An asymptotic characterization of finite degree U-statistics with sample size dependent kernels: applications to nonparametric estimators and test statistics [1]

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: f.yao@cgiar.org
Voice: + 1 202 862 6488

and

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

February, 2013

Abstract. We provide a simple result on the H-decomposition of a U-statistics that allows for easy determination of its magnitude when the statistic's kernel depends on the sample size n. The result provides a direct and convenient method to characterize the asymptotic magnitude of semiparametric and nonparametric estimators or test statistics involving high dimensional sums. We illustrate the use of our result in previously studied estimators/test statistics and in a novel nonparametric R^2 test for overall significance of a nonparametric regression model.

Keywords and phrases: U-statistics, nonparametric R^2, nonparametric testing.

JEL Classifications: C12, C14.

AMS-MS Classification. 62F10, 62G05, 62G08, 62G20.

[1] We thank the Associate Editor and two anonymous reviewers for their comments. Any remaining errors are the authors' responsibility.

1 Introduction

Nonparametric and semiparametric statistical models have gained popularity due to their flexibility in specifying functional forms for moments or distributions under study (Li and Racine (2007), Tsybakov (2009)). In many instances, the asymptotic characterization of estimators and test statistics associated with these models involves the study of U-statistics. For example, consider the estimation of a generalized mean $\theta = E(c(X)m(X))$ (Newey (1994), Imbens and Ridder (2009)) where $m(x) \equiv E(Y|X = x)$ and $c(X)$ is a known function. Given a random sample $\{(y_i, x_i)\}_{i=1}^n$, the estimator for $m(x)$ can usually be written as $\hat{m}(x) = \sum_{i=1}^{n} w_{in}(x) y_i$ for weight functions $w_{in}(x)$ that generally depend on the sample size n. Hence, a nonparametric estimator for θ can be defined as $\hat{\theta}_n = \sum_{j=1}^{n} c(x_j) \sum_{i=1}^{n} w_{in}(x_j) y_i = T_n + \binom{n}{2} u_n$, where $T_n = \sum_{i=1}^{n} c(x_i) w_{in}(x_i) y_i$ and $u_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \phi_n(Z_i, Z_j)$ is a U-statistic of degree 2, symmetric kernel $\phi_n(Z_i, Z_j) = c(x_i) w_{jn}(x_i) y_j + c(x_j) w_{in}(x_j) y_i$ with $Z_i = (x_i, y_i)$.

The asymptotic behavior of T_n can easily be studied by suitable central limit theorems or law of large numbers for triangular arrays, but a characterization of the asymptotic properties of u_n is more involved as the kernel $\phi_n(Z_i, Z_j)$ depends on the sample size n.

There exists, however, a large literature providing various asymptotic results. For non degenerate U-statistics with n-dependent kernels of degree $k \in \mathbb{N}$, Weber (1983) and Rao Jammalamadaka and Janson (1986) ($k = 2$) obtain central limit theorems (CLT) when the sequence $\{Z_i\}_{i=1,2,\cdots}$ is independent and identically distributed (IID) under different assumptions on the order of the sequence of U-statistic variances $V(u_n)$. For example, Weber requires that the variance of the conditional expectation of the kernels depend on n in a specific manner (see condition (i) in Theorem 1) and that the U-statistics projections have nonzero variances. Thus, his results cannot be applied to degenerate U-statistics (see terms A_{2n} or A_{3n} in Lemma 1 of our section 3).

Powell et al. (1989) obtains a CLT for U-statistics of degree $k = 2$ when $E(\phi_n^2(Z_i, Z_j)) = o(n)$ and $\{Z_i\}_{i=1,2,\cdots}$ is IID. Under the same conditions, Martins-Filho and Yao (2006) show that U and V statistics of degree k are $\sqrt{n}$ asymptotically equivalent. Their result, together with a lemma in Lee (1988), implies the $\sqrt{n}$

asymptotic equivalence in probability of a U-statistic projection and a V statistic. These results, however, provide little guidance on analyzing the asymptotic properties of higher degree U-statistic whose magnitude are different from $\sqrt{n}$. These situations occur frequently in nonparametric statistics/econometrics, as with test statistics in Zheng (1996), Fan and Li (1996), Li (1999), Lavergne and Vuong (2000), Gu et al. (2007) and Su and Ullah (2012) which converge to a normal distribution at a rate of $nh_n^{1/2}$, where h_n is a bandwidth sequence used in estimation.

Weber (1980) and van Zwet (1984) obtain Berry-Esseen bounds when $\{Z_i\}_{i=1,2,\cdots}$ is IID and the third order moment of ϕ_n exist and Hall (1984), de Jong (1987) and Fan and Li (1999) have been obtained CLTs for degenerate U-statistics with kernels that have fixed variance.

Applications of the extant results to characterize non/semiparametric estimators or test statistics requires the verification of the order for the variance $V(u_n)$ of their associated U-statistic. It is, therefore, convenient to use Hoeffding (1961) H-decomposition and write u_n as a linear combination of k uncorrelated U-statistics of degree $1, 2, \cdots, k$. Once the order of the variance of the component U-statistics are obtained, the order of $V(u_n)$ can easily be established. Furthermore, since H-decompositions are exact representations of the U-statistic of interest, contrary to the projections used in Hoeffding (1948) of Weber (1983), it is possible to focus on the component terms of u_n with leading variances. Obtaining the order of magnitude of each components in the H-decomposition easily enables us to ignore the degenerate terms and the terms whose orders are negligible, so we can focus on the exact expression of the leading term to perform further asymptotic analysis.

In this paper, following Hoeffding (1961) and Lee (1990), we provide a convenient expression that determines the order of each component of an H-decomposition for a U-statistic of degree k with n-dependent kernel. As expected, since the U-statistic kernel depends on n, the order of each component depends on n explicitly. Furthermore, it depends on the leading variance of the conditional expectation of the U-statistic kernel, which also depends on n. We illustrate the use of our result by applying it to previously studied estimators and test statistics considered by Li (1996) and Lavergne and Vuong (2000). In addition, we propose and apply our result to the study of a novel test statistic for the overall significance of a regression model based on a nonparametric R^2.[1] Besides this introduction, we provide our main result in section 2 and illustrate its use in section 3. A brief conclusion is provided in section 4. All proofs and technical assumptions are relegated to the Appendix.

2 The order of magnitude of U-statistics

Let $\{Z_i\}_{i=1}^n$ be a sequence of independent and identically distributed (IID) random variables and $\phi_n(Z_1, \cdots, Z_k)$ be a symmetric function with $k < n$. We call $\phi_n(Z_1, \cdots, Z_k)$ a kernel function that depends on n and a U-statistic u_n of degree k is defined as

$$u_n = \binom{n}{k}^{-1} \sum_{(n,k)} \phi_n(Z_{i_1}, \cdots, Z_{i_k}), \tag{1}$$

where $\sum_{(n,k)}$ denotes the sum over all subsets $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of $\{1, 2, \cdots, n\}$. Now, let $\phi_{cn}(z_1, \cdots, z_c) = E(\phi_n(Z_1, \cdots, Z_c, Z_{c+1}, \cdots, Z_k)|Z_1 = z_1, Z_2 = z_2, \cdots, Z_c = z_c)$, $\sigma_{cn}^2 = Var(\phi_{cn}(Z_1, \cdots, Z_c))$ and $\theta_n = E(\phi_n(Z_1, \cdots, Z_k))$. In addition, recursively define $h_n^{(1)}(z_1) = \phi_{1n}(z_1) - \theta_n, \cdots, h_n^{(c)}(z_1, \cdots, z_c) = \phi_{cn}(z_1, \cdots, z_c) - \sum_{j=1}^{c-1} \sum_{(c,j)} h_n^{(j)}(z_{i_1}, \cdots, z_{i_j}) - \theta_n$ for $c = 2, \cdots, k$, where the sum $\sum_{(c,j)}$ is over all subsets $1 \leq i_1 < \cdots < i_j \leq c$ of $\{1, \cdots, c\}$. By Hoeffding's H-decomposition we have

$$u_n = \theta_n + \binom{n}{k}^{-1} \sum_{j=1}^{k} \binom{n-j}{k-j} \sum_{(n,j)} h_n^{(j)}(Z_{v_1}, \cdots, Z_{v_j}) = \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(Z_{v_1}, \cdots, Z_{v_j}),$$

where $H_n^{(j)}(Z_{v_1}, \cdots, Z_{v_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(Z_{v_1}, \cdots, Z_{v_j})$. Since u_n can be written as a finite sum of $H_n^{(j)}$, its magnitude can be determined by studying $H_n^{(j)}$. The following result shows that the magnitude of $H_n^{(j)}$ is determined by n and the leading variance σ_{jn}^2 defined above.

Theorem 1. *Let $\{Z_i\}_{i=1}^n$ be an IID sequence and u_n be defined as in equation (1) such that*

$$u_n = \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(z_{v_1}, \cdots, z_{v_j}).$$

Then,

(a) $Var\left(H_n^{(j)}\right) = O\left(n^{-j} \sum_{c=1}^{j} \sigma_{cn}^2\right) = O\left(n^{-j}\sigma_{jn}^2\right)$ *and* $H_n^{(j)} = O_p\left((n^{-j}\sigma_{jn}^2)^{\frac{1}{2}}\right)$*;*

[1] As pointed out be a referee, it is possible to generalize equation (5.13) and Theorem 5.1 in Hoeffding (1948) to sample size dependent kernels to study the asymptotic magnitude of the variance of U-statistics. However, in this case one could not determine the specific component expressions which determine the order of magnitude of the underlying U-statistic.

(b) for $1 \le c \le c' \le k$, we have $\frac{\sigma^2_{cn}}{c} \le \frac{\sigma^2_{c'n}}{c'}$.

Theorem 1 establishes that provided the $H_n^{(j)}$'s are of different magnitudes, the magnitude of u_n can be determined by the largest term among $H_n^{(j)}$'s and θ_n. Instead of a laborious case by case component analysis, the magnitude of u_n can easily be determined using $H_n^{(j)} = O_p((n^{-j}\sigma^2_{jn})^{\frac{1}{2}})$ for $j = 1, \cdots, k$. Since in most instances k is relatively small, our theorem provides a convenient manner to determine the magnitude of u_n by analyzing the leading variance σ^2_{jn}.

We expect that, in general, the order of degeneracy determines the asymptotic distribution of U-statistics. For example, when the U-statistic is degenerate of order d, i.e., $0 = \sigma^2_{1n} = \cdots = \sigma^2_{dn} < \sigma^2_{d+1\,n}$, then $H_n^{(d+1)}$ will determine its asymptotic distribution. We can draw a similar observation for U-statistics whose kernels are independent of sample size. However, we note that since σ^2_{jn}'s all depend on n, we need to argue separately that $H_n^{(j)} = O_p\left((n^{-j}\sigma^2_{jn})^{\frac{1}{2}}\right)$ for $j = d+2, \cdots, k$ is of smaller order than $H_n^{(d+1)}$, so that we can focus on $H_n^{(d+1)}$ to study the asymptotic distribution of u_n. This could entail additional regularity conditions, for example, assumptions on the bandwidth to decay to zero in the examples studied in section 3. There will be, of course, no additional work in U-statistics whose kernel does not depend on n, since σ^2_j's do not depend on n, $H^{(j+1)}$ is automatically of smaller order than $H^{(j)}$.

Theorem 1 is useful in a variety of settings. As mentioned in the introduction, Lemma 3.1 in Powell et al. (1989) is a special case with $k = 2$. Alternatively, if the U-statistic kernel does not depend on n then our theorem reproduces classical results in Hoeffding (1961) and Lee (1990) which give $H_n^{(j)} = O_p((n^{-j})^{\frac{1}{2}})$.

Theorem 1 is also useful in establishing the magnitude of some symmetric statistics. In particular, if $S_n(Z_1, Z_2, \cdots, Z_n)$ is a symmetric statistic of finite order k, then by Theorem 1 in Lee (1990, p. 164), S_n is a U-statistic of degree k with n-dependent kernels given by

$$\psi_n(z_1, \cdots, z_k) = \sum_{j=0}^{k} \left(\begin{array}{c} n \\ j \end{array} \right) \left(\begin{array}{c} k \\ j \end{array} \right)^{-1} \sum_{(k,j)} s_n^{(j)}(z_{i_1}, \cdots, z_{i_j})$$

where $s_n^{(j)}(z_1, \cdots, z_j) = E(S_n(Z_1, Z_2, \cdots, Z_n)|Z_1 = z_1, \cdots, Z_j = z_j) - \sum_{l=0}^{j-1} \sum_{(j,l)} s_n^{(l)}(z_{i_1}, \cdots, z_{i_l})$ and $s_n^{(0)} = E(S_n(Z_1, Z_2, \cdots, Z_n))$. Thus, Theorem 1 can be used to establish bounds and orders for the variance of these symmetric statistics. As such, our Theorem 1 aides the verification of bounds on second moments assumed, for example, in Rubin and Vitale (1980).

Lenth (1983) has shown that symmetric statistics that satisfy

$$S_n(Z_1, \cdots, Z_n) = \frac{1}{n}\sum_{i=1}^{n} S_{n-1}(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) \tag{2}$$

are U-statistics of degree $k < n$. Hence, Theorem 1 can also be applied to establish the magnitude of symmetric statistics that satisfy condition (2).

Although there exists a rich literature on asymptotic characterizations of symmetric statistics based on H-decompositions (Efron and Stein (1981), Karlin and Rinott (1982), Dynkin and Mandelbaum (1983), Takemura (1983), Mandelbaum and Taqqu (1984), van Zwet (1984), Friedrich (1989) and Pecatti (2004)), all of these papers assume *a priori* the existence of higher order moments for the symmetric statistic, failing to provide a route for the determination of the magnitude of u_n as in Theorem 1. To our knowledge, the most promising result in this regard is Theorem 4.1 in Vitale (1992). He provides a useful representation for the variance of conditional expectations of symmetric statistics, but gives no additional insight on how to establish their magnitude as $n \to \infty$.

3 Applications: nonparametric estimator and test statistics

3.1 Previous literature

We consider the application of Theorem 1 to semiparametric estimation of partially linear models. Li (1996) shows that a $\sqrt{n}$ consistent estimator for the coefficients of the parametric part of the regression function can be obtained by using a nonnegative second-order kernel if the dimension of the variables in the nonparametric part is less than or equal to five. The result is obtained by, among other things, determining the order of magnitude of a sum of squared differences between a nonparametric conditional mean estimator and the true conditional mean, which is a U-statistic of finite degree. A specific result is provided in his Lemma 2 for $S_{\hat{g}-g} = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}_i - g_i)^2 I(\hat{f}_i > b)$, where $\hat{g}_i = \frac{1}{(n-1)a^q}\sum_{j\neq i} K_{ij} g_j/\hat{f}_i$, $\hat{f}_i = \frac{1}{(n-1)a^q}\sum_{j\neq i} K_{ij}$ and $K_{ij} = K(\frac{Z_i - Z_j}{a})$. K is assumed to be a v^{th} order kernel function, a is a bandwidth, $I(\cdot)$ is an indicator function and $b \in \mathbb{R}$. We show below that the order of magnitude of $S_{\hat{g}-g}$ can be determined conveniently by applying our theorem.

Using the assumptions and notations in Li (1996) we have,

$$
\begin{aligned}
S_{\hat{g}-g} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\hat{f}_i(n-1)a^q}\sum_{j\neq i}K_{ij}(g_j-g_i)\right)^2 I(\hat{f}_i>b)\\
&\leq \frac{1}{n(n-1)^2a^{2q}b^2}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i}K_{ij}(g_j-g_i)K_{il}(g_l-g_i)\\
&= \frac{1}{n(n-1)^2}\sum\sum_{i\neq j}\underbrace{\frac{1}{a^{2q}b^2}K_{ij}^2(g_j-g_i)^2}_{\psi_{nij}}+\frac{1}{n(n-1)^2}\sum\sum\sum_{i\neq j\neq l}\underbrace{\frac{1}{a^{2q}b^2}K_{ij}K_{il}(g_j-g_i)(g_l-g_i)}_{\psi_{nijl}}\\
&= S_{1n}+S_{2n}.
\end{aligned}
$$

$nS_{1n}=\frac{1}{(n-1)^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i<j}}^{n}\underbrace{(\psi_{nij}+\psi_{nji})}_{\phi_{nij}}=\frac{1}{2}\left(O(n^{-3})+\begin{pmatrix} n\\ 2\end{pmatrix}^{-1}\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ i<j}}^{n}\phi_{nij}$ and $\begin{pmatrix} n\\ 2\end{pmatrix}^{-1}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i<j}}^{n}\phi_{nij}$ is a second degree U-statistic u_n. For an arbitrary constant c, $a^{3q}b^4\sigma_{2n}^2\leq ca^{3q}b^4E(\psi_{nij}^2)=\frac{c}{a^q}E\left(K_{ij}^4(g_j-g_i)^4\right)=O(1)$, thus $H_n^{(2)}=O_p((n^{-2}a^{-3q}b^{-4})^{\frac{1}{2}})$ and $H_n^{(1)}=O_p((n^{-1}a^{-2q}b^{-4})^{\frac{1}{2}})$. By his assumption A1 for $v\geq 1$, $|g(Z_i+a\psi)-g(Z_i)|\leq H_g(Z_i)a\psi$, where $H_g(\cdot)$ has finite fourth moment. Hence, $\theta_n=\frac{1}{a^qb^2}\int K^2(\psi)(g(Z_i+a\psi)-g(Z_i))^2f(Z_i)f(Z_i+a\psi)d\psi dZ_i=O(a^{-q+2}b^{-2})$. Combing the results, we have $nS_{1n}=O_p(a^{-q+2}b^{-2})$, and consequently $S_{1n}=O_p(n^{-1}a^{-q+2}b^{-2})$.

Now, $S_{2n}=\frac{1}{6}\left(O(n^{-4})+\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}\right)\sum\sum\sum_{i<j<l}(\underbrace{\psi_{nijl}+\psi_{nilj}+\psi_{njil}+\psi_{njli}+\psi_{nlij}+\psi_{nlji}}_{\phi_{nijl}})$ and we have that $\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}\sum\sum\sum_{i<j<l}\phi_{nijl}$ is a third degree U-statistic. $a^{2q}b^4\sigma_{3n}^2\leq ca^{2q}b^4E(\phi_{nijl}^2)=\frac{c}{a^{2q}}E(K_{ij}^2K_{il}^2(g_j-g_i)(g_l-g_i))=O(a^2)$. Hence, $\sigma_{3n}^2=O(a^{-2q+2}b^{-4})$ and $H_n^{(3)}=O_p((n^{-3}a^{-2q+2}b^{-4})^{\frac{1}{2}})$. Furthermore, we have $\sigma_{2n}^2\leq cE\left(E^2(\psi_{nijl}|Z_i,Z_j)+E^2(\psi_{nijl}|Z_i,Z_l)+E^2(\psi_{nijl}|Z_j,Z_l)\right)$ and given his assumption A1 we have $E(\psi_{nijl}|Z_i,Z_j)=\frac{K_{ij}}{a^qb^2}(g_j-g_i)\frac{1}{a^q}E(K_{il}(g_l-g_i)|Z_i)=\frac{K_{ij}}{a^qb^2}(g_j-g_i)O(a^v)$ uniformly over Z_i. Thus, $E(E^2(\psi_{nijl}|Z_i,Z_j))=O(a^{-q+2v+2}b^{-4})$. Similar arguments give $E(E^2(\psi_{nijl}|Z_i,Z_l))=O(a^{-q+2v+2}b^{-4})$ and $E(E^2(\psi_{nijl}|Z_i,Z_j))=O(a^{-q+4}b^{-4})$. Thus, $\sigma_{2n}^2=O(a^{-q+4}b^{-4})$ and $H_n^{(2)}=O_p((n^{-2}a^{-q+4}b^{-4})^{\frac{1}{2}})$. In addition, $\sigma_{1n}^2\leq cE(E^2(\psi_{nijl}|Z_i)+E^2(\psi_{nijl}|Z_j)+E^2(\psi_{nijl}|Z_l))$. $E(\psi_{nijl}|Z_i)=\frac{1}{b^2}\int K(\psi_1)K(\psi_2)(g(Z_i-a\psi_1)-g(Z_i))(g(Z_i-a\psi_2)-g(Z_i))f(Z_i-a\psi_1)f(Z_i-a\psi_2)d\psi_1d\psi_2=O(a^{2v}b^{-2})$ uniformly over Z_i. Consequently, $E(E^2(\psi_{nijl}|Z_i))=O(a^{4v}b^{-4})$. Similarly, $E(E^2(\psi_{nijl}|Z_j))$ and $E(E^2(\psi_{nijl}|Z_l)$ are of the same order. Thus, we have $H_n^{(1)}=O(n^{-\frac{1}{2}}a^{2v}b^{-2})$ and $\theta_n=cE\psi_{nijl}=O(a^{2v}b^{-2})$. In all,

$$S_{2n}=O_p(n^{-\frac{3}{2}}a^{-q+1}b^{-2})+O_p(n^{-1}a^{-\frac{q}{2}+2}b^{-2})+O(n^{-\frac{1}{2}}a^{2v}b^{-2})+O(a^{2v}b^{-2}).$$

Combining results for S_{1n} and S_{2n}, we conclude $S_{\hat{g}-g} = O_p(n^{-1}a^{-q+2}b^{-2}) + O_p(a^{2v}b^{-2})$ as in Li (1996).

As a second example, consider the nonparametric significance test proposed in Lavergne and Vuong (2000). The test statistic is similar to that in Fan and Li (1996), but it places less restrictive conditions on the smoothing parameters. The test is properly centered and has smaller bias in finite sample. The asymptotic distribution of the test is studied with eleven terms provided in the proof of their Theorem 1. Here, we focus exclusively on the term $I_{1,3}$ as all other terms are of similar nature. Following their assumptions and notations, we take g and h as bandwidths and $L(\cdot)$ and $K(\cdot)$ as kernel functions. $r_1(X_{1i}) = E(Y|X_{1i})$, $r_2(X_{2i}) = E(Y|X_{2i}) = r_1(X_{1i}) + \delta_n d(X_{2i})$ indicating the local alternative for $\delta_n \in [0,1]$. The regressors X_{1i}, X_{2i} are of dimensions p_1 and p_2 respectively, with $p_1 < p_2$ and the components of X_{1i} are a subset of the components of X_{2i}.

$$I_{1,3} = \frac{(n-3)!}{n!}\sum\sum_{i\neq j\neq l}\sum \underbrace{(Y_i - r_1(X_{1i}))f_1(X_{1i})(Y_l - r_1(X_{1l}))g^{-p_1}L\left(\frac{X_{1i}-X_{1l}}{g}\right)h^{-p_2}K\left(\frac{X_{2i}-X_{2j}}{h}\right)}_{\psi_{nijl}}$$

$$= \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1}\sum\sum_{i<j<l}\sum\left(\underbrace{\psi_{nijl} + \psi_{nilj} + \psi_{njil} + \psi_{njli} + \psi_{nlij} + \psi_{nlji}}_{\phi_{nijl}}\right),$$

which is a U-statistic of degree 3. We determine its order of magnitude using Theorem 1. First, $g^{p_1}h^{p_2}\sigma_{3n}^2 \leq cg^{p_1}h^{p_2}E\psi_{nijl}^2 = \frac{c}{g^{p_1}h^{p_2}}E(u_i^2u_l^2f_1^2(X_{1i})L^2(\frac{X_{1j}-X_{1l}}{g})K^2(\frac{X_{2j}-X_{2j}}{h})) = O(1)$ and consequently we have $\sigma_{3n}^2 = O(g^{-p_1}h^{-p_2})$ and $H_n^{(3)} = O((n^{-3}g^{-p_1}h^{-p_2})^{\frac{1}{2}})$.

Second, $\sigma_{2n}^2 \leq cE\left(E^2(\psi_{nijl}|Z_i,Z_j) + E^2(\psi_{nijl}|Z_i,Z_l) + E^2(\psi_{nijl}|Z_j,Z_l)\right)$. Since $E(u_l|X_{2l}) = 0$, we have $E(\psi_{nijl}|Z_i,Z_j) = 0$. $E(\psi_{nijl}|Z_i,Z_l) = u_if_iu_lE(L_{njl}K_{nij}|X_{2i},X_{1l}) = u_if_iu_lO(1)$ uniformly over X_{1l}, and X_{2i}. $E(\psi_{nijl}|Z_j,Z_l) = u_lL_{njl}E(u_if_iK_{nij}|X_{2j}) = \delta_nu_lL_{njl}E(d(X_{2i})f_iK_{nij}|X_{2j}) = \delta_nu_lL_{njl}O(1)$ uniformly over X_{2j}, where the second to last equality follows from $u_i = \delta_nd(X_{2i}) + Y_i - r_2(X_{2i})$ and $E(Y_i - r_2(X_{2i})|X_{2i}) = 0$. Consequently, $\sigma_{2n}^2 = O(\delta_n^2g^{-p_1}))$ and $H_n(2) = O_p((n^{-2}\delta_n^2g^{-p_1})^{\frac{1}{2}})$.

Third, $\sigma_{1n}^2 \leq cE(E^2(\psi_{nijl}|Z_i) + E^2(\psi_{nijl}|Z_j) + E^2(\psi_{nijl}|Z_l))$. $E(\psi_{nijl}|Z_i) = E(\psi_{nijl}|Z_j) = 0$ and $E(\psi_{nijl}|Z_l) = u_lE(u_if_1(X_{1i})L_{njl}K_{nij}|X_{1l}) = \delta_nu_lE(d(X_{2i})f_1(X_{1i})L_{njl}K_{nij}|X_{1l}) = \delta_nu_lO(1)$ uniformly over X_{1l}. Consequently, $\sigma_{1n}^2 = O(\delta_n^2)$ and $H_n^{(1)} = O_p((n^{-1}\delta_n^2)^{\frac{1}{2}})$.

Finally, given that $\theta_n = 0$ we have $I_{1,3} = O_p((n^{-1}\delta_n^2)^{\frac{1}{2}}) + O_p((n^{-2}\delta_n^2g^{-p_1})^{\frac{1}{2}}) + O_p((n^{-3}g^{-p_1}h^{-p_2})^{\frac{1}{2}})$.

Under the assumptions in Theorem 1 of Lavergne and Vuong (2000) we obtain $nh^{\frac{p2}{2}} I_{1,3} = \delta_n \sqrt{n} h^{\frac{p2}{2}} O_p(1) + o_p(1)$, which coincides with their Proposition 2.

3.2 A new test for overall significance of a nonparametric regression

Here, we use Theorem 1 to establish the asymptotic distribution of a new test for overall significance of a nonparametric regression. For simplicity of exposition we consider the standard univariate nonparametric regression model

$$y_t = m(x_t) + \epsilon_t, \text{ with } t = 1, 2, \cdots, n, \tag{3}$$

where $m(x_t) = E(y_t|x_t), E(\epsilon_t|x_t) = 0, V(\epsilon_t|x_t) = \sigma^2(x_t)$. Pearson's correlation ratio for this model is given by $\frac{V(m(x))}{V(y)} = 1 - \frac{E(y-m(x))^2}{V(y)}$ and can be interpreted as a nonparametric R^2 (Doksum and Samarov (1995)). Under the null hypothesis $H_0 : P(E(y_t|x_t) = \mu) = 1$ we have $R^2 = 0$ with alternative hypothesis given by $H_1 : P(E(y_t|x_t) = \mu) < 1$, where $\mu \in \Re$ is some constant. Thus, we define

$$\hat{R}^2 = 1 - \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2}{s_y^2} \tag{4}$$

where $s_y^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$, $\bar{y} = n^{-1}\sum_{t=1}^{n} y_t$ and $\hat{m}(x)$ is the local linear estimator (Stone (1977), Fan (1992)) for $m(x)$. Specifically, $\hat{m}(x) = \hat{\alpha}$ where $(\hat{\alpha}, \hat{\beta}) = argmin_{\alpha,\beta} \sum_{t=1}^{n}(y_t - \alpha - \beta(x_t - x))^2 K\left(\frac{x_t - x}{h_n}\right)$, $K(\cdot) : \mathbb{R} \to \mathbb{R}$ is a kernel function and $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth. Values of $\hat{R}^2$ in the vicinity of zero are an indication of poor model fit, i.e., an indication that x_t is not a regressor. The following lemma provides the asymptotic distribution of a suitably centered and normalized $\hat{R}^2$. We note that, as is common with these types of test statistics, a bias correction is needed. However, we do not explore this correction here as our purpose is simply to illustrate the use of Theorem 1.

Lemma 1. *Under H_0 and assumptions A1-A7 in the Appendix, we have that*

$$nh_n^{1/2}\left(\hat{R}^2 + \frac{A_{1n} + I_{3n}}{s_y^2}\right) \xrightarrow{d} N(0, E(\sigma^2(x_t))^{-2}V),$$

where $V = 2E\left(\frac{\sigma^4(x)}{f(x)}\right)\int(2K(u) - \int K(x)K(u+x)dx)^2 du$, $I_{3n} = \frac{1}{n^3 h_n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n} K^2(\frac{x_i - x_t}{h_n})\frac{\epsilon_i^2}{f^2(x_t)}$ *and* $A_{1n} = -\frac{2}{n^2 h_n}\sum_{t=1}^{n} K(0)\frac{\epsilon_t^2}{f(x_t)}$.

In essence, the proof of the Lemma (see Appendix) requires the analysis of the asymptotic behavior of s_y^2 and $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2$. Theorem 1 is used repeatedly. First, $\bar{y}^2$ is a second degree U-statistics whose kernel does not depend on n, thus a straightforward application of Theorem 1 gives $\bar{y}^2 = \mu^2 + \frac{2}{n}\sum_{t=1}^{n}\mu\epsilon_t + O_p(n^{-1})$. Hence, under H_0 we have $\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2 = \mu^2+\frac{2}{n}\sum_{t=1}^{n}\mu\epsilon_t+\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2-(\mu^2+\frac{2}{n}\sum_{t=1}^{n}\mu\epsilon_t+O_p(n^{-1})) = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2+O_p(n^{-1})$. Also, under H_0 we have $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + \frac{2}{n}\sum_{t=1}^{n}(\mu - \hat{m}(x_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\mu - \hat{m}(x_t))^2$. The second and third terms on the right side of the last equality can be represented as U-statistics of order up to three. We apply Theorem 1 repeatedly to determine their order of magnitude and obtain converge in distribution at the rate $nh_n^{\frac{1}{2}}$.

4 Summary and conclusion

We provide a simple result that permits the determination of the magnitude of a U-statistics of finite degree k with kernel that depends on n. The order of magnitude depends on the leading variance of the conditional expectation of the kernel function, which depends on n. Our result permits researchers to easily obtain the magnitude of nonparametric and semiparametric estimators and test statistics where high dimensional sums are involved.

Appendix

The proof of Theorem 1: Let $G_x(\cdot)$ be the distribution function of a single point mass at x, and let $F(\cdot)$ be the distribution of the random variable Z_i. As in Lee (1990) we have

$$h_n^{(j)}(z_1, \cdots, z_j) = \int \cdots \int \phi_n(u_1, \cdots, u_k)\prod_{i=1}^{j}(dG_{z_i}(u_i) - dF(u_i)) \prod_{i=j+1}^{k} dF(u_i).$$

Define $h_{c,n}^{(j)}(z_1, \cdots, z_c) = E(h_n^{(j)}(Z_1, \cdots, Z_c, Z_{c+1}, \cdots, Z_j)|Z_1 = z_1, \cdots, Z_c = z_c)$ and $\gamma_{cjn}^2 = Var(h_{c,n}^{(j)}(Z_1, \cdots, Z_c))$. Then, from Theorem 2 in section 1.6 in Lee (1990) we have $h_{j-1,n}^{(j)}(z_1, \cdots, z_{j-1}) = 0$ and for any $1 \leq c \leq j-1$, $E(h_{j-1,n}^{(j)}(z_1, \cdots, z_c, Z_{c+1}, \cdots, Z_{j-1})) = 0$. By the Law of Iterated Expectation, we have $E(h_{c,n}^{(j)}(Z_1, \cdots, Z_c)) = E(h_n^{(j)}(Z_1, \cdots, Z_j)) = 0$. Hence, we have $\gamma_{cjn}^2 = 0$ for all $1 \leq c \leq j-1$ and the only nonzero γ_{cjn}^2 is $\gamma_{jjn}^2 \equiv \gamma_{jn}^2 = Var(h_{j,n}^{(j)}(Z_1, \cdots, Z_j)) = Var(h_n^{(j)}(Z_1, \cdots, Z_j))$. Now, consider the covariance between $h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})$ and $h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j})$ with $c_1=0,\cdots,j$ variables in common. Then,

$$cov(h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j}),h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j}))=E\left(h_n^{(j)}(Z_1,\cdots,Z_j)h_n^{(j)}(Z_1,\cdots,Z_{c_1},Z_{j+1},\cdots Z_{2j-c_1})\right)=$$
$$\int\cdots\int h_n^{(j)}(z_1,\cdots,z_j)h_n^{(j)}(z_1,\cdots,z_{c_1},z_{j+1},\cdots z_{2j-c_1})\prod_{i=1}^{2j-c_1}dF(z_i)=$$
$$\int\cdots\int(\int\cdots\int h_n^{(j)}(z_1,\cdots,z_j)\prod_{i=c_1+1}^{j}dF(z_i)\times(\int\cdots\int h_n^{(j)}(z_1,\cdots,z_{c_1},z_{j+1},\cdots z_{2j-c_1})$$
$$\times\prod_{i=j+1}^{2j-c_1}dF(z_i))\prod_{i=1}^{c_1}dF(z_i)=\int\cdots\int(h_{c_1n}^{(j)}(z_1,\cdots,z_{c_1}))^2\prod_{i=1}^{c_1}dF(z_i)=E(h_{c_1n}^{(j)}(z_1,\cdots,z_{c_1}))^2$$
$$=V(h_{c_1n}^{(j)}(z_1,\cdots,z_{c_1}))=\gamma_{c_1jn}^2\neq 0 \text{ only when } c_1=j.$$

The total number of pairs of $h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})$ and $h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j})$ with c_1 elements in common such that $cov(h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j}),h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j}))=\gamma_{c_1jn}^2$ is $\binom{n}{j}\binom{j}{c_1}\binom{n-k}{k-c_1}$ because there are $\binom{n}{j}$ ways to choose the variables in $h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})$, and there are $\binom{j}{c_1}$ ways to choose c_1 variables among the j variables in $h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})$ so that they appear in $h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j})$. Also, the different variables in $h_n^{(j)}(Z_{u_1},\cdots,Z_{u_j})$ can be chosen in $\binom{n-k}{k-c_1}$ number of ways. Hence, $Var(H_n^{(j)})=\binom{n}{j}^{-1}\sum_{c_1=0}^{j}\binom{j}{c_1}\binom{n-k}{k-c_1}\gamma_{c_1jn}^2=\binom{n}{j}^{-1}\gamma_{jn}^2$ since $\gamma_{c_1jn}^2=0\ \forall\, 0\le c_1\le j-1$. Furthermore, for $j'\ge j$ we have

$$cov(h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j}),h_n^{(j')}(Z_{v_1},\cdots,Z_{v_j},Z_{v_{j+1}},\cdots,Z_{v_{j'}}))=$$
$$E\left(h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})E(h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j},Z_{v_{j+1}},\cdots,Z_{v'_j}|Z_{v_{j+1}}=z_{v_{j+1}},\cdots,Z_{v_{j'}}=z_{v_{j'}})\right)=0.$$

Thus,

$$\sigma_{cn}^2=Var(\phi_{cn}(Z_1,\cdots,Z_c))=Var\left(\sum_{c=1}^{j}\sum_{(c,j)}h_n^{(j)}(Z_{i_1},\cdots,Z_{i_j})+\theta_n\right)$$
$$=\sum_{c=1}^{j}Var\left(\sum_{(c,j)}h_n^{(j)}(Z_{i_1},\cdots,Z_{i_j})\right)=\sum_{c=1}^{j}\binom{c}{j}^2Var(H_{cn}^{(j)})=\sum_{j=1}^{c}\binom{c}{j}\gamma_{jn}^2.$$

Here, $H_{cn}^{(j)}=\binom{c}{j}^{-1}\sum_{(c,j)}h_n^{(j)}(Z_{i_1},\cdots,Z_{i_j})$ and by following similar arguments as above we obtain $Var(H_{cn}^{(j)})=\binom{c}{j}^{-1}\gamma_{jn}^2$.

Since

$$\sum_{c=1}^{d}\left(\begin{array}{c} d \\ c \end{array}\right)(-1)^{d-c}\sigma_{cn}^2 = \sum_{c=1}^{d}\sum_{j=1}^{c}\left(\begin{array}{c} c \\ j \end{array}\right)\left(\begin{array}{c} d \\ c \end{array}\right)(-1)^{d-c}\gamma_{jn}^2 = \sum_{j=1}^{d}\sum_{c=j}^{d}\left(\begin{array}{c} c \\ j \end{array}\right)\left(\begin{array}{c} d \\ c \end{array}\right)(-1)^{d-c}\gamma_{jn}^2$$

and with $c' = c - j$,

$$= \sum_{j=1}^{d}\sum_{c'=0}^{d-j}\left(\begin{array}{c} c'+j \\ j \end{array}\right)\left(\begin{array}{c} d \\ c'+j \end{array}\right)(-1)^{d-c'-j}\gamma_{jn}^2$$

$$= \sum_{j=1}^{d}\sum_{c'=0}^{d-j}\left(\begin{array}{c} d-j \\ c' \end{array}\right)\left(\begin{array}{c} d \\ j \end{array}\right)(-1)^{d-c'-j}\gamma_{jn}^2$$

given $\left(\begin{array}{c} c'+j \\ j \end{array}\right)\left(\begin{array}{c} d \\ c'+j \end{array}\right) = \left(\begin{array}{c} d-j \\ c' \end{array}\right)\left(\begin{array}{c} d \\ j \end{array}\right)$

$$= \sum_{j=1}^{d}(\sum_{c'=0}^{d-j}\left(\begin{array}{c} d-j \\ c' \end{array}\right)(-1)^{-c'})(-1)^{d-j}\left(\begin{array}{c} d \\ j \end{array}\right)\gamma_{jn}^2$$

$$= \gamma_{dn}^2,$$

where the last equality follows from $\sum_{c'=0}^{d-j}\left(\begin{array}{c} d-j \\ c' \end{array}\right)(-1)^{-c'} = \sum_{c'=0}^{d-j}\left(\begin{array}{c} d-j \\ c' \end{array}\right)(-1)^{c'} = 0$ except when $j = d$. Hence, we obtain result (a) that $Var(H_n^{(j)}) = \left(\begin{array}{c} n \\ j \end{array}\right)^{-1}\gamma_{jn}^2 = \left(\begin{array}{c} n \\ j \end{array}\right)^{-1}\sum_{c=1}^{j}\left(\begin{array}{c} j \\ c \end{array}\right)(-1)^{j-c}\sigma_{cn}^2$. Second, since $\sigma_{cn}^2 = Var(\phi_{cn}(Z_1,\cdots,Z_c)) = \sum_{j=1}^{c}\left(\begin{array}{c} c \\ j \end{array}\right)\gamma_{jn}^2 \geq 0$, for $c \leq c'$, we have $c\sigma_{c'n}^2 - c'\sigma_{cn}^2 = c\sum_{j=1}^{c'}\left(\begin{array}{c} c' \\ j \end{array}\right)\gamma_{jn}^2 - c'\sum_{j=1}^{c}\left(\begin{array}{c} c \\ j \end{array}\right)\gamma_{jn}^2 = \sum_{j=1}^{c}(c\left(\begin{array}{c} c' \\ j \end{array}\right) - c'\left(\begin{array}{c} c \\ j \end{array}\right))\gamma_{jn}^2 + \sum_{j=c+1}^{c'} c\left(\begin{array}{c} c' \\ j \end{array}\right)\gamma_{jn}^2 \geq 0$, since $c\left(\begin{array}{c} c' \\ j \end{array}\right) - c'\left(\begin{array}{c} c \\ j \end{array}\right) \geq 0$ for $c' \geq c \geq j \geq 1$, which verifies result (b).

The proof of Lemma1 depends on the following assumptions:

A1. $\{x_t, y_t\}_{t=1}^n$ is an independently and identically distributed sequence.

A2. $E(\epsilon_t|x_t) = 0$, $V(\epsilon_t|x_t) = \sigma^2(x_t) > 0$, $\sigma^2(x)$ is continuous at x and $E\left((\sigma^2(x_t))^2\right) < \infty$.

A3. Denote the marginal density of x_t by f. We have: (1) $0 < \underline{B}_f \leq f(x) \leq \bar{B}_f < \infty$ for all $x \in G$, G a compact subset of $\mathbb{R}$; (2) for all $x, x' \in G$, $|f(x) - f(x')| < m_f|x - x'|$ for some $0 < m_f < \infty$; (3) $f(x)$ is uniformly continuous in G.

A4. $0 < \underline{B}_m \leq m(x) \leq \bar{B}_m < \infty$ for all $x \in G$, where $m(x) : \mathbb{R} \to \mathbb{R}$ is a measurable twice continuously differentiable function in $\mathbb{R}$, $|m^{(2)}(x)| < \bar{B}_{2m} < \infty$ for all $x \in G$.

A5. $nh_n^3 \to \infty$.

A6. $K(\cdot) : S \to \Re$ is a symmetric density function with bounded support $S \subset \mathbb{R}$ such that $\int xK(x)dx = 0$, $\int x^2K(x)dx = \sigma_K^2$, $|K(x)| < B_k < \infty$ for all $x \in \mathbb{R}$ and $|u^jK(u) - v^jK(v)| \leq c_k|u - v|$, for $j = 0, 1, 2, 3$.

A7. $E(\epsilon_t^4|x_t) < \infty$, $f_{x_t|\epsilon_t}(x) < \infty$, $f(x, \epsilon)$ is continuous around x.

The proof of Lemma 1: Given A1 and A2 we have (1) $\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + O_p(n^{-1})$. Under H_0 $\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + \frac{2}{n}\sum_{t=1}^{n}(\mu-\hat{m}(x_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\mu-\hat{m}(x_t))^2$. Letting $e'=(1,0)$ we follow Fan (1992) and write $\frac{2}{n}\sum_{t=1}^{n}(\mu-\hat{m}(x_t))\epsilon_t = -\frac{2}{n^2h_n}\sum_{t=1}^{n}\sum_{i=1}^{n}e'S_n^{-1}(x_t)K(\frac{x_i-x_t}{h_n})\epsilon_i\epsilon_t\begin{pmatrix}1\\ \frac{x_i-x_t}{h_n}\end{pmatrix}$ where $S_n(x)=\begin{pmatrix}S_{0n}(x) & S_{1n}(x)\\ S_{1n}(x) & S_{2n}(x)\end{pmatrix}$ where $S_{jn}(x_t)=\frac{1}{nh_n}\sum_{i=1}^{n}K(\frac{x_i-x_t}{h_n})(\frac{x_i-x_t}{h_n})^j$, $j=0,1,2$. Under A5 and A6 we can use Lemma 1 in Martins-Filho and Yao (2007) to obtain $\sup_{x_t\in G}|S_{jn}(x_t)-ES_{jn}(x_t)| = O_{a.s.}((\frac{nh_n}{lnn})^{-\frac{1}{2}})$. In addition, under A3, $E(S_{jn}(x_t))=\int K(\psi)\psi^j f(x_t+h_n\psi)d\psi \to f(x_t)\int K(\psi)\psi^j d\psi$ uniformly over $x_t\in G$. Hence, $\sup_{x_t\in G}|S_{jn}^{-1}(x_t)-S_j^{-1}(x_t)| = o_{a.s.}(1)$ and

$$\begin{aligned}
\frac{2}{n}\sum_{t=1}^{n}(\mu-\hat{m}(x_t))\epsilon_t &= -\frac{2}{n^2h_n}\sum_{t=1}^{n}\sum_{i=1}^{n}e'\left(S_n^{-1}(x_t)-S^{-1}(x_t)+S^{-1}(x_t)\right)K(\frac{x_i-x_t}{h_n})\epsilon_i\epsilon_t\begin{pmatrix}1\\ \frac{x_i-x_t}{h_n}\end{pmatrix}\\
&= \left(-\frac{2}{n^2h_n}\sum_{t=1}^{n}\frac{1}{f(x_t)}K(0)\epsilon_t^2 - \frac{2}{n^2h_n}\sum_{\substack{t=1\\ t\neq i}}^{n}\sum_{i=1}^{n}\frac{1}{f(x_t)}K(\frac{x_i-x_t}{h_n})\epsilon_i\epsilon_t\right)(1+o_{a.s.}(1))\\
&= (A_{1n}+A_{2n})(1+o_{a.s.}(1)).
\end{aligned}$$

Using similar arguments we have $\frac{1}{n}\sum_{t=1}^{n}(\mu-\hat{m}(x_t))^2 = I_n(1+o_{a.s.}(1))$ where

$$I_n = \frac{1}{n^3h_n^2}\sum_{t=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f^2(x_t)}K\left(\frac{x_i-x_t}{h_n}\right)K\left(\frac{x_j-x_t}{h_n}\right)\epsilon_i\epsilon_j.$$

The magnitude of I_n is obtained by considering the following cases: (a) when $t=i=j$ we have the corresponding terms in I_n being $I_{1n}=\frac{1}{n^3h_n^2}\sum_{t=1}^{n}\frac{1}{f^2(x_t)}K^2(0)\epsilon_t^2$, which by A2, A3 and A5, we have $I_{1n} = O_p((nh_n)^{-2}) = o_p(n^{-1})$; (b) when $t=i$ (or $t=j$) we have the corresponding terms in I_n being

$$I_{2n} = \frac{2}{n^3h_n^{\frac{3}{2}}}\sum_{\substack{t=1\\ t\neq j}}^{n}\sum_{j=1}^{n}\frac{1}{2\sqrt{h_n}f^2(x_t)}K(0)K(\frac{x_j-x_t}{h_n})\epsilon_t\epsilon_j = \frac{2}{n^3h_n^{\frac{3}{2}}}\sum_{\substack{t=1\\ t\neq j}}^{n}\sum_{j=1}^{n}\psi_n(Z_t,Z_j)$$

where $Z_t=(x_t,\epsilon_t)$. Then,

$$nh_n^{\frac{3}{2}}I_{2n} = \frac{1}{n^2}\sum_{\substack{t=1\\ t\neq j}}^{n}\sum_{j=1}^{n}(\psi_n(Z_t,Z_j)+\psi_n(Z_j,Z_t)) = \frac{1}{n^2}\sum_{\substack{t=1\\ t\neq j}}^{n}\sum_{j=1}^{n}\phi_n(Z_t,Z_j) = \frac{2}{n^2}\sum_{\substack{t=1\\ t<j}}^{n}\sum_{j=1}^{n}\phi_n(Z_t,Z_j),$$

which is U-statistic of degree 2. Its magnitude is easily obtained applying Theorem 1. We have $H_n^{(1)}=0$, $\theta_n=0$ as $E(\phi_n(Z_t,Z_j)|Z_t)=0$ and $E(\phi_n(Z_t,Z_j))=0$. Hence, by A2 and A7 we have that $nh_n^{\frac{3}{2}}I_{2n} = O_p((n^{-2}\sigma_{2n}^2)^{\frac{1}{2}}) = O_p\left(n^{-1}\left(E\left(\phi_n^2(Z_t,Z_j)\right)\right)^{\frac{1}{2}}\right)$ and

$$\begin{aligned}
E\left(\phi_n^2(Z_t,Z_j)\right) \le cE\left(\psi_n^2(Z_t,Z_j)\right) &= \frac{c}{4h_n}K^2(0)E\left(\frac{1}{f^4(x_t)}K^2(\frac{x_j-x_t}{h_n})\times\epsilon_t^2\epsilon_j^2\right)\\
&\to \frac{c}{4}K^2(0)E\left(\frac{1}{f^3(x_t)}\sigma^4(x_t)\right)\int K^2(\psi)d\psi < \infty.
\end{aligned}$$

Thus, $nh_n^{\frac{3}{2}} I_{2n} = O_p(n^{-1})$ and $I_{2n} = o_p(n^{-1})$; (c) when $i = j$ but $t \neq i$, we have the corresponding terms in I_n being $I_{3n} = \frac{1}{n^3 h_n^2} \sum_{\substack{t=1 \\ t\neq i}}^{n} \sum_{i=1}^{n} \frac{1}{f^2(x_t)} K^2(\frac{x_i - x_t}{h_n}) \epsilon_i^2$; (d) when t, i and j are distinct we put $\psi_n(Z_t, Z_i, Z_j) = \frac{1}{h_n^2} \frac{1}{f^2(x_t)} K(\frac{x_i - x_t}{h_n}) K(\frac{x_j - x_t}{h_n}) \epsilon_i \epsilon_j$ and the corresponding terms in I_n are

$$
\begin{aligned}
I_{4n} &= \frac{1}{n^3} \sum_{\substack{t=1 \\ t\neq i\neq j}}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_n(Z_t, Z_i, Z_j) = \frac{1}{3n^3} \sum_{\substack{t=1 \\ t\neq i\neq j}}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{(\psi_n(Z_t, Z_i, Z_j) + \psi_n(Z_i, Z_t, Z_j) + \psi_n(Z_j, Z_i, Z_t))}_{\phi_n(Z_t, Z_i, Z_j)} \\
&= \frac{1}{n^3} \sum\sum\sum_{1=t<i<j=n} 2\phi_n(Z_t, Z_i, Z_j), \text{ where } \phi_n(Z_t, Z_i, Z_j) \text{ is symmetric.} \\
&= \frac{1}{3} \left(\frac{6}{n^3} - \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} + \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} \right) \sum\sum\sum_{1=t<i<j=n} \phi_n(Z_t, Z_i, Z_j).
\end{aligned}
$$

Note that $\frac{6}{n^3} - \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} = O(n^{-4})$ and $u_n = \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} \sum\sum\sum_{1=t<i<j=n} \phi_n(Z_t, Z_i, Z_j)$ is a U-statistic of degree three. Again, its magnitude can be easily obtained from Theorem 1. In this case $\theta_n = E(\phi_n(Z_t, Z_i, Z_j)) = H_n^{(1)} = 0$ since the conditional expectation of $\phi_n(Z_t, Z_i, Z_j)$ is zero conditioning on Z_t, Z_i, or Z_j. Now,

$$
\begin{aligned}
E(\phi_n(Z_t, Z_i, Z_j)|Z_t, Z_i) &= E(\psi_n(Z_t, Z_i, Z_j) + \psi_n(Z_i, Z_t, Z_j) + \psi_n(Z_j, Z_i, Z_t)|Z_t, Z_i) \\
&= \frac{\epsilon_i \epsilon_t}{h_n} E\left(\frac{1}{h_n f^2(x_j)} K\left(\frac{x_i - x_j}{h_n}\right) K\left(\frac{x_t - x_j}{h_n}\right) |Z_t, Z_i \right) = \phi_{2n}(Z_t, Z_i),
\end{aligned}
$$

and $u_n = \frac{6}{n(n-1)} \sum_{\substack{t=1 \\ t<i}}^{n} \sum_{i=1}^{n} \phi_{2n}(Z_t, Z_i) + O_p(H_n^{(3)})$, where $Var(H_n^{(3)}) = O(n^{-3}(\sigma_{1n}^2 + \sigma_{2n}^2 + \sigma_{3n}^2))$. Now, $\sigma_{1n}^2 = Var(E(\phi_n(Z_t, Z_i, Z_j)|Z_t)) = 0$ and $\sigma_{2n}^2 = Var(E(\phi_n(Z_t, Z_i, Z_j)|Z_t, Z_i)) \leq E(E(\phi_n(Z_t, Z_i, Z_j)|Z_t, Z_i))^2 \leq E(\phi_n^2(Z_t, Z_i, Z_j)) = \sigma_{3n}^2 \leq 3cE(\psi_n^2(Z_t, Z_i, Z_j))$. Now, under assumptions A3 and A7

$$
\begin{aligned}
h_n^2 E(\psi_n^2(Z_t, Z_i, Z_j)) &= \frac{1}{h_n^2} E\left(\frac{1}{f^4(x_t)} K^2(\frac{x_i - x_t}{h_n}) K^2(\frac{x_j - x_t}{h_n}) \sigma^2(x_i) \sigma^2(x_j) \right) \\
&\to \left(\int K^2(\psi) d\psi \right)^2 E\left(\frac{\sigma^4(x_t)}{f^2(x_t)} \right) < \infty
\end{aligned}
$$

Hence, $Var(H_n^{(3)}) = O(n^{-3} h_n^{-2}) = o(n^{-2})$, and thus $I_{4n} = \frac{1}{n(n-1)} \sum_{\substack{t=1 \\ t<i}}^{n} \sum_{i=1}^{n} (\phi_{2n}(Z_t, Z_i) + \phi_{2n}(Z_i, Z_t)) + o(n^{-1})$.

Therefore, we can conclude that

$$
\frac{1}{n} \sum_{t=1}^{n} (\mu - \hat{m}(x_t))^2 - I_{3n}(1 + o_{a.s.}(1)) = \frac{1}{n(n-1)} \sum_{\substack{t=1 \\ t<i}}^{n} \sum_{i=1}^{n} (\phi_{2n}(Z_t, Z_i) + \phi_{2n}(Z_i, Z_t)) + o(n^{-1}))(1 + o_{a.s.}(1))
$$

$$
\begin{aligned}
&= \frac{1}{n(n-1)} \sum_{\substack{t=1 \\ t<i}}^{n} \sum_{i=1}^{n} [\frac{\epsilon_i \epsilon_t}{h_n} E(\frac{1}{h_n f^2(x_j)} K(\frac{x_i - x_j}{h_n}) K(\frac{x_t - x_j}{h_n}) |Z_t, Z_i) \\
&+ \frac{\epsilon_t \epsilon_i}{h_n} E(\frac{1}{h_n f^2(x_j)} K(\frac{x_t - x_j}{h_n}) K(\frac{x_i - x_j}{h_n}) |Z_i, Z_t)](1 + o_{a.s.}(1)) + o_p(n^{-1}) \\
&= A_{3n}(1 + o_{a.s.}(1)) + o_p(n^{-1}),
\end{aligned}
$$

and consequently we can write

$$\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(x_t))^2-(I_{3n}+A_{1n})(1+o_{a.s.}(1))=\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2+(A_{2n}+A_{3n})(1+o_{a.s.}(1))+o_p(n^{-1}).$$

The terms A_n and A_{1n} are bias terms and we focus on A_{2n} and A_{3n} to determine the asymptotic distribution. Note that

$$\begin{aligned}
A_{3n} &= \left(\frac{1}{n(n-1)}-\frac{1}{n^2}+\frac{1}{n^2}\right)\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}(\phi_{2n}(Z_t,Z_i)+\phi_{2n}(Z_i,Z_t))\\
&= \left(O(n^{-3})+\frac{1}{n^2}\right)\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}(\phi_{2n}(Z_t,Z_i)+\phi_{2n}(Z_i,Z_t))\\
&= \frac{1}{n^2}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}[\phi_{2n}(Z_t,Z_i)+\phi_{2n}(Z_i,Z_t)]+o_p(n^{-1})=A_{31n}+o_p(n^{-1}).
\end{aligned}$$

Since $A_{31n}=\frac{1}{n^2}\sum\limits_{\substack{t=1\\t<i}}^{n}\sum\limits_{i=1}^{n}(\phi_{2n}(Z_t,Z_i)+\phi_{2n}(Z_i,Z_t))$ is a U-statistic of degree 2 we have using Theorem 1 and given A2 that $A_{31n}=O_p(n^{-1}(E(\phi_{2n}^2(Z_t,Z_i)))^{\frac{1}{2}})$. Furthermore,

$$h_nE(\phi_{2n}^2(Z_t,Z_i))\to\int\kappa^2(\psi_1)d\psi_1E\left(\frac{\sigma^4(x_t)}{f(x_t)}\right)<\infty$$

where $\kappa(x)=\int K(u)K(u+x)du$, hence $(E(\phi_{2n}^2(Z_t,Z_i)))^{\frac{1}{2}}=O(h_n^{-\frac{1}{2}})$ and $A_{31n}=O_p((n^2h_n)^{-\frac{1}{2}})$. Now,

$$\begin{aligned}
A_{2n}+A_{31n} &= \frac{1}{n^2h_n}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}\Big[(-2)\frac{1}{f(x_t)}K(\frac{x_i-x_t}{h_n})\epsilon_i\epsilon_t-2\frac{1}{f(x_i)}K(\frac{x_t-x_i}{h_n})\epsilon_t\epsilon_i\\
&+ \epsilon_i\epsilon_tE(\frac{1}{h_nf^2(x_j)}K(\frac{x_i-x_j}{h_n})K(\frac{x_t-x_j}{h_n})|Z_t,Z_i)\\
&+ \epsilon_t\epsilon_iE(\frac{1}{h_nf^2(x_j)}K(\frac{x_t-x_j}{h_n})K(\frac{x_i-x_j}{h_n})|Z_i,Z_t)\Big]\\
&= \frac{1}{n^2h_n}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}[\psi_n(Z_t,Z_i)+\psi_n(Z_i,Z_t)+\psi_n'(Z_t,Z_i)+\psi_n'(Z_i,Z_t)]\\
&= \frac{1}{n^2h_n}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}\phi_n(Z_t,Z_i).
\end{aligned}$$

Since $\phi_n(Z_t,Z_i)$ is symmetric and $E(\phi_n(Z_t,Z_i)|Z_t)=0$ we have that $A_{2n}+A_{31n}$ is a degenerate U-statistic of degree 2. It is easy to show that: (i) $\frac{1}{h_n}E\left(\phi_n^2(Z_t,Z_i)\right)\to 2(2E\left(\frac{\sigma^4(x_t)}{f(x_t)}\right)(4\int K^2(\psi)d\psi+\int\kappa^2(\psi)d\psi-4\int K(\psi)\kappa(\psi)d\psi))=2V$; (ii) For $G_n(Z_1,Z_2)=E(\phi_n(Z_t,Z_1)\phi_n(Z_t,Z_2)|Z_1,Z_2)$, we have $E(G_n^2(Z_1,Z_2))=O(h_n^3)$; (iii) $E\phi_n^4(Z_t,Z_i)=O(h_n)$. From (i)-(iii), we have $\frac{E(G_n^2(Z_1,Z_2))}{(E(\phi_n^2(Z_t,Z_i)))^2}=\frac{O(h_n^3)}{O(h_n^2)}\to 0$, and $\frac{\frac{1}{n}E(\phi_n^4(Z_t,Z_i))}{(E(\phi_n^2(Z_t,Z_i)))^2}=\frac{\frac{1}{n}O(h_n)}{O(h_n^2)}\to 0$. Hence, by the central limit theorem in Hall (1984) $nh_n^{\frac{1}{2}}(A_{2n}+A_{31n})\xrightarrow{d}N(0,V)$ and given that $s_y^2\xrightarrow{p}E(\sigma^2(x_t))>0$ we have $nh_n^{\frac{1}{2}}\left(\hat{R}^2+(s_y^2)^{-1}((I_{3n}+A_{1n})(1+o_{a.s.}(1)))\right)\xrightarrow{d}N(0,(E(\sigma^2(x_t)))^{-2}V)$.

References

de Jong, P., 1987. A central limit theorem for generalized quadratic forms. Probability Thoery and Related Fields 75, 261–277.

Doksum, K., Samarov, A., 1995. Nonparametric estimation of global functionals and a meaure of the explanatory power of covariates in regression. Annals of Statistics 23, 1443–1473.

Dynkin, E. B., Mandelbaum, A., 1983. Symmetric statistics, poisson point processes and multiple wiener integrals. Annals of Statistics, 739–745.

Efron, B., Stein, C., 1981. A jackknife estimate of variance. Annals of Statistics 9, 586–596.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Satistical Association 87, 998–1004.

Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistic of absolutely regular processes with applications to model specification tests. Journal of Nonparametric Statistics 10, 245–271.

Friedrich, K., 1989. A berry-esseen bound for functions of independent variables. The Annals of Statistics 17, 170–183.

Gu, J., Li, D., Liu, D., Aug. 2007. Bootstrap non-parametric significance test. Journal of Nonparametric Statistics 19, 215–230.

Hall, P., 1984. Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. Journal of Multivariate Analysis 14, 1–16.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics 19, 293–325.

Hoeffding, W., 1961. The strong law of large numbers for U-Statistics.

Imbens, G., Ridder, G., 2009. Estimation and Inference for generalized full and partial means and average derivatives.
URL `http://www.economics.harvard.edu/faculty/imbens/files/ir.pdf`

Karlin, S., Rinott, Y., 1982. Applications of anova type decompositions for comparisons of conditional variance statistics including jackknife estimates. Annals of Statistics 10, 485–501.

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. Econometric Theory 16, 576–601.

Lee, A. J., 1990. U-Statistics. Marcel Dekker, New York.

Lee, B., 1988. Nonparametric tests using a kernel estimation method. Doctoral dissertation, University of Wisconsin, Madison.

Lenth, R., 1983. Some properties of u statistics. The American Statistician 37, 311–313.

Li, Q., 1996. On the root-n-consistent semiparametric estimation of partially linear models. Economics Letters 51, 277–285.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101–147.

Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

Mandelbaum, A., Taqqu, M., 1984. Invariance principle for symmetric statistics. Annals of Statistics 12, 483–496.

Martins-Filho, C., Yao, F., 2006. A Note on the use of V and U statistics in nonparametric models of regression. Annals of the Institute of Statistical Mathematics 58, 389–406.

Martins-Filho, C., Yao, F., 2007. Nonparametric frontier estimation via local linear regression. Journal of Econometrics 141, 283–319.

Newey, W. K., 1994. Kernel estimation of partial means and a general variance estimator. Econometric Theory 10, 233–253.

Pecatti, G., 2004. Hoeffding-anova decompositions for symmetric statistics of exchangeable observations. Annals of Statistics 32, 1796–1829.

Powell, J. L., Stock, J. H., Stoker, T. M., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.

Rao Jammalamadaka, S., Janson, S., 1986. Limit theorems for a triangular scheme of u-statistics with applications to inter-point distances. Annals of Probability 14, 1347–1358.

Rubin, H., Vitale, A., 1980. Asymptotic distribution of symmetric statistics. Annals of Statistics 8, 165–170.

Stone, C. J., 1977. Consistent nonparametric regression. The Annals of Statistics 5, 595–620.

Su, L., Ullah, A., 2012. A nonparametric goodness of fit based test for conditional heterocedasticity. Working Paper, Singapore Management University. Econometric Theory forthcoming.

Takemura, A., 1983. Tensor analysis of anova decomposition. Journal of the American Statistical Association 78, 894–900.

Tsybakov, A. B., 2009. Introduction to nonparametric estimation. Springer-Verlag, New York, NY.

van Zwet, W. R., 1984. A berry-esseen bound for symmetric statistics. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 66, 425–440.

Vitale, R., 1992. Covariances of symmetric statistics. Journal of Multivariate Analysis 41, 14–26.

Weber, N. C., 1980. Rates of convergence for u-statistics with varying kernels. Bulletin of the Australian Mathematical Society 21, 1–5.

Weber, N. C., 1983. Central limit theorems for a class of symmetric statistics. Mathematical Proceedings of the Cambridge Philosophical Society 94, 307–313.

Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263–289.