2-1-1997

Some Higher Order Theory for a Consistent Nonparametric Model Specification Test

Yanqin Fan

Oliver B. Linton

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1148

SOME HIGHER ORDER THEORY FOR A CONSISTENT NONPARAMETRIC MODEL SPECIFICATION TEST

Yanqin Fan and Oliver Linton

February 1997

Some Higher Order Theory for a Consistent Nonparametric Model Specification Test

by

Yanqin Fan and Oliver Linton*

University of Windsor and Yale University

Abstract

We provide second order theory for a smoothing-based model specification test. We derive the asymptotic cumulants and justify an Edgeworth distributional approximation valid to order close to n^{-1}. This is used to define size-corrected critical values whose null rejection frequency improves on the normal critical values. Our simulations confirm the efficacy of this method in moderate sized samples.

Key Words: Consistent test; Edgeworth expansion; Kernel estimation; Nonlinear regression.

FEBRUARY 6, 1997

*Address correspondence to: Oliver Linton, Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06520-8281, USA. Phone: (203) 432-3699. Fax: (203) 432-6167. E-mail address: linton@econ.yale.edu

1 Introduction

The costs of misspecifying econometrics models can be considerable in terms of the biased inferences that result. This has motivated much work on producing tests of parametric and semiparametric null hypotheses against general nonparametric alternatives, especially in regression. These tests fall in two general categories: those based on smoothing methods, for example Gozalo (1993), Fan (1994), Hong and White (1995), Fan and Li (1996a), Li and Wang (1996), and Zheng (1996); and those based on 'non-smoothing' techniques such as the empirical distribution function, for example Bierens (1982, 1990), Andrews (1995), Bierens and Ploberger (1996), and Gozalo and Linton (1996). These two classes of tests have different properties: The smoothing based tests typically have asymptotic normal distributions under the null hypothesis, while the 'non-smoothing' based tests have non-normal null asymptotic distributions that depend on the nuisance parameters in the null model; the smoothing based tests are in general more powerful than the non-smoothing based tests against high frequency alternatives, while the non-smoothing based tests are more powerful against low frequency alternatives than the smoothing based tests, see for example Rosenblatt (1975), Fan (1994), Ghosh and Huang (1991), Eubank and LaRiccia (1992), and Fan and Li (1996b).

Smoothing-based tests have the advantage that the null distribution is normal and so it is relatively easy to obtain critical values. The principles involved also extend easily to a large variety of data-generating processes. The simplest version of the smoothing-based test for regression was introduced independently in Li and Wang (1996) and Zheng (1996), see also Fan and Li (1996a). This is simply a kernel-weighted quadratic form in the parametric residuals. It is in the spirit of the Lagrange Multiplier test, since it involves explicit estimation only under the null hypothesis. This quadratic form is asymptotically normal with zero mean, i.e. there are no smoothing bias terms present as in other tests, thus providing a simple to compute test of the parametric model.

Although their null distribution is normal, there has been much concern over the small sample properties of smoothing-based tests. This is in part due to the fact that they more closely resemble parametric *chi-squared* tests than *normal* tests [in fact, Staniswalis and Severini (1991) use a *chi-squared* approximation]. This view is also borne out by monte carlo experiments which have found large biases and skewnesses and both under and overrejection, see for example Härdle and Mammen (1993), Fan (1995), Gozalo (1995), and Hjellvik and Tjøstheim (1996).

We investigate analytically the small sample properties of the specification test in Li and Wang (1996) and Zheng (1996) through Edgeworth expansion. In a related setting, these techniques were used by the second author to determine bandwidth selection methods in various semiparametric models, see Linton (1995,1996); also, see Robinson (1995) for related work on Berry-Esseen bounds. In this paper, we use the expansions to define a method of size correction which is then investigated on simulated data. To justify our formal expansion requires some new work. The mathematical form of the test statistic is a ratio of a degenerate weighted U-statistic to a non-degenerate weighted U-statistic. While Edgeworth theorems exist for U-statistics that are asymptotically normal [Callaert, Jannssen and Verarbereke (1980) and Bickel, Götze and van Zwet (1986)], and even for degenerate unweighted U-statistics that are asymptotically mixtures of chi-squares [Götze (1979)], to our knowledge there is no such existing theory applicable to the particular random variables that constitute our test statistic.

Notation. For any scalar random variable X, let $\kappa_i(X)$, $i = 1, 2, \ldots,$ denote the i'th asymptotic cumulant of X. Let $\Phi(\cdot)$ denote the normal cumulative distribution function.

2 Higher Order Approximations to the Distribution of a Smoothing Test

Consider the following nonparametric regression model:

$$Y_i = g(X_i) + u_i,\ \ i = 1, \ldots, n,$$

where $\{Y_i, X_i'\}_{i=1}^n$ is independently and identically distributed (i.i.d.) as $\{Y, X'\}$, $g(\cdot)$: $\mathbb{R}^d \to \mathbb{R}$ is the true but unknown regression function, and u_i is the error satisfying $E(u_i|X_i) = 0$. The null and alternative hypotheses of interest are

$$\mathbf{H}_0 \;:\; \Pr\left[g(X) = g_0(X, \beta_0)\right] = 1 \quad \text{for some } \beta_0 \in \mathcal{B}$$
$$\mathbf{H}_A \;:\; \Pr\left[g(X) = g_0(X, \beta)\right] < 1 \quad \text{for all } \beta \in \mathcal{B},$$

where $g_0(\cdot, \beta)$ is a known function once the unknown parameter β is given. The parameter space $\mathcal{B}$ is an open subset of $\mathbb{R}^p$.

The test we examine is based on the quadratic form

$$I_n = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} \widehat{u}_i \widehat{u}_j K_{ij}, \tag{1}$$

where $K_{ij} = K((X_i - X_j)/h)$ in which $K(\cdot)$ is a kernel function and $h = h(n)$ is a smoothing parameter, and $\widehat{u}_i = Y_i - g_0(X_i, \widehat{\beta})$ in which $\widehat{\beta}$ is, under the null hypothesis, a root-n consistent estimate of β. Under certain conditions, including $h \to 0$ and $nh^d \to \infty$, it has been shown that $nh^{d/2} I_n \to N(0, \sigma^2)$ in distribution under $\mathbf{H}_0$, where $\sigma^2 = 2 \int K^2(u) du \int f^2(x) \sigma_u^4(x) dx$ with $\sigma_u^2(x) = E(u^2 | X = x)$ and $u = Y - g(X)$, while $f(\cdot)$ is the marginal density of X, see Li and Wang (1996) and Zheng (1996). In addition,

$$\widehat{\sigma}^2 = \frac{2}{n(n-1)h^d} \sum_i \sum_{j \neq i} \widehat{u}_i^2 \widehat{u}_j^2 K_{ij}^2$$

consistently estimates σ^2 under $\mathbf{H}_0$.[1] Therefore, under $\mathbf{H}_0$,

$$\sup_{x \in \mathbb{R}} \left| \Pr(\widehat{T} \leq x) - \Phi(x) \right| = o(1), \tag{2}$$

where $\widehat{T} = \hat{\sigma}^{-1} nh^{d/2} I_n$. The test is implemented one-sided, that is we reject $\mathbf{H}_0$ if $\widehat{T} > z_\alpha$, where $\Phi(z_\alpha) = 1 - \alpha$ for a level α test.

Simulation results on this and other smoothing based tests indicate poor performance of the standard normal distribution in moderately large samples, see the references cited in the introduction.[2]

[1] An alternative estimate of σ^2 can be based on nonparametric residuals although this is less attractive to implement.

[2] Zheng (1996) applied a two-sided test based on $\widehat{T}$ to some simulated data and reported good size performance of his test. However, implementing a one-sided test based on two-sided critical values will incur loss of power.

This is consistent with (2) since the $o(1)$ error could decrease arbitrarily slowly. In the next section, we provide a refined asymptotic approximation for the null distribution of $\widehat{T}$ based on Edgeworth theory that gives an error close to $o(n^{-1})$ in (2).

3 Higher Order Asymptotic Approximation

We now provide a refined asymptotic approximation for the finite sample distribution of $\widehat{T}$ based on Edgeworth theory. Our theorem is stated for the version of the test in which $\widehat{\beta}$ is the nonlinear least squares estimator. Let $G_i = \partial g(X_i, \beta_0)/\partial\beta$, $i = 1, \ldots, n$.

THEOREM 1. *Suppose that Assumptions A given in the appendix hold and that the null hypothesis is true. Then, (a) the asymptotic cumulants of* $\widehat{T}$ *are, to order* $O(h^{3d/2} + h^{d/2}/\sqrt{n})$,

$$\kappa_1(\widehat{T}) = h^{d/2}[m_3 - 2m_2] - \frac{2}{nh^{d/2}}m_4, \tag{3}$$

$$\kappa_2(\widehat{T}) = 1 + 2h^d[2m_{22} + m_{33} - 4m_{12} + 2m_{13} - 4m_{23}], \tag{4}$$

$$\kappa_3(\widehat{T}) = 8h^{d/2}m_{111} + \frac{4}{nh^{d/2}}m_4, \tag{5}$$

$$\kappa_4(\widehat{T}) = 48h^d m_{1111}, \tag{6}$$

where

$$m_2 = \frac{1}{h^d\sigma_n}E[u_2^2\omega_{21}K_{12}], \qquad m_3 = \frac{1}{\sigma_n}E[u_1^2\tau_{11}],$$

$$m_{22} = \frac{1}{\sigma_n^2 h^{2d}}E[u_4^2u_2^2\omega_{41}K_{12}(\omega_{43}K_{32} + \omega_{23}K_{34})], \qquad m_{33} = \frac{1}{\sigma_n^2}E[u_2^2u_1^2\tau_{21}^2],$$

$$m_{12} = \frac{1}{\sigma_n^2 h^{2d}}E[u_1^2u_2^2K_{12}\omega_{13}K_{32}], \qquad m_{13} = \frac{1}{h^d\sigma_n^2}E[u_1^2u_2^2K_{12}\tau_{12}],$$

$$m_{23} = \frac{1}{h^d\sigma_n^2}E[u_3^2u_2^2\omega_{31}K_{12}\tau_{32}], \qquad m_{111} = \frac{1}{\sigma_n^3 h^{2d}}E[u_1^2u_2^2u_3^2K_{12}K_{13}K_{23}],$$

$$m_{1111} = \frac{1}{\sigma_n^4 h^{3d}}E[u_1^2u_2^2u_3^2u_4^2K_{12}K_{13}K_{24}K_{34}], \qquad m_4 = \frac{1}{\sigma_n^3 h^d}E[u_1^3u_2^3K_{12}^3],$$

with $\omega_{ki} = G_k'\Omega^{-1}G_i$, $\mu_\Omega = \Omega^{-1}[E(G_1K_{12}G_2')/h^d]\Omega^{-1}$, $\tau_{kl} = G_k'\mu_\Omega G_l$, $\Omega = E(G_iG_i')$, *and* $\sigma_n^2 = 2h^{-d}E(u_1^2u_2^2K_{12}^2)$.

Furthermore, letting

$$\widetilde{F}_{n1}(x) = \Phi\left(x + \frac{\gamma_{n1} + \gamma_{n2}x^2}{6}\right) \quad ; \quad \widetilde{F}_{n2}(x) = \Phi\left(x + \frac{\gamma_{n1} + \gamma_{n2}x^2}{6} + \frac{\gamma_{n3}x + \gamma_{n4}x^3}{72}\right)$$

with $\gamma_{n1} = \kappa_3(\widehat{T}) - 6\kappa_1(\widehat{T})$, $\gamma_{n2} = -\kappa_3(\widehat{T})$, $\gamma_{n3} = 9\kappa_4(\widehat{T}) - 14\kappa_3^2(\widehat{T}) - 36\left\{\kappa_2(\widehat{T}) - 1\right\} + 24\kappa_1(\widehat{T})\kappa_3(\widehat{T})$, *and* $\gamma_{n4} = 8\kappa_3^2(\widehat{T}) - 3\kappa_4(\widehat{T})$, *under the additional conditions B given in the appendix, (b) we have the formal Edgeworth approximations*

$$\sup_{x\in\mathbb{R}} |\Pr(\widehat{T} \leq x) - \widetilde{F}_{nj}(x)| = o(n^{-\epsilon_j}), \quad j = 1, 2, \tag{7}$$

where $n^{-\epsilon_1} = h^{d/2}$, $n^{-\epsilon_2} = \min\left\{(nh^{d/2})^{-1}, h^d\right\}$.

The first qualitative conclusion we draw from Theorem 1 is that the magnitude of the size distortion can be made as close to order $n^{-1/2}$ by taking the bandwidth as small as possible. However, there is a constraint that $nh^d \to \infty$ which precludes the root-n distortion being reached for a consistent test.

There is a trade-off to be made here between having good size distortion and good power. As was shown in Zheng (1996), the test has power against Pitman local alternatives at distance $(nh^{d/2})^{-1/2}$ from the null hypothesis.[3] At one extreme, this magnitude can be made arbitrarily close to order $n^{-1/2}$ by taking a large bandwidth, while if bandwidth is small this magnitude can be arbitrarily large, i.e. power is essentially zero.[4]

[3] Consider the local alternative hypothesis

$$\mathbf{H}_n: \ g(X) = g_0(X;\beta_0) + \delta_n\gamma(X)$$

for some sequence $\delta_n \to 0$. When $\delta_n = (nh^{d/2})^{-1/2}$, we have, under $\mathbf{H}_n$, $\widehat{T} \to N\left[\int \gamma^2(x)f^2(x)dx/\sigma, 1\right]$ in distribution.

[4] In the extreme case where one takes h as a fixed constant, the test is equivalent to a version of the ICM test of

Figure 0 below gives the trade-off between achievable size distortion and power in the range where there is power against alternatives larger than $n^{-1/4}$.

*** FIGURE 0 HERE ***

The trade-off is unaffected by the dimensions d, since by choosing bandwidth of the form $n^{-\delta/d}$, for positive δ, one gets size distortion of order $n^{-\delta/2}$.[5] For example, at $\delta = 1/2$ one achieves size distortion of magnitude $n^{-1/4}$ and power of magnitude $n^{-3/8}$.[6] See Gourieroux and Tenreiro (1994) for an interesting discussion about the local power of similar smoothing-based tests.

Note that in view of the trade-off between size distortion and power, a simple notion of optimality based on maximizing local power for the size adjusted test [as used in classical parametric statistics, see Rothenberg (1984)], is not well defined here. Really, one needs to have a preference function jointly defined over both size distortion and local power which then enables one to pick some unique point on the 'budget constraint' given in Figure 0.

Regarding the direction of the higher order effects, both the skewness and kurtosis are unambiguously positive, while the mean and variance can take either sign.

EXAMPLE. Under homoskedasticity, the order $h^{d/2}$ mean and skewness corrections simplify. Specifically,

---

Bierens and Ploberger (1996) (see Fan and Li (1996b) for a proof of this result) which has power against Pitman local alternatives at distance $n^{-1/2}$ from the null hypothesis. However the null asymptotic distribution of the test statistic with a fixed h is no longer the standard normal, in fact, it is an infinite sum of weighted independent *chi-squared(1)* distributions. This explains the poor finite sample performance of the standard normal distribution for large smoothing parameter values observed in our simulation studies.

[5]For a broad range of bandwidths, specifically $\delta < 2/3d$, the terms in the cumulant corrections of order $1/nh^{d/2}$ are of smaller order than h^d and can be ignored.

[6]We show below that by using size-adjusted critical values one can reduce the size distortion to magnitude less than order $n^{-\delta}$.

$$\kappa_1(\widehat{T}) = -h^{d/2}\frac{\text{tr}(\Omega_f \Omega^{-1})}{2\nu_2(K)\nu_2(f)},$$

where $\nu_j(g) = \int g^j(s)ds$ and $\Omega_f = E\left[G_1 G_1' f(X_1)\right]$, while

$$\kappa_3(\widehat{T}) = 8h^{d/2}\frac{\nu_3(f)\nu_{3*}(K)}{\{2\nu_2(K)\nu_2(f)\}^{3/2}},$$

where $\nu_{3*}(K) = \int\int K(u-v)K(u)K(v)dudv = \int K(v)K*K(v)dv$. If, furthermore, f were uniform and K were Gaussian, then the mean correction is $-h^{d/2}d/2\sqrt{\pi}$ and the skewness correction is $h^{d/2}8\pi/\sqrt{6}$.

4 Size Correction

We now use the approximations of Theorem 1 to define a method of correcting the test for its deficient size. Similar procedures have been defined for a wide range of parametric testing problems, see Rothenberg (1984).

Let α_n be the true rejection frequency of the test, i.e. $\alpha_n = \Pr\left(\widehat{T} > z_\alpha | \mathbf{H}_0\right)$, then from Theorem 1, we have that

$$\alpha_n = \alpha + O(n^{-\epsilon_1}).$$

Define the corrected critical values $c_{\alpha 1}$ and $c_{\alpha 2}$, where

$$\begin{aligned} c_{\alpha 1} &= z_\alpha - \frac{\gamma_{n1} + \gamma_{n2} z_\alpha^2}{6} \\ c_{\alpha 2} &= z_\alpha - \frac{\gamma_{n1} + \gamma_{n2} z_\alpha^2}{6} - \frac{\gamma_{n3} z_\alpha + \gamma_{n4} z_\alpha^3}{72} + \frac{z_\alpha \gamma_{n2}\left(\gamma_{n1} + \gamma_{n2} z_\alpha^2\right)}{18} \end{aligned}$$

and let α_{nj}^c be the rejection frequency of the test using $c_{\alpha j}$ as critical value. Then,

$$\text{(a) } \alpha_{n1}^c = \alpha + o(n^{-\epsilon_1}) \quad ; \quad \text{(b) } \alpha_{n2}^c = \alpha + o(n^{-\epsilon_2}) \tag{8}$$

which follows from (7). In practice, we must replace γ_{nj} by an estimate, say $\widehat{\gamma}_{nj}$, $j = 1, \ldots, 4$. Letting

$$\begin{aligned} \widehat{c}_{\alpha 1} &= z_\alpha - \frac{\widehat{\gamma}_{n1} + \widehat{\gamma}_{n2} z_\alpha^2}{6} \\ \widehat{c}_{\alpha 2} &= z_\alpha - \frac{\widehat{\gamma}_{n1} + \widehat{\gamma}_{n2} z_\alpha^2}{6} - \frac{\widehat{\gamma}_{n3} z_\alpha + \widehat{\gamma}_{n4} z_\alpha^3}{72} + \frac{z_\alpha \widehat{\gamma}_{n2} \left(\widehat{\gamma}_{n1} + \widehat{\gamma}_{n2} z_\alpha^2\right)}{18}, \end{aligned}$$

we have the following corollary

Corollary. *Provided the estimation error in $\widehat{\gamma}_{nj}$, $j = 1, \ldots, 4$, is small enough, (8) holds for the estimated critical values. Specifically, for (a) we just need consistency, i.e. $(\widehat{\gamma}_{nj} - \gamma_{nj})/\gamma_{nj} = o_p(1)$ for $j = 1, 2$, while for (b) we need that $(\widehat{\gamma}_{nj} - \gamma_{nj})/\gamma_{nj} = o_p(n^{-(\epsilon_2 - \epsilon_1)})$ for $j = 1, 2$ and $(\widehat{\gamma}_{nj} - \gamma_{nj})/\gamma_{nj} = o_p(1)$ for $j = 3, 4$.*

We now discuss how to estimate m_{jkl} and hence γ_{nj}. Note that there are no terms which arise from "smoothing bias" which makes our job easier. The population moments m_{jkl} can be estimated by their sample equivalents. For example,

$$\widehat{m}_3 = \frac{1}{n\widehat{\sigma}} \sum_{i=1}^{n} \widehat{u}_i^2 \widehat{\tau}_{ii} \quad ; \quad \widehat{m}_{111} = \frac{1}{n^3 \widehat{\sigma}^3 h^{2d}} \sum\sum\sum_{i,j,l:\ i \neq j, i \neq l, j \neq l.} \widehat{u}_i^2 \widehat{u}_j^2 \widehat{u}_l^2 K_{ij} K_{il} K_{jl},$$

where $\widehat{\tau}_{ii} = \widehat{G}_i' \widehat{\mu}_\Omega \widehat{G}_i$, with

$$\widehat{\mu}_\Omega = \widehat{\Omega}^{-1} \left[\frac{1}{n^2 h^d} \sum\sum_{i \neq j} \widehat{G}_i K_{ij} \widehat{G}_j' \right] \widehat{\Omega}^{-1},$$

where $\widehat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \widehat{G}_i \widehat{G}_i'$. Here, $\widehat{G}_i = \partial g(X_i, \widehat{\beta}) / \partial \beta$, $i = 1, \ldots, n$.

In view of the U-statistic structure of the above estimates, we expect them to be root-n consistent for the population quantities which would satisfy the conditions of the corollary.

5 Simulation Results

To evaluate the performance of the Edgeworth expansion derived in the previous sections, we applied it to testing two parametric specifications.

The first one (DGP1) is taken from Härdle and Mammen (1993). There is one regressor X_i ($d = 1$) and the parametric specification is $g_0(X_i, \beta) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$. In the simulation experiment, we generated n independent observations X_i from the uniform distribution on $[0, 1]$ and generated u_i independently from the standard normal distribution. As in Härdle and Mammen (1993), we set $\beta_0 = \beta_1 = \beta_2 = 1$, and take the kernel function $K(\cdot)$ to be the quartic kernel

$$K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1).$$

The smoothing parameter is chosen according to $h = c x_{sd} n^{-1/3}$, where c is a constant and x_{sd} is the sample standard deviation of $\{X_i\}_{i=1}^n$.

The second model (DGP2) is the one considered in Zheng (1996). There are two regressors X_{1i} and X_{2i} and the parametric model is specified as $g_0(X_i, \beta) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$. In the experiment, the two regressors are generated as $X_{1i} = Z_{1i}$ and $X_{2i} = (Z_{1i} + Z_{2i})/\sqrt{2}$, where Z_{1i} and Z_{2i} are independent observations from the standard normal distribution. The error term u_i is drawn independently from the standard normal distribution. As in Zheng (1996), $\beta_0 = \beta_1 = \beta_2 = 1$; the kernel function K is the bivariate standard normal density function; and the smoothing parameter h is $h = cn^{-2/5}$ where c is a constant.

For both models, we considered two sample sizes: $n = 100$ and 200, and four values for the constant c in h: $c = 0.5$, 1.0, 1.5, and 2.0. The number of replications is $m = 1,000$.

In principle, the test based on $\widehat{T}$ is a one-sided test. However, Zheng (1996) used two-sided critical values from the standard normal distribution. Hence, in our experiment, for DGP1, we computed rejection rates of $\widehat{T}$ based on one-sided critical values from the asymptotic normal distribution (z_α), the first order Edgeworth expansion ($\hat{c}_{\alpha 1}$), and the second order Edgeworth expansion ($\hat{c}_{\alpha 2}$); for DGP2, we computed rejection rates of $\widehat{T}$ based on both one-sided and two-sided critical values from the asymptotic normal distribution, the first order Edgeworth expansion, and the second order Edgeworth expansion. The rejection rates for DGP1 are reported in Tables 1 and 2, and those for DGP2 are reported in Tables 5 and 6. In addition, we also computed the first four corrected cumulants, their means and standard deviations over 1000 replications, as well as the first four empirical (estimated) cumulants of $\widehat{T}$. These are reported in Tables 3 and 4 (for DGP1) and Tables 7 and 8 (for DGP2). To examine the overall performance of the asymptotic normal distribution and the second order Edgeworth expansion, we have also drawn QQ plots for both DGP1 and DGP2 when $n = 100$. These are given in Figures 1 (DGP1) and 2 (DGP2). Both Figure 1 and Figure 2 have four graphs corresponding to $c = 0.5$, 1.0, 1.5 and 2.0 respectively. In each graph, the standard normal critical values (z_α) and the Edgeworth corrected critical values ($\hat{c}_{\alpha 2}$) are plotted against the empirical critical values of $\widehat{T}$ generated from 1,000 replications. As reference, the 45 degree line is also drawn. The distribution that generates a QQ plot closer to the 45 degree line provides a more accurate approximation to the true distribution of $\widehat{T}$. The two vertical lines in each figure corresponds to the empirical 95% and 99% critical values of $\widehat{T}$. Detailed results are summarized below.

*** Tables and Figures here ***

For DGP1, both Tables 1 and 2 as well as Figure 1 show clearly that the standard normal distribution performs poorly as reported in Härdle and Mammen (1993), and its performance is very sensitive to the value of the smoothing parameter. The second order Edgeworth expansion, on the other hand, significantly improves on the asymptotic normal distribution and is relatively stable over the values of the smoothing parameter considered. The QQ plots in Figure 1 indicate that the second order Edgeworth expansion over-corrects slightly on both tails for $c = 0.5$, and is very accurate for $c = 1.0$, 1.5, and 2.0. Tables 3 and 4 show that the mean of $\widehat{T}$ is always negative and becomes smaller as h increases; The variance of $\widehat{T}$ is smaller than one and decreases as h increases; Overall the first two corrected cumulants are very close to the corresponding empirical (estimated) cumulants; The third and fourth corrected cumulants are not as close to the corresponding empirical cumulants as the first two.

For DGP2, Tables 5 and 6 show that the asymptotic distribution provides a more accurate approximation than for DGP1. For almost all the values of the smoothing parameter considered, the test based on two-sided asymptotic critical values has a better size than that based on one-sided asymptotic critical values. However, both are sensitive to the values of the smoothing parameter. As h increases, the rejection rate based on asymptotic critical values (both one-sided and two-sided) decreases. As n increases, the rejection rates are closer to the nominal sizes as predicted by the asymptotic theory. The rejection rates based on Edgeworth corrected critical values are very stable over the smoothing parameter values considered, and they are closer to the nominal size than based on asymptotic critical values except at 1% level. The latter may be due to the small sample size or the small number of replications in the experiment. Figure 2 shows that the standard normal distribution provides an accurate approximation at $c = 0.5$, and is better than the Edgeworth expansion in the middle but worse in the right tails. However the standard normal distribution gets worse when c increases, and the Edgeworth expansion is very stable. Tables 7 and 8 indicate that the test statistic $\widehat{T}$ still has a negative mean but closer to zero than for DGP1; it's variance is closer to one; It is positively skewed. Although the test based on two-sided normal critical values has an accurate size, it loses power in comparison with the one based on one-sided critical values. Hence the test based on Edgeworth corrected one-sided critical values is recommended.

6 Conclusion

Both our theoretical and simulation results indicate superior performance of the Edgeworth expansion to the standard normal distribution in approximating the finite sample distribution of $\widehat{T}$ under the null hypothesis. This is consistent with the superior performance of the bootstrap reported in Li and Wang (1996) whose validity is predicated on Edgeworth expansions similar to those established in this paper.

7 Appendix

Let

$$\bar{\sigma}^2 = \frac{2}{n(n-1)h^d}\sum_i\sum_{j\neq i}u_i^2u_j^2K_{ij}^2, \tag{9}$$

$$\overline{T} = \frac{nh^{d/2}I_n}{\bar{\sigma}}, \tag{10}$$

$$T_n = \frac{nh^{d/2}I_n}{\sigma_n}, \tag{11}$$

where $\sigma_n^2 = E[\bar{\sigma}^2]$. Let $\theta_n = E[u_1^2u_2^2K_{12}^2]$. Then $\sigma_n^2 = 2\theta_n \Big/ h^d$.

We assume that $\mathbf{H}_0$ is true and that the following conditions hold:

Assumption A.

1. *The parametric regression function* g_0 *is twice continuously differentiable with respect to both* β *and* x. *The estimator* $\widehat{\beta}$ *is root-n consistent, specifically,*

$$n^{1/2}(\widehat{\beta}-\beta_0) = \Omega^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^n G_iu_i + o_p(1)$$

where $\Omega = E(G_iG_i')$ *is finite and non-singular.*

2. *The estimator* $\widehat{\beta}$ *has the moderate deviation property: for some* $c > 0$,

$$\Pr\left[\left\|n^{1/2}(\widehat{\beta}-\beta_0)\right\| > c\log n\right] = o(n^{-\epsilon_2}),$$

where $\|x\| = (x'x)^{1/2}$ *is Euclidean norm.*

3. *The marginal density* $f(\cdot)$ *is strictly positive, bounded, and continuously differentiable on its support.*

4. *The conditional moments* $\mu_j(x) = E(|u_i|^j \,|X_i = x)$ *exist and satisfy* $\int \mu_j(x) f(x) dx < \infty$ *for* $j = 1, 2, \ldots$.

5. *The kernel* K *is a symmetric, bounded, differentiable density function, and satisfies* $\int K^2(u) du < \infty$.

6. *The smoothing parameter satisfies* $h \to 0$ *and* $nh^d \to \infty$, *as* $n \to \infty$.

REMARK. The precise number of moments in A4 depends on the rate of convergence (i.e. the bandwidth and dimensionality); we have chosen to implicitly assume that all moments exist. The moderate deviation assumption A2 has been verified for maximum likelihood estimators under a variety of conditions, see Pfanzagl (1980).

PROOF OF THEOREM 1. The proof is divided into three parts. In part A we outline the asymptotic expansions establishing the order in probability of the remainder terms. In part B below we justify dropping of remainder terms as far as the order $n^{-\epsilon_2}$ distributional approximation is concerned, while in part C we give an Edgeworth theorem.

A Asymptotic Expansion

As in Zheng (1996),

$$\widehat{\sigma}^2 = \bar{\sigma}^2[1 + O_p(n^{-1})],$$

whence

$$\widehat{T} = \overline{T}(\overline{\sigma}\widehat{\sigma}^{-1}) = \overline{T} + O_p(n^{-1}), \tag{12}$$

where we have used the fact that $\bar{\sigma}^2 = O_p(1)$. In fact, it will follow from the calculations given in Appendix B that the distribution of $\widehat{T}$ is the same as that of $\overline{T}$ up to order n^{-1}.

We write

$$\bar{\sigma}^2 - \sigma_n^2 = \frac{2}{n(n-1)h^d} \sum_i \sum_{j \neq i} [u_i^2 u_j^2 K_{ij}^2 - \theta_n]. \tag{13}$$

Since $\bar{\sigma}^2/2$ is a non-degenerate U-statistic, one can show by using Lemma 3.1 in Powell, Stock, and Stoker (1989) that $\bar{\sigma}^2 - \sigma_n^2 = O_p(n^{-1/2})$. Therefore, we get

$$\bar{\sigma}^{-1} = \sigma_n^{-1}[1 - \frac{1}{2\sigma_n^2}(\bar{\sigma}^2 - \sigma_n^2) + O_p(n^{-1})], \tag{14}$$

and consequently

$$\overline{T} = T_n - \frac{1}{2\sigma_n^2} T_n(\bar{\sigma}^2 - \sigma_n^2) + O_p(n^{-1}). \tag{15}$$

We now simplify the expression for T_n. Using assumption A1, we have that $\widehat{u}_i = u_i - G_i'(\widehat{\beta} - \beta) + o_p(n^{-1/2})$. Thus,

$$\begin{aligned} I_n = & \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} u_i u_j K_{ij} - \frac{2}{n^2(n-1)h^d} \sum_k \sum_i \sum_{j \neq i} u_k \omega_{ki} u_j K_{ij} \\ & + \frac{1}{n^3(n-1)h^d} \sum_k \sum_i \sum_{j \neq i} \sum_l u_k \omega_{ki} K_{ij} \omega_{jl} u_l + O_p(n^{-3/2}). \end{aligned} \tag{16}$$

Let $U = \frac{1}{n(n-1)h^d} \sum_i \sum_{j \neq i} G_i K_{ij} G_j'$. Then, noting that U is a non-degenerate U-statistic, one can show by using Lemma 3.1 in Powell, Stock, and Stoker (1989) that $U = E(G_1 K_{12} G_2'/h^d) + O_p(n^{-1/2})$. Thus, the third term on the right hand side of (16) equals

$$\frac{1}{n^2} \sum_k \sum_l u_k G_k' \Omega^{-1} U \Omega^{-1} G_l u_l = \frac{1}{n^2} \sum_k \sum_l u_k G_k' \mu_\Omega G_l u_l + O_p(\frac{1}{n^{3/2}}). \tag{17}$$

Substituting (16) into (11) and using (17), one gets

$$\begin{aligned} T_n = & \frac{1}{(n-1)h^{d/2}\sigma_n} \sum_i \sum_{j \neq i} u_i u_j K_{ij} - \\ & \frac{2}{n(n-1)h^{d/2}\sigma_n} \sum_k \sum_i \sum_{j \neq i} u_k \omega_{ki} u_j K_{ij} \\ & + \frac{h^{d/2}}{n\sigma_n} \sum_k \sum_l u_k \tau_{kl} u_l + O_p(\frac{h^{d/2}}{\sqrt{n}}) \\ \equiv & \; T_{n1} - 2T_{n2} + T_{n3} + O_p(\frac{h^{d/2}}{\sqrt{n}}). \end{aligned} \tag{18}$$

Now consider the second term on the right hand side of (15). From the facts that: $T_{n2} = O_p(h^{d/2})$, $T_{n3} = O_p(h^{d/2})$, and $\bar{\sigma}^2 - \sigma_n^2 = O_p(n^{-1/2})$, it follows that

$$T_n(\bar{\sigma}^2 - \sigma_n^2) = T_{n1}(\bar{\sigma}^2 - \sigma_n^2) + O_p(\frac{h^{d/2}}{\sqrt{n}}). \tag{19}$$

Equations (15), (18), and (19) yield the following asymptotic expansion for $\overline{T}$:

$$\overline{T} = T_{n1} - 2T_{n2} + T_{n3} - 2T_{n4} + O_p(\frac{h^{d/2}}{\sqrt{n}}), \tag{20}$$

where T_{n1}, T_{n2}, T_{n3} are defined in (18), and

$$\begin{aligned} T_{n4} &= \frac{1}{4\sigma_n^2} T_{n1}(\bar{\sigma}^2 - \sigma_n^2) \\ &= \frac{1}{2n(n-1)^2 h^{3d/2}\sigma_n^3} \sum_i \sum_{j\neq i} \sum_k \sum_{l\neq k} u_i u_j K_{ij} [u_k^2 u_l^2 K_{kl}^2 - \theta_n]. \end{aligned}$$

In conclusion, the asymptotic expansion of $\overline{T}$ up to order $O_p(\frac{h^{d/2}}{\sqrt{n}})$ is given by (20) whose properties we record below in more detail.

Lemma A1.

(i) T_{n1} *defined below satisfies* $T_{n1} \to N(0,1)$ *in distribution,*

$$T_{n1} = \frac{1}{(n-1)h^{d/2}\sigma_n} \sum_i \sum_{j\neq i} u_i u_j K_{ij}; \tag{21}$$

(ii) T_{n2} *satisfies*

$$\begin{aligned} T_{n2} = & \frac{1}{n(n-1)h^{d/2}\sigma_n} \sum\sum\sum_{i\neq j\neq k} u_k \omega_{ki} u_j K_{ij} \\ & + \frac{1}{n(n-1)h^{d/2}\sigma_n} \sum\sum_{i\neq j} u_j^2 \omega_{ji} K_{ij} \\ & + O_p(n^{-1}). \end{aligned} \tag{22}$$

In addition, $h^{-d/2}T_{n2} \to N(\mu^*, \sigma^{2*})$, *where* $\mu^* \neq 0$ *and* $\sigma^{2*} > 0$ *are constants;*

(iii) T_{n3} *satisfies*

$$T_{n3} = \frac{h^{d/2}}{n\sigma_n} \sum\sum_{k \neq l} u_k \tau_{kl} u_l + \frac{h^{d/2}}{n\sigma_n} \sum_k u_k^2 \tau_{kk}. \tag{23}$$

In addition, $h^{-d/2} T_{n3} \to \sum_{i=1}^{\infty} \lambda_i \chi^2_{i,[1]} + \mu^{**}$, *where* λ_i *and* μ^{**} *are constants, and* $\chi^2_{i,[1]}$*'s are independent* χ^2 *random variables with degree* 1;

(iv) T_{n4} *is given by*

$$T_{n4} = \frac{1}{2n(n-1)^2 h^{3d/2} \sigma_n^3} \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} u_i u_j K_{ij} [u_k^2 u_l^2 K_{kl}^2 - \theta_n]. \tag{24}$$

In addition, $n^{1/2} T_{n4} \to N(0, \sigma^{2**})$ *in distribution, where* σ^{2**} *is a positive constant.*

Lemma A1 states that up to order $O_p(h^{d/2} n^{-1/2})$, $\widehat{T}$ is the same as the truncated statistic

$$T_* = T_{n1} - 2T_{n2} + T_{n3} - 2T_{n4}.$$

We are now ready to derive the first four moments of T_*. First we introduce some additional notations. Let

$$\begin{aligned} \mu_i &= E(T_{ni}), \qquad i = 1,2,3,4, \\ \mu_{ij} &= E(T_{ni}T_{nj}), \qquad i,j = 1,2,3,4, \\ \mu_{ijk} &= E(T_{ni}T_{nj}T_{nk}), \qquad i,j,k = 1,2,3,4, \\ \mu_{ijkl} &= E(T_{ni}T_{nj}T_{nk}T_{nl}), \qquad i,j,k,l = 1,2,3,4. \end{aligned}$$

$$\begin{aligned} \mu_2 &= \frac{1}{h^{d/2}\sigma_n} E[u_2^2 \omega_{21} K_{12}] = h^{d/2} m_2, \quad \mu_3 = \frac{h^{d/2}}{\sigma_n} E[u_1^2 \tau_{11}] = h^{d/2} m_3, \\ \mu_4 &= \frac{1}{\sigma_n^3 (n-1) h^{3d/2}} E[u_1^3 u_2^3 K_{12}^3] = \frac{1}{(n-1)h^{d/2}} m_4. \end{aligned}$$

To evaluate the second moment of T_*, we note from Lemma A1 (ii), (iii), and (iv) that $\mu_{14} = 0$, $\mu_{44} = O(n^{-1})$, $\mu_{24} = O(h^{d/2}/\sqrt{n})$, and $\mu_{34} = O(h^{d/2}/\sqrt{n})$. Hence,

$$E[T_*^2] = \mu_{11} + 4\mu_{22} + \mu_{33} - 4\mu_{12} + 2\mu_{13} - 4\mu_{23} + O(\frac{h^{d/2}}{\sqrt{n}}), \tag{25}$$

where using the expressions for T_{n1}, T_{n2}, and T_{n3} given in equations (21)-(23), one can easily show that

$$
\begin{aligned}
\mu_{11} &= 1+O(n^{-1}),\\
\mu_{22} &= \frac{1}{\sigma_n^2 h^d}E[u_4^2u_2^2\omega_{41}K_{12}(\omega_{43}K_{32}+\omega_{23}K_{34})]+\mu_2^2+O(n^{-1}) = h^d[m_{22}+m_2^2]+O(n^{-1}),\\
\mu_{33} &= \frac{2h^d}{\sigma_n^2}E[u_2^2u_1^2\tau_{21}^2]+\mu_3^2+O(n^{-1}) = h^d[2m_{33}+m_3^2]+O(n^{-1}),\\
\mu_{12} &= \frac{2}{\sigma_n^2 h^d}E[u_1^2u_2^2K_{12}\omega_{13}K_{32}] = 2h^d m_{12},\\
\mu_{13} &= \frac{2}{\sigma_n^2}E[u_1^2u_2^2K_{12}\tau_{12}] = 2h^d m_{13},\\
\mu_{23} &= \frac{2}{\sigma_n^2}E[u_3^2u_2^2\omega_{31}K_{12}\tau_{32}]+\mu_2\mu_3+O(n^{-1}) = h^d[2m_{23}+m_2m_3]+O(n^{-1}).
\end{aligned}
$$

Similarly, Lemma A1 implies that $\mu_{222} = O(h^{3d/2})$, $\mu_{333} = O(h^{3d/2})$, $\mu_{444} = O(n^{-3/2})$, $\mu_{144} = O(n^{-1})$, $\mu_{124} = O(h^{d/2}/\sqrt{n})$, and $\mu_{134} = O(h^{d/2}/\sqrt{n})$. Thus, ignoring terms of order $O(h^{3d/2} + h^{d/2}/\sqrt{n})$ or smaller, we get

$$E[T_*^3] = \mu_{111} - 6\mu_{112} - 6\mu_{114} + 3\mu_{113} + 12\mu_{122} + 3\mu_{133} - 12\mu_{123}. \tag{26}$$

There are only three terms on the right hand side of (26) that are of order larger than $O(h^{3d/2})$. They are given by

$$
\begin{aligned}
\mu_{111} &= \frac{8}{\sigma_n^3 h^{3d/2}}E[u_1^2u_2^2u_3^2K_{12}K_{13}K_{23}] + \frac{4}{nh^{3d/2}\sigma_n^3}E[u_1^3u_2^3K_{12}^3]\\
&= 8h^{d/2}m_{111} + \frac{4}{nh^{d/2}}m_4,\\
\mu_{112} &= \frac{1}{\sigma_n h^{d/2}}E[u_4^2\omega_{43}K_{34}] + O(h^{3d/2}) = h^{d/2}m_2 + O(h^{3d/2}),\\
\mu_{113} &= \frac{h^{d/2}}{\sigma_n}E[u_3^2\tau_{33}] + O(h^{3d/2}) = h^{d/2}m_3 + O(h^{3d/2}).
\end{aligned}
$$

To evaluate $E[T_*^4]$, again we first use Lemma A1 to single out terms of order $O(h^{2d} + h^{d/2}/\sqrt{n})$ or smaller and ignore them. This gives

$$E[T_*]^4 = \mu_{1111} - 8\mu_{1112} + 4\mu_{1113} - 8\mu_{1114} + 24\mu_{1122} + 6\mu_{1133} - 24\mu_{1123}, \tag{27}$$

where

$$
\begin{aligned}
\mu_{1111} &= 3 + 48h^d m_{1111} + O(h^{2d}), \\
\mu_{1112} &= 2h^d[3m_{12} + 4m_{111}m_2] + O(h^{2d}), \\
\mu_{1113} &= 2h^d[3m_{13} + 4m_{111}m_3] + O(h^{2d}), \\
\mu_{1114} &= O(1/n), \\
\mu_{1122} &= h^d[m_{22} + m_2^2], \\
\mu_{1133} &= h^d[2m_{33} + m_3^2], \\
\mu_{1123} &= h^d[2m_{23} + m_2 m_3].
\end{aligned}
$$

■

B Error Term Calculation

We establish the following result:

LEMMA B. *The distribution of $\widehat{T}$ is the same as the distribution of T_* [defined after Lemma A1] to order $n^{-\epsilon_2}$.*

PROOF OF LEMMA B. This argument is similar to that given in Linton (1995) but is included for completeness. We use the following result of Sargan and Mikhail (1971). For any random variables T, T^*, R with $T = T^* + R$, we have for all x and ζ,

$$
\left|\Pr\left(T \leq x\right) - \Pr\left(T^* \leq x\right)\right| \leq \Pr\left(|R| > \zeta\right) + \Pr\left(|T^* - x| < \zeta\right), \tag{28}
$$

see Rothenberg (1984) and Robinson (1988). Provided T^* has a bounded density, the last term is $O(\zeta)$ as $\zeta \to 0$. We therefore choose $\zeta = O(n^{-\epsilon_2} \log^{-1} n)$ and show that

$$\Pr\left(n^{\epsilon_2} \log n \, |R| > c\right) = o(n^{-\epsilon_2}) \tag{29}$$

for some positive constant c. It follows from (29) that T and T^* have the same distribution to order $n^{-\epsilon_2}$. This property (over and above the in probability equivalence) is important to establish for the interpretation of the asymptotic cumulants, see for example Srinavasan (1970) and Rothenberg (1984).

We must check that (29) holds for all the various remainder terms in the expansion; we just sketch the argument for (12). Let $T = \widehat{T}$, $T^* = T_*$, and R being the difference between them. Thus,

$$\begin{aligned} R &= -T_* \times \frac{1}{\widehat{\sigma}}(\widehat{\sigma} - \overline{\sigma}) \\ &= -T_* \times \frac{1}{2\widehat{\sigma}\sigma^*}(\widehat{\sigma}^2 - \overline{\sigma}^2), \end{aligned}$$

where the second line follows by the mean value theorem with σ^* intermediate between $\widehat{\sigma}$ and $\overline{\sigma}$. Let $\mathcal{E} = \{n^{\epsilon_2} \log n \, |R| > c\}$, $\mathcal{E}_0 = \{|T_* / 2\widehat{\sigma}\sigma^*| \leq c_0\}$, $\mathcal{E}_1 = \{n^{a_1} |\widehat{\sigma}^2 - \overline{\sigma}^2| \leq c_1\}$, and $\mathcal{E}_2 = \{n^{a_2} |\overline{\sigma}^2 - \sigma_n^2| \leq c_2\}$ for positive constants c_j and a_j with $a_1 < 1$ and $a_2 < 1/2$ [recall that $n(\widehat{\sigma}^2 - \overline{\sigma}^2) = O_p(1)$ and $n^{1/2}(\overline{\sigma}^2 - \sigma_n^2) = O_p(1)$]. We exploit the following inequality for events $\mathcal{E}$, $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E} \cap \mathcal{E}_1 \cap \mathcal{E}_2) + \Pr(\mathcal{E}_1^c) + \Pr(\mathcal{E}_2^c)$$

and then use the fact that $\mathcal{E}_1 \cap \mathcal{E}_2 \subset \mathcal{E}_0$ for certain c_j [since $\liminf_{n\to\infty} \sigma_n^2 > 0$] to end up with

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E} \cap \mathcal{E}_0) + \Pr(\mathcal{E}_1^c) + \Pr(\mathcal{E}_2^c).$$

By the Markov inequality

$$\Pr(\mathcal{E}_1^c) \leq \frac{E\left[\{n |\widehat{\sigma}^2 - \overline{\sigma}^2|\}^{r_1}\right]}{n^{r_1(1-a_1)} c_1^{r_1}} \quad ; \quad \Pr(\mathcal{E}_2^c) \leq \frac{E\left[\left\{n^{1/2} |\overline{\sigma}^2 - \sigma_n^2|\right\}^{r_2}\right]}{n^{r_2(1/2-a_2)} c_1^{r_2}}$$

for any r_1 and r_2. Thus, provided the relevant moments exist and $r_1(1-a_1) > \epsilon_2$ and $r_2(1/2-a_2) > \epsilon_2$, we have

$$\Pr(\mathcal{E}_\ell^c) = o(n^{-\epsilon_2}), \quad \ell = 1, 2.$$

Finally,

$$\begin{aligned} \Pr(\mathcal{E} \cap \mathcal{E}_0) &\leq \Pr\left[n^{\epsilon_2} \log n \left|\widehat{\sigma}^2 - \overline{\sigma}^2\right| \geq c\right] \\ &= o(n^{-\epsilon_2}) \end{aligned}$$

by the same arguments. In this case, we take $a_1 < 1 - \epsilon_2$ and $a_2 < 1/2 - \epsilon_1$, with r_1 and r_2 correspondingly large. The moments of $\overline{\sigma}^2 - \sigma_n^2$ exist by standard arguments based on assumptions A4 and A5 [note that $\int K^2(u)du < \infty$ and K bounded imply that $\int K^j(u)du < \infty$, $j \geq 2$], see Robinson (1995). The moments of $\widehat{\sigma}^2 - \overline{\sigma}^2$ do not necessarily exist unless the corresponding moments of $n^{1/2}(\widehat{\beta} - \beta_0)$ do. Therefore, it is necessary to expand out $\widehat{\sigma}^2 - \overline{\sigma}^2$ in terms of U-statistics [whose moments do exist] and $n^{1/2}(\widehat{\beta} - \beta_0)$; assumption A2 is used to ensure that contribution from $n^{1/2}(\widehat{\beta} - \beta_0)$ is sufficiently small.

■

C Distributional approximation

The truncated test statistic T_* is a linear combination of certain U-statistics. In this section we establish the validity of an Edgeworth expansion for the first of these four terms, i.e. T_{n1}. The remaining terms are of smaller order and we expect that with considerably more work one could establish the validity of the distributional approximation for T_* itself using the techniques of Linton (1996). In any case, we shall restrict our attention to the degenerate weighted U-statistic[7]

[7] The upper triangular form is just for convenience and all results likewise hold for the full quadratic form.

$$U_n = \frac{1}{nh^{d/2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} u_i u_j K_{ij},$$

and we shall suppose for notational simplicity that $u_i|X_i$, $i = 1, \ldots, n$ are independent with variance such that U_n has mean zero and variance one (and is thus asymptotically standard normal). Consistent with the above moment calculations, we shall suppose that U_n has third and fourth cumulants $\varkappa_3 n^{-\epsilon_1}$ and $\varkappa_4 n^{-\epsilon_2}$, say, to order $n^{-\epsilon_2}$ [here, $\epsilon_2 = 2\epsilon_1$]. In the sequel all calculations are made in the conditional distribution [given $X_1, \ldots, X_n$].

We make use of the following proposition.

PROPOSITION C1: *Let F and G be two signed measures with Fourier transforms γ and ψ, where $\gamma(0) = 1$, while ψ is continuously differentiable with $\psi(0) = 1$ and $\psi'(0) = 0$. Suppose also that G is differentiable and $\int |x||G'(x)|dx < \infty$. Then, for all x and all $T > 0$, there is a constant m such that*

$$|F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^{T} \left| \frac{\psi(s) - \gamma(s)}{s} \right| ds + \frac{24m}{\pi T}.$$

The so-called smoothing lemma is proved in Bhattacharya and Rao (1976, Lemma 12.1 and Lemma 12.2). For our application we identify G with $\widetilde{F}_{n2}$ [or rather the corresponding Edgeworth measure based on the cumulants of U_n], i.e.

$$\widetilde{F}_{n2}(x) = \Phi(x) - \phi(x) \left\{ \frac{\varkappa_3 H_2(x)}{6n^{\epsilon_1}} + -\frac{3\varkappa_4 H_3(x) + \varkappa_3^2 H_5(x)}{72n^{\epsilon_2}} \right\},$$

where $H_j(x)$ are the Hermite polynomials, e.g. $H_2(x) = x^2 - 1$, and F with the distribution function of U_n. We choose $T = n^{\epsilon_2} \log n$, in which case

$$\sup_{-\infty < x < \infty} \left| \Pr[U_n \leq x] - \widetilde{F}_{n2}(x) \right| \leq \frac{2}{\pi} \int_0^{n^{\epsilon_2} \log n} \left| \frac{\psi_n(s) - \widetilde{\psi}_n(s)}{s} \right| ds + o(n^{-\epsilon_2}),$$

where $\widetilde{\psi}_n(s) = \int e^{\mathbf{i}sx} d\widetilde{F}_{n2}(x)$ is the Fourier transform of $\widetilde{F}_{n2}(x)$, while $\psi_n(s) = E\left[e^{\mathbf{i}sU_n}\right]$ is the characteristic function of U_n.

We split the range of integration into several different parts and show that each sub-integral is $o(n^{-\epsilon_2})$. We have

$$\begin{aligned}\int_0^{n^{\epsilon_2}\log n}\left|\frac{\psi_n(s)-\widetilde{\psi}_n(s)}{s}\right|ds &\leq \int_0^{n^{\epsilon_1/2}/\log n}\left|\frac{\psi_n(s)-\widetilde{\psi}_n(s)}{s}\right|ds+\\ &\quad \int_{n^{\epsilon_1/2}/\log n}^{n^{\epsilon_1}}\left|\frac{\psi_n(s)}{s}\right|ds+\\ &\quad \int_{n^{\epsilon_1}}^{n^{\epsilon_2}\log n}\left|\frac{\psi_n(s)}{s}\right|ds+\\ &\quad \int_{n^{\epsilon_1}}^{\infty}\left|\frac{\widetilde{\psi}_n(s)}{s}\right|ds\\ &= I_1+I_2+I_3+I_4.\end{aligned}$$

For large s we rely on crude bounds for the magnitude of $|\psi_n(s)|$, which hold when s is kept at a distance from the origin. Our proof technique in this follows closely that used in Callaert, Janssen and Verarbereke (1980). The last integral, I_4, is $o(n^{-\epsilon_2})$ because of the form of the Edgeworth characteristic function $\widetilde{\psi}_n$. The main difficulty here is in establishing what happens when s is very small.

We first establish the behaviour of the characteristic function for large values of its argument, i.e. I_3. Let

$$H_{\ell N}=\frac{1}{nh^{d/2}}\sum_{j=N+1}^{n}u_\ell u_j K_{\ell j},\quad \ell=1,\ldots,N,$$

where $6<N(n)<n$ is an integer, and

$$U_{N-1,m}=\frac{1}{nh^{d/2}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{m}u_iu_jK_{ij}$$

for integers $N,m\leq n$, with $U_n=U_{n-1,n}$. Note that

$$\begin{aligned} U_{N,n} - U_{N-1,N} &= \frac{1}{nh^{d/2}} \left\{ \sum_{i=1}^{N} \sum_{j=i+1}^{n} u_i u_j K_{ij} - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} u_i u_j K_{ij} \right\} \\ &= \sum_{\ell=1}^{N} H_{\ell N}. \end{aligned}$$

We make the additional assumption that

Assumption B1. *For any* $a > 0$, *there exists a constant* $c < 1$, *such that for* $N(n) = O(n^{\delta})$, *with* $0 < \delta < (4 - 5\epsilon_2)/4$,

$$\Pr\left[\left|E\left\{\exp(\mathbf{i}sH_{1N})\left|\mathcal{X}_{N,n}\right.\right\}\right| \geq c\right] = o\left(\frac{1}{n^{\epsilon_2}\log n}\right),$$

uniformly for $s \in \left[an^{\epsilon_1}, n^{\epsilon_2}\log n\right]$, *where* $\mathcal{X}_{N,n} = \left\{u_{N+1}, u_{N+2}, \ldots, u_n\right\}$.

Remark. Verification of this condition [for similar random variables] was provided in Linton (1996) when u_i are normally distributed.

Lemma C1: *For some* $a > 0$,

$$\int_{an^{\epsilon_1}}^{n^{\epsilon_2}\log n} \left|\frac{\psi_n(s)}{s}\right| ds = o(n^{-\epsilon_2}). \tag{30}$$

Proof. The proof of Lemma C1 is completed using the following lemma which is proved below.

Lemma C2: *For all* s *and for all* n *and* N, *with* $6 < N < n$, *there exists* $\lambda < \infty$, *such that*

$$|\psi_n(s)| \leq E\left[\left|E\left\{\exp(\mathbf{i}sH_{1N})\left|\mathcal{X}_{N,n}\right.\right\}\right|^{N-6}\right]\left\{1 + \lambda\sum_{k=0}^{3}|s|^k\left[\frac{N^{2k}h^{dk/2}}{n^k}\right]\right\} + \lambda|s|^4\frac{N^4}{n^4}. \tag{31}$$

Then, by assumption B1 there exists a constant $c < 1$ such that

$$|E[\exp\mathbf{i}sH_{1N})|u_{N+1}, \ldots, u_n]| \leq c,$$

uniformly for $s \in [an^{\epsilon_1},\, n^{\epsilon_2}\log n]$ with probability $1 - o\left(\frac{1}{n^{\epsilon_2}\log n}\right)$. This gives us an exponential bound for

$$E[|E[\exp(\mathbf{i}sH_N)|u_{N+1},\ldots,\, u_n]|^{N-6}]\left\{1+\lambda\sum_{k=0}^{3}|s|^k\left[\frac{N^{2k}h^{dk/2}}{n^k}\right]\right\}.$$

The second term on the right hand side of (31) contributes a term of magnitude $O(n^{4\epsilon_2}\log n \times N^4/n^4)$ to (30); this is $o(n^{-\epsilon_2})$, provided N is small enough; $N(n) = o(n^{(4-5\epsilon_2)/4})$ will work.

■

We now examine I_1 and I_2 which concern the case that s is small, i.e. less than n^{ϵ_1}. We first show that in the relevant range the characteristic function of the quadratic form $U_n(u_1,\ldots,u_n)$ is close to the characteristic function of $U_n(\zeta_1,\ldots,\zeta_n)$, where ζ_j, $j=1,\ldots,n$ are independent random variables with the same moments as u_j to order p. We will find it convenient to replace u_j by normal [or scale mixtures of normal] random variables ζ_j, because quadratic forms in these latter variables are much easier to handle. Let $\lambda_{\max} = \max_{1\le j\le n}\lambda_{nj}$, where λ_{nj}, $j=1,\ldots,n$, are the eigenvalues of the smoother matrix $\mathcal{K} = n^{-1}h^{-d/2}\left(K_{ij}\right)_{i,j}$. To establish this result we use a modification of Proposition 1.4 of Mikosch (1991):

Proposition C2. *Assume that* $\sup_{j\ge 1} E|u_j|^p < \infty$ *for some integer* $p > 2$. *Suppose also that* $E(\zeta_i^k) = E(u_i^k)$, $k = 2,\ldots,p-1$. *If* $\lambda_{\max} \to 0$, *then*

$$\left|E\left[e^{itU_n(u_1,\ldots,u_n)}\right] - E\left[e^{itU_n(\zeta_1,\ldots,\zeta_n)}\right]\right| \le |t|^p O(L_{p,n}), \tag{32}$$

where $L_{p,n} = \sum_{i=1}^n\left(\sum_j K_{ij}^2\right)^{p/2}\Big/\left(nh^{d/2}\right)^p$.

Remark. Mikosch's result is stated for i.i.d random variables but extends straightforwardly to heterogeneous random variables under our assumptions. The eigenvalue condition holds in probability by the following argument. By definition,

$$\lambda_{\max} = \max_{\theta\in\Theta_n} Y_n(\theta) \quad ; \quad Y_n(\theta) = \theta'\mathcal{K}\theta \quad ; \quad \Theta_n = \{\theta\in\mathbb{R}^n; \quad \theta'\theta = 1\}.$$

We have that $Y_n(\theta_n) \rightarrow_p 0$ for any fixed or random sequence $\theta_n \in \Theta_n$ by calculation of the first moment [in fact, $Y_n(\theta_n) = O_p(h^{d/2})$]. In particular it is true for $\widehat{\theta}_n = \arg\max_{\theta \in \Theta_n} \theta' \mathcal{K} \theta$ [which exists by the compactness of Θ_n].

We have

$$\begin{aligned} L_{p,n} &= \sum_{i=1}^{n} \left(\sum_j K_{ij}^2 \right)^{p/2} \Big/ \left(n^2 h^d \right)^{p/2} \\ &= \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{nh^d} \sum_j K_{ij}^2 \right)^{p/2} \right\} \Big/ n^{(\frac{p}{2}-1)} \\ &= O_p(n^{-(\frac{p}{2}-1)}) \end{aligned}$$

under our conditions. Therefore, (32)$\leq |t|^p O(n^{-(p-2)/2})$ and

$$\int_0^{n^{\epsilon_1}} \left| \frac{E\left[e^{itU_n(u_1,\ldots,u_n)}\right] - E\left[e^{itU_n(\zeta_1,\ldots,\zeta_n)}\right]}{t} \right| dt = O(n^{p\epsilon_1} n^{-(p-2)/2}). \tag{33}$$

We shall use the additional assumption:

Assumption B2. *The random variables u_i are symmetric about zero and have positive kurtosis. The bandwidth h is of larger magnitude than $n^{-1/2d}$*

which guarantees that we can choose ζ_j to be scale mixtures of normals [that is $\zeta | \sigma^2 \sim N(0, \sigma^2)$, where σ has some distribution P_σ] whose moments through $p - 1 = 5$ agree with u_j, and, since $\epsilon_1 < 1/4$, (33)$= o(n^{-1/2})$ [$= o(n^{-\epsilon_2})$] as required.

Remark. This assumption is clearly restrictive relative to other assumptions; it really reflects our poor proof technique rather than actual constraints on the sampling scheme. In fact, inspection of the cumulants in Theorem 1 show that [apart from one term of order $n^{-1}h^{-d/2}$ which can be ignored for a much wider range of bandwidths] no moments of u_i higher than the second appear, i.e. u_i might as well be normally distributed under a much wider range of circumstances.

We can now assume without loss of generality that u_i are scale mixtures of normals. For any random variables u_j we can write

$$U_n = \sum_{j=1}^{n} \lambda_{nj}(v_{nj}^2 - \sigma_{nj}^2) \quad (34)$$

$$\equiv \sum_{j=1}^{n} Y_{nj}, \quad (35)$$

where v_{nj}, $j = 1, \ldots, n$, are uncorrelated random variables. However, in the special case that u_j are conditionally normally distributed, we have $v_{nj} = u_j$ and the corresponding Y_{nj}, $j = 1, \ldots, n$, are independent random variables [at least conditional on the scale parameters]. The proof now proceeds as for the standard case of independent random variables with n replaced by h^{-d}; everything just depends on the magnitudes of averages of various moments of Y_{nj}. See Feller (1966, p 521) for an early treatment. Define the Lyapunov coefficients

$$l_{r,n} = \sum_{j=1}^{n} E\left|Y_{nj}\right|^r, \quad r = 3, 4, \ldots$$

and note that $l_{r,n} = O(n^{-(r-2)\epsilon_1})$, $r = 3, 4, \ldots$, by direct calculation.

LEMMA C3. *For some* $a > 0$,

$$\int_{n^{\epsilon_1/2}/\log n}^{an^{\epsilon_1}} \left|\frac{\psi_n(s)}{s}\right| ds = o(n^{-\epsilon_2}).$$

Proof of Lemma C3. We use the following result for sums of independent random variables which follow by identical arguments to Bhattacharya and Rao (1976, Lemma 8.9) after replacing their definition of $l_{r,n}$ by ours.

Proposition C3. *For all positive* $\delta < 1 \cdot 5$,

$$|\psi_n(t)| \leq \exp\left\{-\frac{\delta^2}{3}t^2\right\},$$

for all t *satisfying* $|t| \leq (1 \cdot 5 - \delta) l_{3,n}^{-1}$.

Then,

$$\int_{n^{\epsilon_1/2}/\log n}^{an^{\epsilon_1}} \left|\frac{\psi_n(s)}{s}\right| ds \;\leq\; n^{-\epsilon_1/2} \log n \int_{n^{\epsilon_1/2}/\log n}^{\infty} \exp\left\{-\frac{\delta^2}{3}s^2\right\} ds$$

$$= \; o(n^{-\epsilon_2}),$$

in fact $o(n^{-k})$ for any positive k, by a standard application of Hôpital's rule.

Lemma C4.

$$\int_0^{n^{\epsilon_1/2}/\log n} \left|\frac{\psi_n(s) - \tilde{\psi}_n(s)}{s}\right| ds = o(n^{-\epsilon_2}).$$

Proof of Lemma C4. We use the following result for sums of independent random variables which follow by identical arguments to Bhattacharya and Rao (1976, Theorem 9.12) after replacing their definition of $l_{r,n}$ by ours.

PROPOSITION C4. *There exists a positive constant* c_1 *and a positive sequence* $\delta(n) \to 0$, *such that for all* t *satisfying* $|t| \leq c_1 n^{-\epsilon_1/2}$, *one has*

$$\left|\psi_n(t) - \widetilde{\psi}_n(t)\right| \leq \frac{\delta(n)}{n^{\epsilon_2}} P(t) \exp\left(-\frac{t^2}{4}\right),$$

for some polynomial $P(t)$ *with constant coefficients and zero intercept.*

Now integrate over t and use the fact that $\int_{-\infty}^{\infty} |t|^{-1} P(t) \exp\left(-\frac{t^2}{4}\right) dt < \infty$ [there is no intercept in P].

■

PROOF OF LEMMA C2. The proof follows by the same argument as given in Lemma 5 of Callaert et al. (1980). The different magnitudes are a consequence of the bandwidth h.

Firstly, by rewriting U_n using addition and subtraction, we obtain

$$\psi_n(s) = E\left(E\left[\exp\{\mathbf{i}s(U_{N,n} - U_{N-1,N})\}\exp(\mathbf{i}sU_{N-1,N})\exp\{\mathbf{i}s(U_{n-1,n} - U_{N,n})\}\,|\mathcal{X}_{N,n}\right]\right).$$

Then, because $|\exp(\mathbf{i}s)| \leq 1$ for any real s,

$$\begin{aligned} |\psi_n(s)| &\leq E\left(\left|E\left[\exp\{\mathbf{i}s(U_{N,n} - U_{N-1,N})\}\exp(\mathbf{i}sU_{N-1,N})\,|\mathcal{X}_{N,n}\right]\right|\right) \\ &\leq \sum_{k=0}^{3} |s|^k E\left(\left|E\left[\exp\{\mathbf{i}s(U_{N,n} - U_{N-1,N})\}\, U_{N-1,N}^k\,|\mathcal{X}_{N,n}\right]\right|\right) + |s|^4 E\left(U_{N-1,N}^4\right), \qquad (36) \end{aligned}$$

by the fact that for any real t and positive integer k, $\left|e^{\mathbf{i}t} - \sum_{j=0}^{k-1} \frac{(\mathbf{i}t)^j}{j!}\right| \leq \frac{t^k}{k!}$.

We deal with the terms $k = 0, \ldots, 3$, in turn.

(i) For $k=0$, we have

$$E\left(\left|E\left[\exp\left\{\mathbf{i}s(U_{N,n}-U_{N-1,N})\right\}|\mathcal{X}_{N,n}\right]\right|\right) = E\left[\left|E\left(\exp\mathbf{i}s\sum_{\ell=1}^{N}H_{\ell N}\,|\mathcal{X}_{N,n}\right)\right|\right]$$

$$\leq E\left[\left|E\left(\exp\mathbf{i}sH_{1N}\,|\mathcal{X}_{N,n}\right)\right|^{N}\right],$$

since, conditional on $\mathcal{X}_{N,n}$, the $H_{\ell N}$ are mutually independent random variables. We can clearly replace N by $N-6$.

(ii) For $k=1$, we have $E\left[\exp\left\{\mathbf{i}s(U_{N,n}-U_{N-1,N})\right\}U_{N-1,N}\,|\mathcal{X}_{N,n}\right]=$

$$\frac{1}{nh^{d/2}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}E\left[u_iu_jK_{ij}\exp\left\{\mathbf{i}s(H_{iN}+H_{jN}\right\}|\mathcal{X}_{N,n}\right]E\left[\exp\left\{\mathbf{i}s\sum_{\substack{\ell=1\\ \ell\neq i,j}}^{N}H_{\ell N}\right\}|\mathcal{X}_{N,n}\right].$$

Since

$$E\left[h^{-d}u_iu_jK_{ij}\exp\left\{\mathbf{i}s(H_{iN}+H_{jN}\right\}|\mathcal{X}_{N,n}\right]=O(1)$$

and we have N^2 such terms (to be multiplied by $h^{d/2}/n$), we get the stated coefficient on $|s|$.

(iii) For $k=2$, we have, similarly, N^4 terms (to be multiplied by h^d/n^2) of the type

$$E\left[h^{-2d}u_{i_1}u_{j_1}u_{i_2}u_{j_2}K_{i_1j_1}K_{i_2j_2}\exp\left\{\mathbf{i}s\left(H_{i_1N}+H_{i_2N}+H_{j_1N}+H_{j_2N}\right)\right\}|\mathcal{X}_{N,n}\right]=O(1).$$

(iv) For $k=3$, we have, similarly, N^6 terms (to be multiplied by $h^{3d/2}/n^3$) of the type

$$E\left[h^{-3d}u_{i_1}u_{j_1}u_{i_2}u_{j_2}u_{i_3}u_{j_3}K_{i_1j_1}K_{i_2j_2}K_{i_3j_3}\exp\left\{\mathbf{i}s\sum_{\substack{j=1,2,3\\ \ell=i,j}}H_{\ell_jN}\right\}|\mathcal{X}_{N,n}\right]=O(1).$$

In conclusion, we have established the magnitudes of the coefficients on $|s|^k$ in the first term on the right hand side of (36).

Finally, we have for any m,N, $E\left[|U_{N-1,m}|^2\right]=O(Nm/n^2)$, and therefore

$$E\left[|U_{N-1,N}|^j\right] = O(N^j/n^j), \quad j = 2, 3, \ldots$$

which establishes the magnitude of the second term on the right hand side of (36).

■

Acknowledgements

We would like to thank Pedro Gozalo, Joel Horowitz, and Qi Li for helpful discussions. Research of the first author was supported by the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and the Humanities Research Council of Canada, while research of the second author was supported by the National Science Foundation. Gauss code is available from the first author for implementing the size correction routines proposed in this paper.

References

Andrews, D.W.K. (1995): "A Conditional Kolmogorov Test," Cowles Foundation Discussion Paper, no 1111R.

Bhattacharya, R. N. and Ranga Rao (1976): *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

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.

Bierens, H.J., (1982): "Consistent Model Specification Tests," *Journal of Econometrics* **20**, 105-134.

BIERENS, H.J., (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica* **58**, 1443-1458.

BIERENS, H.J., AND W. PLOBERGER (1996): "Asymptotic Theory of Integrated Conditional Moment Tests," Forthcoming in *Econometrica.*

CALLAERT, H., P. JANSSEN AND N. VERAVERBEKE (1980): "An Edgeworth Expansion for U-Statistics," *Annals of Statistics* **8**, 299–312.

EUBANK, R.L., AND V.N. LARICCIA (1992): "Asymptotic Comparison of Cramér-von Mises and Nonparametric Function Estimation Techniques for Testing Goodness-of-Fit," *The Annals of Statistics* **20**, 2071-2086.

FAN, Y. (1994): "Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method," *Econometric Theory* **10**, 316-356.

FAN, Y. (1995): "Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test," *Econometric Reviews* **14**, 367-382.

FAN, Y., AND Q. LI (1996a): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica* **64**, 865-890.

FAN, Y., AND Q. LI (1996b): "Consistent Model Specification Tests: Kernel Based Tests Versus Bierens' ICM Tests," Unpublished manuscript, University of Windsor, Ontario.

FELLER, W. (1966): *An Introduction to Probability Theory and its Applications.* Volume 2. John Wiley: New York.

GHOSH, B.K., AND W. HUANG (1991): "The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness-of-Fit," *The Annals of Statistics* **19**, 999-1009.

GOURIEROUX, C., AND C. TENREIRO (1994): "A Comparison of Kernel Based Goodness of Fit Tests," CREST working paper.

GOZALO, P.L. (1993) "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, **9**, 451-477.

GOZALO, P.L. (1995): "Nonparametric Bootstrap Analysis with Applications to Engel Demand Curves," Working paper 95-7, Brown University. Forthcoming in *Journal of Econometrics.*

GOZALO, P.L., AND O.B. LINTON (1996): "Conditional independence restrictions: Testing and Estimation," *Cowles Foundation Discussion Paper* no. 1140.

GÖTZE, F. (1979): "Asymptotic Expansions for Bivariate von Mises Functionals," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 333-355.

GÖTZE, F. (1987): "Approximations for Multivariate U Statistics," *Journal of Multivariate Analysis* **22**, 212–229.

HÄRDLE, W. AND E. MAMMEN (1993) "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, **21**, 1926-1947.

HJELLVIK, V., AND D. TJØSTHEIM (1996): "Nonparametric Statistics for Testing of Linearity and Serial independence," *Journal of Nonparametric Statistics* **6**, 223-251.

HONG, Y., AND H. WHITE (1995): "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica* **63**, 1133-1159.

HONG, Y., (1996): "Consistent Specification Testing Using Optimal Nonparametric Kernel Estimation", Unpublished manuscript, Cornell University.

LEE, A.J., (1990): *U-Statistics: Theory and Practice.* Marcel Dekker.

LI, Q., AND S. WANG (1996): "A Simple Consistent Test for a Parametric Regression Function," Unpublished manuscript, University of Guelph, Ontario.

LINTON, O.B. (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica* **63**, 1079-1112.

LINTON, O.B. (1996): "Edgeworth Approximation for MINPIN Estimators in Semiparametric Regression Models," *Econometric Theory* **12**, 30-60.

PFANZAGL, J., (1980): "Asymptotic Expansions in Parametric Statistical Theory," in: *Developments in Statistics*, vol3. ed. P.R.Krishnaiah. Academic Press.

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

ROBINSON, P. M. (1988): "The Stochastic Difference between Econometric Statistics," *Econometrica*, 56, 531-548.

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

ROSENBLATT, M.A. (1975): "Quadratic Measure of Deviation of two-dimensional Density Estimates and a Test of Independence," *The Annals of Statistics* **3**, 1-14.

ROTHENBERG, T., (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics." Ch.14 in: *Handbook of Econometrics*, vol 2, ed. Z. Griliches and M.Intriligator. North Holland.

SARGAN, J. D. AND E. MIKHAIL (1971): "A General Approximation to the Distribution of Instrumental Variable Estimates," *Econometrica* **39**, 131–169.

SRINAVASAN, T.N. (1970): "Approximations to Finite Sample Moments of Estimators whose Exact Sampling Distributions are unknown," *Econometrica* **38**, 533-541.

STANISWALIS, J. G. AND T.A. SEVERINI (1991) "Diagnostics for Assessing Regression Models," *Journal of the American Statistical Association* **86**, 684-692.

Zheng, J.X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics* **75**, 263-289.

Table 1: Estimated Size (DGP1, $n = 100$)

	Second Corr. cri. values				Asymptotic critical values				First corrected			
h_n	1%	5%	10%	50%	1%	5%	10%	50%	1%	5%	10%	50%
0.5	0.012	0.059	0.121	0.469	0.007	0.027	0.045	0.284	0.008	0.043	0.104	0.469
1.0	0.009	0.052	0.105	0.488	0.003	0.016	0.031	0.231	0.005	0.032	0.076	0.488
1.5	0.008	0.047	0.099	0.491	0.002	0.011	0.018	0.183	0.002	0.021	0.058	0.491
2.0	0.006	0.043	0.100	0.499	0.001	0.008	0.014	0.145	0.002	0.017	0.047	0.499

Table 2: Estimated Size (DGP1, $n = 200$)

	Second Corr. cri. values				Asymptotic critical values				First corrected			
h_n	1%	5%	10%	50%	1%	5%	10%	50%	1%	5%	10%	50%
0.5	0.013	0.054	0.092	0.492	0.007	0.029	0.056	0.315	0.010	0.048	0.081	0.492
1.0	0.05	0.041	0.090	0.495	0.004	0.016	0.033	0.238	0.004	0.030	0.069	0.495
1.5	0.005	0.041	0.091	0.511	0.004	0.009	0.022	0.191	0.003	0.024	0.064	0.511
2.0	0.005	0.034	0.083	0.517	0.004	0.007	0.014	0.156	0.004	0.015	0.050	0.517

Table 3: Cumulants (DGP1, $n = 100$)

	Empirical				Random draw				Mean			
h_n	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
0.5	−0.446	0.899	0.473	0.138	−0.483	0.847	0.258	0.197	−0.415	0.851	0.387	0.275
sd					0.069	0.062	0.212	0.251				
1.0	−0.599	0.765	0.446	0.289	−0.664	0.717	0.539	0.436	−0.579	0.723	0.612	0.625
sd					0.067	0.077	0.155	0.340				
1.5	−0.715	0.645	0.421	0.373	−0.762	0.614	0.737	0.749	−0.693	0.607	0.774	0.973
sd					0.067	0.081	0.137	0.395				
2.0	−0.802	0.556	0.399	0.401	−0.828	0.523	0.870	1.042	−0.780	0.502	0.907	1.320
sd					0.066	0.081	0.129	0.437				

Table 4: Cumulants (DGP1, $n = 200$)

	Empirical				Random draw				Mean			
h_n	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
0.5	−0.392	0.882	0.439	0.309	−0.297	0.934	0.225	0.160	−0.380	0.892	0.401	0.263
sd					0.039	0.032	0.102	0.142				
1.0	−0.544	0.754	0.384	0.255	−0.428	0.846	0.559	0.591	−0.529	0.792	0.587	0.551
sd					0.042	0.040	0.080	0.177				
1.5	−0.644	0.658	0.369	0.260	−0.530	0.762	0.795	1.066	−0.635	0.697	0.727	0.841
sd					0.044	0.045	0.072	0.202				
2.0	−0.730	0.573	0.381	0.430	−0.618	0.679	0.945	1.464	−0.718	0.610	0.844	1.129
sd					0.045	0.048	0.070	0.230				

Table 5: Estimated Size (DGP2, $n = 100$)

	Second Corr. cri. values				Asymptotic critical values				First corrected			
h_n	1%	5%	10%	50%	1%	5%	10%	50%	1%	5%	10%	50%
						One-sided Test						
0.5	0.014	0.035	0.087	0.481	0.004	0.027	0.067	0.418	0.008	0.033	0.090	0.481
1.0	0.015	0.046	0.087	0.469	0.010	0.033	0.052	0.347	0.010	0.038	0.079	0.469
1.5	0.011	0.048	0.088	0.466	0.009	0.023	0.045	0.265	0.009	0.034	0.069	0.466
2.0	0.008	0.044	0.082	0.488	0.006	0.015	0.033	0.224	0.006	0.028	0.060	0.488
						Two-sided Test						
0.5	0.062	0.050	0.084	0.506	0.002	0.035	0.087	0.519	0.007	0.033	0.075	0.530
1.0	0.020	0.051	0.108	0.505	0.005	0.036	0.099	0.526	0.008	0.041	0.077	0.474
1.5	0.018	0.060	0.115	0.482	0.005	0.032	0.088	0.539	0.008	0.032	0.066	0.436
2.0	0.030	0.075	0.124	0.494	0.004	0.024	0.078	0.560	0.005	0.021	0.051	0.394

Table 6: Estimated Size (DGP2, $n = 200$)

h_n	Second Corr. cri. values				Asymptotic critical values				First corrected			
	1%	5%	10%	50%	1%	5%	10%	50%	1%	5%	10%	50%
					One-sided Test							
0.5	0.014	0.061	0.113	0.539	0.009	0.043	0.103	0.486	0.012	0.059	0.114	0.539
1.0	0.009	0.052	0.119	0.531	0.008	0.035	0.078	0.443	0.009	0.047	0.108	0.531
1.5	0.010	0.057	0.114	0.520	0.008	0.034	0.066	0.368	0.008	0.048	0.101	0.520
2.0	0.010	0.060	0.113	0.513	0.009	0.029	0.060	0.320	0.008	0.043	0.091	0.513
					Two-sided Test							
0.5	0.026	0.051	0.113	0.494	0.010	0.043	0.100	0.499	0.016	0.047	0.106	0.504
1.0	0.016	0.045	0.103	0.517	0.010	0.039	0.092	0.491	0.013	0.041	0.087	0.504
1.5	0.013	0.043	0.095	0.519	0.007	0.032	0.080	0.530	0.008	0.033	0.073	0.485
2.0	0.013	0.048	0.106	0.522	0.006	0.028	0.077	0.540	0.005	0.028	0.062	0.466

Table 7: Cumulants (DGP2, $n = 100$)

h_n	Empirical				Random draw				Mean			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
0.5	−0.177	0.900	0.097	−0.339	0.042	1.022	−0.536	0.004	−0.130	0.968	0.113	0.044
sd					0.145	0.024	0.697	0.116				
1.0	−0.316	0.873	0.390	0.220	−0.102	0.878	0.214	0.615	−0.257	0.897	0.354	0.269
sd					0.075	0.037	0.332	0.272				
1.5	−0.441	0.794	0.512	0.561	−0.309	0.798	0.690	1.276	−0.375	0.795	0.580	0.652
sd					0.060	0.048	0.230	0.411				
2.0	−0.538	0.689	0.493	0.697	−0.449	0.697	0.884	10663	−0.478	0.681	0.790	1.158
sd					0.059	0.056	0.199	0.537				

Table 8: Cumulants (DGP2, $n = 200$)

	Empirical				Random draw				Mean			
h_n	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
0.5	−0.045	0.989	0.029	−0.094	−0.011	0.984	−0.142	0.096	−0.101	0.985	0.127	0.031
sd					0.090	0.009	0.453	0.065				
1.0	−0.133	0.959	0.189	−0.017	−0.181	0.937	0.345	0.231	−0.201	0.943	0.313	0.172
sd					0.039	0.015	0.191	0.121				
1.5	−0.242	0.924	0.426	0.222	−0.256	0.879	0.543	0.597	−0.294	0.878	0.486	0.427
sd					0.032	0.022	0.128	0.205				
2.0	−0.342	0.871	0.552	0.488	−0.326	0.819	0.708	0.980	−0.381	0.800	0.651	0.769
sd					0.033	0.029	0.115	0.288				