}$,

$$\hat{F}(z) = \Phi(z) - \left[C_2^L \kappa_1 - \frac{\kappa_2 z}{C_2^{d+2}} - \frac{4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}}{3} \right] \frac{\varphi(z)}{n^{1/2}} \{1 + o(1)\}$$

(c) If $n^{1/2} h^{d+2} \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1 \varphi(z) n^{1/2} h^L \{1 + o(1)\} .$$

III. $L > 2(d+2)$.

(a) If $n h^L + \frac{1}{n^{1/2} h^{d+2}} \to 0$,

$$\hat{F}(z) = \Phi(z) + 4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\} \frac{\varphi(z)}{3n^{1/2}} \{1 + o(1)\} .$$

(b) If $h \sim C_3 n^{-\frac{1}{L}}$,

$$\hat{F}(z) = \Phi(z) - \left[C_3^L \kappa_1 - \frac{4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}}{3} \right] \frac{\varphi(z)}{n^{1/2}} \{1 + o(1)\} .$$

(c) If $h \sim C_4 n^{-\frac{1}{2(d+2)}}$,

$$\hat{F}(z) = \Phi(z) + \left[\frac{\kappa_2 z}{C_4^{d+2}} + \frac{4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}}{3} \right] \frac{\varphi(z)}{n^{1/2}} \{1 + o(1)\} .$$

(d) If $n^{1/2} h^{d+2} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z \varphi(z)}{n h^{d+2}} \{1 + o(1)\} .$$

(e) If $n h^L \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1 \varphi(z) n^{1/2} h^L \{1 + o(1)\} .$$

In each of the seven cases I(a)-(c), II(a), II(c), III(d), and III(e), the correction term in the expansion is of larger order than $n^{-1/2}$. In the other four cases it is of exact order $n^{-1/2}$, but of these the cases I(b), II(b), III(b), and III(c), which involve a knife-edge choice of bandwidth, include κ_1 or κ_2 (which depend on the kernel K) or both in the correction term. It is case III(a) which corresponds in detail to the "parametric" situation in the sense that K is not involved, and κ_3 and κ_4 are the limits of $E(V_1^3)$ and $E(W_{12} V_1 V_2)$ (see Appendix

A for the definitions of V_1 and W_{12}). The term involving κ_3 and κ_4 is analogous to the $n^{-1/2}$ correction term in the Edgeworth expansion of studentized ordinary U-statistics (see Helmers (1991)). κ_3 and κ_4 are related to the third moment of *U*. κ_1 and κ_2 are respectively limits of $\sigma_\nu^{-1}\nu^\tau(EU-\overline{\mu})/h^L$ and $h^{d+2}E(W_{12}^2)$ (see Lemmas 11 and 12) so that κ_1 and κ_2 are related to first and second moments of *U*. In standard parametric higher-order theory κ_1 and κ_2 do not arise since unbiased statistics with variance $O(n^{-1})$ are typically considered, not, as here, $O(n^{-1}h^{-d-2})$. We can also derive analogous expressions based on Theorem 1. For

example, for (*L*, *d*, *h*) satisfying III(a), we have

(4.1) $$F(z)=\Phi(z)-\frac{4(\kappa_3+3\kappa_4)}{3n^{1/2}}(z^2-1)\varphi(z)\{1+o(1)\}\ .$$

For *U* and related statistics, Härdle, Hart, Marron, and Tsybakov (1992), Härdle and Tsybakov (1993), and Powell and Stoker (1996) derived h that are optimal in the sense of asymptotically minimizing leading terms in the mean squared error (MSE). These optimal *h* are of form

(4.2) $$h^*=C^*n^{-2/(2L+d+2)},\quad 0<C^*<\infty,$$

where we are in one of the cases I(c), II(c) or III(e), in each of which the leading correction term is $-\kappa_1\varphi(z)n^{1/2}h^L$, so that bias correction has the greatest impact in improving the quality of the normal approximation. However, the conventional approach of relating choice of *h* to MSE is not directed towards producing a version of the statistic which, in some sense, makes the normal approximation especially good, and in the context of the present paper the latter goal is relevant. Under (4.2)

(4.3) $$F(z)=\Phi(z)-C^{*L}\kappa_1\varphi(z)n^{-\frac{2L-d-2}{2(2L+d+2)}}\{1+o(1)\}\ .$$

Here, the order of the correction term can be as large as $n^{-1/2(2d+5)}$ when $L=(d+3)/2$ (see Assumption (iv)) and tends to $n^{-1/2}$ only as $L/d\to\infty$, so (4.2) is certainly not optimal in the sense of minimizing the error in the normal approximation. The h which minimizes the integrated MSE of nonparametric derivative-of-density estimates is of form $h^+=C^+n^{-1/(2L+d+2)}$, for $0<C^+<\infty$, but this is even larger than (4.2) and thus provides an even larger correction term than (4.3). Robinson (1995) calculated the rate of decay of h that minimizes the order of the normal approximation error. This exceeds $n^{-1/2}$ due to choosing $L<2(d+2)$, and the more detailed information provided by our Edgeworth expansion allows us to discuss the choice of h itself. In particular, the optimal rate of h here is that in I(b) as described by Robinson (1995), but we would like to know how to choose C_1 in

$$(4.4) \qquad h = C_1 n^{-\frac{3}{2(L+d+2)}} .$$

One possibility is to minimize the maximal deviation from the normal approximation, by

$$C_1^A = \underset{C}{argmin}\underset{z \varepsilon R}{max}\left|\left(C^L\kappa_1 - \frac{\kappa_2 z}{C^{d+2}}\right)\varphi(z)\right| .$$

Because $\kappa_2>0$ this equals

$$\underset{C}{argmin}\underset{z\varepsilon R}{max}\left(C^L|\kappa_1| + \frac{\kappa_2 z}{C^{d+2}}\right)\varphi(z)$$
$$= \underset{C}{argmin}\left\{C^L|\kappa_1| + \frac{\kappa_2 z^*(C)}{C^{d+2}}\right\}\varphi(z^*(C)) ,$$

where

$$z^*(C) = C^{d+2}\{(C^{2L}\kappa_1^2 + 4\kappa_2^2/C^{2d+4})^{1/2} - C^L|\kappa_1|\}/2\kappa_2 .$$

Using the envelope theorem, the first order condition of minimization with respect to C is

$$(4.5) \qquad \left\{LC^{L-1}|\kappa_1| - \frac{(d+2)\kappa_2}{C^{d+3}}z^*(C)\right\}\varphi(z^*(C)) = 0 .$$

Solving (4.5), we derive

$$(4.6) \qquad C_1^A = \left\{\frac{(d+2)^2\kappa_2^2}{4L(L+d+2)\kappa_1^2}\right\}^{\frac{1}{2(L+d+2)}} .$$

The second order condition is easily verified using (4.5) and $z^{*\prime}(C)<0$. Though (4.6) is infeasible since it involves unknown κ_1 and κ_2, we can replace κ_1 and κ_2 by their estimates $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ in Section 2 to give the feasible version

$$(4.7) \qquad \tilde{C}_1^A = \left\{\frac{(d+2)^2\tilde{\kappa}_2^2}{4L(L+d+2)\tilde{\kappa}_1^2}\right\}^{\frac{1}{2(L+d+2)}} .$$

The estimates $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$, introduced to provide empirical Edgeworth expansions (Theorems 2 and 4), are consistent under the conditions stated there, so that $\tilde{C}_1^A$ is consistent for the optimal C_1^A.

One could consider variants of this idea for bandwidth choice, for example maximizing with respect to z over some desired proper subset of R, such as $\{z : |z| > a\}$ for some $a>0$, perhaps to stress one of the usual critical regions. However, the simple forms (4.6) and (4.7) seem appealing. Hall and Sheather (1988) (see also Hall, 1992, p.321) used an Edgeworth expansion for studentized sample quantiles, especially the median, to determine a choice of the bandwidth employed in the studentization. In their problem, the basic $n^{1/2}$ - consistent statistic of interest, the sample quantile, does not involve a bandwidth. In our case, on the other hand, though we also consider studentization involving a bandwidth, it is the bandwidth in the basic statistic of interest, the averaged derivative, that is to be chosen using the Edgeworth expansion. Moreover, unlike us, Hall and Sheather (1988) did not maximize over the argument z, but simply balanced the mean and variance terms of the expansion for given z, so that their data-dependent bandwidth is z-dependent (and thus a 'local' bandwidth). It might be anticipated that the step of maximizing over z, which is incorporated in our procedure, would lead to a more complicated, perhaps only implicitly-defined, formula for the optimal C, and the emergence nevertheless of the simple closed form (4.6) is of some interest. We believe our 'global' approach could be employed in choosing the bandwidth in other semiparametric and nonparametric problems involving smoothing.

5. A MONTE CARLO STUDY

We report results from a Monte Carlo study for the Tobit model $Y_i = (\beta^\tau X_i + \varepsilon_i) I(\beta^\tau X_i + \varepsilon_i \geq 0)$ where $X_i = (X_{1i}, X_{2i})^\tau$ is bivariate. We took $(X_i^\tau, \varepsilon_i) \sim N(0, I_3)$ so that $g(x) = \beta^\tau x\{1 - \Phi(-\beta^\tau x)\} + \varphi(-\beta^\tau x)$ and $\overline{\mu} = -\beta/(8\pi)$. We took $\beta = (1,1)'$. There is no closed form formula for Σ, $\kappa_1, \kappa_2, \kappa_3, \kappa_4$, the first being needed in the expansions of Theorems 1 and 2, and the last four in the expansions of Theorems 1 and 3, so they were calculated by simulation, with 100,000 replications, to be $\Sigma = \begin{pmatrix} 0.00887 & 0.00458 \\ 0.00458 & 0.00887 \end{pmatrix}$, $\kappa_1 = 0.397$, $\kappa_2 = 1.724$, $\kappa_3 = -0.144$

and $\kappa_4 = -0.266$, for example $\Sigma = 10^{-5} \sum_{i=1}^{10^5} 4\{\mu(X_i, Y_i) - \bar{\mu}\}\{\mu(X_i, Y_i) - \bar{\mu}\}^{\tau}$ where $(X_i^{\tau}, Y_i),\ i = 1, ..., 10^5$ are generated independently and identically following the above Tobit model. We employed three values of *L*, *L*=4, 8 and 10 which respectively correspond to the cases I, II and III in Section 4 (and easily satisfy assumptions (iv) and (v)), using normal density-based multiplicative *L*-th order bivariate kernel functions proposed in Robinson (1988), $K(u_1, u_2) = K_L(u_1)K_L(u_2)$, where

$$K_L(u) = \sum_{j=0}^{(L-2)/2} c_j u^{2j} \varphi(u) \ ,$$

such that

$$\sum_{j=0}^{(L-2)/2} c_j m_{2(i+j)} = \delta_{i0},\ i = 0, 1, \cdots, (L-2)/2 \ , \qquad (5.1)$$

$$m_{2j} = \int u^{2j} \varphi(u) du \ ,$$

and δ_{i0} is Kronecker's delta. The values of c_j calculated from these simultaneous equations are in Table 1. We chose $H(u_1, u_2) = \varphi(u_1)\varphi(u_2)$ in estimation of κ_1 in the

TABLE 1

L-th order kernel functions.

L	c_0	c_1	c_2	c_3	c_4
4	1.5	-0.5	–	–	–
8	2.185	-2.185	0.4375	-0.02083	–
10	1.924	-1.347	0.1230	0.00698	-0.000489

empirical Edgeworth expansions. We considered inference on the two elements of $\bar{\mu}$ individually, but since the results for these are very similar we report them for the first only.

Figures 1-7 compare approximations to the distribution of the unstudentized statistic $(U_{(1)} - \bar{\mu}_{(1)})/\sigma$, where $U_{(1)}$ and $\bar{\mu}_{(1)}$ are the first elements of U, $\bar{\mu}$, and $\sigma^2 = 0.00887$. We used h = 1, 0.8, 0.6 and 0.4 for n=100, and h=0.8, 0.6 and 0.4 for n=400, with 600 replications, and we set b=1.2h following the discussion in Section 2. We report results for only *L*=4 because the results for *L*=8 and 10 are qualitatively much the same while exhibiting less bias. The solid line is the empirical distribution function of Z, while the dotted, broken, and broken-and-dotted lines are the standard normal distribution function Φ , the empirical Edgeworth expansion (Theorem 2), and the theoretical Edgeworth expansion (Theorem 1) respectively. The empirical Edgeworth correction results from averaging $\tilde{\kappa}_i$ across 600 replications for each sample size, bandwidth choice and kernel order. The two empirical Edgeworth expansions in each Figure involve respectively all three correction terms (shorter broken line) and one correction term of order $n^{-1/2}$ (longer broken line) in (2.6), which corresponds to the feasible version of (4.1). We examine the "one-term" case because this is the one we would hope to be able to recommend, since it involves just the "parametric" $n^{-1/2}$ correction and, depending only on κ_3 and κ_4 but not on κ_1 and κ_2, is free of K.

We first compare the "three-term" empirical Edgeworth expansion (EE3) with the empirical distribution (ED) and the normal approximation (N), finding a range of n and h where EE3 well approximates ED, and better than N, for example, see Figures 1, 2, 3, and 4. It emerges that h=1.0 (Figure 1) is too large in that neither N nor EE3 performs well, but when h=0.8 or 0.6 (Figures 2, 3) EE3 is satisfactory, and better than N, whereas when h=0.4 (Figure 4), the opposite outcome is observed. It is not surprising that N sometimes outperforms EE3 since n is finite (see Hall (1992), p.45) and the $\tilde{\kappa}_i$ are subject to sampling error. We also considered, but have not included, the case h=0.1 with n=100, where the variance in the empirical distribution is very large, and both N and EE3 performed poorly. Neither N nor EE3 could be expected to work well for sufficiently large or small h. Comparing Figure 6 with Figures 2, 3, say, EE3 appears to improve with increasing n.

It might then come as something as a surprise that in most cases the figures reveal that EE3 approximates ED better than the "three-term" theoretical Edgeworth expansion (TE3). A possible explanation is as follows. The proof of Theorem 1 (see (A.13)) implies that an alternative theoretical Edgeworth approximation to (2.3) is

$$(5.2)\quad \Phi(z)-\varphi(z)\left\{n^{1/2}\sigma_\nu^{-1}\nu^\tau(EU-\bar{\mu})+\frac{E(W_{12}^2)}{n}z+\frac{4E(V_1^3)+12E(W_{12}V_1V_2)}{3n^{1/2}}(z^2-1)\right\} .$$

The expectations are untidy, depending on n so the proof goes on to obtain the simpler and more elegant $\tilde{F}(z)$, involving the n-free κ_i. However, in comparing (5.2) with the $\tilde{\kappa}_i$ EE3 might seem to most directly estimate (5.2), which might be a more accurate approximation to ED than TE3, (2.3).

Comparing shorter broken and longer broken lines, the "one-term" empirical Edgeworth expansion (EE1) is better for some values of h depending on n than EE3, in particular when (n,h)=(100,0.6) and (400,0.4) (Figures 3 and 7). These are the cases of relatively small h, so that the bias is small but $n^{-1}h^{-d-2}$ is relatively large, namely the $\tilde{\kappa}_1$ correction is negligible but the $\tilde{\kappa}_2$ one tends to be too large, having the effect of pushing the curve up and down around -1 and 1 respectively. It is clear from the discussion in Section 4 that one expects the choice of h to be especially crucial where "one-term" expansions are concerned.

Figures 8-23 compare approximations to the distribution of studentized statistics $n^{1/2}(U_{(1)}-\overline{\mu}_{(1)})/\hat{\sigma}$, where $\hat{\sigma}^2$ is the leading element of $\hat{\Sigma}$, based on Theorem 4. U, $\tilde{\kappa}_i$, and $\hat{\Sigma}$ involved in Figures 8-15, 16-19, and 20-23 used respectively kernel functions of orders L=4, 8, and 10; see (5.1) and Table 1. We took h=0.2, 0.4, 0.6, 0.8 for each of n=100, 400 with 600 replications similarly to the unstudentized case above. Moreover, for L=8, 10 we include only results for n=100 because these for n=400 are very similar. Because the theoretical Edgeworth expansions (Theorem 3) performed less well than in the unstudentized cases featured in Figures 1-7, and because they are in any case of less practical interest than empirical expansions, we exclude the former cases from Figures 8-23 for ease of reading.

Making broad comparisons across the three groups of figures, 8-15, 16-19, and 20-23, we find that bias tends to vary inversely with L, keeping n and h fixed. This is consistent with the theoretical (asymptotic) bias-reducing properties motivating higher-order kernels, but Monte Carlo studies of semiparametric estimates employing such kernels (see e.g. Robinson (1988)) have found that these properties are not necessarily mirrored in finite samples, so these results of ours are rather pleasing. Generally in Figures 8-23, we observe that EE3 approximates ED very well except for largish h, see e.g. Figure 12, where N also performs poorly. Comparing Figures 8-15 with Figures 1-7 for the unstudentized statistic (with L=4 throughout), EE3 is seen to work better for the studentized statistic. The reason may be similar to the one we offered for the apparent superiority of EE3 over TE3 in the unstudentized case, namely, $Var(U_1)$ can better normalize U than Σ, and $\hat{\Sigma}$, in view of its construction, more directly estimate $Var(U_1)$. When L=10, the "parametric" case III(a) is justified theoretically, and EE1 performs satisfactorily for certain (n,h), in particular for (n,h)= (100,0.8) (Figure 20).

We next consider interval estimation. A $100(1-\alpha)\%$ confidence interval based on N is

$$(5.3) \qquad \left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} z_{\frac{\alpha}{2}} \, , \, U_{(1)} + \frac{\hat{\sigma}}{n^{1/2}} z_{\frac{\alpha}{2}} \right) \quad ,$$

where z_γ satisfies $\int_{z_\gamma}^{\infty} \varphi(z)dz = \gamma$. We can correct this interval using Theorem 4. Inverting the empirical Edgeworth expansion there, we have the Cornish-Fisher expansion (see Hall (1992), p.88),

$$\begin{aligned} w_\gamma &= z_\gamma + n^{1/2}h^L \tilde{\kappa}_1 - \frac{\tilde{\kappa}_2}{nh^{d+2}} z_\gamma - \frac{4}{3n^{1/2}} \{(2z_\gamma^2+1)\tilde{\kappa}_3 + 3(z_\gamma^2+1)\tilde{\kappa}_4\} \\ &\qquad + o\,(n^{1/2}h^L + n^{-1}h^{-d-2} + n^{-1/2}) \\ &= \tilde{w}_\gamma + o(n^{-1/2} + n^{1/2}h^L + n^{-1}h^{-d-2}) \quad , \end{aligned}$$

where w_γ is the $100\gamma\%$ quantile of the sampling distribution. Then the corrected interval estimate is

$$(5.4) \qquad \left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} \tilde{w}_{1-\frac{\alpha}{2}} \, , \, U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} \tilde{w}_{\frac{\alpha}{2}} \right) \quad .$$

Note that $\tilde{w}_{1-\alpha/2} \neq -\tilde{w}_{\alpha/2}$ in general so that (5.4) is not symmetric around the point estimate $U_{(1)}$, unlike (5.3). According to our interval estimation in the current Tobit example, this correction is supported when the Edgeworth expansion approximates well the empirical distribution function, which is mostly the case for the studentized statistic. We report two typical cases where the correction appears effective. One is when N fails to well approximate ED due to the large bias of U, and the other is when $\hat{Z}$ has variance significantly less than unity. Figures 24-27 show the "true" 80% and 90% confidence intervals derived from ED (solid line), the corresponding interval estimates obtained from N, see (5.3) (dotted line) and from EE3, see (5.4) (broken line) for (n,h,L)=(100,0.6,4), (400,0.2,4). The vertical closely-spaced dotted line indicates the true parameter value $\bar{\mu}_{(1)} = -1/(8\pi)$. The true interval is derived like (5.3) or (5.4) as

(5.5) $$\left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} t_{1-\frac{\alpha}{2}},\ U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} t_{\frac{\alpha}{2}} \right),$$

where t_γ denotes the $100\gamma\%$ quantile of ED. Both estimates (5.3) and (5.4) include the true value in all four figures. In Figures 24 and 26, we observe that they are of similar length, though (5.3) is typically biased to the right and it does not cover the left part of the "true" interval, while (5.4) covers almost the whole true interval. In Figures 25 and 27, we observe that (5.3) clearly overestimates (5.5), while (5.4) performs satisfactorily. When (*n*, *h*, *L*)=(100, 0.6, 4), N is biased to the left (Figure 9) and when (*n*, *h*, *L*)=(400, 0.2, 4), it has larger variance than ED (Figure 15) so that (5.3) estimates the confidence interval as described. Our experiment demonstrates that the Cornish-Fisher expansion can produce better interval estimates than N.

We proposed optimal bandwidth choices which minimize the error of the normal approximation in Section 4. (4.4), (4.6) with *L*=4 and κ_i described above yield the optimal bandwidth as *h*=0.445 and 0.343 for *n*=100 and 400 respectively. ED with these values of *h*, as well as *h*=0.2 and 0.6, is compared in Figures 28 and 29 with N, which seems to best approximate ED with optimal *h*.

As discussed in Section 4, Theorems 1 and 3 also imply that bias correction should have the greatest influence in improving the second order properties of *U* when the minimum MSE bandwidth is used. In view of Theorem 2 and Lemma 11, $h^{*L}\tilde{\kappa}_1$ estimates the bias $\sigma_\nu^{-1}\nu^\tau(EU-\bar{\mu})$ consistently and so $\sigma_\nu^{-1}\nu^\tau U - h^{*L}\tilde{\kappa}_1$ is a bias-corrected estimate of $\sigma_\nu^{-1}\nu^\tau\bar{\mu} = -1/(8\pi\sigma)$. Table 2 shows the average estimates of $\sigma_\nu^{-1}\nu^\tau(U-\bar{\mu})$, $h^{*L}\tilde{\kappa}_1$, and $\sigma_\nu^{-1}\nu^\tau(U-\bar{\mu}) - h^{*L}\tilde{\kappa}_1$ for each *n* from 600 replications when *L*=4 and the (infeasible) minimum MSE bandwidth choice of Powell and Stoker (1996) was used. The bandwidth was calculated by means of Monte Carlo simulation to be h^* = 0.9048, 0.8061 and 0.7128 for *n*=100, 200, and 400 respectively. We used $b=1.2h^*$ in estimating $\tilde{\kappa}_1$. Comparing the first and the third column of Table 2, the bias-corrected estimate is seen to perform much better than the uncorrected one, especially for *n*=400. Powell and Stoker (1996) also proposed a feasible minimum-MSE bandwidth $\tilde{h}^*$, which depends on two user-specified parameters h_0 and τ (see (4.35), (4.38), and (4.40) of Powell and Stoker (1996)). On the basis of our calculations, though both absolute bias $|E(\tilde{h}^*-h^*)|$ and MSE $E(\tilde{h}^*-h^*)^2$ were relatively insensitive to h_0 (while exhibiting some tendency to decrease in h_0), they were highly sensitive to τ.

The Figures and Table 2 are not available in this electronic version.

σ^-