Yale University
EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers
Cowles Foundation

11-1-1994

# Edgeworth Approximation for MINPIN Estimators in Semiparametric Regression Models

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. 1086



# EDGEWORTH APPROXIMATION FOR MINPIN ESTIMATORS IN SEMIPARAMETRIC REGRESSION MODELS

Oliver Linton

November 1994

# EDGEWORTH APPROXIMATION FOR MINPIN ESTIMATORS IN SEMIPARAMETRIC REGRESSION MODELS

By Oliver Linton
Department of Economics
Yale University, New Haven, CT 06520

We examine the higher order asymptotic properties of semiparametric regression estimators that were obtained by the general MINPIN method described in Andrews [1]. We derive an order $n^{-1}$ stochastic expansion and give a theorem justifying order $n^{-1}$ distributional approximation of the Edgeworth type.

This paper is based on Chapter II of my Ph.D. thesis. I would like to thank T. J. Rothenberg and P. J. Bickel for useful discussions and a co-editor for helpful comments. I would also like to thank Glena Ames for retyping this manuscript.

## 1. Introduction

Semiparametric methods allow one to obtain precise estimates of certain key quantities without fully specifying the probability distribution of the data, see for example Andrews [1,2], Bickel [5], Bickel et al. [7], Härdle and Stoker [17], Linton [24], Manski [26], Newey [31], Powell, Stock and Stoker [37], and Robinson [38,39,42,43]; for reviews, see Newey [32] and Robinson [40].

The notion of precise used by these authors rests on first order asymptotic theory for large sample sizes. There are two main problems with this theory. Firstly, the asymptotic distribution can provide a poor approximation to the actual distribution of the statistic considered. Practical experience suggests that estimator performance deteriorates with the number of nuisance parameters being estimated. Since semiparametric estimators take account of an infinite dimensional nuisance parameter, one must expect a potentially large small sample cost that is not reflected in the limiting distribution[1]. Secondly, a large number of choices have to be made in constructing the estimators about which this asymptotic theory provides little guidance. These choices can substantially affect the magnitude of estimators and test statistics.

We suggest using higher order approximations to overcome these problems. These methods have found considerable application in econometrics, see *inter alia*: Sargan [47], Phillips [35,36], and Rothenberg [44,45]. In other work, Linton [22,23], we have developed detailed "second order" moment approximations for various semiparametric estimators in regression models, including: the partially linear model considered in Robinson [39] and the heteroskedastic linear regression considered in Robinson [38]. These approximations quantify the small sample cost of the semiparametric estimation strategy in terms of a few interpretable quantities. This provides information about the choice of method, e.g. kernel versus nearest neighbor, and the choice of smoothing parameter.

We show that these approximations can be developed in a more general framework, and indeed that distributional approximations can also be validated. We examine the

*MINPIN* estimator $\widetilde{\tau}$ that minimizes the random criterion $\Psi_n(\tau, \widehat{G})$, where $\widehat{G}$ is a preliminary nonparametric estimate of a vector of regression functions $G$ and $n$ is sample size. The contrast $\Psi_n$ is chosen to ensure that $\widetilde{\tau}$ so defined is consistent: for example, $\Psi_n(\tau, G)$ could be the negative of the sample log likelihood function. This method of generating estimators of $\tau$ is considered in Andrews [1] and Bickel et al. [7], wherein conditions are given under which $\widetilde{\tau}$ is $\sqrt{n}$ consistent and asymptotically normal.

We make two contributions. Firstly, we derive an order $n^{-1}$ stochastic expansion for the standardized estimator $\widetilde{t} = \sqrt{n}(\widetilde{\tau} - \tau)$, i.e. we find random variables $\widetilde{t}^{**}$ and $R$ such that $\widetilde{t} = \widetilde{t}^{**} + R$, where the reminder term $R$ does not affect the distribution of $\widetilde{t}$ to order $n^{-1}$. The leading term $\widetilde{t}^{**}$ is $O_p(1)$ and has bounded moments to some suitably high order. In fact, $\widetilde{t}^{**}$ is a polynomial in a vector of weighted $U$-statistics of order up to $m \geq 2$, where the order $m$ that is needed is determined partly by the rate of convergence of $\widehat{G}$. A weighted U-statistic of order $m$ is a random variable of the form

$$\mathcal{U}_m = \sum_{i_1,..,i_m=1}^{n} \chi_{ni_1..i_m}\varphi(X_{i_1},..,X_{i_m}), \tag{1.1}$$

where $\{\chi_{ni_1..i_m}\}$ are deterministic weights satisfying certain order of magnitude conditions specified below, while $X_{i_1},..,X_{i_m}$ are independent random variables; see Lee [20] for a review of $U$-statistic material. Our second contribution is to justify Edgeworth type distributional approximations for the distribution of $\widetilde{t}^{**}$, and hence of $\widetilde{t}$. More precisely, we give conditions under which certain $\mathcal{U}_3$ possess an order $n^{-1}$ Edgeworth distributional approximation. A delta-method argument then ensures that $\widetilde{t}^{**}$ itself has an order $n^{-1}$ Edgeworth approximation. Our work on U-statistics extends Callaert, Janssen and Verarbereke [8] who deal only with unweighted U-statistics of second order.

For $\sqrt{n}$ consistency, we generally[2] need $\widehat{G}$ to be consistent at a rate better than $n^{1/4}$. This requirement is satisfied by a number of nonparametric estimators of regression functions or their derivatives under smoothness and dimensionality restrictions, see Müller [29] and Härdle [18]. We do not go into the details of the nonparametric estimation, and refer the reader to these references for more discussion.

We work throughout with a fixed design for convenience. This may often be justified by ancillarity considerations.

In Section 2 we define the sampling scheme we shall be analyzing, and give a number of examples. In Section 3 we develop the stochastic expansion and verify that the polynomial statistic $\widetilde{t}^{**}$ has the same distribution, to order $n^{-1}$, as $\widetilde{t}$. In Section 4 we prove that weighted $U$-statistics of order 3 have a valid order $n^{-1}$ Edgeworth approximation. All proofs are contained in Appendices I and II.

Notations. For any vector $x = (x_1, .., x_k)^T$, let $|x| = [\sum_{j=1}^k x_j^2]^{1/2}$ be its Euclidean norm, while for any real symmetric matrix $A$, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the smallest and largest eigenvalues respectively. We use $\phi(.)$ and $\Phi(.)$ for the standard normal density and cdf respectively. Finally, for a finite set $S$, let $\#S$ be its cardinality.

## 2. Assumptions and Examples

The observed data $\{X_i\}_{i=1}^n$ are partitioned as $X_i = (Y_i^T,\ Z_i^T)^T$, where the dependent variables $Y_i$ are $K_y \times 1$, and the regressors $Z_i$ are $K_z \times 1$. The data are independent across $i$, and are described by the following conditional Lebesgue density functions:

$$f_{Y_i|Z_i}(y;\ \tau,\ G(Z_i)),\ i = 1, 2,\ ...,\ n,$$

where $\tau$ is a $P \times 1$ vector of unknown parameters, while $G = (G^1, G^2,\ ...,\ G^L)^T$ is an $L \times 1$ vector of unknown regression functions. Our objective is to obtain estimates of $\tau$ given preliminary nonparametric estimates $\widehat{G}$ of $G$.

We consider an estimator of $\tau$ that minimizes the criterion $\Psi_n(\tau, \widehat{G})$ in the special case where $\Psi_n(\tau, G)$ is a sample average – which it would be under the independent sampling assumption. In fact, we actually work with solutions $\widetilde{\tau}$ of the associated first order conditions:

$$n^{-1} \sum_{i=1}^n \frac{\partial \Psi}{\partial \tau_\pi}(X_i; \widetilde{\tau}, \widehat{G}(Z_i)) = 0,\ \pi = 1,\ 2,\ ...,\ P. \tag{2.1}$$

Andrews [1] and Bickel et al. [7] both derive the first order limiting distribution of $\widetilde{\tau}$, in slightly different settings. Pfanzagl [33] analyzes the higher order properties of such estimators in a parametric context, i.e. where $G$ is known. We extend his theory to encompass the semiparametric case.

*Examples*

The following examples fit in to our framework and are of particular interest for econometric applications:

**Example 1.** *Homoskedastic Nonlinear Regression*

$$Y_i = g(Z_i) + U_i,\ i = 1, 2, ..., n,$$

where $U_i$ is a vector of iid stochastic errors with mean zero and covariance matrix $\Sigma$, while $Z_i$ are fixed regressors. The finite dimensional parameter of interest is $\Sigma$, while $g(.)$ is of unknown functional form. A scalar version of this estimation problem is analyzed in Carroll and Hall [10]. Estimates of $\Sigma$ are provided by any solution of

$$n^{-1}\sum_{i=1}^{n}\Xi[\{Y_i - \widehat{g}(Z_i)\}\{Y_i - \widehat{g}(Z_i)\}^T - \Sigma] = 0,$$

where the known $\Xi(.)$ is such that $E[\Xi(U_iU_i^T - \Sigma)] = 0$.

**Example 2.** *Partially Linear Regression*

$$y_{1i} = \beta^T Y_{2i} + \theta(Z_i) + \varepsilon_i;\ Y_{2i} = g_2(Z_i) + \eta_i,\ i = 1, 2, ..., n,$$

where $Z_i$ are fixed in repeated samples, while $\varepsilon_i$ and $\eta_i$ are mean zero, independent and mutually independent random errors. The finite dimensional parameters of primary interest are $\beta$, while $g_1(.)$, where $g_1(Z_i) \equiv E[y_{1i}]$, and the vector $g_2(.)$ are unknown regression functions. Robinson [39] and Chen [12] established the first order asymptotic theory for various estimators of $\beta$ under the alternative sampling scheme in which $(Y_i^T, Z_i^T)^T$ are iid. Suitable estimators of $\beta$ solve $n^{-1}\sum_{i=1}^{n}\Xi[\widehat{\eta}_i(\widehat{\nu}_i - \widehat{\eta}_i^T\beta)] = 0$, where $E[\Xi(\eta_i\varepsilon_i)] = 0$, while $\widehat{\eta}_i = Y_{2i} - \widehat{g}_2(Z_i)$ and $\widehat{\nu}_i = y_{1i} - \widehat{g}_1(Z_i)$, where $\widehat{g}_1$ and $\widehat{g}_2$ are preliminary nonparametric estimates of $g_1$ and $g_2$.

**Example 3.** *Heteroskedastic Linear Regression*

$$y_i = \beta^T Z_i + \varepsilon_i \sigma(Z_i),\ i = 1,\, 2,\, ...,\, n,$$

where $\varepsilon_i$ are iid mean zero and variance one, while $Z_i$ are fixed in repeated samples. The first order asymptotic theory for feasible GLS estimators of $\beta$ was established by Carroll [9], and subsequently advanced by Robinson [38] under the assumption that $(y_i,\ Z_i^T)^T$ are iid. We can consider $\widehat{\beta}$ that satisfies: $n^{-1}\sum_{i=1}^n \Xi(\widehat{\sigma}_i^{-2} Z_i(y_i - \widehat{\beta}^T Z_i)) = 0$, where $E[\Xi(\sigma_i^{-1} Z_i \varepsilon_i)] = 0$, while $\widehat{\sigma}_i^2 = \widehat{g}_2(Z_i) - \widehat{g}_1^2(Z_i)$ estimates $\sigma_i^2$, where $\widehat{g}_2(Z_i)$ and $\widehat{g}_1(Z_i)$ are nonparametric estimates of $E[y_i^2|Z_i]$ and $E[y_i|Z_i]$.

*Nonparametric Estimation*

We use nonparametric estimators that are linear in some dependent variable, i.e.

$$\widehat{G}^k(Z_i) = \sum_j w_{ij}^k \varphi^k(Y_j),\ i = 1, .., n\ ; k = 1, .., L, \tag{2.2}$$

for some functions $\varphi^k(.)$ such that $E[\varphi^k(Y_i)] = G^k(Z_i)$, where $\{w_{ij}^k\}$ are deterministic smoothing weights depending only on $Z_1,\ Z_2,\ ...,\ Z_n$. This includes most commonly employed estimators of regression functions and their derivatives, see Härdle and Linton [16]. It is convenient to rearrange (2.2) as

$$\widehat{\Gamma}_i^k \equiv \widehat{G}^k(Z_i) - G^k(Z_i) = V_{ki} + B_{ki},$$

where $V_{ki} = \sum_j w_{ij}^k U_j^k$ with $U_j^k = \varphi^k(Y_j) - G^k(Z_j)$, and $B_{ki} = \sum_j w_{ij}^k G^k(Z_j) - G^k(Z_i)$. We make the following assumptions:

**A1**

(i) *For each* $k$, *there exists* $b < \infty$ *and* $\alpha_k > 0$, *such that* $\max_{i \le n} |B_{ki}| < bn^{\alpha_k}$.

(ii) *For each* $k$, *there exists* $\overline{w}_k$ *and* $\mu_k > 0$ *such that* $\#\{(i, j)\ :\ w_{ij}^k \neq 0\} = O(n^{1+2\mu_k})$ *and* $n^{2\mu_k}|w_{ij}^k| < \overline{w}_k$.

(iii) *Let* $\zeta < \min[\min_k\{\mu_k\}, \min_k\{\alpha_k\}]$, *with* $\zeta > 1/4$.

**A2** *The sequence of* $L \times 1$ *vectors* $\{U_i\}_{i=1}^n$ *are independent and mean zero. Let* $\Sigma_i \equiv E[U_i U_i^T]$, *then* $\Sigma_i$ *is bounded away from zero and infinity. Furthermore,* $\sup_{i \geq 1} E[|U_i|^J] < \infty$ *for some* $J > 4$.

**Remark:** Conditions A1$(i) - (iii)$ are satisfied for kernel weights when $Z$ has bounded support and $G$ is sufficiently smooth. See for example Müller [29].

Let $\mathcal{G}$ be the space of allowable function valued nuisance parameters and define the norm $\| \bullet \|_n$ on $\mathcal{G}$ by

$$\|H\|_n = \max_{j \leq L} \max_{i \leq n} |H_j(Z_i)|.$$

We make use of the following lemma:

**Lemma 2.1:** *Let* $1/4 \leq \theta < \zeta$. *Assume that A1–A2 hold, and that* $J$, $\zeta$, *and* $\phi$ *satisfy* $J > (1 + 2\phi)/(\zeta - \theta)$. *Then, for some* $\delta > 0$,

$$\Pr[n^{\theta+\delta} \|\widehat{G} - G\|_n > c\log n] = o(n^{-2\phi}), \qquad \forall c > 0.$$

For example: when $\theta = 1/4$, $\phi = 1/2$ and $\zeta = 1/3$, we must have $J \geq 18$. Lemma 2.1 is proved in Appendix I. See Robinson [41] for similar results in a parametric context. The importance of this lemma is to establish when one can restrict attention to a small neighborhood of $G$.

*Parametric Estimation*

We now consider the properties that $\Psi$ must possess. We need some more notation. For two-dimensional arrays (matrices) $A = (a_{jk})$, we denote the $j, k$'th element of the inverse of $A$ by $a^{jk}$, when it is defined. We shall denote derivatives of $\Psi_n(\tau, G)$ with respect to the elements of $\tau$ and to elements of $G$ by Greek and Roman subscripts respectively. Thus, for each $\pi$, $\alpha$, and $k$, $\Psi_{n\pi\alpha;k} = n^{-1} \sum_{i=1}^n \Psi_{\pi\alpha i;k}$, where $\Psi_{\pi\alpha i;k}(\tau, G) = \frac{\partial^3 \Psi}{\partial \tau^\pi \partial \tau^\alpha \partial G^k}(X_i;\, \tau,\, G(Z_i))$ for each $i$. We use the affix $(\widehat{\ })$ to imply dependency on $\widehat{G}$, i.e. $\widehat{\Psi}_{n\pi\alpha} = n^{-1} \sum_{i=1}^n \widehat{\Psi}_{\pi\alpha i}$, where $\widehat{\Psi}_{\pi\alpha i} = \frac{\partial^2 \Psi}{\partial \tau^\pi \partial \tau^\alpha}(X_i;\, \tau,\, \widehat{G}(Z_i))$, and the affix $(\ -\ )$ to denote mean values, i.e. $\overline{\Psi}_{n\pi\alpha} = n^{-1}\sum_{i=1}^{n} E[\Psi_{\pi\alpha i}]$. Let $\omega = (\tau,\, G) \in \mathcal{T} \times \mathcal{G} \equiv \Omega$ be a typical parameter value, and let $\omega_i = (\tau^T,\, G(Z_i)^T)^T \in \mathcal{T} \times \mathbb{R}^L$, where $\mathcal{T}$ is an open subset of $\mathbb{R}^P$. Finally, we define a neighborhood $\mathcal{N}_\omega$ of $\omega$ using the norm $\|\bullet\|_n$, and the corresponding neighborhood of $\omega_i$ by $\mathcal{N}_{\omega_i}$.

We assume that the true parameter $\tau$ is in the interior of $\mathcal{T}$, and that

**B1** *For all $i$ and for some $p, m \geq 4$, all the partial derivatives $\Psi_{\pi_1\pi_2\ldots\pi_p i;k_1k_2\ldots k_m}(\omega^*)$ exist and are continuous functions of $\omega^*$ in a neighborhood $\mathcal{N}_{\omega_i}$ of the true parameter value.*

**B2** *There exists a finite integer $n_0$ such that $\inf_{n\geq n_0} \lambda_{\min}(\overline{\Psi}_{n\pi\alpha}) > 0$.*

**B3** *For every $\omega \in \Omega$ there exists a neighborhood $\mathcal{N}_\omega$ of $\omega$ and an $\varepsilon > 0$, such that for all $\pi_1$, ..., $\pi_p$, $k_1$, ..., $k_m$, with $p$, $m \geq 4$, we have for some $d \geq 4$:*

$$\sup_{n\geq 1} E_\omega[\Psi^{*\varepsilon}_{n\pi_1\pi_2\ldots\pi_p;k_1k_2\ldots k_m}]^d < \infty,$$

*where $\Psi^{*\varepsilon}_{n\pi_1\pi_2\ldots\pi_p;k_1k_2\ldots k_m} = \sup_{\omega^*\in\mathcal{N}_\omega}[n^{-1}\sum_{i=1}^{n} |\Psi_{\pi_1\pi_2\ldots\pi_p i;k_1k_2\ldots k_m}(\omega^*)|^{1+\varepsilon}]^{1/1+\varepsilon}$.*

**B4** $\forall \pi, k, i,\ \forall \omega \in \Omega$: (1) $E_\omega[\Psi_{\pi i}(\omega_i)] = 0$; (2) $E_\omega[\Psi_{\pi i;k}(\omega_i)] = 0$.

**B5** *For some $p$, $m \geq 4$, all the standardized arrays*

$$n^{-1/2}\sum_{i=1}^{n}[\Psi_{\pi_1\pi_2\ldots\pi_p i;k_1k_2\ldots k_m} - \overline{\Psi}_{\pi_1\pi_2\ldots\pi_p;k_1k_2\ldots k_m}]$$

*possess uniformly, in a neighborhood of the true $\tau$, valid order $n^{-1}$ Edgeworth approximations at the true $G$.*

**Remarks:**

1. Bhattacharya and Ghosh [3], Pfanzagl [33], and Skovgaard [48] give primitive conditions under which B5 is satisfied.

2. Assumption B4(1) is necessary for consistency of $\tilde{\tau}$, while Assumption B4(2) ensures that there is no "information loss" resulting from estimation of $G$. B4(2) is intuitively of the same form as Bickel's [5] orthogonality condition, but is expressed in

terms of partial derivatives of $\Psi$ with respect to the components of $G$, rather than in the language of functional derivatives of the likelihood as used in Bickel et al. [7]. When $\Psi$ is the likelihood function, this condition corresponds to the information matrix being block diagonal between $\tau$ and $G$. B4(2) is not necessary for the $\sqrt{n}$ consistency of the estimator, but the assumption simplifies some derivatives, and in practice one would choose a $\Psi$ to satisfy this condition.

3. The quantities $m$, $d$, $J$, and $\zeta$ must satisfy many additional restrictions which are not given here. Instead, we refer the reader to Appendix I where the conditions, we denote by the letter $C^*$, are stated and used in proving Theorems 3.1–3.3. These conditions can be traded off against each other; for example, allowing $m$ to be very large one can weaken the remaining conditions.

4. Condition B3 is a local regularity condition similar to the stochastic equicontinuity property used in Andrews [1] and Bhattacharya and Ghosh [3], assumption A2. In some cases it is quite easy to verify. Consider Example 3, and take the normal based likelihood function; in this case $\Psi_{\pi i} = \sigma_i^{-2} Z_{\pi i}(y_i - Z_i^T \beta)$. Provided $\sigma_i^2 > \underline{\sigma}^2 > 0$ and $d = J$, condition B3 follows.

## 3. Expansions

We adopt the tensor notation frequently used in the Edgeworth approximation literature, see McCullagh [25]. Apart from the index $i$, which will be reserved for observations, when an index is repeated in an expression, it is to be summed over. Unless otherwise stated, the ranges of the indices are as follows: subscript $i$ runs through 1, 2, ..., $n$, Greek subscripts run through 1, 2, ..., $P$, while Roman subscripts other than $i$ run through 1, 2, ..., $L$.

For any $\tau$ we can expand each equation in (2.1) about the true value $\tau_0$, obtaining for each $\pi$,

$$Q_{n\pi}(t) = \sqrt{n}\widehat{\Psi}_{n\pi}(\tau_0)+\widehat{\Psi}_{n\pi\alpha}(\tau_0)t^{\alpha}+\frac{1}{2}n^{-1/2}\widehat{\Psi}_{n\pi\alpha\gamma}(\tau_0)t^{\alpha}t^{\gamma}+\frac{1}{3!}n^{-1}\widehat{\Psi}_{n\pi\alpha\gamma\delta}(\tau_0)t^{\alpha}t^{\gamma}t^{\delta}+R_{Q\pi}, \tag{3.1}$$

where $t^{\alpha} = \sqrt{n}(\tau^{\alpha} - \tau_0^{\alpha})$, while $R_{Q\pi} = \frac{1}{4!}n^{-3/2}\widehat{\Psi}_{n\pi\alpha\gamma\delta\rho}(\tau^*)t^{\alpha}t^{\gamma}t^{\delta}t^{\rho}$, with $|\tau_{\alpha}^* - \tau_{0\alpha}| \le |\tau_{\alpha} - \tau_{0\alpha}|$ for each $\alpha$. The solutions of $Q_n(t) = 0$ depend on: $\widehat{\Psi}_{n\pi}$, $\widehat{\Psi}_{n\pi\alpha}$, $\widehat{\Psi}_{n\pi\alpha\gamma}$, and $\widehat{\Psi}_{n\pi\alpha\gamma\delta}$, all of which depend on $\widehat{G}$ in a potentially non-linear fashion. Therefore, we expand each of these hatted quantities in a Taylor series about $G$. The non-leading terms, $\widehat{\Psi}_{n\pi\alpha}$, $\widehat{\Psi}_{n\pi\alpha\gamma}$, and $\widehat{\Psi}_{n\pi\alpha\gamma\delta}$, all expand in a similar fashion: for example,

$$\widehat{\Psi}_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha} = \widehat{\Psi}^{\#m}_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha} + \widehat{R}^{\#m}_{\pi\alpha},$$

in which the truncated array is

$$\widehat{\Psi}^{\#m}_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha} = \Psi_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha} + n^{-1}\sum_{i=1}^{n}\Psi_{\pi\alpha i;k}\widehat{\Gamma}_i^k + .. + n^{-1}\sum_{i=1}^{n}\frac{1}{m!}\Psi_{\pi\alpha i;k_1k_2\dots k_m}\widehat{\Gamma}_i^{k_1}\ \dots\ \widehat{\Gamma}_i^{k_m}, \tag{3.2}$$

while the remainder is

$$\widehat{R}^{\#m}_{\pi\alpha} = n^{-1}\sum_{i=1}^{n}\frac{1}{(m+1)!}\Psi_{\pi\alpha i;k_1k_2\dots k_{m+1}}(X_i;\ \tau,\ G^*(Z_i))\widehat{\Gamma}_i^{k_1}\ \dots\ \widehat{\Gamma}_i^{k_{m+1}}. \tag{3.3}$$

Here, for each $i$, $G^*(Z_i)$ is a vector of intermediate values such that $|G^{*k}(Z_i)-G^k(Z_i)| \le |\widehat{G}^k(Z_i)-G^k(Z_i)|$ for each $k$. Each term in (3.2) can be rewritten as a sum of weighted U-statistics of orders up to $j$. For example, $n^{-1}\sum_{i=1}^n \Psi_{\pi\alpha i;kl}\widehat{\Gamma}_i^k\widehat{\Gamma}_i^l =$

$$\begin{aligned}
&n^{-1}\textstyle\sum_{i=1}^n E(\Psi_{\pi\alpha i;kl})B_{li}B_{ki} + n^{-1}\sum_{i=1}^n E(\Psi_{\pi\alpha i;kl})\{\sum_{j=1}^n w_{ij}^k w_{ij}^l E[U_j^k U_j^l]\}+\\
&n^{-1}\textstyle\sum_{i=1}^n B_{li}B_{ki}[\Psi_{\pi\alpha i;kl}-E(\Psi_{\pi\alpha i;kl})]+n^{-1}\sum_{j=1}^n\{\sum_{i=1}^n E(\Psi_{\pi\alpha i;kl})w_{ij}^k w_{ij}^l\}[U_j^kU_j^l-E(U_j^kU_j^l)]+\\
&2n^{-1}\textstyle\sum_{j=1}^n\{\sum_{i=1}^n E(\Psi_{\pi\alpha i;kl})B_{li}w_{ij}^k\}U_j^k+\\
&n^{-1}\textstyle\sum_{j_1,j_2=1}^n\{\sum_{i=1}^n E(\Psi_{\pi\alpha i;kl})w_{ij_1}^k w_{ij_2}^l\}U_{j_1}^kU_{j_2}^l+n^{-1}\sum_{j,i=1}^n w_{ij}^k w_{ij}^l[\Psi_{\pi\alpha i;kl}-E(\Psi_{\pi\alpha i;kl})]U_j^kU_j^l+\\
&n^{-1}\textstyle\sum_{j_1,j_2,i=1}^n w_{ij_1}^k w_{ij_2}^l[\Psi_{\pi\alpha i;kl} - E(\Psi_{\pi\alpha i;kl})]U_{j_1}^kU_{j_2}^l,
\end{aligned}$$

where the first line contains only deterministic constants, the second line has only single sums of independent random variables, while the third and fourth lines are weighted $U$-statistics of orders 2 and 3 respectively. This structure is exploited in Section 4 below. By direct calculation, $\Psi_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha} = O_p(n^{-1/2})$, while

$$n^{-1}\sum_{i=1}^{n}\frac{1}{j!}\Psi_{\pi\alpha i;k_1k_2...k_j}\widehat{\Gamma}_i^{k_1}\ ...\ \widehat{\Gamma}_i^{k_j} \leq\ O_p(n^{-j\zeta}),\ j=2,..,m,$$

using Lemma 2.1 and the Cauchy–Schwarz inequality.

The leading statistics $n^{1/2}\widehat{\Psi}_{n\pi}$, $\pi = 1,..,P$, also expand in a Taylor series, but with terms of different magnitude. Firstly, $n^{1/2}\Psi_{n\pi} = O_p(1)$, while in Appendix I we show that $n^{-1/2}\sum_{i=1}^{n}\Psi_{\pi i;k}\widehat{\Gamma}_i^k = O_p(n^{-\zeta})$ (this holds because $\Psi_{\pi i;k}$ are zero mean random variables). When $E(\Psi_{\pi i;k_1k_2...k_j}) = 0$, which happens in many adaptive situations, $n^{-1/2}\sum_{i=1}^{n}\frac{1}{j!}\Psi_{\pi i;k_1k_2...k_j}\widehat{\Gamma}_i^{k_i}\ ...\ \widehat{\Gamma}_i^{k_j} = O_p(n^{-j\zeta})$, $j = 2,..,m$. However, in general, $E(\Psi_{\pi i;k_1k_2...k_j}) \neq 0$, and we can only establish, using the crude method of Cauchy-Schwarz bounding and Lemma 2.1, that $n^{-1/2}\sum_{i=1}^{n}\frac{1}{j!}\Psi_{\pi i;k_1k_2...k_j}\widehat{\Gamma}_i^{k_i}\ ...\ \widehat{\Gamma}_i^{k_j} \leq O_p(n^{1/2}n^{-j\zeta})$, $j = 2,..,m$.

We draw a number of conclusions from the above analysis. Firstly, at this level of generality a necessary condition for $\widetilde{\tau}$ to be $\sqrt{n}$ consistent is that $\widehat{G}$ converges to $G$ at a rate faster than $n^{1/4}$, because $n^{-1/2}\sum_{i=1}^{n}\Psi_{\pi i;kl}\widehat{\Gamma}_i^k\widehat{\Gamma}_i^l$ may not be $o_p(1)$ otherwise. Secondly, even if this condition is satisfied, to approximate the distribution of $\widetilde{\tau}$ to order $n^{-1}$ we may need to include many terms from the expansion of $\sqrt{n}\widehat{\Psi}_{n\pi}(\tau,\widehat{G})$. The number of terms which must be included depends on $\zeta$. Since the $m$'th term in the expansion of $n^{1/2}\widehat{\Psi}_{n\pi}(\tau,\widehat{G})$ is $O_p(n^{1/2}n^{-m\zeta})$, we should include $m$ terms, where $m+1 > 3/2\zeta$. When $\zeta > 3/8$, we only need include three terms from both the expansion around $\tau$ and the expansion about $G$ in order to get $o_p(n^{-1})$ remainder terms.

Define the following $O_p(1)$ arrays

$$\widehat{S}_{n\alpha} = n^{1/2}\widehat{\Psi}_{n\alpha};\ \widehat{S}_{n\pi\alpha} = n^{\zeta}(\widehat{\Psi}_{n\pi\alpha} - \overline{\Psi}_{n\pi\alpha});\ \widehat{S}_{n\pi\alpha\gamma} = n^{\zeta}(\widehat{\Psi}_{n\pi\alpha\gamma} - \overline{\Psi}_{n\pi\alpha\gamma}),$$

and the corresponding truncated statistics: $\widehat{S}^{\#}_{n\alpha}$, $\widehat{S}^{\#}_{n\pi\alpha}$ and $\widehat{S}^{\#}_{n\pi\alpha\gamma}$, where we drop the $m$ superscript for convenience[3]. We shall refer to the collection of all standardized arrays of the above form, up to fourth order partials, as $\widehat{S}$ and $\widehat{S}^{\#}$ respectively.

We first establish the existence of a consistent root. The MINPIN estimator $\widetilde{\tau}$ is defined to be any solution of (2.1) if one exists, and zero otherwise.

**Theorem 3.1:** *Assume that conditions A1–A2, B1–B5 and C* hold. Then, with probability* $1 - o(n^{-1})$*, the random variable* $\widetilde{\tau}$ *solves the quasi-likelihood equations (2.1) in the set* $\{\tau : |\tau - \tau_0| \leq \frac{c\log}{\sqrt{n}}\}$ *for some* $c > 0$.

We now replace the arrays $\widehat{S}$ by their truncations $\widehat{S}^{\#}$. Let $\widetilde{\tau}^* = \sqrt{n}(\widetilde{\tau}^* - \tau_0)$ solve the truncated equations $Q_n^*(t) = 0$, where

$$Q_{n\pi}^*(t) = \widehat{S}_{n\pi}^{\#} + \overline{\Psi}_{n\pi\alpha}t^\alpha + n^{-\zeta}\widehat{S}_{n\pi\alpha}^{\#}t^\alpha + \frac{1}{2}n^{-1/2}\overline{\Psi}_{n\pi\alpha\gamma}t^\alpha t^\gamma + \frac{1}{2}n^{-(1/2+\zeta)}\widehat{S}_{n\pi\alpha\gamma}^{\#}t^\alpha t^\gamma + \frac{1}{6}n^{-1}\overline{\Psi}_{n\pi\alpha\gamma\delta}t^\alpha t^\gamma t^\delta.$$

The following theorem establishes that $\widetilde{t}$ can be approximated in distribution by $\widetilde{t}^*$:

**Theorem 3.2:** *Assume that conditions A1–A2, B1–B5 and C* hold. Then, there exists a positive constant c such that*

$$\Pr\left[|\widetilde{t}^* - \widetilde{t}| > \frac{c}{n\log n}\right] = o(n^{-1}).$$

The truncated equations $Q_n^*$ are polynomials in the vector $\widetilde{t}^*$, and, by the implicit function theorem, $\widetilde{t}^*$ is a smooth function of the elements of $\widehat{S}^{\#}$. We can invert this function to write $\widetilde{t}_\pi^*$ as a power series in the random variables $\widehat{S}^{\#}$. The power series expansion is truncated at a suitable point, and the truncated expansion, denoted $\widetilde{t}^{**}$, is then used as an approximation to $\widetilde{t}$.

Skovgaard [48] shows how to find this polynomial approximation when $G$ is known and the estimator is the MLE. The same techniques can be used here. Let

$$\widehat{Y}_\pi = -\overline{\Psi}_n^{\pi\alpha}\widehat{S}_{n\alpha}^{\#};\ \widehat{Y}_{\pi\gamma} = -\overline{\Psi}_n^{\pi\alpha}\widehat{S}_{n\alpha\gamma}^{\#};\ \widehat{Y}_{\pi\gamma\delta} = -\overline{\Psi}^{\pi\alpha}\widehat{S}_{n\alpha\gamma\delta}^{\#};$$
$$M^\pi = \overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\alpha};\ M^{\pi\gamma} = \overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\alpha\gamma};\ M^{\pi\gamma\delta} = \overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\alpha\gamma\delta};\ M^{\pi\gamma\delta\rho} = \overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\alpha\gamma\delta\rho},$$

and similarly define the infeasible arrays: $Y_\pi$, $Y_{\pi\alpha}$ and $Y_{\pi\alpha\gamma}$. In general, $\widetilde{t}^{**}$ involves polynomials up to the order $[\zeta^{-1}]$, where $[\bullet]$ denotes least dominating integer, in the above arrays. Therefore, provided $\zeta > 1/3$, the polynomial approximation to $\widetilde{t}^*$ is

$$\widetilde{t}^{**} = \widehat{Y}_\pi + A(\widehat{Y}_{\pi\alpha}, \widehat{Y}_\alpha) + B(\widehat{Y}_{\pi\alpha\gamma}, \widehat{Y}_{\pi\alpha}, \widehat{Y}_\alpha), \tag{3.4}$$

where $A$ and $B$ are homogeneous polynomials in the arrays $\widehat{Y}_\pi$, $\widehat{Y}_{\pi\alpha}$, and $\widehat{Y}_{\pi\alpha\gamma}$, given by

$$\begin{aligned} A &= n^{-\zeta}\widehat{Y}_{\pi\alpha}\widehat{Y}_\alpha - \tfrac{1}{2}n^{-1/2}M^{\pi\alpha\gamma}\widehat{Y}_\alpha\widehat{Y}_\gamma; \\ B &= n^{-2\zeta}\widehat{Y}_{\pi\alpha}\widehat{Y}_{\alpha\gamma}\widehat{Y}_\gamma - \tfrac{1}{2}n^{-(\zeta+1/2)}\widehat{Y}_{\pi\alpha}M^{\pi\gamma\delta}\widehat{Y}_\gamma\widehat{Y}_\delta - n^{-(\zeta+1/2)}M^{\pi\alpha\gamma}\widehat{Y}_\alpha\widehat{Y}_{\gamma\delta}\widehat{Y}_\delta + \tfrac{1}{2}n^{-(\zeta+1/2)}\widehat{Y}_{\pi\alpha\gamma}\widehat{Y}_\alpha\widehat{Y}_\gamma \\ &\quad - \tfrac{1}{6}n^{-1}M^{\pi\alpha\gamma\delta}\widehat{Y}_\alpha\widehat{Y}_\gamma\widehat{Y}_\delta + \tfrac{1}{2}n^{-1}M^{\pi\alpha\rho}\widehat{Y}_\alpha M^{\gamma\delta\rho}\widehat{Y}_\delta\widehat{Y}_\rho. \end{aligned}$$

We now prove that the distribution of $\widetilde{t}$ can be approximated by the distribution of $\widetilde{t}^{**}$ with error of order $n^{-1}$:

**Theorem 3.3:** *Assume that A1–A2, B1–B5 and C* hold. Then for some $c > 0$, we have*

$$\Pr\left[|\widetilde{t}^{**} - \widetilde{t}^*| > \frac{c}{n\log n}\right] = o(n^{-1}).$$

In a number of leading examples, the quasi-likelihood equations (2.1) can be solved to define $\widetilde{\tau}$ explicitly; for example, when $\Psi_{n\pi\alpha\gamma} = 0$. In this case, $\widetilde{t}$ is a ratio $\widetilde{t}_\pi = \widehat{\Psi}_n^{\pi\alpha}\sqrt{n}\widehat{\Psi}_{n\alpha}$, and the polynomial approximation to $\widetilde{t}$ is of the simpler form

$$\widetilde{t}^{**} = \widehat{Y}_\pi + n^{-\zeta}\widehat{Y}_{\pi\alpha}\widehat{Y}_\alpha + n^{-2\zeta}\widehat{Y}_{\pi\alpha}\widehat{Y}_{\pi\gamma}\widehat{Y}_\alpha. \tag{3.5}$$

We now investigate further the examples of Section 2, in the special cases where the simpler expansion (3.5) suffices.

*Homoskedastic Nonlinear Regression*

Consider the scalar version of model 1 in which the errors $u_i$ are iid with zero mean and variance $\sigma^2$. We consider the standard estimator of $\sigma^2$, i.e. $\widetilde{\sigma}^2 = n^{-1}\sum_{i=1}^n [y_i - \widehat{g}(Z_i)]^2$, where $\widehat{g}(Z_i) = \sum_{i=1}^n w_{ij} y_j$. By straightforward manipulation

$$\widetilde{t} \equiv \sqrt{n}(\widetilde{\sigma}^2 - \sigma^2) = n^{-1/2}\sum_{i=1}^n \delta_i [u_i^2 - \sigma^2] + \sum\sum_{i \neq j} \rho_{ij} u_i u_j + n^{-1/2}\sum_{i=1}^n \xi_i u_i + b, \tag{3.6}$$

where: $\rho_{ij} = n^{-1/2}[\sum_k w_{kj} w_{ki} - 2w_{ij}]$, $\xi_i = 2\{\sum_k w_{ki} B_i - B_i\}$, $\delta_i = 1 + \sum_k w_{ki}^2$, and $b = n^{-1/2}\sum_i B_i^2 + n^{-1/2}\sum\sum_{i\neq j} w_{ij}^2 - 2\sigma^2 n^{-1/2} w_{ii}$, with $B_i = \sum_j w_{ij} g(Z_j) - g(Z_i)$. From A1(ii): $n^{-1/2}\sum_{i=1}^n \delta_i [u_i^2 - \sigma^2] = O_p(1)$, $b = O(n^{1/2-2\zeta})$, $\sum\sum_{i\neq j} \rho_{ij} u_i u_j = O_p(n^{-\zeta})$, while $n^{-1/2}\sum_{i=1}^n \xi_i u_i = O_p(n^{-\zeta})$. In this example, it is not necessary to employ a truncation argument as contained in Theorems 3.1–3.3, since (3.6) is exact.

*Partially Linear Regression*

Consider the scalar version of model 2 in which both errors $\varepsilon_i$ and $\eta_i$ are iid with zero mean and variance $\sigma_\varepsilon^2$ and $\sigma_\eta^2$ respectively. The standardized Robinson estimator of $\beta$ is

$$\widetilde{t} \equiv \sqrt{n}(\widetilde{\beta} - \beta) = \left[n^{-1}\sum_{i=1}^n \widetilde{\eta}_i^2\right]^{-1}\left[n^{-1/2}\sum_{i=1}^n \widetilde{\eta}_i \widetilde{\varepsilon}_i\right],$$

where $\widetilde{\eta}_i = \eta_i - (B_{\eta i} + V_{\eta i})$ and $\widetilde{\varepsilon}_i = \varepsilon_i - (B_{\varepsilon i} + V_{\varepsilon i})$ are nonparametric residuals, with $B_{\eta i} = \sum_j w_{ij} g_2(Z_j) - g_2(Z_i)$ and $B_{\varepsilon i} = \sum_j w_{ij}\theta(Z_j) - \theta(Z_i)$ deterministic, while $V_{\eta i} = \Sigma_j w_{ij}\eta_j$ and $V_{\varepsilon i} = \sum_j w_{ij}\varepsilon_j$ are both zero mean weighted sums of iids. See Linton [22] for further discussion. The standardized estimator is in closed form and the expansions leading up to Theorem 3.1 and Theorem 3.2 are unnecessary, although Theorem 3.3 is required. Let $\mathcal{X} = n^{-1/2}\sum_{i=1}^n \widetilde{\eta}_i \widetilde{\varepsilon}_i$ and $\mathcal{Y} = n^{-1/2}\sum_{i=1}^n [\widetilde{\eta}_i^2 - \sigma_\eta^2]$, then

$$\widetilde{t}^{**} = \sigma_\eta^{-2}\mathcal{X} - \sigma_\eta^{-4}\frac{\mathcal{X}\mathcal{Y}}{\sqrt{n}} + \sigma_\eta^{-6}\frac{\mathcal{X}\mathcal{Y}^2}{n}, \tag{3.7}$$

where

$\mathcal{X} = n^{-1/2}\sum_{i=1}^n \eta_i\varepsilon_i + n^{-1/2}\sum_{i=1}^n \xi_{1i}\varepsilon_i\eta_j + n^{-1/2}\sum_{i=1}^n \xi_{2i}\eta_i + \sum\sum_{i\neq j} \rho_{ij}\varepsilon_i\eta_j + n^{-1/2}\sum_{i=1}^n B_{\eta i}B_{\varepsilon i}$,
$\mathcal{Y} = n^{-1/2}\sum_{i=1}^n(\eta_i^2 - \sigma_\eta^2) + n^{-1/2}\sum_{i=1}^n \xi_{2i}\eta_i + n^{-1/2}\sum_{i=1}^n \rho_{ij}\eta_i\eta_j + n^{-1/2}\{\sum_{i=1}^n B_{\eta i}^2 + \sigma_\eta^i \sum\sum_{i\neq j}$
$w_{ij}^2 - 2\sigma_\eta^2 w_{ii}\}$,

with: $\rho_{ij} = n^{-1/2}\{\sum_k w_{kj}w_{ki} - 2w_{ij}\}$, $\xi_{1i} = n^{-1/2}\{\sum_k w_{ki}w_{\eta k} - B_{\eta i}\}$, and $\xi_{2i} = n^{-1/2}\{\sum_k w_{ki}B_{\varepsilon k} - B_{\varepsilon i}\}$. Thus, the truncation (3.7) is a polynomial in weighted U-statistics of order 2.

## 4. Distributional Approximation

We have just shown that $\widetilde{t}^{**}$ has the same distribution as $\widetilde{t}$ to order $n^{-1}$. Here we derive and justify approximations to the distribution of $\widetilde{t}^{**}$.

The first order behavior of $\widetilde{t}^{**}$ is determined by dropping everything in (3.4) but the leading term: provided B2 holds, $\widetilde{t}_\pi = -\overline{\Psi}_n^{\pi\alpha} S_{n\alpha} + o_p(1)$, and for any $P$-vector $c$,

$$\Pr[(c_\alpha c_\beta \Omega_{\alpha\beta})^{-1/2} c_\pi \widetilde{t}_\pi \leq x] - \Phi(x) = o(1),$$

where $\Omega_{\alpha\beta}(\tau_0) = \lim_{n\to\infty}[\overline{\Psi}_n^{\alpha\gamma}\frac{1}{n}\sum_{i=1}^n E[\Psi_{\gamma i}\Psi_{\vartheta i}]\overline{\Psi}_n^{\vartheta\beta}]$. This result is proved in Andrews [1] under weaker conditions than ours.

We now turn to the higher order properties of $\widetilde{t}^{**}$. To approximate the distribution of $\widetilde{t}^{**}$ to order $n^{-1}$ we use the well established Edgeworth method, see Rothenberg [45] for a review. In parametric settings, the first four (asymptotic) cumulants of a statistic are sufficient to determine the order $n^{-1}$ approximation to its distribution the form of which is given by an Edgeworth measure which depends on these cumulants, see (4.3) below. We calculate order $n^{-1}$ approximations to the cumulants of $\widetilde{t}^{**}$ and substitute these into the Edgeworth measure. Getting the cumulant approximations involves straightforward but tedious computation which is carried out in Linton [22] for the Robinson estimator in the partially linear model, and in Linton [23] for the semiparametric GLS estimator in the heteroskedastic linear regression model. In fact, in these papers only the mean and variance of the standardized estimator were calculated, and only to order $n^{-2\mu}$, for some $\mu < 1/2$. The resulting approximations we described as "second order." Our purpose here is to justify an order $n^{-1}$ distributional approximation based on order $n^{-1}$ approximations to all the relevant cumulants. In fact, we work with the special case for which the first four cumulants suffice. See the examples below for an explanation of the plausibility of these conditions.

The truncated statistic $\widetilde{t}^{**}$ is a polynomial in the array $\widehat{S}^{\#}$. In our special case, $\widehat{S}^{\#}_{n\pi}$, $\widehat{S}^{\#}_{n\pi\alpha}$ and $\widehat{S}^{\#}_{n\pi\alpha\gamma}$ are vectors of zero mean weighted $U$-statistics of orders up to 3. Apart from a deterministic bias term that can be handled by analytic methods, these can be written in the form

$$\mathbf{Q} = n^{-1/2}\sum_{i=1}^{n} g(X_i) + n^{-\zeta}\sum\sum_{j>i} \rho_{ij}\varphi(X_i,\ X_j) + n^{-\xi}\sum\sum\sum_{k>j>i} \pi_{ijk}\nu(X_i,\ X_j,\ X_k), \quad (4.1)$$

where $\xi > \zeta$. Both $\varphi$ and $\nu$ are permutation invariant, i.e. $\varphi(x,y) = \varphi(y,x)$ and $\nu(x,y,z) = \nu(z,x,y)$, with $E[\varphi(X_i,X_j)]|\mathcal{F}] = E[\nu(X_i,X_j,X_k)|\mathcal{F}] = 0$, where $\mathcal{F}$ denotes any proper subset of either $\{X_i,X_j\}$ or $\{X_i,X_j,X_k\}$, while $E[g(X_i)] = 0$, $i,j,k = 1,2,..,n$. Thus, these three separate pieces are zero mean and mutually orthogonal by construction. Here, $\{\rho_{ij}\}$ and $\{\pi_{ijk}\}$ are sequences of non-random weights depending on $\{w^k_{ij}\}$ in such a way that $\sum\sum_{j>i}\rho_{ij}\varphi(X_i,\ X_j)$ and $\sum\sum\sum_{k>j>i}\pi_{ijk}\nu(X_i,\ X_j,\ X_k)$ are both $O_p(1)$. In this representation, the double and triple sums are known as degenerate weighted $U$-statistics of orders two and three respectively.

Our argument is based on establishing the following:

*(EDGE1) The vector* $\mathbf{Q} = (\mathbf{Q}_1,\ \mathbf{Q}_2,\ ...,\ \mathbf{Q}_r)^T$ *possesses an order* $n^{-1}$ *Edgeworth approximation.*

*(EDGE2) If* $\mathbf{Q}$ *possess an order* $n^{-1}$ *Edgeworth approximation, then* $\mathbf{P}(\mathbf{Q})$ *does too, where* $\mathbf{P}(\cdot)$ *is a polynomial with bounded coefficients.*

Essentially this strategy pursued in Bhattacharya and Ghosh [3]. However, they deal only with $U$-statistics of order 1, i.e. single sums, for which (*EDGE1*) has been established under weak conditions. They also establish (*EDGE2*). Their argument justifying

(*EDGE2*) can be employed here, although a rigorous proof of this would require considerably more work.

We prove (*EDGE1*) for our more general class of statistics. Random variables of the type (4.1) are related to standard $U$-statistics whose properties are well established. In particular, Bickel, Götze, and van Zwet [6] and Callaert, Janssen and Verarbereke [8] give conditions under which the formal Edgeworth approximation is valid for standard $U$-statistics of order 2. Standard $U$-statistics have correction terms that are $O_p(n^{-1/2})$, while our weighted $U$-statistics have correction terms that are $O_p(n^{-\zeta})$, where in general $\zeta < 1/2$ and hence the correction term in (4.1) is of larger order. Therefore, we must extend this previous work.

Below we establish that the formal order $n^{-1}$ Edgeworth approximation to the distribution function of $\mathbf{Q}$ indeed has error of order $n^{-1}$. We prove our theorem for scalar valued iid random variables. The extension to non-identically distributed and multivariate statistics is also valid under the high level assumption D5. The proof of this, however, requires somewhat more notation than we already have, and we leave this for future work[4].

Let $\psi_n(s) = E[e^{\mathbf{i}s\mathbf{Q}}]$ and $\eta(s) = E[e^{\mathbf{i}sg(X_1)}]$ where $\mathbf{i} = \sqrt{-1}$, and assume

**D1** $\#\{(i,j) : \rho_{ij} \neq 0\} = O(n^{1+2\zeta})$; $\#\{(i,j,k) : \pi_{ijk} \neq 0\} = O(n^{1+2\xi})$. *There is a finite constant* $\delta$ *such that* $n^{1/2+\zeta}|\rho_{ij}| < \delta$ *and* $n^{1/2+\xi}|\pi_{ijk}| < \delta$. *We require that* $\zeta > 3/8$ *and* $\xi > 5/8$.

**D2** $E[|g(X_1)|^6] < \infty$; $E[\varphi(X_1, X_2)|^6] > \infty$; $E[|\nu(X_1,\, X_2,\, X_3)|^6]$.

**D3** $\limsup_{|s|\to\infty} |\eta(s)| < 1$.

**D4** $\int_{n^{1/4+\zeta}/\log n}^{n \log n} \left|\frac{\psi_n(s)}{s}\right| ds = o(n^{-1})$.

**Remarks.**

1. The order of magnitude assumption about $\zeta$ and $\xi$ is for convenience. Undoubtedly, approximations can be developed and justified for whatever values of these constants are appropriate. We focus on this special case to make the analysis tractable.

It can occur in all of the examples we considered. A consequence of D1 is that the following quantities are $O(1)$:

$$\mathcal{W}_1 = n^{-(\zeta+1/2)} \sum\sum_{j>i} \rho_{ij},\ \mathcal{W}_2 = \sum\sum_{j>i} \rho_{ij}^2,\ \mathcal{W}_3 = n^{-2\zeta} \sum\sum\sum_{k>j>i} \rho_{ij}\rho_{jk},$$
$$\mathcal{W}_4 = n^{-(\xi+1/2)} \sum\sum\sum_{k>j>i} \pi_{ijk},\ \mathcal{W}_5 = n^{-(1+2\zeta)} \sum\sum\sum_{l>k>j>i} \rho_{ij}\rho_{kl}.$$

2. The moment assumptions D2 are slightly stronger than those in Callaert, Janssen and Verarbereke [8].

We prove the theorem for the standardized (to have mean zero and variance one) statistic $\overline{\mathbf{Q}}$, where $\overline{\mathbf{Q}} = \sigma_{\mathbf{Q}}^{-1}\mathbf{Q}$, with

$$\sigma_{\mathbf{Q}}^2 \equiv E[g(X_1)^2] + n^{-2\zeta}E[\varphi(X_1,X_2)^2]\mathcal{W}_2 \ = \mathrm{var}[\mathbf{Q}]+o(n^{-1}).$$

When $\zeta > 3/8$ and $\xi > 5/8$, the third and fourth cumulants of $\overline{\mathbf{Q}}$ are

$$\kappa_3 = \kappa_{3g} + n^{-1/2}6\sigma_g^{-3}E[g(X_1)g(X_2)\varphi(X_1,X_2)]\mathcal{W}_1 + o(n^{-1}),$$
$$\kappa_4 = \kappa_{4g}+n^{-1}\sigma_g^{-4}\{24E[g(X_1)g(X_2)\varphi(X_1,X_2)]\mathcal{W}_1+24E[g(X_1)g(X_3)\varphi(X_1,X_2)\varphi(X_2,X_3)]\mathcal{W}_3$$
$$+6E[g(X_1)g(X_2)g(X_3)\nu(X_1,X_2,X_3)]\mathcal{W}_4\} + o(n^{-1}), \tag{4.2}$$

where $\sigma_g^2 = E[g(X_1)^2]$, $\kappa_{3g} = n^{-1/2}\sigma_g^{-3}E[g(X_1)^3]$ and $\kappa_{4g} = n^{-1}\{\sigma_g^{-4}E[g(X_1)^4]-3\}$. Under D1–D2: $\kappa_3 = \overline{\kappa}_3 n^{-1/2}$ and $\kappa_4 = \overline{\kappa}_4 n^{-1}$, where $\overline{\kappa}_3$ and $\overline{\kappa}_4$ are both $O(1)$. Let also $\widetilde{F}_{n2}(x)$ be the standard Edgeworth measure

$$\widetilde{F}_{n2}(x) = \Phi(x) - \phi(x)\left[\frac{\overline{\kappa}_3}{6\sqrt{n}}(x^2-1) + \frac{\overline{\kappa}_4}{24n}(x^3-3x) + \frac{\overline{\kappa}_3^2}{72n}(x^5-10x+15x)\right]. \tag{4.3}$$

In Appendix II we prove

**Theorem 4.1:** *Assume that D1–D4 hold. Then*

$$\sup_x |\Pr[\overline{\mathbf{Q}} \le x] - \widetilde{F}_{n2}(x)| = o(n^{-1}).$$

Thus the formal Edgeworth approximation for $\overline{\mathbf{Q}}$ is valid, and, according to the argument we have laid out, so is one for $\widetilde{t}$. What is the form of the corresponding distributional approximation for $\tilde{t}$? Let $\kappa_1(\tilde{t}^{**})$ and $\kappa_2(\tilde{t}^{**})$ be order $n^{-1}$ approximations to the mean and variance of $\tilde{t}^{**}$ respectively, and define the restandardized estimator $\bar{\tilde{t}}^{**} = [\tilde{t}^{**} - \kappa_1(\tilde{t}^{**})]/\kappa_2(\tilde{t}^{**})^{1/2}$, which has mean zero and variance one to order $n^{-1}$. Now let $\kappa_3(\bar{\tilde{t}}^{**}) = \overline{\kappa}_3 n^{-1/2}$ and $\kappa_4(\bar{\tilde{t}}^{**}) = \overline{\kappa}_4 n^{-1}$ be order $n^{-1}$ approximations to the third and fourth cumulants of $\bar{\tilde{t}}^{**}$ respectively, where $\overline{\kappa}_3$ and $\overline{\kappa}_4$ are $O(1)$. Then, $\bar{\tilde{t}}^{**}$ has an approximate distribution function of the form $\widetilde{F}_{n2}$, with these $\overline{\kappa}_3$ and $\overline{\kappa}_4$ constants. Therefore,

$$\Pr[\tilde{t}^{**} \le x] = \widetilde{F}_{n2}(\kappa_1 + x\sqrt{\kappa_2}) + o(n^{-1}).$$

## 5. Examples

We now verify the conditions of Theorem 4.1 for some of the examples of Section 2. D1–D4 are straightforward to establish in most cases; the main problem arises in verifying D4. For notational simplicity we restrict our attention to quadratic statistics

$$\mathbf{Q} = n^{-1/2}\sum_{i=1}^{n} g(X_i) + n^{-\zeta}\sum\sum_{j>i} \rho_{ij}\varphi(X_i, X_j). \tag{5.1}$$

We replace D4 by a condition E1 that can be directly verified under iid sampling.

Firstly, we rewrite (5.1) as $\sum_{i,j}\theta_{ij}$, where

$$\theta_{ij} = \frac{1}{2n\sqrt{n}}[g(X_i) + g(X_j)] + n^{-\zeta}\rho_{ij}\varphi(X_i, X_j)\mathbf{I}(j > i),$$

with $\mathbf{I}(\cdot)$ the indicator function. Let $H_N = \sigma_{\mathbf{Q}}^{-1}\sum_{j=N+1}^{n}\theta_{1j}$, where $6 < N(n) < n$ is an integer, and

**E1** *There exists a constant $c < 1$, such that for $N(n) = O(n^{\delta})$, with $0 < \delta < 1/6$,*

$$\Pr[|E[\exp(\mathbf{i}sH_N)|X_{N+1}, X_{N+2}, ..., X_n| \ge c] = o\left(\frac{1}{n\log n}\right),$$

*uniformly for $s \in \left[\frac{n^{1/4+\zeta}}{\log n}, n\log n\right]$.*

Then

**Theorem 5.1:** *Assume that D1-D4 and E1 hold. Then the conclusion of Theorem 4.1 holds.*

In many cases of interest (see below) condition E1 is easy to verify.

*Homoskedastic Nonlinear Regression*

As far as justifying the Edgeworth approximation for (3.5) is concerned, we can restrict attention to $\mathbf{Q} = n^{-1/2}\sum_{i=1}^{n}[u_i^2 - \sigma^2] + \sum\sum_{j>i}\rho_{ij}u_iu_j$, which is of the form (5.1) with $g(u_1) = u_1^2 - \sigma^2$ and $\varphi(u_1, u_2) = u_1u_2$. In this case $H_N = \delta_{n1}(u_1^2 - \sigma^2) + \delta_{n2} + \delta_{n3}u_1$, where $\delta_{n1}$, $\delta_{n2}$ and $\delta_{n3}$ depend on $u_j$, $j > N$. Assume also that $u_i$ are normally distributed. In this case, the random variables $u_1^2 - \sigma^2$ and $\delta_{n3}u_1$ are mutually independent given $u_{N+1}$, ..., $u_n$. Therefore,

$$E_N\left[e^{\mathbf{i}sH_N}\right] = E_N\left[e^{\mathbf{i}s\delta_{n1}(u_1^2-\sigma^2)}\right]\left[e^{\mathbf{i}s\delta_{n2}}\right]E_N\left[e^{\mathbf{i}s\delta_{n3}u_1}\right],$$

where $E_N$ denotes expectation conditional on $u_{N+1}$, ..., $u_n$. Since $|e^{\mathbf{i}sx}| < 1\ \forall x$, we have

$$\left|E_N\left[e^{\mathbf{i}sH_N}\right]\right| = \left|E_N\left[e^{\mathbf{i}s\delta_{n1}(u_1^2-\sigma^2)}\right]\right|,$$

while

$$E_N\left[e^{\mathbf{i}s^*(u_1^2-\sigma^2)}\right] = \mathrm{mgf}_{\chi^2}(-\mathbf{i}s^{*2}/2) = \left[1 - 2\mathbf{i}s^{*2}\right]^{-1/2},$$

where $s^* = s\delta_{n1} = s(n-N)/n\sqrt{n}$. Provided only $s \geq n^{1/2+\delta}$ for some $\delta > 0$, we have $s^* \to \infty$, and condition E1 is satisfied.

In conclusion, the Edgeworth approximation of Theorem 4.1 is justified for the standardized estimator. The relevant cumulants can be calculated from (3.5) and an extension of (4.2). In particular,

$$\kappa_1(\widetilde{t}^{**}) = b\ ;\ \kappa_2(\widetilde{t}^{**}) = [E(\varsigma_i^2)]n^{-1}\sum_{i=1}^{n}\delta_i^2 + \sigma^4\sum\sum_{j\neq i}\rho_{ij}^2 + \sigma^2 n^{-1}\sum_{i=1}^{n}\xi_i^2.$$

where $\varsigma_i = u_i^2 - \sigma^2$. In the special case that $u_i$ are symmetric about zero, the skewness and kurtosis of the order $n^{-1}$ recentered and rescaled estimator are (to order $n^{-1}$):

$$\kappa_3(\widetilde{t}^{**}) = \left\{E(\varsigma_i^3)/[E(\varsigma_i^2)]^{3/2}\right\} n^{-1/2} \equiv \overline{\kappa}_3 n^{-1/2}; \; \kappa_4(\widetilde{t}^{**}) = \left\{E(\varsigma_i^4)/[E(\varsigma_i^2)]^2 - 2\right\} n^{-1} \equiv \overline{\kappa}_4 n^{-1},$$

and do not depend on the precise way in which $g$ is estimated. The mean and variance formulas can be further specialized for specific nonparametric estimators as in Linton [22,23].

*Partially Linear Regression*

In this case, it is sufficient to verify that the quadratic statistics: $n^{-1/2}\sum_{i=1}^n \eta_i\varepsilon_i + \sum\sum_{j>i}\rho_{ij}\varepsilon_i\eta_j$ and $n^{-1/2}\sum_{i=1}^n[\eta_i^2 - \sigma_\eta^2] + \sum\sum_{j>i}\rho_{ij}\eta_i\eta_j$ satisfy the conditions of Theorem 5.1; these can be verified using similar arguments to those given above.

## 6. Conclusions

Our main conclusion is that the formal order $n^{-1}$ Edgeworth approximation to the distribution of the standardized semiparametric estimator is valid under sufficient smoothness and moment conditions. In some special cases the first four cumulants are sufficient to determine the approximating distribution. Even so, the computations required to construct these approximations are burdensome. In practice, second order approximations that involve only the first two moments and in which the error is of order $n^{-2\mu}$, for some $\mu < 1/2$, may be sufficiently illuminating. These second order approximations can feasibly be constructed, see Linton [21,22,23] and provide useful information about the small sample behavior of the semiparametric estimator that is left out of the first order theory.

Our work demonstrates that the distributional approximation derived from these asymptotic moments is valid. The distributional approximations can be employed to improve critical values, see Chesher and Spady [13]. One can also discuss higher order efficiency in a more rigorous setting than is provided by the asymptotic moments alone. On this point, we mention that the classical results of Pfanzagl [34] regarding the relationship between first, second, and third order efficiency in parametric models do not hold when $G$ has to be estimated. The second order properties of $\tilde{\tau}$ are dominated by $\widehat{G}$. We cannot even uniformly rank different regression function estimators according to mean squared error. Therefore, there will not be a uniformly (in $G$) best estimator of $\tau$.

## FOOTNOTES

1. Robinson [43] establishes Berry-Essen bounds for a class of semiparametric estimators. These show that the (worst possible) rate at which the normal approximation is approached is slower than the $n^{-1/2}$ rate usual for parametric procedures. The Monte Carlo evidence presented in Hsieh and Manski [19] and Newey [31] suggests that performance of these estimators may be quite good in a range of circumstances, but see also Stock [49] and Stoker [50] for less flattering results.

2. Although this is often not required in certain *adaptive* situations.

3. Note that we make expansions of different lengths for $\widehat{S}^{\#}_{n\alpha}$, $\widehat{S}^{\#}_{n\pi\alpha}$ and $\widehat{S}^{\#}_{n\pi\alpha\gamma}$.

4. I am aware of only one paper, Götze [15], that deals with Edgeworth approximation for multivariate $U$-statistics under primitive conditions. In fact, this paper only establishes the validity of an order $n^{-1/2}$ Edgeworth approximation.

## APPENDIX I

**Proof of Lemma 2.1:** By the Bonferroni inequality,

$$\Pr[n^{\vartheta}\|\widehat{G}-G\|_n > c\log n] \le \sum_{k=1}^{L}\sum_{i=1}^{n}\Pr[n^{\vartheta}|\widehat{G}^k(Z_i)-G^k(Z_i)| > c\log n],$$

for any $\vartheta > 0$. But

$$\Pr[n^{\vartheta}|\widehat{G}^k(Z_i)-G^k(Z_i)| > c\log n] \le \Pr[n^{\vartheta}|V_{ki}| > c\log n/2] + \Pr[n^{\vartheta}|B_{ki}| > c\log n/2],$$

by the triangle inequality. Then, by assumption A2(ii), the second term is zero for large $n$. Also, $V_{ki} = O_p(n^{-\mu_k})$, and by the Markov inequality,

$$\Pr[n^{\vartheta}|V_{ki}| > c\log n] \le \frac{E[|V_{ki}|^J]}{(c\log n)^J} n^{J\vartheta}.$$

By the Marcinkiewicz–Zygmund inequality, $E[|V_{ki}|^J]$ is uniformly of order $n^{-\mu_k J}$ – see Robinson [38], Lemma 7. Therefore

$$\sum_{k=1}^{L}\sum_{i=1}^{n}\Pr[n^{\vartheta}|\widehat{G}^k(Z_i)-G^k(Z_i)| > c\log n] = o(n^{-2\phi}),$$

provided $1 + J\vartheta - J\zeta < -2\phi$ which holds under the stated conditions. ■

**Proof of Theorem 3.1:** In the sequel let $c$ and $\lambda$ be generic constants. The condition C* consists of several parts which are stated in the proof.

We may write $Q_n(t) = 0$, as

$$t = \sum_j A_j(t) \equiv \varphi(t), \qquad \text{(A1.1)}$$

where $\{A_j(t)\}$ consists of $-\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\alpha}$, $-n^{-\zeta}\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\pi\alpha}t^{\alpha}$, $-n^{-1/2}\overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\pi\alpha\gamma}t^{\alpha}t^{\gamma}$, $-\frac{1}{2}n^{(1/2+\zeta)}\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\pi\alpha\gamma}t^{\alpha}t^{\gamma}$, and $-\frac{1}{3!}n^{-1}\overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\pi\alpha\gamma\delta}t^{\alpha}t^{\gamma}t^{\delta}$, as well as linear combinations of $R_{Q\pi}$, $\widehat{R}^{\#}_{\alpha}$, $\widehat{R}^{\#}_{\pi\alpha}$, $\widehat{R}^{\#}_{\pi\alpha\gamma}$ and $\widehat{R}^{\#}_{\pi\alpha\gamma\delta}$.

Let $H_n(\tau_0; c) = \{t \ : \ |t| < c\log n\}$, and assume that $t \in H_n(\tau_0; c)$ for some $c$. We show below that

$$\Pr[|\varphi(t)| < c\log n] = 1 - o(n^{-1}), \qquad \text{(A1.2)}$$

i.e. $\varphi$ maps $H_n(\tau_0;\, c)$ into itself with high probability. In fact, we establish that for some constant $c$,

$$\Pr[|A_j(t)| > c\log n] = o(n^{-1}), \tag{A1.3}$$

for each $j$. Since $\Pr[|\sum_j A_j(t)| > c\log n] \leq \sum_j \Pr[|A_{nj}(t)| > c'\log n]$, for some constant $c'$ by repeated application of the triangle inequality, (A1.3) is sufficient for (A1.2) to hold.

Then, since $\varphi(t)$ is a continuous function of $t$ that maps a compact subset of $\mathbb{R}^p$ into itself, we can apply Brouwer's fixed point theorem: with probability $1 - o(n^{-1})$ there exists a solution $\widetilde{t}$ to the system of equations (A1.1) on $H_n(\tau_0;\, c)$.

**Proof of (A1.3):** We must show that $\Pr[|A_j(t)| > c\log n] = o(n^{-1})$, where $A_j$ come from:

(a) $-\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\alpha}$ ; (b) $-\sqrt{n}\,\overline{\Psi}_n^{\pi\alpha}\widehat{R}^{\#}_{n\alpha}$ ; (c) $n^{-\zeta}\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\pi\alpha}t^{\alpha}$ ; (d) $\overline{\Psi}_n^{\pi\alpha}\widehat{R}^{\#}_{n\pi\alpha}t^{\alpha}$; (e) $n^{-1/2}\overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\pi\alpha\gamma}t^{\alpha}t^{\gamma}$; (f) $\frac{1}{2}n^{-(1/2+\zeta)}\overline{\Psi}_n^{n\pi\alpha}\widehat{S}^{\#}_{n\pi\alpha\gamma}t^{\alpha}t^{\gamma}$ ; (g) $\frac{1}{2}n^{-1/2}\overline{\Psi}_n^{n\alpha}\widehat{R}^{\#}_{n\pi\alpha\gamma}t^{\alpha}t^{\gamma}$ ; (h) $\frac{1}{3!}n^{-1}\overline{\Psi}_n^{\pi\alpha}\overline{\Psi}_{n\pi\alpha\gamma\delta}t^{\alpha}t^{\gamma}t^{\delta}$ ; (i) $\frac{1}{3!}n^{-1}\overline{\Psi}_n^{\pi\alpha}\widehat{R}^{\#}_{n\pi\alpha\gamma\delta}t^{\alpha}t^{\gamma}t^{\delta}$ ; (j) $\overline{\Psi}_n^{\pi\alpha}R_{Q\alpha}$.

(a) We first consider terms due to $-\overline{\Psi}_n^{\pi\alpha}\widehat{S}^{\#}_{n\alpha}$. This is of the form $Hx$ where $H$ is a $P\times P$ matrix and $x$ a $P\times 1$ vector. Since $|Hx| \leq \lambda_{\max}(H)\max_{1\leq k\leq P}|x_k|$, where $\lambda_{\max}(\overline{\Psi}_n^{\pi\alpha})$ is uniformly bounded by Assumption B2, we can restrict ourselves to examining $\Pr[|\widehat{S}^{\#}_{n\alpha}| > c\log n]$ for each $\alpha$, where

$$\widehat{S}^{\#}_{n\alpha} = S_{n\alpha} + n^{-1/2}\sum_{i=1}^{n}\Psi_{\alpha i;k}\widehat{\Gamma}_i^k + \cdots + \frac{1}{m!}n^{-1/2}\sum_{i=1}^{n}\Psi_{\alpha i;k_1k_2\ldots k_m}\widehat{\Gamma}_i^{k_1}\widehat{\Gamma}_i^{k_2}\cdots\widehat{\Gamma}_i^{k_m}. \tag{A1.4}$$

(i) The leading term $S_{n\alpha}$ is a sum of independent random variables, which by Assumption B5 possesses a valid order $n^{-1}$ Edgeworth approximation such that

$$\sup_x |\Pr[S_{n\alpha} \leq x] - \widetilde{F}_{n2}(x)| = o(n^{-1}),$$

where $\widetilde{F}_{n2}(x) = \Phi(x) - \phi(x) \sum_{j=1}^{2} n^{j/2} Q_j(x)$ and where $Q_j$ are polynomials with coefficients determined by the first four cumulants of $S_{n\alpha}$. By the properties of the Edgeworth measure, $\widetilde{F}_{n2}(-c \log n) = o(n^{-1})$ for any positive constant $c$.

(ii) We now examine $\Pr[|n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} \widehat{\Gamma}_i^k| > c \log n]$. By construction, $n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} \widehat{\Gamma}_i^k$ $= n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} [B_{ki} + V_{ki}]$, where $\Psi_{\alpha i;k}$ are zero mean and independent (across $i$) random variables.

(ii.1) Therefore, $n^{\zeta}[n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} B_{ki}] = O_p(1)$. By Markov's inequality

$$\Pr[|n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} B_{ki}| > c \log n] \leq \frac{1}{[c \log n]^d} \frac{1}{n^{d\zeta}} E[|n^{\zeta} n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} B_{ki}|^d],$$

where $E[|n^{\zeta} n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} B_{ki}|^d]$ is bounded by Assumption B3. Therefore,

$$\Pr[|n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} B_{ki}| > c \log n] = o(n^{-1}),$$

provided

**C*(1)** $d\zeta \geq 1$.

(ii.2) Rewriting $n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;k} V_{ki} = n^{-\zeta} \sum_{i \neq j} \rho_{ij} \varphi(X_i,\, X_j)$, where $\rho_{ij} = n^{\zeta} n^{-1/2} w_{ij}$ and $\varphi(X_i,\, X_j) = \Psi_{\alpha i;k}(X_i) U_{kj}$, and applying the Markov's inequality as above:

$$\Pr[|\sum_{i \neq j} \rho_{ij} \varphi(X_i,\, X_j)| > c n^{\zeta} \log n] = o(n^{-1}),$$

provided $E[|\sum_{i \neq j} \rho_{ij} \varphi(X_i,\, X_j)|^d] < \lambda < \infty$ for some $d > 0$ such that $d\zeta \geq 1$. The moment exists provided $\sup_{i \geq 1} E[|\Psi_{\alpha i;k}(X_i)|^{2d}] < \infty$ and $\sup_{j \geq 1} E[|U_{kj}|^{2d}] < \infty$, see Mikosch [27], Lemma 1.3.

(iii) We now examine $n^{-1/2} \sum_{i=1}^{n} \Psi_{\alpha i;kl} \widehat{\Gamma}_i^k \widehat{\Gamma}_i^l$, which is bounded by

$$n^{-2\delta} [n^{1/4+\delta} \|\widehat{G} - G\|_n]^2 \left[ n^{-1} \sum_{i=1}^{n} |\Psi_{\alpha i;kl}|^{1+\varepsilon} \right]^{1/(1+\varepsilon)}$$

for any $\varepsilon > 0$, by the Hölder inequality. We now use the fact that

$$\Pr(A) \leq \Pr(A \cap B) + \Pr(B^c)$$

for any events $A$ and $B$, where $B^c$ denotes the complement of $B$. We take $B = \{G \in \mathcal{N}_G^{1/4}\}$, where $\mathcal{N}_G^{\nu} = \{n^{\nu}\|H - G\| \le c\log n\}$ for any $\nu$, and

$$A = \left\{ n^{-2\delta}[n^{1/4+\delta}\|\widehat{G} - G\|_n]^2 \left[ n^{-1}\sum_{i=1}^{n} |\Psi_{\alpha i;kl}|^{1+\varepsilon} \right]^{1/(1+\varepsilon)} > c\log n \right\}.$$

Note that

$$A \cap B \subseteq \left\{ n^{-1}\sum_{i=1}^{n} |\Psi_{\alpha i;kl}|^{1+\varepsilon} > cn^{2\delta(1+\varepsilon)}(\log n)^{-(1+\varepsilon)} \right\},$$

for some $c$ and for some $\delta,\ \varepsilon > 0$. Since $\Pr[\widehat{G} \notin \mathcal{N}_G^{1/4}] = o(n^{-1})$, by Lemma 2.1, it suffices to establish that

$$\Pr\left[ n^{-1}\sum_{i=1}^{n} |\Psi_{\alpha i;kl}|^{1+\varepsilon} > cn^{2\delta(1+\varepsilon)}(\log n)^{-(1+\varepsilon)} \right] = o(n^{-1}). \tag{A1.5}$$

Finally, by Markov's inequality, the left hand side of (A1.4) is bounded by

$$O(1)E\left[ \left\{ n^{-1}\sum_{i=1}^{n} |\Psi_{\alpha i;kl}|^{1+\varepsilon} \right\}^r \right] n^{-2\delta(1+\varepsilon)r}(\log n)^{(1+\varepsilon)r}$$

for any $r$; this is $o(n^{-1})$ provided $\sup_{i\ge 1} E[|\Psi_{\alpha i;kl}|^{(1+\varepsilon)r}] < \infty$, where

**C*(2)** $2\delta(1+\varepsilon)r > 1$.

The same argument applies to $n^{-1/2}\sum_{i=1}^{n} \Psi_{\alpha i;k_1\ldots k_j}\widehat{\Gamma}_i^{k_1}\cdots\widehat{\Gamma}_i^{k_j}$, for $j = 3$, ..., $m$.

(b) The remainder term $R_{\pi}^{\#}$ can be dealt with in the same way

$$\Pr\left[ |n^{-1/2}\sum_{i=1}^{n} \Psi_{\pi i;k_1k_2\ldots k_{m+1}}(X_i;\tau, G^*(Z_i))\widehat{\Gamma}_i^{k_1}\cdots\widehat{\Gamma}_i^{k_{m+1}}| > c\log n \right]$$

is bounded by

$$O(1)E\left[ \left\{ \Psi^{*\varepsilon}_{n\alpha;k_1\ldots k_{m+1}} \right\}^r \right] (\log n)^{(m+1)r} n^{-r\{(m+1)(\theta+\delta)-1/2\}} + o(n^{-1}).$$

Therefore, provided $\sup_{n\ge 1} E[\{\Psi^{*\varepsilon}_{n\alpha;k_1\ldots k_{m+1}}\}^r] < \infty$, with

C*(3) $r\{(m+1)(\theta+\delta) - 1/2\} > 1$.

the result follows.

Terms (c), ..., (j) require less stringent moment restrictions, and their analysis is omitted. ■

**Proof of Theorem 3.2** We have to show that there is a constant $c$ such that $\Pr[|\widehat{\widehat{R}} > \frac{c}{n \log n}] = o(n^{-1})$ on $|t| \le c \log n$, where $\widehat{\widehat{R}}$ consists of:

$$(a)\ \sqrt{n}\widehat{R}^{\#}_{\alpha}\ ;\ (b)\ \widehat{R}^{\#}_{\pi\alpha}\ ;\ (c)\ \frac{1}{2}n^{-1/2}\widehat{R}^{\#}_{\pi\gamma\alpha}t^{\alpha}t^{\gamma}\ ;\ (d)\ \frac{1}{3!}n^{-1}\widehat{R}^{\#}_{\pi\gamma\alpha\delta}t^{\alpha}t^{\gamma}t^{\delta}\ ;\ (e)\ \frac{1}{4!}n^{-3/2}\widehat{\Psi}_{n\pi\alpha\gamma\delta\rho}(\tau^{*})t^{\alpha}t^{\gamma}t^{\delta}t^{\rho}.$$

The argument is essentially the same as that used to show Theorem 3.1.

(a) We require $\sup_{n\ge1} E[\{\Psi^{*\varepsilon}_{n\alpha;k_1\dots k_{m+1}}\}^r \ < \infty$, with

**C*(4)** $r\{(m+1)(\theta+\delta) - 3/2\} > 1.$

(b) We require $\sup_{n\ge1} E[\{\Psi^{*\varepsilon}_{n\pi\alpha;k_1\dots k_{m+1}}\}^r] < \infty$, with $r\{(m+1)(\theta+\delta)-1\} > 1$, which is implied by **C*(4).**

(c) We require $\sup_{n\ge1} E[\{\Psi^{*\varepsilon}_{n\pi\alpha\gamma;k_1\dots k_{m+1}}\}^r] < \infty$, with $r\{(m+1)(\theta+\delta) - 1/2\} > 1$, which is implied by **C*(4).**

(d) We require $\sup_{n\ge1} E[\{\Psi^{*\varepsilon}_{n\pi\alpha\gamma\delta;k_1\dots k_{m+1}}\}^r] < \infty$, with $r(m+1)(\theta+\delta) > 1$, which is implied by **C*(4).**

(e) We require $\sup_{n\ge1} E[\{\Psi^{*\varepsilon}_{n\pi\alpha\gamma\delta\rho}\}^{2+\eta}] < \infty$, for some $\eta > 0$. ■

**Proof of Theorem 3.3:** It is sufficient to show

$$\Pr[|Q^{*}_{n}(\tilde{t}^{**})| > n^{-1}c\log n] = o(n^{-1}), \qquad \text{(A1.6)}$$

where $Q^{*}_{n}(\tilde{t}^{**})$ consists of a sum of terms each of which is a fourth order homogeneous polynomial in the standardized arrays $\widehat{S}^{\#}$. In fact $Q^{*}_{n}(\tilde{t}^{**})$ is $O_p(n^{-4\zeta})$, and is $o_p(n^{-1})$ when $\zeta > 1/3$. To prove (A1.6) we use the same techniques as in the previous theorem. We omit the details. ■

# APPENDIX II

**Proof of Theorem 4.1:** The following proposition is instrumental in establishing the validity of Theorem 4.1:

**Proposition:** *Let $F$ and $G$ be two signed measures with Fourier transforms $\gamma$ and $\psi$, where $\gamma'(0) = 0$, while $\psi$ is continuously differentiable, $\psi(0) = 1$ and $\gamma'(0) = 0$. 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*

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

The so-called smoothing lemma is proved in Bhattacharya and Rao [4]. For our application we identify $G$ with $\widetilde{F}_{n2}$, and $F$ with the distribution function of $\overline{\mathbf{Q}}$. We choose $T = n\log n$, in which case

$$\sup_x |\Pr[\overline{\mathbf{Q}} \leq x] - \widetilde{F}_{n2}(x) \leq \frac{2}{\pi}\int_0^{n\log n}\left|\frac{\psi_n(s)-\widetilde{\psi}_n(s)}{s}\right| ds + o(n^{-1}),$$

where

$$\widetilde{\psi}_n(s) = \exp\left(-\frac{s}{2}\right)^2\left[1 + \frac{\overline{\kappa}_3}{6\sqrt{n}}(\mathbf{i}s)^3 + \frac{\overline{\kappa}_4}{24n}(\mathbf{i}s)^4 + \frac{\overline{\kappa}_3^2}{72n}(\mathbf{i}s)^6\right] = \int e^{\mathbf{i}sx}d\widetilde{F}_{n2}(x) \quad \text{(A2.1)}$$

is the Fourier transform of the signed measure $\widetilde{F}_{n2}$.

We split the range of integration into several different parts and show that each subintegral is $o(n^{-1})$. We have

$$\int_0^{n\log n}\left|\frac{\psi_n(s)-\widetilde{\psi}_n(s)}{s}\right| ds \leq$$

$$\int_0^{n^\chi/\log n}\left|\frac{\psi_n(s)-\widetilde{\psi}_n(s)}{s}\right| ds + \int_{n^\chi/\ \log n}^{n^{1/4+\zeta}/\log n}\left|\frac{\psi_n(s)}{s}\right| ds + \int_{n^{1/4+\zeta}/\log n}^{n\ \log n}\left|\frac{\psi_n(s)}{s}\right| ds + \int_{\log n}^{\infty}\left|\frac{\widetilde{\psi}_n(s)}{s}\right| ds$$

$$= \mathbf{I} + \mathbf{II} + \mathbf{III} + \mathbf{IV}$$

for any $\chi \in (0,\ \zeta + 1/4)$. We choose $\chi$ in the sequel. This is a convenient decomposition, because different methods are applicable for each range. When $s$ is very small, we can rely solely on a Taylor series expansion, while for larger $s$ we also rely on crude bounds for the magnitude of $|\psi_n(s)|$, which hold when $s$ is kept at a distance from the origin. The last integral, **IV**, is $o(n^{-1})$ because of the form of the Edgeworth characteristic function $\widetilde{\psi}_n$, while **III** $= o(n^{-1})$ by D4. Our proof technique follows closely that used in Callaert, Janssen and Verarbereke [8], henceforth CJV, and we omit many details.

We use the following facts:

**(F1)** There exists an $\varepsilon > 0$ such that

$$|\eta(s)| \leq \exp\left(-\frac{1}{3}s^2\sigma_g^2\right),$$

for $|s| < \varepsilon/\sigma_g$ — see CJV, P304.

**(F2)** For any $\varepsilon > 0$ there exists a $\theta(\varepsilon) > 0$ such that

$$\left|\eta\left(\frac{s}{\sigma_g\sqrt{n}}\right)\right| \leq e^{-\theta},$$

for $s \leq \varepsilon\sqrt{n}$. This is a consequence of Cramer's condition D4.

**(F3)** For any random variables $X$ and $Y$, and for any $K$:

$$\left|E\left\{\left[e^{\mathbf{i}sX} - \sum_{j=0}^{K}\frac{(\mathbf{i}s)^j}{j!}X^j\right]Y\right\}\right| \leq \frac{2}{K!}s^K E[|X^{K+1}||Y|],$$

by Taylor expansion.

**(F4)** The following quantities are $O(1)$, by D1,

$$n^{(\zeta+1/2)}\sum\sum\sum_{j\neq i,k\neq j}\rho_{ij}^2\rho_{ki}\ ;\ n^{-(1/2+\zeta)}\sum\sum\sum\sum_{j\neq i,k\neq l}\rho_{ij}^2\rho_{kl}$$

$$n^{-(1+3\zeta)}\sum\sum\sum\sum\sum_{j\neq i,k\neq l,m\neq l}\rho_{ij}\rho_{kl}\rho_{lm};\ n^{-(3/2+3\zeta)}\sum\sum\sum\sum\sum_{j\neq i,k\neq l,m\neq p}\rho_{ij}\rho_{kl}\rho_{mp}.$$

**(F5)** Let $n^{-\varepsilon}\delta(n) \to \infty$, for some $\varepsilon > 0$, then

$$\int_{\delta(n)}^{\infty} t^k e^{-\lambda t^2}dt = O(\delta^{k+1}e^{-\lambda\delta^2}) = o(n^{-1}).$$

**I.**

Let $A_n = \frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}} \sum_{i=1}^n g(X_i)$, $B_n = \sigma_{\mathbf{Q}}^{-1} \sum\sum_{j>i} \rho_{ij}\varphi(X_i,\, X_j)$, and $C_n = \sigma_{\mathbf{Q}}^{-1} \sum\sum\sum_{k>j>i} \pi_{ijk}\nu(X_i,\, X_j,\, X_k)$. Then

$$\psi_n(s) = E[\exp(\mathbf{i}sA_n)\exp\{\mathbf{i}s(n^{-\zeta}B_n + n^{-\xi}C_n)\}]. \tag{A2.2}$$

We must show that

$$\int_0^{n^{\chi}/\log n} \left| \frac{\psi_n(s) - \widetilde{\psi}_n(s)}{s} \right| ds = o(n^{-1}),$$

where $\widetilde{\psi}_n$ is defined in (A2.1).

We first Taylor expand the second exponential in (A2.2), drop the remainder, and define the truncated term

$$\psi_{1,n}(s) = E\{\exp(\mathbf{i}sA_n)[1 + \mathbf{i}sn^{-\zeta}B_n + \frac{(\mathbf{i}s)^2}{2}n^{-2\zeta}B_n^2 + \mathbf{i}sn^{-\xi}C_n]\} \equiv \psi_{1,n}^b(s) + \psi_{1,n}^c(s),$$

with $\psi_{1,n}^c(s) = \mathbf{i}sn^{-\xi}E[\exp(\mathbf{i}sA_n)C_n]$.

Note that

$$\psi_{1,n}(s) = I_0^* + \mathbf{i}s\{n^{1/2}I_2^*E_2^*\mathcal{W}_1 + n^{1/2}I_3^*E_3^{*\prime}\mathcal{W}_4\} + \frac{(\mathbf{i}s)^2}{2}\{n^{-2\zeta}I_2^*E_2^{*\prime}\mathcal{W}_2 + I_3^*I_3^*\mathcal{W}_3 + nI_4^*[E_2^*]^2\mathcal{W}_5\},$$

where

$$\begin{aligned}
I_k^*(s) &= \eta\left(\tfrac{s}{\sigma_{\mathbf{Q}}\sqrt{n}}\right)^{n-k} \\
E_2^* &= \sigma_{\mathbf{Q}}^{-1}E\left\{\exp\left[\tfrac{\mathbf{i}s}{\sigma_{\mathbf{Q}}\sqrt{n}}\{g(X_1)+g(X_2)\}\right]\varphi(X_1,X_2)\right\} \\
E_2^{*\prime} &= \sigma_{\mathbf{Q}}^{-2}E\left\{\exp\left[\tfrac{\mathbf{i}s}{\sigma_{\mathbf{Q}}\sqrt{n}}\{g(X_1)+g(X_2)\}\right]\varphi(X_1,X_2)^2\right\} \\
E_3^* &= \sigma_{\mathbf{Q}}^{-2}E\left\{\exp\left[\tfrac{\mathbf{i}s}{\sigma_{\mathbf{Q}}\sqrt{n}}\{g(X_1)+g(X_2)+g(X_3)\}\right]\varphi(X_1,X_2)\varphi(X_1,X_3)\right\} \\
E_3^{*\prime} &= \sigma_{\mathbf{Q}}^{-1}E\left\{\exp\left[\tfrac{\mathbf{i}s}{\sigma_{\mathbf{Q}}\sqrt{n}}\{g(X_1)+g(X_2)+g(X_3)\}\right]\nu(X_1,X_2,X_3)\right\}.
\end{aligned}$$

In the next section we demonstrate that $\psi_{1,n}(s)$ can be approximated, in the range $s \in [0,\, n^{\chi}/\log n]$, by the following simpler function:

$$\widetilde{\psi}_{1,n}(s) = I_0 + \mathbf{i}s\{n^{1/2}I_2E_2\mathcal{W}_1 + n^{1/2}I_3E_3'\mathcal{W}_4\} + \frac{(\mathbf{i}s)^2}{2}\{n^{-2\zeta}I_2E_2'\mathcal{W}_2 + I_3E_3\mathcal{W}_3 + nI_4E_4\mathcal{W}_1^2\},$$

where

$$\begin{aligned}
E_2 &= \sigma_g^{-3}n^{-1}(\mathbf{i}s)^2E[g(X_1)g(X_2)\varphi(X_1,X_2)] + \sigma_g^{-4}n^{-3/2}(\mathbf{i}s)^3E[g^2(X_1)g(X_2)\varphi(X_1,\,X_2)]\} \\
E_2' &= \sigma_g^{-2}E[\varphi(X_1,\,X_2)^2] \\
E_3 &= \sigma_g^{-4}n^{-1}(\mathbf{i}s)^3E[g(X_2)g(X_3)\varphi(X_1,\,X_2)\varphi(X_1,\,X_3)] \\
E_3' &= \sigma_g^{-4}n^{-3/2}(\mathbf{i}s)^3E[g(X_1)g(X_2)g(X_3)\nu(X_1,\,X_2,\,X_3)] \\
E_4 &= \sigma_g^{-6}n^{-2}(\mathbf{i}s)^4E^2[g(X_1)g(X_2)\varphi(X_1,\,X_2)] \\
I_k &= e^{-s^2/2}\left\{1 - \tfrac{(\mathbf{i}s)^2}{2}\sigma_{n,k}^2 + \tfrac{(\mathbf{i}s)^3}{6\sqrt{n}}\kappa_{3g} + \tfrac{(\mathbf{i}s)^4}{24n}\kappa_{4g} + \tfrac{(\mathbf{i}s)^6}{72n}\kappa_{3g}^2\right\},
\end{aligned}$$

where $I_k$ is an $o(n^{-1})$ approximation to the characteristic function of $\frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}}\sum_{l=k+1}^{n} g(X_l)$ in which $\sigma_{n,k}^2 = \text{var}\left[\frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}}\sum_{l=1}^{k} g(X_l)\right]$.

By the triangle inequality

$$\begin{aligned}
\mathbf{I} &\le \int_0^{n^\chi/\log n}\left|\tfrac{\psi_n(s)-\psi_{1,n}(s)}{s}\right|ds + \int_0^{n^\chi/\log n}\left|\tfrac{\psi_{1,n}(s)-\widetilde{\psi}_{1,n}(s)}{s}\right|ds + \int_0^{n^\chi/\log n}\left|\tfrac{\widetilde{\psi}_n(s)-\widetilde{\psi}_{1,n}(s)}{s}\right|ds \\
&\equiv \mathbf{I.1} + \mathbf{I.2} + \mathbf{I.3}
\end{aligned}$$

We show **I.1**, **I.2**, and **I.3** below.

**I.1.** We start by introducing

$$\psi_{1,n}^*(s) = E[\exp\{\mathbf{i}s(A_n + n^{-\zeta}B_n)\}(1 + \mathbf{i}sn^{-\xi}C_n)] \equiv \psi_{1,n}^{*b}(s) + \psi_{1,n}^{*c}(s),$$

where $\psi_{1,n}^{*c}(s) = \mathbf{i}sn^{-\xi}E[\exp\{\mathbf{i}s(A_n + n^{-\zeta}B_n)\}C_n]$. Then,

$$|\psi_n(s) - \psi_{1,n}(s)| \le |\psi_n(s) - \psi_{1,n}^*(s)| + |\psi_{1,n}^*(s) - \psi_{1,n}(s)|.$$

By Taylor expansion (F3),

$$|\psi_{1,n}^*(s) - \psi_n(s)| \le \lambda n^{-2\xi}|s|^2E[|C_n|^2]$$

for some $\lambda < \infty$. Therefore

$$\int_0^{n^{\chi}/\log n} \left| \frac{\psi_n(s) - \psi^*_{1,n}(s)}{s} \right| ds = o(n^{-2(\xi - \chi)}),$$

which is $o(n^{-1})$, when $\chi < 1/8$, since $\xi \geq 5/8$.

We now examine $\int_0^{n^{\chi}/\log n} \left| \frac{\psi_{1,n}(s) - \psi_{1,n}(s)}{s} \right| ds$. By Taylor expansion $|\psi^b_{1,n}(s) - \psi^{*b}_{1,n}(s)|$ is bounded by $n^{-3\zeta}|s|^3 E[|B_n|^3]$. For $\chi$ very small we have

$$\int_0^{n^{\chi}/\log n} \left| \frac{\psi^b_{1,n}(s) - \psi^{*b}_{1,n}(s)}{s} \right| ds \leq n^{-3\zeta} E[|B_n|^3] \int_0^{n^{\chi}/\log n} s^2 ds = o(n^{3(\chi - \zeta)}),$$

which is $o(n^{-1})$, when $3(\chi - \zeta) \leq -1$. Therefore, we choose $\chi = \zeta - 1/3$. For larger $\chi$ we use the bound

$$n^{-3\zeta}|s|^3 |E[\exp(\mathbf{i}sA_n)B_n^3]| + n^{-4\zeta}|s|^4 E[B_n^4] \tag{A2.5}$$

for $|\psi^b_{1,n}(s) - \psi^{*b}_{1,n}(s)|$, and so

$$\int_0^{n^{\chi}/\log n} \left| \frac{\psi^b_{1,n}(s) - \psi^{*b}_{1,n}(s)}{s} \right| ds = o(n^{4(\chi - \zeta)}),$$

which is $o(n^{-1})$, when $\chi = \zeta - 1/4$. We now bound the first term in (A2.5) in the range $s \in [n^{\chi'}/\log n,\, n^{\chi}/\log n]$. Note that

$$B_n^3 = \sum\sum_{i \neq j} \rho^3_{ij} \varphi(X_i,\, X_j)^3 + \cdots + \sum\sum\sum\sum\sum\sum_{i \neq j \neq k \neq l \neq m \neq p} \rho_{ij}\rho_{kl}\rho_{mp} \varphi(X_i,\, X_j)\varphi(X_k,\, X_l)\varphi(X_m,\, X_p),$$

where the intermediate sums have 3, 4 and 5 different indices respectively. By the triangle inequality we must bound $|\sum\sum \rho^3_{ij} E[\exp(\mathbf{i}sA_n)\varphi(X_i,\, X_j)^3|$, ..., $|\sum\sum\sum\sum\sum\sum \rho_{ij}\rho_{kl}\rho_{mp} E[\exp(\mathbf{i}sA_n)\varphi(X_i,\, X_j)\varphi(X_k,\, X_l)\varphi(X_m,\, X_p)]|$.

We only show the calculation for the six-fold sum; the same method applies to the other terms. We decompose $A_n$ into orthogonal parts so that $A_n = \frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}} \sum_{i=1}^6 g(X_i) + \frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}} \sum_{i=7}^n g(X_i)$, and then use a conditioning argument to obtain

$$E[\exp(\mathbf{i}sA_n)\varphi(X_i,\, X_j)\varphi(X_k,\, X_l)\varphi(X_m, X_p)] =$$
$$I_6^* E[\exp\{\mathbf{i}s\tfrac{1}{\sigma_{\mathbf{Q}}\sqrt{n}} \textstyle\sum_{t=1}^6 g(X_t)\}\varphi(X_i,\, X_j)\varphi(X_k,\, X_l)\varphi(X_m,\, X_p)].$$

Choosing $K = 2$ in F3 we have

$$\left| E\left[\exp\left\{\mathbf{i}s(\sigma_{\mathbf{Q}}\sqrt{n})^{-1} \sum_{t \in \Delta_6} g(X_t)\right\} \varphi(X_i,\, X_j)\varphi(X_k,\, X_l)\varphi(X_m,\, X_p)\right]\right| \leq \lambda \frac{s}{\sqrt{n}}$$

for some constant $\lambda < \infty$. Furthermore, by F1 there is a positive constant $\lambda$, such that $|I_6^*| \leq e^{-\lambda s^2}$, for $s < \varepsilon\sqrt{n}\sigma_g/\sigma_{\mathbf{Q}}$. Then, applying F4 we have

$$\int_{n^{\chi'}/\log n}^{n^{\chi}/\log n} n^{-3\zeta} s^2 E[\exp(\mathbf{i}sA_n)B_n^3]ds \leq \lambda n^k \int_{n^{\chi'}/\log n}^{\infty} s^3 e^{-\lambda s^2} ds,$$

for some integer $k$, which is $o(n^{-1})$ by F5.

**I.2.** We estimate the error in approximating $\psi_{1,n}$ by $\widetilde{\psi}_{1,n}$ using the triangle inequality:

$$|\psi_{1,n}(s) - \widetilde{\psi}_{1,n}(s)| \leq |I_*^0 - I^0| + |s|n^{1/2}|I_2^* E_2^* - I_2 E_2||\mathcal{W}_1| + |s|n^{1/2}|I_3^* E_3^{*\prime} - I_3 E_3'||\mathcal{W}_4| + \frac{s^2}{2}\{n^{-2\zeta}|I_2^* E_2^{*\prime} - I_2 E_2'||\mathcal{W}_2| + |I_3^* E_3^* - I_3 E_3||\mathcal{W}_3| + n|I_4^*[E_2^*]^2 - I_4 E_4||\mathcal{W}_5|\},$$

where, for example, $|I_2^* E_2^* - I_2 E_2| \leq \lambda\{|I_2^* - I_2| + |E_2^* - E_2|\}$, for some $\lambda < \infty$. Therefore, **I.2** is a consequence of the following two lemmas:

**Lemma 4.1.1.** *For $0 \leq s \leq \varepsilon\sqrt{n}$, we have for any $k$, there exists $\lambda < \infty$, such that*

$$|I_k^* - I_k| \leq \lambda \delta_n n^{-1} s P(s) e^{ct^2},$$

*where $\delta_n \to 0$, $P(s)$ is a polynomial in $s$ with bounded coefficients, and $c$ is a positive constant.*

**Lemma 4.1.2.** *For all $s$, we have for some $\lambda < \infty$:*

$$|E_2^* - E_2| \leq \lambda \left[\frac{s^4}{n^2} + \frac{s^2}{n^{1+2\zeta}} + \frac{|s^3|}{n^{3/2+2\zeta}}\right]; \ |E_2^{*\prime} - E\prime_2| \leq \lambda \frac{|s|}{\sqrt{n}}; |E_3^* - E_3| \leq \lambda \left[\frac{|s|^3}{n\sqrt{n}} + \frac{s^2}{n^{1+2\zeta}}\right];$$
$$|E_3^{*\prime} - E_3'| \leq \lambda \left[\frac{|s^4|}{n^2} + \frac{|s|^3}{n^{3/2+2\zeta}}\right] |[E_2^*]^2 - E_4| \leq \lambda \left[\frac{|s|^5}{n^{5/2}} + \frac{s^4}{n^{2+2\zeta}}\right].$$

Then, for example

$$\sqrt{n}|\mathcal{W}_1| \int_0^{n^{\chi}/\log n} |E_2^*(s) - E_2(s)|ds \leq \lambda\sqrt{n} \int_0^{n^{\chi}/\log n} \left[\frac{s^4}{n^2} + \frac{s^2}{n^{1+2\zeta}} + \frac{|s^3|}{n^{3/2+2\zeta}}\right] ds$$
$$= \delta_n \left[\frac{n^{5\chi}}{n^{3/2}} + \frac{n^{3\chi}}{n^{1/2+2\zeta}} + \frac{n^{4\chi}}{n^{1+2\zeta}}\right]$$

for some sequence $\delta_n \to 0$. This is $o(n^{-1})$, provided: $\chi \leq 1/10$ and $\chi \leq 2\zeta/3 - 1/1/6$. These conditions are sufficient for the integrals involving $E$'s.

The integrals involving $|I_k^* - I_k|$ are $o(n^{-1})$ because of the exponential bound established in Lemma 4.1.1. In sum, $\mathbf{I.2} = o(n^{-1})$.

**I.3** is $o(n^{-1})$ by direct computation — both $\widetilde{\psi}_n$ and $\widetilde{\psi}_{1,n}$ involve polynomials in $s$ and $e^{-s^2/2}$ — and are suitably well behaved. Truncate further and check that the polynomials coefficients agree to $o(n^{-1})$.

**II.**

For $6 < N(n) < n$, define the partial sums: $A_N = \frac{1}{\sigma_{\mathbf{Q}}\sqrt{n}} \sum_{i=1}^N g(X_i)$, $B_{N,n} = \sigma_{\mathbf{Q}}^{-1} \sum_{j>i} \sum_{i=1}^N \rho_{ij}\varphi(X_i,\, X_j)$, and $C_{N,n,n} = \sigma_{\mathbf{Q}}^{-1} \sum\sum_{k>j>i} \sum_{i=1}^N \pi_{ijk}\nu(X_i,\, X_j,\, X_k)$. Then

$E[|A_N|^k] = O\left(\left[\frac{N}{n}\right]^{k/2}\right)$, $E[|B_{N,n}|^k] = O\left(\left[\frac{N}{n}\right]^{k/2}\right)$, $E[|C_{N,n,n}|^k] = O\left(\left[\frac{N}{n}\right]^{k/2}\right)$.

Let

$$D_{N,n} = A_n - A_N + n^{-\zeta}(B_n - B_{N,n}) + n^{-\xi}(C_n - C_{N,n,n}),$$

then

$$\psi_n(s) = E[\exp(\mathbf{i}sA_n)\exp(\mathbf{i}sD_{N,n})\exp(\mathbf{i}sn^{-\zeta}B_{N,n})\exp(\mathbf{i}sn^{-\xi}C_{N,n,n})],$$

where $D_{N,n}$ is independent of $A_n$. Furthermore, we can rewrite

$$\exp(\mathbf{i}sn^{-\zeta}B_{N,n}) \equiv 1 + \mathbf{i}sn^{-\zeta}B_{N,n} + \tfrac{(\mathbf{i}s)^2}{2}n^{-2\zeta}B_{N,n}^2 + \tfrac{(\mathbf{i}s)^3}{3!}n^{-3\zeta}B_{N,n}^3 + R_B$$
$$\exp(\mathbf{i}sn^{-\xi}C_{N,n,n}) \equiv 1 + \mathbf{i}sn^{-\xi}C_{N,n,n} + R_C$$

for implicit remainders $R_B$ and $R_C$. Therefore,

$$\begin{aligned}
\psi_n(s) = {} & E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})] \\
& + \mathbf{i}sn^{-\zeta}E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})B_{N,n}] + \cdots + E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})R_B] \\
& + \mathbf{i}sn^{-\xi}E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})C_{N,n,n}] + E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})R_C] \\
& + (\mathbf{i}s)^2 n^{-(\zeta+\xi)}E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})B_{N,n}C_{N,n,n}] + \cdots \\
& + E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})R_B R_C].
\end{aligned}$$

There are ten terms to examine. We first subdivide the range of integration further into (a) $[n^{\chi}/\log n, \varepsilon\sqrt{n}]$ and (b) $[\varepsilon\sqrt{n},\, n^{1/4+\zeta}/\log n]$.

**REGION (A)**

For some $\varepsilon > 0$, $\left|\eta\left(\frac{s}{\sigma_{\mathbf{Q}}\sqrt{n}}\right)\right| \leq \exp\left[-\frac{s^2}{3}n^{-1}(\sigma_g^2/\sigma_{\mathbf{Q}}^2)\right]$ for $s \leq \varepsilon\sqrt{n}$.

(i) Since $|\exp(\mathbf{i}sD_{n,N})| \leq 1$, we have

$$|E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{n,N})]| \leq |E[\exp\mathbf{i}sA_N)]| \leq \exp\left[-\frac{s^2}{3}Nn^{-1}(\sigma_g^2/\sigma_{\mathbf{Q}}^2)\right].$$

(ii)

$$E[\exp(\mathbf{i}A_N)\exp(\mathbf{i}sD_{N,n})B_{N,n}] = \sum\sum \rho_{ij}E[\varphi(X_i,\, X_j)\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})]. \tag{A2.6}$$

We work on each pair of indices $\{i, j\}$ separately. Decompose $A_N$ into $A_{N1}$ and $A_{N2}$, where $A_{N2}$ does not depend on $X_i$ or $X_j$. Therefore,

$$E[\varphi(X_i,\, X_j)\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})] = E[\varphi(X_i,\, X_j)\exp(\mathbf{i}sA_{N1})\exp(\mathbf{i}sD_{N,n})]E[\exp(\mathbf{i}sA_{N2})].$$

Then use the fact that $|\exp(\mathbf{i}sA_{N1})\exp(\mathbf{i}sD_{N,n})| \leq 1$ to find

$$|E[\varphi(X_i,\, X_j)\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})]| \leq E[|\varphi(X_i,\, X_j)|]E[|\exp(\mathbf{i}sA_{N2})|],$$

where $E[|\exp(\mathbf{i}sA_{N2})|] \leq \exp[-\lambda s^2(N/n)]$ for some $\lambda > 0$. Therefore

$$|E[\exp(\mathbf{i}sA_N)\exp(\mathbf{i}sD_{N,n})B_{N,n}]| \leq A_1(n)P(s)\exp[-\lambda s^2(N/n)],$$

where $A_1 = O(n^\alpha)$ for some real number $\alpha$, while $P(s)$ is a polynomial in $s$. The same reasoning applies to the other terms depending on $B_{N,n}^2$, $B_{N,n}^3$, and $C_{N,n,n}$, provided the relevant absolute moments are bounded. A sufficient condition for

$$\int_{n^\chi/\log n}^{\infty} A_1(n)P(s)\exp[-\lambda s^2(N/n)] = o(n^{-1})$$

is $Nn^{2\chi-1-\varepsilon} \to \infty$, for some $\varepsilon > 0$.

(iii) **REMAINDER TERMS:** Provided the relevant moments exist:

(R1) $E[|R_B|] \leq \frac{2}{4!} s^4 n^{-4\zeta} E[B_{N,n}^4] \leq \lambda s^4 n^{-4\zeta} \left(\frac{N}{n}\right)^2$

(R2) $E[|R_C|] \leq s^2 n^{-2\xi} E[C_{N,n,n}^2] \leq \lambda s^2 n^{2\xi} \left(\frac{N}{n}\right)$

(R3) $E[|R_B R_C|] \leq \left\{E[|R_B|^{3/2}]\right\}^{2/3} \left\{E[|R_C|^3]\right\}^{1/3} \leq \lambda s^6 n^{-(4\zeta+2\xi)} \left(\frac{N}{n}\right)^3.$

We subdivide the range of integration further, so that

$$\int_{n^\chi/\log n}^{\varepsilon\sqrt{n}} \left|\frac{\psi_n(s)}{s}\right| ds = \int_{n^\chi/\log n}^{n^{\chi_1}/\log n} \left|\frac{\psi_n(s)}{s}\right| ds + \int_{n^{\chi_1}/\log n}^{n^{\chi_2}/\log n} \left|\frac{\psi_n(s)}{s}\right| ds + \cdots + \int_{n^{\chi_q}/\log n}^{\varepsilon\sqrt{n}} \left|\frac{\psi_n(s)}{s}\right| ds,$$

where the number of integrals, $q$, depends on $\zeta$. Note that the contribution of the remainder terms to the integral $\int_{n^{\chi_{j-1}}/\log n}^{n^{\chi_j}/\log n} \left|\frac{\psi_n(s)}{s}\right| ds$ are: (R1) $O\left(n^{-4\zeta}\left(\frac{N}{n}\right)^2 n^{4\chi_j/\log n}\right)$, (R2) $O\left(n^{-2\xi}\left(\frac{N}{n}\right) n^{2\chi_j/\log n}\right)$, and (R3) $O\left(n^{-(4\zeta+2\xi)}\left(\frac{N}{n}\right)^3 n^{6\chi_j/\log n}\right)$. We choose $N$ differently for each subintegral so as to balance the remainder terms with the main terms. We illustrate the argument for the special case that $\zeta = 3/8$ and $\xi$ is large enough to be ignored (e.g. $\xi = 7/8$) so that $q = 3$ and:

(I1) $N = O(n^{3/4})$ on $[n^{1/8}/\log n,\, n^{2/8}/\log n]$

(I2) $N = O(n^{1/2})$ on $[n^{2/8}/\log n,\, n^{3/8}/\log n]$

(I3) $N = O(n^{1/4})$ on $[n^{3/8}/\log n,\, \varepsilon/\sqrt{n}]$.

**REGION (B)**. When $s$ is large, we use the bound F2 instead of F1, and then apply the same method as for moderate $s$. In fact when $s \in [\varepsilon\sqrt{n},\, n^{1/4+\zeta}]$ we choose $N = [2\log n/\chi_1]$. Provided $\zeta > 1/4$, the same argument works. ∎

**THEOREM 5.1:** *We have to show that* $\int_{n^{1/4+\zeta}}^{n\log n} \left|\frac{\psi_n(s)}{s}\right| ds = o(n^{-1})$. *We use the following lemma:*

**LEMMA 5.1.1:** *For all* $s$ *and for all* $n$ *and* $N$, *with* $6 < N < n$, *there exists*

$\lambda < \infty$, *such that*

$$\psi_n(s) \leq E[|E[\exp(\mathbf{i}sH_N)|X_{N+1},\, ...,\, X_n]|^{N-6}]\left\{1+\lambda\sum_{k=0}^{3}|s|^k\left[\frac{N}{n}\right]^{3k/2}\right\}+\lambda|s|^4\frac{N^6}{n^6}.$$

Then, by assumption E1 there exists a constant $c < 1$ such that

$$|E[\exp \mathbf{i}sH_N)|X_{N+1},\, ...,\, X_n]| \leq c,$$

uniformly for $s \in \left[\frac{n^{1/4+\zeta}}{\log n}, n\log n\right]$ with probability $1 - o\left(\frac{1}{n\log n}\right)$. This gives us an exponential bound for

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

The second term contributes $o(n^{-1})$, provided $N$ is small enough; $N(n) = o(n^{1/6})$ will work. ■

## PROOFS OF LEMMAS

**Proof of LEMMA 4.1.1.** Exactly the same as CJV, Lemma 2. ■

**Proof of LEMMA 4.1.1.** Consider the first inequality. Let $\widetilde{E}_2$ be $E_2$ with $\sigma_{\mathbf{Q}}$ replacing $\sigma_g$. Then $|E_2 - E_2^*| \leq |\widetilde{E}_2 - E_2^*| + |\widetilde{E}_2 - E_2|$, and $|\widetilde{E}_2 - E_2^*| \leq \lambda\frac{s^4}{n^2}$ for some $\lambda < \infty$, by the general inequality F3 with $K = 3$. Also by Taylor expanding $\sigma_{\mathbf{Q}}$ in terms of $\sigma_g$ we find

$$|\widetilde{E}_2 - E_2| \leq \lambda\frac{s^2}{n^{1+2\zeta}} + \frac{|s^3|}{n^{3/2+2\zeta}}.$$

The remaining inequalities are established similarly. ■

**Proof of LEMMA 5.1.1.** The proof follows by the same argument as given in Lemma 5 of CJV. Let $\widetilde{B}_{N-1,N} = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\theta_{ij}$, then $E[|\widetilde{B}_{N-1,N}|^j] = O[(N^2n^{-(1+4\zeta)})^{j/2}] + O[(N^3n^{-3})^{j/2}]$, then $N$ is sufficiently small. ■

## REFERENCES


1. Andrews, D. W. K. (1989): "Semiparametric Econometric Models: I. Estimation," Cowles Foundation Discussion Paper No. 908.

2. Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Models," *Econometrica* 59, 307–346.

3. Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics* 6, 434–451.

4. Bhattacharya, R. N. and Ranga Rao (1976): *Normal approximation and asymptotic expansions.* Wiley, New York.

5. Bickel, P. J. (1982): "On Adaptive Estimation," *Annals of Statistics* 14, 1463–1484.

6. 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.

7. Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1994): *Efficient and Adaptive Inference in Semiparametric Models.* Baltimore: Johns Hopkins University Press.

8. Callaert, H., P. Janssen and N. Veraverbeke (1980): "An Edgeworth Expansion for U-Statistics," *Annals of Statistics* 8, 299–312.

9. Carroll, R. J. (1982): "Adapting for Heteroscedasticity in Linear Models," *Annals of Statistics* 10, 1224–1233.

10. Carroll, R. J. and P. Hall (1989): "Variance Function Estimation in Regression: The Effect of Estimating the Mean," *Journal of the Royal Statistical Society, Series B* 51, 3–14.

11. Cavanagh, C. L. (1990): "The Cost of Adapting for Heteroskedasticity in Linear Regression Models." Unpublished manuscript, Harvard University.

12. Chen, H. (1988): "Convergence Rates for Parametric Components in a Partly Linear Model," *Annals of Statistics* 16, 136–146.

13. Chesher, A. and R. Spady (1991): "Asymptotic Expansion of the Information Test," *Econometrica* 59, 787–816.

14. Chibisov, D. M. (1980): "An Asymptotic Expansion for the Distribution of a Statistic Admitting an Stochastic Expansion," *Theory of Probability and its Applications* 25, 732–744.

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

16. Härdle, W. and O. B. Linton (1994): "Applied Nonparametric Methods," in R. F. Engle III and D. McFadden, eds., *The Handbook of Econometrics, Volume 4.*

17. Härdle, W. and T. M. Stoker (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association* 84, 986–995.

18. Härdle, W. (1990): *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

19. Hsieh, D. A. and C. F. Manski (1987): "Monte Carlo Evidence on Adaptive Maximum Likelihood Estimation of a Regression," *Annals of Statistics* 15, 541–551.

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

21. Linton, O. B. (1991): "Edgeworth Approximation in Semiparametric Regression Models," Ph.D. Thesis, Department of Economics, UC Berkeley.

22. Linton, O. B. (1992): "Second Order Approximation in the Partially Linear Model," Cowles Foundation Discussion Paper No. 1065. Forthcoming in *Econometrica.*

23. Linton, O. B. (1993): "Second Order Approximation in a Linear Regression with Heteroskedasticity of Unknown Form." Forthcoming in *Econometric Reviews.*

24. Linton, O. B. (1993): "Adaptive Estimation in ARCH Models." *Econometric Theory.* 9, 539-570.

25. McCullagh, P. (1989): *Tensor Methods in Statistics.* Chapman and Hall.

26. Manski, C. F. (1984): “Adaptive Estimation of Non-linear Regression Models,” *Econometric Reviews* 3, 145–194.

27. Mikosch, T. (1991): “Functional Limit Theorems for Random Quadratic Forms,” *Stochastic Processes and their Application* 37, 81–98.

28. Mitrofanova, N. M. (1967): “An Asymptotic Expansion for the Maximum Likelihood Estimator of a Vector Parameter,” *Theory of Probability and its Applications* 12, 364–372.

29. Müller, H. G. (1988): “Nonparametric Regression Analysis of Longitudinal Data,” *Lecture Notes in Statistics*, Vol. 46. Heidelberg/New York: Springer Verlag.

30. Müller, H. G. and U. Stadtmüller (1987): “Estimation of Heteroskedasticity in Regression Analysis,” *Annals of Statistics* 15, 610–625.

31. Newey, W. K. (1988): “Adaptive Estimation of Regression Models via Moment Restrictions,” *Journal of Econometrics* 38, 301–339.

32. Newey, W.K. (1990): "Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics*, 5, 99-135.

33. Pfanzagl, J. (1973): “Asymptotic Expansions Related to Minimum Contrast Estimators,” *Annals of Statistics* 1, 993–1026.

34. Pfanzagl, J. (1980): “Asymptotic Expansions in Parametric Statistical Theory,” in P. R. Krishnaiah, ed., *Developments in Statistics*, Vol. 3. Academic Press.

35. Phillips, P. C. B. (1977): “A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators,” *Econometrica* 45, 463–485.

36. Phillips, P. C. B. (1977): “An Approximation to the Finite Sample Distribution of Zellner’s Seemingly Unrelated Regression Estimator,” *Journal of Econometrics* 6, 147–164.

37. Powell, J. L., J. H. Stock and T. M. Stoker (1989): “Semiparametric Estimation of Index Coefficients,” *Econometrica* 57, 1403–1430.

38. Robinson, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica* 56, 875–891.

39. Robinson, P. M. (1988): "Root-$N$-Consistent Semiparametric Regression," *Econometrica* 56, 931–954.

40. Robinson, P. M. (1988): "Semiparametric Econometrics: A Survey," *Journal of Applied Econometrics* 3, 35–51.

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

42. Robinson, P. M. (1991): "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica* 59, 1329–1364.

43. Robinson, P. M. (1991): "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica* 59, 755–786.

44. Rothenberg, T. J. (1984): "Approximate Normality of Generalized Least Squares Estimates," *Econometrica* 52, 811–825.

45. Rothenberg, T. J. (1986): "Approximating the Distribution of Econometric Estimators and Test Statistics," in Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 2, Ch. 15. Amsterdam: North–Holland.

46. Sargan, J. D. and E. Mikhail (1971): "A General Approximation to the Distribution of Instrumental Variable Estimates," *Econometrica* 39, 131–169.

47. Sargan, J. D. (1976): "Econometric Estimators and the Edgeworth Expansion," *Econometrica* 44, 421–448.

48. Skovgaard, I. M. (1981): "Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators in the General Case," *Scandinavian Journal of Statistics* 10, 227–236.

49. Stock, J. H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association* 84, 567–576.

50. Stoker, T. M. (1991): "Smoothing Bias in Derivative Estimation," forthcoming in *Journal of the American Statistical Association*.