AISM (2006) 58: 389–406
DOI 10.1007/s10463-006-0034-z

Carlos Martins-Filho · Feng Yao

A note on the use of V and U statistics in nonparametric models of regression

Received: 8 September 2004 / Revised: 12 April 2005 / Published online: 1 June 2006

Abstract We establish the $\sqrt{n}$ asymptotic equivalence of V and U statistics when the statistic's kernel depends on n. Combined with a lemma of B. Lee this result provides conditions under which U statistics projections and V statistics are $\sqrt{n}$ asymptotically equivalent. The use of this equivalence in nonparametric regression models is illustrated with several examples; the estimation of conditional variances, skewness, kurtosis and the construction of a nonparametric R-squared measure.

Keywords U statistics · V statistics · local linear estimation

1 Introduction

The use of U statistics (see Hoeffding 1948), as a means of obtaining asymptotic properties of kernel based estimators in semiparametric and nonparametric regression models is now common practice in econometrics. For example, Powell, Stock and Stoker (1989) use U statistics to obtain the asymptotic properties of their semiparametric index model estimator. Ahn and Powell (1993) use U statistics to investigate the properties of a two stage semiparametric estimation of censored selection models where the first stage estimator involves a nonparametric regression estimator for the selection variable. Fan and Li (1996) use U statistics to study

C. Martins-Filho (✉)
Department of Economics, Oregon State University,
Ballard Hall 303, Corvallis, OR 97331-3612 USA
E-mail: carlos.martins@orst.edu
Tel.:+1-541-737-1476

F. Yao
Department of Economics, University of North Dakota,
P.O. Box 8369, Grand Forks, ND 58203 USA
E-mail: feng.yao@mail.business.und.edu
Tel.: +1-701-7773360

a set of specification tests for nonparametric regression models and Zheng (1998) uses U-statistics to provide a nonparametric kernel based test for parametric quantile regression models. See also Kemp (2000) and D'Amico (2003) for more recent uses.

The appeal for the use of U statistics derives from the fact that often estimators or statistics of interest – T_n – are expressed as linear combinations of nonparametric regression estimators. Consider for example a nonparametric regression model $E(Y|X = x) = m(x)$ with observations $\{(y_i, x_i)\}_{i=1}^n$, $T_n = \sum_{i=1}^n c_i \hat{m}(x_i)$ where $c_i \in \Re$ are nonstochastic and $\hat{m}(x)$ is an arbitrary nonparametric estimator for $m(x)$. Since $\hat{m}(x)$ can usually be written as $\hat{m}(x) = \sum_{j=1}^n w_{jn}(x) y_j$, we can write

$$T_n = \sum_{i=1}^n c_i w_{in}(x_i) y_i + \binom{n}{2} u_n = T_{1n} + T_{2n},$$

where $u_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \psi_n(Z_i, Z_j)$ is a U statistic with kernel $\psi_n(Z_i, Z_j) = c_i w_{jn}(x_i) y_j + c_j w_{in}(x_j) y_i$ and $Z_i = (x_i, y_i)$. The asymptotic characterization of T_n therefore depends on two terms. T_{1n} is normally handled by a suitable central limit theorem or law of large numbers. T_{2n}, the U statistic component, is studied using Hoeffding's (1961) H-decomposition which breaks it into an average of independent and identically distributed terms (projection) and a remainder term that is orthogonal to the space in which the projection lies and is of order smaller than that of the projection (see, e.g., Lee, 1990; Serfling, 1980).

In this note we show that a convenient approach to establishing the asymptotic properties of T_n is to study it directly via *von Mises'* V statistics. To this end we show that under suitable conditions V statistics are $\sqrt{n}$ asymptotically equivalent to U statistics. Combined with the use of Hoeffding's H-decomposition our results establish the $\sqrt{n}$ asymptotic equivalence of V statistics and the corresponding U statistic projection. In contrast with Serfling (1980) our results allow the statistics' kernel to depend on n. The remainder of the paper is structured as follows. We introduce the asymptotic equivalence result of V and U statistics in Sect. 2. Applications to the estimation of conditional variance, skewness and kurtosis in nonparametric regression models and constructing nonparametric R-square are provided in Sect. 3. A brief conclusion is given in Sect. 4.

2 Asymptotic Equivalence of U and V statistics

Let $\{Z_i\}_{i=1}^n$ be a sequence of independent and identically distributed (i.i.d.) random variables and $\psi_n(Z_1, \ldots, Z_k)$ be a symmetric function with $k \le n$. We call $\psi_n(Z_1, \ldots, Z_k)$ a kernel function and a k-dimensional U statistic will be denoted by u_n which is defined as,

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

where $\sum_{(n,k)}$ denotes a sum over all subsets $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of $\{1, 2, \ldots, n\}$. A k-dimensional V statistic, denoted by v_n, is defined as

$$v_n = n^{-k} \sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} \psi_n(Z_{i_1}, \ldots, Z_{i_k}). \tag{2}$$

The following Theorem establishes the $\sqrt{n}$ asymptotic equivalence of U and V statistics under suitable conditions.

Theorem 1 *Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and u_n and v_n be U and V statistics with kernel function $\psi_n(Z_1, \ldots, Z_k)$. If $E\left(\psi_n^2(Z_{i_1}, \ldots, Z_{i_k})\right) = o(n)$ for all $1 \leq i_1, \ldots, i_k \leq n$, $k \leq n$, then $u_n - v_n = o_p(n^{-1/2})$.*

Proof From Bönner and Kirschner (1977) we have

$$v_n = \frac{1}{n^k} \sum_{\kappa=1}^{k} \sum_{(n,\kappa)} \sum_{\{\nu_1,\ldots,\nu_\kappa,k\}} \frac{k!}{\prod_{j=1}^{\kappa} \nu_j!} \psi_n(\underbrace{Z_{i_1}, \ldots, Z_{i_1}}_{\nu_1-\text{ times}}, \ldots, \underbrace{Z_{i_\kappa}, \ldots, Z_{i_\kappa}}_{\nu_\kappa-\text{ times}}),$$

where $\sum_{\{\nu_1,\ldots,\nu_\kappa,k\}}$ denotes the sum over all $\nu_1, \ldots, \nu_\kappa$ such that $k = \sum_{j=1}^{\kappa} \nu_j$, $\nu_j \geq 1$, $\kappa \in \{1, \ldots, k\}$. Now let $\phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa}) = \sum_{\{\nu_1,\ldots,\nu_\kappa,k\}} \frac{k!}{\prod_{j=1}^{\kappa} \nu_j!} \psi_n (\underbrace{Z_{i_1}, \ldots, Z_{i_1}}_{\nu_1-\text{ times}}, \ldots, \underbrace{Z_{i_\kappa}, \ldots, Z_{i_\kappa}}_{\nu_\kappa-\text{ times}})$ so that

$$v_n = \frac{1}{n^k} \sum_{(n,k)} \phi_{k,n}(Z_{i_1}, \ldots, Z_{i_k}) + \frac{1}{n^k} \sum_{\kappa=1}^{k-1} \sum_{(n,\kappa)} \phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa}). \tag{3}$$

Note that

$$\begin{aligned} \sum_{(n,k)} \phi_{k,n}(Z_{i_1}, \ldots, Z_{i_k}) &= k! \sum_{(n,k)} \psi_n(Z_{i_1}, \ldots, Z_{i_k}) \\ &= \frac{n!}{(n-k)!} \binom{n}{k}^{-1} \sum_{(n,k)} \psi_n(Z_{i_1}, \ldots, Z_{i_k}) = P_k^n u_n, \end{aligned}$$

where $P_k^n = \frac{n!}{(n-k)!}$. Hence $v_n = \frac{1}{n^k} P_k^n u_n + \frac{1}{n^k} \sum_{\kappa=1}^{k-1} \sum_{(n,\kappa)} \phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa})$. Letting $u_n^{(\kappa)} = \binom{n}{\kappa}^{-1} \sum_{(n,\kappa)} \phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa})$ we write $u_n - v_n = \left(1 - \frac{P_k^n}{n^k}\right) u_n - \frac{1}{n^k} \sum_{\kappa=1}^{k-1} \binom{n}{\kappa} u_n^{(\kappa)}$. By Chebyshev's inequality $P\left(n^{1/2}|u_n - v_n| > \epsilon\right) \leq \frac{nE((u_n-v_n)^2)}{\epsilon^2}$ for all $\epsilon > 0$ and consequently to reach the desired conclusion it suffices to show that $nE((u_n - v_n)^2) = o(1)$. By the c_r-inequality we can write

$$nE((u_n - v_n)^2) \leq kn \left(1 - \frac{P_k^n}{n^k}\right)^2 E(u_n^2) + kn \sum_{\kappa=1}^{k-1} \binom{n}{\kappa}^2 n^{-2k} E((u_n^{(\kappa)})^2). \tag{4}$$

We will deal with the two terms on the righthand side of inequality (4) separately. We first note that by the representation theorem for U statistics (Serfling, 1980; p. 180) $u_n^{(\kappa)} = \frac{1}{n!}\sum_p W_n^{(\kappa)}(Z_{i_1}, \ldots, Z_{i_n})$ where

$$W_n^{(\kappa)}(Z_{i_1}, \ldots, Z_{i_n}) = \frac{1}{\gamma_\kappa}(\phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa}) + \phi_{\kappa,n}(Z_{i_{\kappa+1}}, \ldots, Z_{i_{2\kappa}}) + \cdots \\ +\phi_{\kappa,n}(Z_{i_{\gamma_\kappa \kappa-\kappa+1}}, \ldots, Z_{i_{\gamma_\kappa \kappa}})),$$

$\gamma_\kappa = [n/\kappa]$ is the greatest integer $\leq n/\kappa$ and $\sum_p$ denotes the sum over all permutations $(i_1, \ldots, i_n)$ of $\{1, 2, \ldots, n\}$ and similarly $u_n = \frac{1}{n!}\sum_p W_n(Z_{i_1}, \ldots, Z_{i_n})$ where $W_n(Z_{i_1}, \ldots, Z_{i_n}) = \frac{1}{\gamma}(\psi_n(Z_{i_1}, \ldots, Z_{i_k}) + \psi_n(Z_{i_{k+1}}, \ldots, Z_{i_{2k}}) + \cdots + \psi_n(Z_{i_{\gamma k-k+1}}, \ldots, Z_{i_{\gamma k}}))$ and $\gamma = [n/k]$ is the greatest integer $\leq n/k$. Now, for the first term on the righthand side of inequality (4), since $n^k - P_k^n = O(n^{k-1})$ we have $n\left(1 - \frac{P_k^n}{n^k}\right)^2 = O(n^{-1})$ and $n\left(1 - \frac{P_k^n}{n^k}\right)^2 E(u_n^2) = O(n^{-1})E\left(\left(\frac{1}{n!}\sum_p W_n(Z_{i_1}, \ldots, Z_{i_n})\right)^2\right)$. By Minkowski's inequality

$$E\left(\left(\frac{1}{n!}\sum_p W_n(Z_{i_1}, \ldots, Z_{i_n})\right)^2\right) \leq \frac{1}{(n!)^2}\left(\sum_p \left(E(|W_n(Z_{i_1}, \ldots, Z_{i_n})|^2)\right)^{1/2}\right)^2 \\ = E(|W_n(Z_{i_1}, \ldots, Z_{i_n})|^2),$$

where the last equality follows since $\{Z_i\}_{i=1,2,\ldots}$ is an i.i.d. sequence. By definition of W_n and another use of Minkowski's inequality we have

$$E(|W_n(Z_{i_1}, \ldots, Z_{i_n})|^2) \leq \gamma^{-2}\left(\sum_{j=1}^{\gamma}\left(E(\psi_n^2(Z_{i_{jk-k+1}}, \ldots, Z_{i_{jk}}))\right)^{1/2}\right)^2 \\ = E(\psi_n(Z_{i_1}, \ldots, Z_{i_k})) = o(n),$$

where the last equality follows by assumption. We therefore conclude that $n\left(1 - \frac{P_k^n}{n^k}\right)^2 E(u_n^2) = o(1)$.

For the second term on inequality (4), note first that $\binom{n}{\kappa} n^{-k} = O(n^{\kappa-k})$ and since $\kappa \in \{1, \ldots, k-1\}$, $\max_\kappa n^{\kappa-k} = n^{-1}$ and therefore $\left(\binom{n}{\kappa} n^{-k}\right)^2 = O(n^{-2})$. Once again, by Minkowski's inequality we have $E((u_n^{(\kappa)})^2) \leq E(|W_n^{(\kappa)}(Z_{i_1}, \ldots, Z_{i_n})|^2)$. Now, given the definition of $W_n^{(\kappa)}$ and the fact that $\{\phi_{\kappa,n}(Z_{i_{j\kappa-\kappa+1}}, \ldots, Z_{i_{j\kappa}})\}_{j=1,2,\ldots}$ is an i.i.d. sequence of random variables we have by another application of Minkowski's inequality that $E(|W_n^{(\kappa)}(Z_{i_1}, \ldots, Z_{i_n})|^2) \leq E((\phi_{\kappa,n}(Z_{i_1}, \ldots, Z_{i_\kappa}))^2)$. Let l_κ denote the number of terms in $\sum_{\{\nu_1, \ldots, \nu_\kappa, k\}}$, then by the c_r inequality

$$E((\phi_{\kappa,n}(Z_{i_1},\ldots,Z_{i_\kappa}))^2)$$
$$= E\left(\left(\sum_{\{\nu_1,\ldots,\nu_\kappa,k\}} \frac{k!}{\prod_{j=1}^{\kappa}\nu_j!}\psi_n \times (\underbrace{Z_{i_1},\ldots,Z_{i_1}}_{\nu_1-\text{ times}},\ldots,\underbrace{Z_{i_\kappa},\ldots,Z_{i_\kappa}}_{\nu_\kappa-\text{ times}})\right)^2\right)$$
$$\le l_\kappa \sum_{\{\nu_1,\ldots,\nu_\kappa,k\}} \frac{(k!)^2}{(\prod_{j=1}^{\kappa}\nu_j!)^2} E\left(\psi_n^2(\underbrace{Z_{i_1},\ldots,Z_{i_1}}_{\nu_1-\text{ times}},\ldots,\underbrace{Z_{i_\kappa},\ldots,Z_{i_\kappa}}_{\nu_\kappa-\text{ times}})\right).$$

Now, since by assumption $E\left(\psi_n^2(Z_{i_1},\ldots,Z_{i_k})\right) = o(n)$ for all $1 \le i_1,\ldots,i_k \le n, k \le n,$, then it is true that for all $\kappa \in \{1,\ldots,k-1\}$ and all $\nu_j \ge 1$ such that $k = \sum_{j=1}^{\kappa}\nu_j$ we have $E\left(\psi_n^2(Z_{i_1},\ldots,Z_{i_1},\ldots,Z_{i_\kappa},\ldots,Z_{i_\kappa})\right) = o(n)$. Then $E(|W_n^{(\kappa)}(Z_{i_1},\ldots,Z_{i_n})|^2) = o(n)$ and consequently $E((u_n^{(\kappa)})^2) = o(n)$. As a result,

$$n\sum_{\kappa=1}^{k-1}\binom{n}{\kappa}^2 n^{-2k}E((u_n^{(\kappa)})^2) \le nO(n^{-2})o(n)\sum_{\kappa=1}^{k-1} l_\kappa \sum_{\{\nu_1,\ldots,\nu_\kappa,k\}} \frac{(k!)^2}{\left(\prod_{j=1}^{\kappa}\nu_j!\right)^2} = o(1),$$

which completes the proof. □

Theorem 1 can be proved using the following two conditions which are implied by $E\left(\psi_n^2(Z_{i_1},\ldots,Z_{i_k})\right) = o(n)$ for all $1 \le i_1,\ldots,i_k \le n$, $k \le n$: (a) $E\left(\psi_n^2(Z_{i_1},\ldots,Z_{i_k})\right) = o(n)$ for all $1 \le i_1 < i_2 < \cdots < i_k \le n$, $k \le n$; (b) for all $\kappa \in \{1,\ldots,k-1\}$ and all $\nu_j \ge 1$ such that $k = \sum_{j=1}^{\kappa}\nu_j$ that define $\underbrace{Z_{i_1},\ldots,Z_{i_1}}_{\nu_1-\text{ times}},\ldots,\underbrace{Z_{i_\kappa},\ldots,Z_{i_\kappa}}_{\nu_\kappa-\text{ times}}$ $E\left(\psi_n^2(Z_{i_1},\ldots,Z_{i_1},\ldots,Z_{i_\kappa},\ldots,Z_{i_\kappa})\right)$ $= o(n^{2(k-\kappa)-1})$. The next corollary establishes the $\sqrt{n}$ asymptotic equivalence between V statistics and U statistics projections.

Corollary 1 *Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and u_n and v_n be U and V statistics with kernel function $\psi_n(Z_1,\ldots,Z_k)$. In addition, let $\hat{u}_n = \frac{k}{n}\sum_{i=1}^n(\psi_{1n}(Z_i)-\theta_n)+\theta_n$, where $\psi_{1n}(Z_i) = E(\psi_n(Z_1,\ldots,Z_k)|Z_i)$ and $\theta_n = E(\psi_n(Z_1,\ldots,Z_k))$. If $E\left(\psi_n^2(Z_1,\ldots,Z_k)\right) = o(n)$ then $\sqrt{n}(v_n-\hat{u}_n) = o_p(1)$.*

Proof From Theorem 1 we have $\sqrt{n}(v_n-u_n) = o_p(1)$, and from Lemma 2.1 in Lee (1988) we have that $\sqrt{n}(u_n-\hat{u}_n) = o_p(1)$. Hence, $\sqrt{n}(v_n-\hat{u}_n) = o_p(1)$. □

3 Some applications in regression models

We provide applications of Theorem 1 and its corollary around the following regression model,

$$y_t = m(x_t) + \varepsilon_t, \tag{5}$$

where, $\{(y_t, x_t)\}_{t=1}^n$ is a sequence of i.i.d. random variables, $y_t, x_t \in \Re$ with joint density $q(y, x)$, $E(\varepsilon_t|x_t) = 0$ and $V(\varepsilon_t|x_t) = \sigma^2(x_t)$. This is similar to the regression model considered by Fan and Yao (1998), with the exception that here the observations are i.i.d. rather than strictly stationary. Our applications all involve establishing the asymptotic distribution of statistics that are constructed as linear combinations of a first stage nonparametric estimator of $m(x)$. As the first stage estimator for $m(x)$ we consider the local linear estimator of Fan (1992), i.e., for $x \in \Re$, we obtain $\hat{m}(x) = \hat{\alpha}$ where

$$(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{\alpha,\beta} \sum_{t=1}^{n} (y_t - \alpha - \beta(x_t - x))^2 K\left(\frac{x_t - x}{h_n}\right),$$

where $K(\cdot)$ is a density function and $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth. We make the following assumptions.

A1. (1) $0 < \underline{B}_g \le g(x) \le \bar{B}_g < \infty$ for all $x \in G$, G a compact subset of $\Re$, where g is the common marginal density of x_t. (2) For all $x, x' \in \Re$, $|g(x) - g(x')| < m_g|x - x'|$ for some $0 < m_g < \infty$. (3) $q(y, x)$ is continuous everywhere.

A2. (1) $0 < \underline{B}_\sigma \le \sigma(x) \le \bar{B}_\sigma < \infty$ for all $x \in \Re$, $\sigma^2(\cdot) : \Re \to \Re$ is a measurable twice continuously differentiable function in $\Re$ with $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all $x \in \Re$. (2) $0 < \underline{B}_m \le m(x) \le \bar{B}_m < \infty$ for all $x \in \Re$, $m(\cdot) : \Re \to \Re$ is a measurable twice continuously differentiable function in $\Re$ with $|m^{(2)}(x)| < \bar{B}_{2m}$ for all $x \in \Re$.

A3. $K(\cdot) : \Re \to \Re$ is a symmetric density function with bounded support $S_K \in \Re$ satisfying (1) $\int xK(x)\mathrm{d}x = 0$; (2) $\int x^2K(x)\mathrm{d}x = \sigma_K^2$; (3) for all $x \in \Re$, $|K(x)| < B_K < \infty$; (4) for all $x, x' \in \Re$, $|K(x) - K(x')| < m_K|x - x'|$ for some $0 < m_K < \infty$;

A4. $nh_n^3 \to \infty$, and $nh_n^3(\ln(h_n))^{-1} \to -\infty$.

Lemma 1 provides auxiliary results that are useful in the applications that follow. We note that the proof of the Lemma is itself facilitated by using Theorem 1 and its corollary. In addition, the proof relies on repeated use of a version of Lebesgue's dominated convergence theorem which can be found in Prakasa-Rao (1983), p. 35. Henceforth, we refer to this result as *the proposition of Prakasa-Rao*. In addition, the proofs rely on Lemma 2 and Theorem 1 in Martins-Filho and Yao (2003).

Lemma 1 *Let $f(x_t, y_t)$ be continuous at x_t and assume the regression model in (5), assumptions A1–A4 and that $E(f^2(x_t, y_t)) < \infty$. Define $I_n = \frac{1}{n}\sum_{t=1}^n(\hat{m}(x_t) - m(x_t))f(x_t, y_t)$, $J_n = \frac{1}{n}\sum_{t=1}^n(\hat{m}(x_t) - m(x_t))^2 f(x_t, y_t)$, $L_n = \frac{1}{n}\sum_{t=1}^n(\hat{m}(x_t) - m(x_t))^3 f(x_t, y_t)$ and $M_n = \frac{1}{n}\sum_{t=1}^n(\hat{m}(x_t) - m(x_t))^4 f(x_t, y_t)$.*

(a) Given $E(y_t^2 f^2(x_t, y_t)) < \infty$ then

$$I_n = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \int f(x_t, y)g(y|x_t)\mathrm{d}y \; + \frac{1}{2}h_n^2\sigma_K^2 E(m^{(2)}(x_t)f(x_t, y_t)) + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2), \tag{6}$$

where $g(y|x)$ denotes the conditional density of y given x.

(b) Given $E(y_t^4 f^2(x_t, y_t)) < \infty$ then

$$\begin{aligned} J_n = \frac{\sigma_K^2 h_n^2}{n} & \sum_{t=1}^{n} \varepsilon_t m^{(2)}(x_t)E(f(x_t, y_t)|x_1, \ldots, x_n) \\ & + \frac{1}{4}h_n^4\sigma_K^4 E((m^{(2)}(x_t))^2 f(x_t, y_t)) \\ & + o_p(n^{-\frac{1}{2}}) + o_p(h_n^3), \end{aligned} \tag{7}$$

(c) Given $E(y_t^6 f^2(x_t, y_t)) < \infty$ then

$$L_n = o_p(h_n^3) \tag{8}$$

(d) Given $E(y_t^8 f^2(x_t, y_t)) < \infty$ then

$$M_n = o_p(h_n^4). \tag{9}$$

Proof Here we prove parts (a)–(c). The proof of (d) is similar to that of (c) and is omitted. Let $\sum_{(p)} h_{i_1\ldots i_p}$ denote the sum of $h_{i_1\ldots i_p}$ over the permutations of $i_1 \ldots i_p$, i.e., $\sum_{(p)} h_{i_1\ldots i_p}$ is the sum of $p!$ terms given by $h_{i_1 i_2\ldots i_p} + h_{i_2 i_1\ldots i_p} + \cdots + h_{i_p i_{p-1}\ldots i_1}$.

(a) Note that,

$$\begin{aligned} \hat{m}(x_t) - m(x_t) = & \frac{1}{nh_n g(x_t)} \sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k \\ & + \frac{1}{2nh_n g(x_t)} \sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right) m^{(2)}(x_{kt})(x_k - x_t)^2 + w(x_t), \end{aligned}$$

where $w(x_t) = \hat{m}(x_t) - m(x_t) - \frac{1}{nh_n g(x_t)} \sum_{k=1}^{n} K(\frac{x_k - x_t}{h_n}) y_k^*$ and $y_k^* = \varepsilon_k + \frac{1}{2}m^{(2)}(x_{kt})$ $(x_k - x_t)^2$ with $x_{kt} = \lambda x_t + (1-\lambda)x_k$ for some $\lambda \in (0, 1)$. Hence, $I_n = I_{1n} + I_{2n} + I_{3n}$, where

$$I_{1n} = \frac{1}{n^2 h_n} \sum_{t=1}^{n}\sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k f(x_t, y_t)\frac{1}{g(x_t)}$$

$$I_{2n} = \frac{h_n}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt}) f(x_t, y_t)\frac{1}{g(x_t)}$$
$$I_{3n} = \frac{1}{n}\sum_{t=1}^{n} w(x_t) f(x_t, y_t).$$

We treat each term separately. Observe that I_{1n} can be written as

$$\begin{aligned} I_{1n} &= \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\left(\frac{1}{h_n}K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k f(x_t, y_t)\frac{1}{g(x_t)}\right. \\ &\qquad\left. + \frac{1}{h_n}K\left(\frac{x_t - x_k}{h_n}\right)\varepsilon_t f(x_k, y_k)\frac{1}{g(x_k)}\right) \\ &= \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\left(\sum_{(2)} h_{tk}\right) = \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\psi_n(Z_t, Z_k) = \frac{1}{2}v_n, \end{aligned}$$

where v_n is a V statistic with $Z_t = (x_t, y_t)$ and $\psi_n(Z_t, Z_k)$ a symmetric kernel. Hence, by Corollary 1, if $E\left(\psi_n^2(Z_t, Z_k)\right) = o(n)$ for all t, k then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. First, note that for $t \neq k$ $\frac{1}{n}E\left(\psi_n^2(Z_t, Z_k)\right) = \frac{2}{n}E(h_{tk}^2) + \frac{2}{n}E(h_{tk}h_{kt})$ and since $E(\varepsilon_k^2|x_1, \ldots, x_n) = \sigma^2(x_k)$, we have from the proposition of Prakasa-Rao that

$$\frac{1}{n}E(h_{tk}^2) = \frac{1}{nh_n^2}E\left(K^2\left(\frac{x_k - x_t}{h_n}\right)\sigma^2(x_k)E(f^2(x_t, y_t)|x_1, \ldots, x_n)\frac{1}{g^2(x_t)}\right) \to 0,$$

provided that $nh_n \to \infty$. Similarly it can be shown that $E(\frac{1}{n}h_{tk}h_{kt}) = o(1)$ and therefore $\frac{1}{n}E(\psi_n^2(Z_t, Z_k)) = o(1)$. For $t = k$ it suffices to verify that $\frac{4K^2(0)}{nh_n^2}$ $E\left(\sigma^2(x_t)\frac{f^2(x_t, y_t)}{g^2(x_t)}\right) = o(1)$, but this follows directly given our assumptions. Now, since $E(\varepsilon_t|x_1, \ldots, x_n) = 0$, $\theta_n = E\left(\psi_n(Z_t, Z_k)\right) = 0$ and $\hat{u}_n = \frac{2}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t)$. Therefore,

$$\begin{aligned} \psi_{1n}(Z_t) &= \frac{1}{h_n}\varepsilon_t \int\int K\left(\frac{x_k - x_t}{h_n}\right) f(x_k, y_k) g(y_k|x_k)\mathrm{d}x_k \mathrm{d}y_k \\ &= \varepsilon_t \int f(x_t, y) g(y|x_t)\mathrm{d}y + o(1), \end{aligned}$$

where $g(y|x) = \frac{q(y,x)}{g(x)}$. Consequently,

$$I_{1n} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t \int f(x_t, y) g(y|x_t)\mathrm{d}y + o_p(n^{-\frac{1}{2}}) \tag{10}$$

given that $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t = O_p(n^{-\frac{1}{2}})$. Now, note that

$$I_{2n} = \frac{1}{4n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{(2)} h_{tk} = \frac{1}{4n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\psi_n(Z_t, Z_k) = \frac{1}{4}v_n,$$

where $h_{tk} = h_n K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt}) f(x_t, y_t) \frac{1}{g(x_t)}$. By Corollary 1 if $E(\psi_n^2(Z_t, Z_k)) = o(n)$ for all t, k, then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. Hence we focus on $\hat{u}_n$. Note that, $\hat{u}_n = \frac{2}{n}\sum_{t=1}^n \psi_{1n}(Z_t) - \theta_n$, where $\psi_{1n}(Z_t) = E(h_{tk}|Z_t) + E(h_{kt}|Z_t)$ and $E(\hat{u}_n) = \theta_n = 2E(h_{tk})$. Hence, by the proposition of Prakasa-Rao,

$$\begin{aligned} E\left(\frac{\hat{u}_n}{h_n^2}\right) &= \frac{2}{h_n}\int\int\int K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt}) \\ &\qquad \times f(x_t, y_t)\frac{g(x_k)}{g(x_t)} q(x_t, y_t) \mathrm{d}x_k \mathrm{d}x_t \mathrm{d}y_t \\ &\longrightarrow 2\sigma_K^2 E(m^{(2)}(x_t) f(x_t, y_t)). \end{aligned}$$

In addition, note that

$$\begin{aligned} V\left(\frac{1}{h_n^2}\hat{u}_n\right) &= \frac{4}{n^2 h_n^4} n V(E(h_{tk}|Z_t) + E(h_{kt}|Z_t)) \\ &= \frac{4}{n h_n^4}\big(E(E(h_{tk}|Z_t))^2 + E(E(h_{kt}|Z_t))^2 \\ &\quad + 2E(E(h_{tk}|Z_t)E(h_{kt}|Z_k)) - \theta_n^2\big). \end{aligned}$$

Under our assumptions, it is straightforward to show that

$$\begin{aligned} &\frac{1}{n h_n^4} E(E(h_{tk}|Z_t))^2 \\ &\quad = \frac{1}{n h_n^2} E\left(f(x_t, y_t)^2 \frac{1}{g^2(x_t)} E^2\left(K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})|Z_t\right)\right) \to 0 \end{aligned}$$

as $n \to \infty$. Similarly, $\frac{1}{n h_n^4} E(E(h_{kt}|Z_t))^2 \to 0$ and $\frac{1}{n h_n^4} E(E(h_{tk}|Z_t)E(h_{kt}|Z_t)) \to 0$ as $n \to \infty$. Hence, $V\left(\frac{1}{h_n^2}\hat{u}_n\right) \to 0$. By Markov's inequality we conclude that

$$\hat{u}_n = 2h_n^2 \sigma_K^2 E(m^{(2)}(x_t) f(x_t, y_t)) + o_p(h_n^2) \text{ and by Corollary 1,}$$

$$I_{2n} = \frac{1}{2} h_n^2 \sigma_K^2 E(m^{(2)}(x_t) f(x_t, y_t)) + o_p(h_n^2). \tag{11}$$

Finally, verification that $E(\psi_n^2(Z_t, Z_k)) = 2E(h_{tk}^2) + 2E(h_{tk}h_{kt}) = o(n)$ for all $t \neq k$ follows directly from our assumptions and use of the proposition of Prakasa-Rao, and for the case where $t = k$ we have that $E(\psi_n^2(Z_t, Z_k)) = 0$. From Martins-Filho and Yao's (2003) Lemma 2 and Theorem 1

$$|w(x_t)| = \left|\hat{m}(x_t) - m(x_t) - \frac{1}{n h_n g(x_t)} \sum_{k=1}^n K\left(\frac{x_k - x_t}{h_n}\right) y_k^*\right| = O_p(R_{n,2}(x_t))$$

and $\sup_{x_t \in G} |R_{n,2}(x_t)| = o_p(h_n^2)$, hence $\sup_{x_t \in G} |w(x_t)| = o_p(h_n^2)$. As a result we have that $|I_{3n}| \leq o_p(h_n^2)\frac{1}{n}\sum_{t=1}^n |f(x_t, y_t)| = o_p(h_n^2)$, since $E(f^2(x_t, y_t)) < \infty$. Therefore, $I_{3n} = \frac{1}{n}\sum_{t=1}^n w(x_t) f(x_t, y_t) = o_p(h_n^2)$. Combining this result with

(10) and (11) proves part (a).
For part (b) note that

$$J_n = \frac{1}{n^3h_n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f(x_t, y_t)\frac{1}{g^2(x_t)}K\left(\frac{x_k - x_t}{h_n}\right)K\left(\frac{x_l - x_t}{h_n}\right)\varepsilon_k\varepsilon_l$$
$$+\frac{h_n^2}{4n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{g^2(x_t)}f(x_t, y_t)K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2$$
$$\times K\left(\frac{x_l - x_t}{h_n}\right)\left(\frac{x_l - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})m^{(2)}(x_{lt})$$
$$+\frac{1}{n}\sum_{t=1}^{n} w^2(x_t)f(x_t, y_t)$$
$$+\frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{g^2(x_t)}f(x_t, y_t)K\left(\frac{x_k - x_t}{h_n}\right)$$
$$\times K\left(\frac{x_l - x_t}{h_n}\right)\left(\frac{x_l - x_t}{h_n}\right)^2 m^{(2)}(x_{lt})\varepsilon_k$$
$$+\frac{2}{n^2h_n}\sum_{t=1}^{n}\sum_{k=1}^{n} w(x_t)\frac{1}{g(x_t)}f(x_t, y_t)K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k$$
$$+\frac{h_n}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{n} w(x_t)\frac{1}{g(x_t)}f(x_t, y_t)K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})$$
$$= J_{1n} + J_{2n} + J_{3n} + J_{4n} + J_{5n} + J_{6n}.$$

We examine each term separately.

$$J_{1n} = \frac{1}{6}\frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{(3)} h_{tkl} = \frac{1}{6}\frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\psi_n(Z_t, Z_k, Z_l) = \frac{1}{6}v_n,$$

where $h_{tkl} = \frac{1}{h_n^2}K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})\varepsilon_k\varepsilon_l f(x_t, y_t)\frac{1}{g^2(x_t)}$. By Corollary 1, if $E(\psi_n^2(Z_t, Z_k, Z_l)) = o(n)$ for all t, k, l then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. Since $\psi_{1n}(Z_t) = 0$ and $\theta_n = 0$, we have that $\hat{u}_n = 0$ and therefore $\sqrt{n}I_{1n} = o_p(1)$. To verify that $\frac{1}{n}E(\psi_n^2(Z_t, Z_k, Z_l)) = o(1)$ for $t \neq k \neq l$ we note that $\frac{1}{n}E(\psi_n^2(Z_t, Z_k, Z_l)) = \frac{4}{n}(3E(h_{tkl}^2) + 2E(h_{tkl}h_{ktl}) + 2E(h_{tkl}h_{ltk}) + 2E(h_{ktl}h_{ltk}))$. Under A1–A4, repeated application of the proposition of Prakasa-Rao to each of these expectations shows that each approaches zero as $n \to \infty$. Also, if $t = k$ but $t \neq l$ we have that $E(\psi_n^2(Z_t, Z_t, Z_l)) = o(n)$ and when $t = k = l$, $E(\psi_n^2(Z_t, Z_t, Z_t)) = o(n^2)$ directly from our assumptions provided $nh_n^3 \to 0$.

$$J_{2n} = \frac{1}{24n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{(3)} h_{tkl} = \frac{1}{24n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\psi_n(Z_t, Z_k, Z_l) = \frac{1}{24}v_n,$$

where $h_{tkl} = h_n^2K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2m^{(2)}(x_{kt})m^{(2)}(x_{lt})\frac{1}{g^2(x_t)}f(x_t, y_t)$. By Corollary 1 $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$ provided that $E(\psi_n(Z_t, Z_k, Z_l)) = o(n)$

for all t, k, l. As above, verification of this condition is straightforward given our assumptions. Note that $\hat{u}_n = \frac{3}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t) - 2\theta_n$ with $E(\hat{u}_n) = \theta_n$. Hence,

$$\frac{1}{h_n^4}E(\hat{u}_n) = \frac{6}{h_n^2}E\left(K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 K\left(\frac{x_l - x_t}{h_n}\right)\left(\frac{x_l - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})\right.$$
$$\left.\times\, m^{(2)}(x_{lt})\frac{1}{g^2(x_t)}f(x_t, y_t)\right)$$

and by the proposition of Prakasa-Rao we have, $\frac{1}{h_n^4}E(\hat{u}_n) \to 6\sigma_K^4 E((m^{(2)}(x_t))^2 f(x_t, y_t))$ and $V(\frac{1}{h_n^4}\hat{u}_n) \to 0$. By Markov's inequality we conclude that $J_{2n} = \frac{1}{4}h_n^4\sigma_K^4$ $E((m^{(2)}(x_t))^2 f(x_t, y_t)) + o_p(h_n^4) + o_p(n^{-1/2})$. $J_{3n} = \frac{1}{n}\sum_{t=1}^{n} w^2(x_t)f(x_t, y_t) = o_p(h_n^4)$ follows directly from the analysis of the I_{3n} in part (a). Now,

$$J_{4n} = \frac{1}{6n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{(3)} h_{tkl} = \frac{1}{6n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\psi_n(Z_t, Z_k, Z_l) = \frac{1}{6}v_n,$$

where $h_{tkl} = K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 m^{(2)}(x_{lt})\varepsilon_k\frac{1}{g^2(x_t)}f(x_t, y_t)$. Once again given our assumptions it can be verified that $E(\psi_n^2(Z_t, Z_k, Z_l)) = o(n)$ for all t, k, l, therefore we have $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$, where $\hat{u}_n = \frac{3}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t) - 2\theta_n$. Given that in this case $\theta_n = 0$, we have

$$\hat{u}_n = \frac{6}{n}\sum_{t=1}^{n}\varepsilon_t E\left(K\left(\frac{x_t - x_k}{h_n}\right)K\left(\frac{x_l - x_k}{h_n}\right)\left(\frac{x_l - x_k}{h_n}\right)^2 m^{(2)}(x_{lk})\frac{1}{g^2(x_k)}f(x_k, y_k)|Z_t\right),$$

Under A1–A4, the proposition of Prakasa Rao gives

$$E\left(\frac{1}{h_n^2}K\left(\frac{x_t - x_k}{h_n}\right)K\left(\frac{x_l - x_k}{h_n}\right)\left(\frac{x_l - x_k}{h_n}\right)^2 m^{(2)}(x_{lk})\frac{1}{g^2(x_k)}f(x_k, y_k)|Z_t\right) \to$$
$$\sigma_K^2 m^{(2)}(x_t)E(f(x_t, y)|x_1, \ldots, x_n).$$

Hence, $\hat{u}_n = \frac{6h_n^2}{n}\sum_{t=1}^{n}\varepsilon_t m^{(2)}(x_t)\sigma_K^2 E(f(x_t, y)|x_1, \ldots, x_n) + o_p(n^{-\frac{1}{2}})$, given that $\frac{6}{n}\sum_{t=1}^{n}\varepsilon_t = O_p(n^{-\frac{1}{2}})$. Consequently, $J_{4n} = \frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\varepsilon_t m^{(2)}(x_t)E(f(x_t, y)$ $|x_1, \ldots, x_n) + o_p(n^{-\frac{1}{2}})$.

$|J_{5n}| = \left|\frac{2}{n}\sum_{t=1}^{n} w(x_t)\left(\frac{1}{nh_n g(x_t)}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})\varepsilon_k\right)f(x_t, y_t)\right|$ and from Martins-Filho and Yao (2003) Theorem 1 we have $\sup_{x_t\in G}|w(x_t)| = o_p(h_n^2)$ and $\sup_{x_t\in G}\left|\frac{1}{nh_n g(x_t)}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})\varepsilon_k\right| = o_p(h_n)$. Hence, $|J_{5n}| \le \frac{2}{n}\sum_{t=1}^{n}|f(x_t, y_t)|o_p(h_n^3) = O_p(1)o_p(h_n^3) = o_p(h_n^3)$ since $\frac{2}{n}\sum_{t=1}^{n}|f(x_t, y_t)| = O_p(1)$, which gives $J_{5n} = o_p(h_n^3)$. Finally, from Martins-Filho and Yao (2003) Theorem 1 we have

$$\sup_{x_t\in G}\left|\frac{h_n}{ng(x_t)}\sum_{k=1}^{n}K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})\right| = O_p(h_n^2) \qquad (12)$$

and $|J_{6n}| \leq \frac{1}{n}\sum_{t=1}^{n}|f(x_t, y_t)|o_p(h_n^2)O_p(h_n^2) = O_p(1)o_p(h_n^4) = o_p(h_n^4)$. Combining the orders of all six terms gives the desired result.

(c) Let $a_{tn} = \frac{1}{nh_n g(x_t)}\sum_{k=1}^{n} K\left(\frac{x_k-x_t}{h_n}\right)\varepsilon_k, b_{tn} = \frac{1}{2nh_n g(x_t)}\sum_{k=1}^{n} K\left(\frac{x_k-x_t}{h_n}\right) m^{(2)}(x_{kt})$ $(x_k - x_t)^2$, $c_{tn} = w(x_t)$. Then $(\hat{m}(x_t) - m(x_t))^3 = a_{tn}^3 + b_{tn}^3 + c_{tn}^3 + 3a_{tn}b_{tn}^2 + 3a_{tn}c_{tn}^2 + 3a_{tn}^2 b_{tn} + 3a_{tn}^2 c_{tn} + 3b_{tn}^2 c_{tn} + 3b_{tn}c_{tn}^2 + 6a_{tn}b_{tn}c_{tn}$ and $n^{-1}\sum_{t=1}^{n}(\hat{m}(x_t) - m(x_t))^3 f(x_t, y_t) = \sum_{i=1}^{10} L_{in}$, where $L_{1n} = n^{-1}\sum_{t=1}^{n} a_{tn}^3 f(x_t, y_t), \ldots, L_{10n} = n^{-1}\sum_{t=1}^{n} 6a_{tn}b_{tn}c_{tn} f(x_t, y_t)$. Since

$$\left|\sup_{x_t}\frac{1}{nh_n g(x_t)}\sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k\right| = o_p(h_n),$$

then $|L_{1n}| \leq o_p(h_n^3) \times \frac{1}{n}\sum_{t=1}^{n}|f(x_t, y_t)| = o_p(h_n^3)$. By (12), $|L_{2n}| \leq O_p(h_n^6)n^{-1}\sum_{t=1}^{n}|f(x_t, y_t)| = O_p(h_n^6)$ since $E(f^2(x_t, y_t)) < \infty$. $|L_{3n}| \leq o_p(h_n^6)$ follows directly from $\sup_{x_t \in G}|w(x_t)| = o_p(h_n^2)$.

$$\begin{aligned}
|L_{4n}| \leq \frac{3}{n}\sum_{t=1}^{n}&\left|\frac{1}{nh_n g(x_t)}\sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)\varepsilon_k\right| \\
&\times\left|\left(\frac{1}{2nh_n g(x_t)}\sum_{k=1}^{n} K\left(\frac{x_k - x_t}{h_n}\right)(x_k - x_t)^2 m^{(2)}(x_{kt})\right)^2\right| \\
&\times|f(x_t, y_t)| \leq o_p(h_n)O_p(h_n^4)\frac{3}{n}\sum_{t=1}^{n}|f(x_t, y_t)| = o_p(h_n^5)
\end{aligned}$$

from (12) and the analysis of J_{5n} in part (b). Similarly, we obtain $|L_{5n}|, |L_{10n}| \leq o_p(h_n^5)$, $|L_{6n}|, |L_{7n}| \leq o_p(h_n^4)$ and $|L_{8n}|, |L_{9n}| \leq o_p(h_n^6)$. Hence, we have $L_n = o_p(h_n^3)$. □

We now use the results in the Lemma to establish the asymptotic distribution of several statistics of interest.

3.1 Estimating conditional variance

Consider first a special case of the model (5), where $\varepsilon_t|x_t \sim N(0, \sigma^2)$. Then a natural estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\varepsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\varepsilon_t + (m(x_t) - \hat{m}(x_t))^2\right).$$

The difficulty in dealing with such expressions lies in the average terms involving $(m(x_t) - \hat{m}(x_t))$ and $(m(x_t) - \hat{m}(x_t))^2$, since the first term is an average of an i.i.d. sequence. By using Corollary 1 and Lemma 1 we have a convenient way to establish their asymptotic properties. This is shown in the next theorem.

Theorem 2 *Assume that in model (5), $\varepsilon_t|x_t \sim N(0, \sigma^2)$ and $E(y_t^4) < \infty$. Under assumptions A1–A4 we have $\sqrt{n}(\hat{\sigma}^2 - \sigma^2 - b_{1n}) \xrightarrow{d} N(0, 2\sigma^4)$, where $b_{1n} = o_p(h_n^2)$.*

Proof Note that $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\varepsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\varepsilon_t + (m(x_t) - \hat{m}(x_t))^2\right)$. Now, letting $f(x_t, y_t) = -2\varepsilon_t$ in part (a) of Lemma 1 we have $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t \int f(x_t, y_t)$ $g(y_t|x_t)\mathrm{d}y_t = 0$, since $E(\varepsilon_t|x_1, \ldots, x_n) = 0$ and $\frac{1}{2}h_n^2\sigma_K^2 E\left(m^{(2)}(x_t)f(x_t, y_t)\right) = 0$. Hence, $\frac{1}{n}\sum_{t=1}^{n} 2(m(x_t) - \hat{m}(x_t))\varepsilon_t = o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. By part (b) of Lemma 1 with $f(x_t, y_t) = 1$ we have

$$\frac{1}{4}h_n^4\sigma_K^4 E((m^{(2)}(x_t))^2 f(x_t, y_t)) = O_p(h_n^4)$$

and

$$\frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\varepsilon_t m^{(2)}(x_t)E(f(x_t, y)|x_1, \ldots, x_n)$$
$$= \frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\varepsilon_t m^{(2)}(x_t) = O_p(n^{-\frac{1}{2}}h_n^2) = o_p(n^{-\frac{1}{2}})$$

by the central limit theorem for i.i.d. sequences. Hence, $\frac{1}{n}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^2 = o_p(n^{-\frac{1}{2}}) + o_p(h_n^3)$ provided $E(y_t^4) < \infty$. Finally, note that $\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2) = \frac{1}{n}\sum_{t=1}^{n}\zeta_t$, then $E(\zeta_t) = 0$, $V(\zeta_t) = E(\zeta_t^2) = 2\sigma^4$, given conditional normality. Hence, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varepsilon_t^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$. Combining the results for each term proves the theorem. □

Theorem 2 can be generalized by relaxing the conditional normality assumption and allowing $V(y_t|x_t) = \sigma^2(x_t)$. This generalization was done by Martins-Filho and Yao (2003) but their proof can be conveniently simplified by using Theorem 1 and its corollary.

3.2 Estimating conditional skewness and kurtosis

We now consider the estimation of Pearson's conditional skewness $\alpha_3 = \mu_3/\mu_2^{3/2}$ and kurtosis $\alpha_4 = \nu_4/\mu_2^2$ for the regressand in (5), where $\mu_r = E\left((y_t - m(x_t))^r|x_t\right)$ for $r = 2, 3, 4$. We assume for simplicity that the these higher order conditional centered moments do not depend on x_t, $t = 1, 2, \ldots$ although the underlying conditional density clearly does. We now define the estimators

$$\hat{\alpha}_3 = \frac{n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^3}{(n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)^{3/2}}, \hat{\alpha}_4 = \frac{n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^4}{(n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)^2}$$

and establish the following theorem.

Theorem 3 *Assume A1–A4 are holding for model (5) and let* $\mu_r = E((y_t - m(x_t))^r|x_t)$ *for* $r = 2, 3, 4$.

(a) If $E(y_t^6) < \infty$ *then* $\sqrt{n}(\hat{\alpha}_3 - \alpha_3 - B_{1n}) \xrightarrow{d} N(0, \mu_2^{-3}V(Z_t))$ *where*

$$B_{1n} = -3/2\mu_2^{-1/2}E\left(m^{(2)}(x_t)\right)h_n^2\sigma_K^2 + o_p(h_n^2)$$

and $Z_t = \varepsilon_t^3 + 1/2\mu_3 - 3\mu_2\varepsilon_t - 3/2\frac{\mu_3}{\mu_2}\varepsilon_t^2$

(b) If $E(y_t^8) < \infty$ *then* $\sqrt{n}(\hat{\alpha}_4 - \alpha_4 - B_{2n}) \xrightarrow{d} N(0, \mu_2^{-4} V(W_t))$ *where*

$$B_{2n} = -2h_n^2 \mu_2^{-2} \mu_3 \sigma_K^2 E(m^{(2)}(x_t)) + o_p(h_n^2)$$

and $W_t = \varepsilon_t^4 + \mu_4 - 4\varepsilon_t \mu_3 - 2\frac{\mu_4}{\mu_2}\varepsilon_t^2$

Proof We first write

$$\hat{\alpha}_3 - \alpha_3 = \left(n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2\right)^{-3/2} \left(b_{1n} - b_{2n}\mu_3\mu_2^{-3/2}\right), \tag{13}$$

$$\hat{\alpha}_4 - \alpha_4 = \left(n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2\right)^{-2} \left(b_{3n} - b_{4n}\mu_4\mu_2^{-2}\right), \tag{14}$$

where $b_{1n} = n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^3 - \mu_3$, $b_{2n} = (n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)^{3/2} - \mu_2^{3/2}$, $b_{3n} = n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^4 - \mu_4$ and $b_{4n} = (n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)^2 - \mu_2^2$. By Theorem 2, and adopting the notation $\sigma^2 \equiv \mu_2$ we have that $n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 = n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 + o_p(n^{-1/2}) + o_p(h_n^2)$. But since $n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 - \mu_2 = o_p(1)$ by Kolmogorov's law of large numbers, we have that

$$n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 = \mu_2 + o_p(1). \tag{15}$$

(a) By the mean value theorem for random variables we have that for $\lambda \in (0, 1)$,

$$b_{2n} = \frac{3}{2}\left(\mu_2 + \lambda(n^{-1}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 - \mu_2)\right)^{1/2} \times \left(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 - \mu_2\right) \tag{16}$$

$$= \frac{3}{2}\mu_2^{1/2}\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - \mu_2) + o_p(n^{-1/2}) + o_p(h_n^2). \tag{17}$$

Also,

$$b_{1n} = n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^3 + 3n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^2\varepsilon_t$$
$$+3n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))\varepsilon_t^2 + n^{-1}\sum_{t=1}^{n}\varepsilon_t^3 - \mu_3$$
$$= p_{n1} + p_{n2} + p_{n3} + p_{n4} - \mu_3.$$

Since $E(y_t^6) < \infty$, by part (c) of Lemma 1 $p_{n1} = o_p(h_n^3)$. By part (b) $p_{n2} = o_p(n^{-1/2}) + o_p(h_n^3)$ and by part (a) $p_{n3} = -3n^{-1}\mu_2 \sum_{t=1}^n \varepsilon_t - 3/2 h_n^2 \sigma_K^2 \mu_2 E(m^{(2)}(x_t)) + o_p(n^{-1/2}) + o_p(h_n^2)$, hence

$$b_{1n} = n^{-1}\sum_{t=1}^{n} \varepsilon_t^3 - \mu_3 - 3n^{-1}\mu_2 \sum_{t=1}^{n} \varepsilon_t - \frac{3}{2} h_n^2 \mu_2 \sigma_K^2 E(m^{(2)}(x_t)) + o_p(n^{-1/2}) + o_p(h_n^2). \tag{18}$$

Combining (17) and (18) we have that $b_{1n} - b_{2n}\mu_3\mu_2^{-3/2} = n^{-1}\sum_{t=1}^n Z_t + A_{1n}$, where $Z_t = \varepsilon_t^3 + 1/2\mu_3 - 3\mu_2\varepsilon_t - 3/2\frac{\mu_3}{\mu_2}\varepsilon_t^2$, $A_{1n} = -\frac{3}{2}\mu_2 E\left(m^{(2)}(x_t)\right) h_n^2 \sigma_K^2 + o_p(n^{-1/2}) + o_p(h_n^2)$. Note that $E(Z_t) = 0$ and by the c_r-inequality $V(Z_t) = E(Z_t^2) < \infty$ provided $E(y_t^6) < \infty$. Hence, by the central limit theorem for i.i.d. random variables

$$\sqrt{n}\left(b_{1n} - b_{2n}\frac{\mu_3}{\mu_2^{3/2}} - A_{1n}\right) \xrightarrow{d} N(0, V(Z_t)) \tag{19}$$

and consequently, combining (15) and (19) we have $\sqrt{n}\left(\hat{\alpha}_3 - \alpha_3 - B_{1n}\right) \xrightarrow{d} N(0, \mu_2^{-3} V(Z_t))$ where $B_{1n} = -3/2\mu_2^{-1/2} E\left(m^{(2)}(x_t)\right) h_n^2 \sigma_K^2 + o_p(h_n^2)$.

(b) Again for $\lambda \in (0, 1)$ we have by the mean value theorem that

$$b_{4n} = 2\mu_2 \frac{1}{n}\sum_{t=1}^{n} (\varepsilon_t^2 - \mu_2) + o_p(n^{-1/2}) + o_p(h_n^2). \tag{20}$$

Also,

$$\begin{aligned} b_{3n} &= n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^4 + 4n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^3\varepsilon_t \\ &\quad + 4n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))\varepsilon_t^3 \\ &\quad + 6n^{-1}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^2\varepsilon_t^2 + n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 - \mu_4 \\ &= q_{n1} + q_{n2} + q_{n3} + q_{n4} + q_{n5} - \mu_4. \end{aligned}$$

By part (d) of Lemma 1 with $f(x_t, y_t) = 1$ we have $q_{n1} = o_p(h_n^4)$, and by part (c) with $f(x_t, y_t) = -4\varepsilon_t$, $q_{n2} = o_p(h_n^3)$. By part (b) with $f(x_t, y_t) = 6\varepsilon_t^2$

$$q_{n4} = \frac{6}{n}\sigma_K^2\sigma^2 h_n^2 \sum_{t=1}^{n} \varepsilon_t m^{(2)}(x_t) + \frac{3}{2} h_n^4 \sigma_K^4 \sigma^2 E\left((m^{(2)}(x_t))^2\right) + o_p(n^{-1/2}) + o_p(h_n^3). \tag{21}$$

By part (a) with $f(x_t, y_t) = -4\varepsilon_t^3$ we have

$$q_{n3} = -\frac{4\mu_3}{n}\sum_{t=1}^{n} \varepsilon_t - 2\sigma_K^2 h_n^2 \mu_3 E\left(m^{(2)}(x_t)\right) + o_p(n^{-1/2}) + o_p(h_n^2). \tag{22}$$

Hence,

$$b_{3n} = n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 - \mu_4 - \frac{4\mu_3}{n}\sum_{t=1}^{n}\varepsilon_t - 2\sigma_K^2 h_n^2 \mu_3 E\left(m^{(2)}(x_t)\right)$$
$$+o_p(n^{-1/2}) + o_p(h_n^2). \tag{23}$$

Combining (20) and (23), we write $b_{3n} - b_{4n}\frac{\mu_4}{(\mu_2)^2} = n^{-1}\sum_{t=1}^{n} W_t + A_{2n}$ where $W_t = \varepsilon_t^4 + \mu_4 - 4\varepsilon_t\mu_3 - 2\frac{\mu_4}{\mu_2}\varepsilon_t^2$ and $A_{2n} = -2h_n^2\sigma_K^2\mu_3 E(m^{(2)}(x_t)) + o_p(n^{-1/2}) + o_p(h_n^2)$. Note that $E(W_t) = 0$ and by the c_r-inequality, $V(W_t) = E(W_t^2) < \infty$. Hence, given that $E(y_t^8) < \infty$ we have the central limit theorem for i.i.d. random variables that

$$\sqrt{n}\left(b_{3n} - b_{4n}\frac{\mu_4}{\mu_2^2} - A_{2n}\right) \xrightarrow{d} N(0, V(W_t)). \tag{24}$$

Combining (15) and (24), we have that

$$\sqrt{n}\left(\hat{\alpha}_4 - \alpha_4 - B_{2n}\right) \xrightarrow{d} N(0, \mu_2^{-4}V(W_t)), \quad \text{where}$$
$$B_{2n} = -2h_n^2\mu_2^{-2}\sigma_K^2\mu_3 E(m^{(2)}(x_t)) + o_p(h_n^2). \tag{25}$$

□

3.3 Estimating a nonparametric R^2 measure

In regression analysis we are usually interested in Pearson's correlation ratio,

$$\eta^2 = \frac{V(E(y|x))}{V(y)} = \frac{V(m(x))}{V(y)}$$

where, y is a regressand and x is the regressor in the context of model (5). Since $V(y) = V(E(y|x))) + E(V(y|x)) = V(m(x)) + E(\sigma^2(x))$, η^2 gives the fraction of the variability of y which is explained with the best predictor $m(x)$. This can be interpreted as a nonparametric coefficient of determination or a nonparametric R^2 measure. Estimation of a nonparametric R^2 measure has been studied by Doksum and Samarov (1995) using a Nadaraya–Watson estimator. A similar topic – estimation of noise to signal ratios – has been considered by Yao and Tong (2000). Here, given that $V(m(x)) = V(y) - E(y - m(x))^2$ we have $\eta^2 = 1 - \frac{E(y-m(x))^2}{V(y)}$ and we consider

$$\hat{\eta}^2 = 1 - \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad \text{where } \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

The following theorem gives the asymptotic characterization of our proposed nonparametric R^2 measure.

Theorem 4 *Assume A1–A4 are holding for model (5) and $E(y_t^4) < \infty$, then we have*

$\sqrt{n}(\hat{\eta}^2-\eta^2-b_{2n}) \xrightarrow{d} N(0, V(\xi_t))$ where $\xi_t = \frac{1}{V(y_t)}\left(\varepsilon_t^2-(1-\eta^2)(y_t-E(y_t))^2\right)$ *and* $b_{2n} = o_p(h_n^2)$.

Proof $\hat{\eta}^2 - \eta^2 = -\frac{1}{\frac{1}{n}\sum_{t=1}^n (y_t-\bar{y})^2}\left(\beta_1 - \beta_2 \frac{E(y_t-m(x_t))^2}{V(y_t)}\right)$ where $\beta_1 = \frac{1}{n}\sum_{t=1}^n (y_t - \hat{m}(x_t))^2 - E(y_t - m(x_t))^2$, $\beta_2 = \frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2 - V(y_t)$. First, note that by the law of large numbers $\frac{1}{n}\sum_{t=1}^n y_t^2 \xrightarrow{p} E(y_t^2)$ and $\bar{y}^2 \xrightarrow{p} E(y_t)^2$, hence

$$\left(\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2\right)^{-1} \xrightarrow{p} V(y_t)^{-1}. \tag{26}$$

For β_2, $\bar{y}^2 - (E(y_t))^2 = \frac{1}{2}\frac{1}{n^2}\sum_{t=1}^n\sum_{k=1}^n (y_t y_k + y_k y_t) - E(m(x_t))^2 = \frac{1}{2}\upsilon_n - (E(m(x_t)))^2$. By Corollary 1, since $\frac{1}{n}E(\psi_n^2(y_t, y_k)) = \frac{4}{n}E((y_t^2))^2 = o(1)$ for all $t, k, \sqrt{n}(\upsilon_n - \hat{u}_n) = o_p(1)$ where $\hat{u}_n = \frac{2}{n}\sum_{t=1}^n(\psi_{1n}(Z_t)) - \theta_n = \frac{4}{n}\sum_{t=1}^n y_t E(m(x_t)) - 2(E(m(x_t)))^2$. Hence, $\bar{y}^2 - (E(y_t))^2 = \frac{2}{n}\sum_{t=1}^n y_t E(m(x_t)) - 2(E(m(x_t)))^2 + o_p(n^{-\frac{1}{2}})$, and

$$\begin{aligned}\beta_2\frac{E(y_t - m(x_t))^2}{V(y_t)} &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{E(\sigma^2(x_t))}{V(y_t)}(y_t^2 - E(y_t^2))\right.\\ &\quad \left. -2\frac{E(\sigma^2(x_t))E(m(x_t))}{V(y_t)}(y_t - E(m(x_t)))\right\} + o_p(n^{-\frac{1}{2}}).\end{aligned}$$

For β_1, $\frac{1}{n}\sum_{t=1}^n(y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^n\left(\varepsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\varepsilon_t + (m(x_t) - \hat{m}(x_t))^2\right)$. Similarly to the proof of Theorem 2, we use part (a) of Lemma 1 with $f(x_t, y_t) = -2\varepsilon_t$ to obtain $\frac{1}{n}\sum_{t=1}^n 2(m(x_t) - \hat{m}(x_t))\varepsilon_t = o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. By part (b) with $f(x_t, y_t) = 1$ we obtain $\frac{1}{n}\sum_{t=1}^n(m(x_t) - \hat{m}(x_t))^2 = o_p(n^{-\frac{1}{2}}) + o_p(h_n^3)$. Therefore, $\frac{1}{n}\sum_{t=1}^n(y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^n \varepsilon_t^2 + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. Hence,

$$\begin{aligned}\beta_1 - \beta_2\frac{E(y_t - m(x_t))^2}{V(y_t)} &= \frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - E(\sigma^2(x_t)) - \frac{E(\sigma^2(x_t))}{V(y_t)}(y_t^2 - E(y_t^2))\\ &\quad + \frac{2E(\sigma^2(x_t))E(m(x_t))}{V(y_t)}(y_t - E(m(x_t)))\\ &\quad + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)\\ &= \frac{1}{n}\sum_{t=1}^{n}\zeta_t + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2).\end{aligned}$$

Since $E(\zeta_t) = 0$, $V(\zeta_t) = V(\varepsilon_t^2 - (1-\eta^2)(y_t - E(y_t))^2) < \infty$ and ζ_t forms an i.i.d. sequence, then by the central limit theorem we have

$$\sqrt{n}\left(\beta_1 - \beta_2\frac{E(y_t - m(x_t))^2}{V(y_t)} - o_p(h_n^2)\right) \xrightarrow{d} N(0, V(\zeta_t)). \tag{27}$$

Given that $\hat{\eta}^2 - \eta^2 = -\frac{1}{\frac{1}{n}\sum_{t=1}^n (y_t-\bar{y})^2}\left(\beta_1 - \beta_2 \frac{E(y_t-m(x_t))^2}{V(y_t)}\right)$ together with (26) and (27) gives the desired result. □

4 Conclusion

We have established the $\sqrt{n}$ asymptotic equivalence between V and U statistics when their kernels depend on n, the sample size. We combine our result with a result of Lee (1988) to obtain the $\sqrt{n}$ asymptotic equivalence between V statistics and U statistics projections. We provide a number of examples and illustrations on how our results can be used in nonparametric kernel estimation. The list of examples is obviously not exhaustive as our results can be used in much broader contexts.

Acknowledgements

We would like to thank Professor Katuomi Hirano and two anonymous referees for comments and suggestions.

References

Ahn, H., Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, *58*, 3–29.

Bönner, N., Kirschner, H.-P. (1977). Note on conditions for weak convergence of Von Mises' differentiable statistical functions. *The Annals of Statistics*, *5*, 405–407.

D'Amico, S. (2003). Density estimation and combination under model ambiguity via a pilot nonparametric estimate: an application to stock returns. Working paper, Department of Economics, Columbia University (http://www.columbia.edu/~sd445/).

Doksum, K., Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, *23*(5), 1443–1473.

Fan, J. (1992). Design adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*, 998–1004.

Fan, J., Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*, 645–660.

Fan, Y., Li, Q. (1996). Consistent model specification test: omitted variables and semiparametric functional forms. *Econometrica*, *64*(4), 865–890.

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

Hoeffding, W. (1961). The strong law of large numbers for U statistics, Institute of Statistics, Mimeo Series 302, University of North Carolina, Chapel Hill, NC.

Kemp, G. (2000). Semiparametric estimation of a logit model. Working paper, Department of Economics, University of Essex (http://www.essex.ac.uk /~ kempgcr/ papers/logitmx.pdf).

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

Lee, B. (1988). Nonparametric tests using a kernel estimation method, Doctoral dissertation, Department of Economics, University of Wisconsin, Madison, WI.

Martins-Filho, C., Yao, F. (2003). A nonparametric model of frontiers, working paper, Department of Economics, Oregon State University (http://oregonstate.edu/~martinsc/martins-filho-yao(03).pdf).

Powell, J., Stock, J., Stoker, T. (1989). Semiparametric estimation of index coefficients. *Econometrica*, *57*, 1403–1430.

Prakasa-Rao, B. L. S. (1983). Nonparametric functional estimation. Orlando, FL: Academic

Serfling, R. J. (1980). Approximation theorems of mathematical statistics. New York: Wiley.

Yao, Q., Tong, H. (2000). Nonparametric estimation of ratios of noise to signal in stochastic regression. *Statistica Sinica*, *10*, 751–770.

Zheng, J. X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, *14*, 123–138.