Statistica Sinica **25** (2015), 1145-1161
doi:http://dx.doi.org/10.5705/ss.2014.008

USING A BIMODAL KERNEL FOR A NONPARAMETRIC REGRESSION SPECIFICATION TEST

Cheolyong Park[1], Tae Yoon Kim[1], Jeongcheol Ha[1], Zhi-Ming Luo[1] and Sun Young Hwang[2]

[1]*Keimyung University and* [2]*Sookmyung Womens University*

Abstract: For a nonparametric regression model with a fixed design, we consider the model specification test based on a kernel. We find that a bimodal kernel is useful for the model specification test with a correlated error, whereas a conventional unimodal kernel is useful only for an iid error. Another finding is that the model specification test suffers from a convergence rate change depending on whether the errors are correlated or not. These results are verified by deriving an asymptotic null distribution and asymptotic (local) power, and by performing a simulation. The validity of the bimodal kernel for testing is demonstrated with the "drum roller" data (see Laslett (1994) and Altman (1994)).

Key words and phrases: bimodal kernel, convergence rate change, correlated error, nonparametric specification test.

1. Introduction

Suppose that we are concerned with the nonparametric regression function specification test given

$$Y_i = m(x_i) + \eta_i \quad (i = 1, \ldots, n), \tag{1.1}$$

where m is a smooth function defined on [0,1], $x_i = i/n$, and $\{\eta_i\}$ is a zero-mean, covariance stationary process. Here, the design points grow closer as the sample size increases; while, the error process remains the same. Refer to Kim et al. (2009) and references therein for detailed discussions about this model. The testing problem under consideration is whether $m(x)$ belongs to a specific parametric family. This can be described as

$$H_0 : m(x) = g(x, \gamma_0) \text{ for all } x \in [0,1] \text{ with some } \gamma_0 \in \mathcal{B} \subset R^q$$

versus

$$H_1 : m(x) \neq g(x, \gamma) \text{ for some } x \in [0,1] \text{ with all } \gamma \in \mathcal{B} \subset R^q.$$

For testing H_0 nonparametrically, we consider a kernel-based test statistic

$$T_n = (n^2h)^{-1} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K\Big(\frac{x_i - x_j}{h}\Big)\hat{\epsilon}_i\hat{\epsilon}_j, \tag{1.2}$$

where K is a kernel satisfying (C1) below, $\epsilon_i = Y_i - g(x_i, \gamma)$ and $\hat{\epsilon}_i = Y_i - g(x_i, \hat{\gamma})$, with a consistent estimator $\hat{\gamma}$ of γ_0. Here, T_n is based on the average squared error

$$d_A(\hat{m}, m) = n^{-1} \sum_{i=1}^{n} \left(\hat{m}(x_i) - m(x_i)\right)^2, \tag{1.3}$$

where $\hat{m}(x) = (nh)^{-1} \sum_{i=1}^{n} K((x - x_i)/h) Y_i$. In recent years, much research has been performed on applying T_n to the model specification test for a random design regression model. See, for example, Fan and Li (1999) Zheng (1996), Luo, Kim, and Song (2011), and Khmaladze and Koul (2004). For the model specification test for the fixed design regression model, not many results are available. See, e.g., Eubank and Spiegelman (1990) or the monograph by Hart (1997), which studies nonparametric lack-of-fit test with iid errors. We study T_n as a nonparametric regression specification test for the fixed design regression model, particularly when errors are correlated. The major strengths of T_n is its consistency, since the existing parametric tests fail to be consistent against all deviations from the null.

It is well known that when a nonparametric method such as $\hat{m}$ is used to recover m, correlated errors cause trouble. See Opsomer, Wang, and Yang (2001) for a detailed discussion of this. We demonstrate that an analogous size distortion problem arises for a nonparametric specification test when errors are correlated. As a possible solution, we recommend the use of bimodal kernel K with $K(0) = 0$. In addition, we find that T_n shows a power rate change, that yields a continuous but non-monotonic power function over the hypothesis domain. We proceed as follows. Section 2 proposes the specification test with bimodal kernel and demonstrates its usefulness with the "drum roller" data (Laslett (1994) and Altman (1994)). Section 3 discusses the power rate change of T_n and its impact. Some other tests are discussed there. Section 4 reports on simulations that check our theoretical results. All proofs are deferred to the Appendix.

2. Size Distortion and Bimodal Kernel

We need the following assumptions.

(C1) K is a square integrable symmetric probability density function with support $[-\kappa, \kappa]$ for some $\kappa > 0$, and K is Lipschitz continuous.

(C2) Errors are a geometrically strong mixing sequence with mean zero and $E|\eta_i|^r < \infty$ for some $r > 4$.

(C3) $nh^{3/2} \to \infty$ and $h \to 0$.

(C4) (i) $g^{(1)}(x, \cdot)$ and $g^{(2)}(x, \cdot)$ are continuous in $x \in [0, 1]$ and dominated by a bounded function $M_g(x)$, where $g^{(1)}(x, \cdot)$ and $g^{(2)}(x, \cdot)$ are the first and second partial derivatives with respect to γ, respectively. (ii) $|g^{(1)}(x,\gamma)^2| \neq 0$ for γ in a neighborhood of $\gamma_* = lim_p\hat{\gamma}$.

(C5) The mean function m supported on the interval [0,1], and has a uniformly continuous and square integrable second derivative $m''(x)$ on the interval (0,1).

Here, (C1) is a standard assumption on the kernel. If, in addition, $K(0) = 0$, the kernel is bimodal. For iid error, $r = 4$ suffices for (C2). If $\mathcal{M}_a^b$ be the σ-field generated by $\{\xi(t) : a \leq t \leq b\}$, then $\{\xi(t) : t \in R\}$ is strong mixing if

$$\alpha(\tau) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{M}_{-\infty}^0 \text{ and } B \in \mathcal{M}_\tau^\infty\} = O(\rho^\tau)$$

for some $0 < \rho < 1$ when $\tau \to \infty$. If (C4) is a standard assumption adopted in non-linear regression models: If $\gamma_* = \text{argmin}_{\gamma\in\mathcal{B}} E(Y_i - g(x_i,\gamma))^2$ and $\hat{\gamma} = \text{argmin}_{\gamma\in\mathcal{B}} \sum_{i=1}^n (Y_i - g(x_i,\gamma))^2$, then under $H_0, \gamma_* = \gamma_0$. And, uder (C4), $\gamma_* = lim_p\hat{\gamma}$ and $\hat{\gamma} - \gamma_* = O_p(n^{-1/2})$ under both H_0 and H_1. One may refer to Fan and Li (1999) for these results. (C5) is needed when H_1 holds.

Theorem 1. *Let* (C1)−(C4) *and* H_0 *hold. If* K *is a bimodal kernel with* $K(0) = 0$, *then*

$$\frac{nh^{1/2}T_n}{\hat{\sigma}_0} \to N(0,1) \tag{2.1}$$

in distribution, where $\hat{\sigma}_0^2$ *is a consistent estimator of*

$$\sigma_0^2 = \int K^2(u)du\Big[\big(E(\epsilon)^2\big)^2 + \Big(\sum_{j=-\infty}^{\infty} E(\epsilon_0\epsilon_j)\Big)^2\Big]. \tag{2.2}$$

Remark 1. Theorem 1 suggests an asymptotic one-sided test for H_0 versus H_1: Reject H_0 at the significance level α if $nh^{1/2}T_n/\hat{\sigma}_0 > z_\alpha$, where z_α is the upper α-percentile of the standard normal distribution. Because

$$Var(\sum_{i=1}^n \epsilon_i) = n(E\epsilon_0^2 + 2\sum_{j=1}^\infty E(\epsilon_0\epsilon_j)) + o(n),$$

one possible estimator of σ_0^2 is to use the block bootstrap variance estimator of $V_n = \sum_{i=1}^n \hat{\epsilon}_i$, or $Var^*(V_n)$, where $\hat{\epsilon}_i = Y_i - \hat{g}(x_i,\hat{\gamma})$:

$$\hat{\sigma}_0^2 = \int K^2(u)du[(Var^*\frac{V_n}{n})^2 + (\sum_{i=1}^n \frac{\hat{\epsilon}_i^2}{n})^2]. \tag{2.3}$$

Table 1. Testing results for drum roller data.

	T^u_{230}	T^b_{230}	T^u_{1150}	T^b_{1150}
p-value	0.088	0.988	0	0.55
Z score	1.34	-2.26	25.7792	-0.126

Remark 2. Correlated error causes size distortion to T_n when a unimodal kernel is employed and a bimodal kernel can correct it. Indeed, we can show that

$$P[nh^{1/2}T_n/\hat{\sigma}_0 > z_\alpha] = P\Big(Z \geq z_\alpha - 2h^{-1/2}K(0)\sum_{i=1}^{n-1}\frac{E\epsilon_0\epsilon_i}{\hat{\sigma}_0}\Big) + o(1), \qquad (2.4)$$

where Z is $N(0,1)$. Verification of (2.4) is given in the Appendix. It indicates that the size inflated (deflated) by $\sum_{i=1} E\epsilon_0\epsilon_i > 0 (< 0)$ leads to a frequent (infrequent) rejection of H_0 unless a bimodal kernel with $K(0) = 0$ is used, and that when a unimodal kernel is used, the size distortion problem may be avoided by over-smoothing (large h). For iid error, Theorem 1 holds trivially for a unimodal kernel because $E(\eta_0\eta_i) = 0$ for any $i \neq 0$.

In order to demonstrate the usefulness of a bimodal kernel for T_n, the "drum roller" data analyzed by Laslett (1994) and Altman (1994) are considered (the data are available on *Statlib*). As noted there, the data appear to exhibit a significant short-range positive correlation; so testing using a bimodal kernel is warranted. Figure 1 shows two fits to the two datasets using cubic B spline basis functions of order 5, where $n = 1{,}150$ represents the full dataset and $n = 230$ uses every fifth observation. We took the fit with $n = 230$ as $\hat{g}_{230}(x)$ for $x \in [0,1]$, and then, use it as $g_0(x,\gamma) = \sum_{j=1}^{5}\gamma_j B_j(x) = \hat{g}_{230}(x)$, where $B_j(x)$ is the j-th cubic B-spline basis function. Testing $H_0 : m(x) = g_0(x,\gamma)$ for $x \in [0,1]$ with $n = 230$ or $n = 1{,}150$, we employed T^u_n with the unimodal kernel and T^b_n with the bimodal kernel. The testing results are summarized in Table 1. Table 1 shows that T^u_n rejects H_0 and T^b_n accepts it regardless of size n, and that the effect of a bimodal kernel is much more significant for n =1,150 than for n = 230, indicating that the size distortion is more severe for a large n in our case.

3. Power Rate Change

We show that we reveal that T_n has different convergence rates under H_1, and that such rate change complicates power under the local alternatives near H_0.

Theorem 2. *Let* (C1)−(C5) *and* H_1 *hold. If* K *is a bimodal kernel with* $K(0) = 0$*, then*

$$\frac{n^{1/2}(T_n - \mu)}{2^{1/2}\sigma_1} \to N(0,1) \qquad (3.1)$$

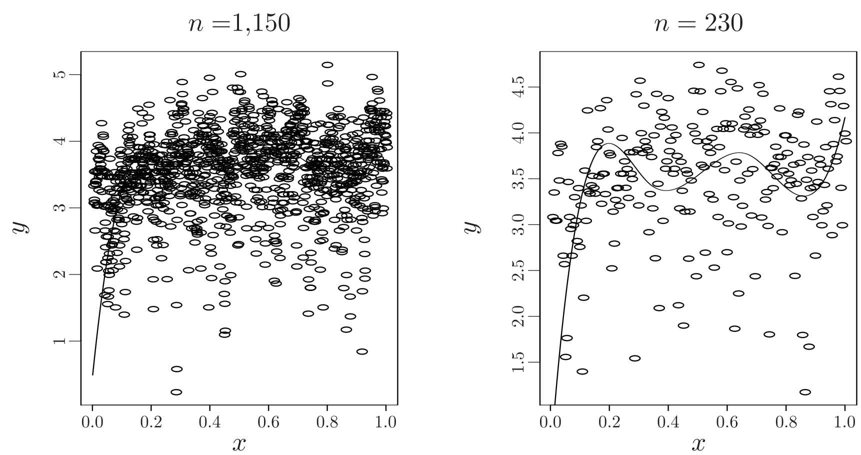

Figure 1. Scatterplots of drum roller data with fitted lines based on five bases of cubic B-splines for n =1,150 (left) and $n = 230$ (right).

in distribution, where $\mu = \int (m(x) - g(x, \tilde{\gamma}))^2 dx$ *and*

$$\sigma_1^2 = \sum_{i=-\infty}^{\infty} Cov[(Y_0 - m(x_0))(g(x_0, \tilde{\gamma}) - m(x_0)), (Y_i - m(x_i))(g(x_i, \tilde{\gamma}) - m(x_i))].$$

Theorems 1 and 2 have the test convergence rate at $nh^{1/2}$ under H_0, and $n^{1/2}$ under H_1. From Theorem 2, the test is consistent with rate $n^{1/2}$, not $nh^{1/2}$, as suggested by Theorem 1. As $n \to \infty$, $P\left(Z \geq n^{1/2}(2^{1/2}\sigma_1)^{-1}(z_\alpha \hat{\sigma_0}(nh^{1/2})^{-1} - \mu)\right) \to 1$. This is related to the fact that T_n is a good approximation to $d_A(\hat{m}, m)$. Since d_A is decomposed as bias and variance components, and bias part goes to zero under H_0 but remains as a constant under H_1, the constant dominating d_A (and hence T_n) under H_1 forces the rate change.

We investigate the power rate change issue at a finer level of the local alternatives. Assume a sequence of local alternatives $H_{1n} : m(x_t) = g(x_t, \gamma_0) + \delta_n l(x_t)$, where the known function $l(\cdot)$ is continuously differentiable and bounded by an integrable function $M(\cdot)$. We have $\sigma_1 \sim \delta_n$ under H_{1n}.

Theorem 3. *Let* (C1)−(C5) *and* H_{1n} *hold. If* K *is a bimodal kernel with* $K(0) = 0$*, then*

(i) *If* $nh\delta_n^2 \to 0$, $(nh^{1/2})(T_n - \delta_n^2 \mu_1)/\sigma_0 \to N(0, 1)$, *where* $\mu_1 = \int l^2(x)dx$.

(ii) *If* $nh\delta_n^2 \to \infty$, $n^{1/2}\delta_n^{-1}(T_n - \delta_n^2\mu_1)/(2^{1/2}\sigma_2) \to N(0, 1)$, *where* $\sigma_2^2 = \sum_{i=-\infty}^{\infty} Cov[(Y_0 - m(x_0))l(x_0), (Y_i - m(x_i))l(x_i)]$.

(iii) *If* $nh\delta_n^2 \to \lambda > 0$, $nh^{1/2}(T_n - \lambda(nh)^{-1}\mu_1)/(\sigma_0^2 + 2\sigma_2^2/\lambda)^{1/2} \to N(0, 1)$.

Remark 3. From Theorem 3, when the local alternative δ_n is of the rate $n^{-1/2}h^{-1/4}$, the test starts to have (local) power. The power function or rate, say

s_n, is a continuous, but not monotonic, function of δ_n. This indicates that no particular asymptotic discontinuity is caused by the power rate change. Verification of these facts is given in the Appendix.

Remark 4. Hart (1997) considered an analogy to T_n with iid error and established a result corresponding to our Theorem 1. He also verified that when local alternative δ_n is of the rate $n^{-1/2}h^{-1/4}$, the test starts to have (local) power, but did not address the rate change and size distortion due to dependence. According to Hart (1997) (see Section 6.2.1 there), if one employs the sup norm, it would be possible to consider the test statistic

$$R_n = \sup_{x\in[0,1]} \left| (nh\hat{\sigma}_r^2)^{-1/2} \sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big)\hat{\epsilon}_i \right|, \tag{3.2}$$

where $\hat{\sigma}_r^2$ is an estimator of $\sigma_r^2 = \int K^2(u)du \sum_{-\infty}^{\infty} E(\epsilon_0\epsilon_j)$, but R_n would suffer from the inflated size distortion when errors are correlated. Glesser and Moore (1983) mentioned that tests based on empiric distribution function are subject to size distortion due to positive dependence, for example. Some numerical work on this point is given in Table 5 of Section 4.

Remark 5. Following the approach of Khmaladze and Koul (2004), it would be possible to consider a goodness-of-fit test for our problem,

$$\xi_n(y) = n^{-1/2} \sum_{i=1}^{n} I(x_i \in B)[I\{\hat{\epsilon}_i \le y\} - F_\epsilon(y)],$$

where $B \subset [0,1]$ and F_ϵ is the distribution function of ϵ if the underlying distribution of F_ϵ. They show that ξ_n is an asymptotic distribution-free goodness-of-fit test and derive Brownian motion as its asymptotic distribution under iid conditions. If one replaces the indicator I by the kernel K in $\xi_n(y)$ above, then we have

$$W_n(y) = (nh)^{-1/2} \sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big)\Big[L\Big(\frac{y-\hat{\epsilon}_i}{h}\Big) - F_\epsilon(y)\Big],$$

where $L(y) = \int_{-\infty}^{y} K(x)dx$. If the sup norm is applied to $W_n(y)$ for obtaining goodness-of-fit test under positive dependence, the test would again suffer from size distortion.

4. Simulation

In this section, we show through simulations that the size and power of T_n are affected by correlated error and furthermore, that such distortions of T_n might be effectively handled by a bimodal kernel. We recommend the use of such a kernel for a nonparametric regression specification test. As before, T_n^u and T_n^b denote the test statistic T_n with a unimodal kernel and a bimodal kernel, respectively.

The simulated regression setting was concerned with testing

$$H_0 : m(x) = 300x^3(1-x)^3 \text{ for all } x \in [0,1] \text{ versus,}$$
$$H_1 : m(x) \neq \gamma x^3(1-x)^3 \text{ for some } x \in [0,1] \text{ with all } \gamma \in R.$$

The regression errors for the model $Y_j = m(x_j) + \epsilon_j$ for $j = 1, \ldots, n$ were produced by an $AR(1)$ process $\epsilon_j = \phi\epsilon_{j-1} + \sqrt{1-\phi^2}Z_j$, where $x_j = j/n$ with $n = 100, 400, 800$, Z_j's was pseudo iid normal random variables $N(0,1)$, and ϵ_1 was $N(0,1)$. The $AR(1)$ parameters $\phi = -0.95, -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9, 0.95$. Here $\phi = -0.0.95$ and 0.95 were added in order to consider severely correlated errors. The kernel functions used were $K(x) = 630(4x^2-1)^2x^4I(-1/2 \leq x \leq 1/2)$ as a bimodal kernel for T_n^b, and $K(x) = (15/16)(1-x^2)^2I(-1 \leq x \leq 1)$ as a unimodal kernel for T_n^u. In addition, block bootstrap estimate $\hat{\sigma}_0^2$ with block length $n^{1/3}$ (refer to (2.3)) was employed. Thus, our test rejected H_0 if $nh^{1/2}T_n/\hat{\sigma}_0 > z_\alpha$. Table 2 provides the simulation results regarding the size distortion and its correction by a bimodal kernel. Tables 3 and 4 provide simulation results regarding how the power of T_n is affected by correlated errors when $m(x) = 300x^3(1-x)^3 + 0.5$ or $m(x) = 300x^3(1-x)^3 + 1$.

Table 2 calculates the size of T_n^u and T_n^b for various values of ϕ when H_0 is true. Indeed, we generated data from $Y_j = \gamma x_j^3(1-x_j)^3 + \epsilon_j$ for $j = 1, \ldots, n$, where $\gamma = 300$ and γ is estimated via least squares, assuming $m(x) = \gamma x^3(1-x)^3$. From Table 2, one sees that T_n^u accepts H_0 almost always when the errors are correlated negatively (i.e., $\phi < 0$), whereas it rejects H_0 too often when $\phi > 0.3$. Further, the size distortion of T_n^u becomes more severe as the errors becomes more severely correlated. This verifies the size distortion of T_n^u due to the correlated errors. Table 2 also confirms that T_n^b corrects the size distortion reasonably well when the errors are positively correlated ($\phi > 0.3$), as suggested by Theorem 1 and (2.4). When errors are negatively correlated or severely positively correlated at small n, T_n^b does not provide a sufficient correction to the size distortion. In addition, there is not so much difference between T_n^u and T_n^b in simulated size for iid errors or $\phi = 0$.

Table 3 calculates the powers of T_n^u and T_n^b for various values of ϕ when H_1 is true or when $m(x) = 300x^3(1-x)^3 + 0.5$. From Table 3, one may observe that T_n^u rejects $H_0 : m(x) = 300x^3(1-x)^3$ less frequently and T_n^b improves it when $\phi \leq -0.7$ at $n = 100$. Such improvements disappear as n increases or ϕ increases to zero. When ϕ is positive, T_n^u outdoes T_n^b significantly in power, which suggests that T_n^b corrects the inflated power of T_n^u at the cost of the reduced power. One sees that such ineludible adjustments of T_n^b remain strong across n as ϕ gets close to 1. Table 4 calculates the powers of T_n^u and T_n^b by considering a more distant

Table 2. Simulated size (%) for T_n^u with a unimodal kernel and T_n^b with a bimodal kernel at size $\alpha = 0.05$ when $m(x) = 300x^3(1-x)^3$ for $x \in [0,1]$ and $h = n^{-1/5}$.

ϕ	T_{100}^u	T_{100}^b	T_{200}^u	T_{200}^b	T_{400}^u	T_{400}^b	T_{800}^u	T_{800}^b
-0.95	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.5	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
-0.4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.3	0.001	0.001	0.000	0.005	0.000	0.003	0.000	0.000
-0.2	0.002	0.002	0.004	0.002	0.003	0.001	0.002	0.002
-0.1	0.004	0.011	0.010	0.008	0.005	0.007	0.007	0.007
0.0	0.017	0.022	0.019	0.009	0.015	0.013	0.022	0.022
0.1	0.042	0.026	0.047	0.034	0.051	0.040	0.068	0.037
0.2	0.077	0.040	0.083	0.057	0.088	0.037	0.120	0.064
0.3	0.121	0.072	0.132	0.072	0.170	0.103	0.202	0.092
0.4	0.215	0.088	0.242	0.120	0.250	0.125	0.306	0.135
0.5	0.307	0.138	0.373	0.166	0.372	0.149	0.467	0.174
0.6	0.460	0.181	0.476	0.188	0.531	0.190	0.598	0.232
0.7	0.606	0.216	0.658	0.222	0.713	0.253	0.769	0.268
0.8	0.785	0.301	0.817	0.282	0.848	0.280	0.905	0.342
0.9	0.928	0.494	0.941	0.429	0.957	0.360	0.982	0.342
0.95	0.974	0.734	0.995	0.691	0.997	0.513	0.995	0.412

Table 3. Simulated power (%) for T_n^u with a unimodal kernel and T_n^b with a bimodal kernel at size $\alpha = 0.05$ when $m(x) = 300x^3(1-x)^3 + 0.5$ for $x \in [0,1]$ and $h = n^{-1/5}$.

ϕ	T_{100}^u	T_{100}^b	T_{200}^u	T_{200}^b	T_{400}^u	T_{400}^b	T_{800}^u	T_{800}^b
-0.95	0.458	0.553	0.928	0.969	1.000	1.000	1.000	1.000
-0.9	0.409	0.563	0.966	0.992	1.000	1.000	1.000	1.000
-0.8	0.368	0.491	0.981	0.994	1.000	1.000	1.000	1.000
-0.7	0.363	0.433	0.991	0.992	1.000	1.000	1.000	1.000
-0.6	0.426	0.426	0.985	0.984	1.000	1.000	1.000	1.000
-0.5	0.403	0.425	0.970	0.961	1.000	1.000	1.000	1.000
-0.4	0.457	0.393	0.968	0.948	1.000	1.000	1.000	1.000
-0.3	0.484	0.373	0.952	0.891	1.000	0.999	1.000	1.000
-0.2	0.536	0.363	0.937	0.875	0.999	0.999	1.000	1.000
-0.1	0.495	0.358	0.933	0.829	0.999	0.996	1.000	1.000
0.0	0.566	0.366	0.928	0.796	1.000	0.994	1.000	1.000
0.1	0.573	0.367	0.928	0.737	1.000	0.976	1.000	1.000
0.2	0.578	0.348	0.913	0.707	0.995	0.960	1.000	0.999
0.3	0.662	0.352	0.895	0.660	0.991	0.946	1.000	1.000
0.4	0.641	0.351	0.891	0.597	0.993	0.888	1.000	0.995
0.5	0.684	0.352	0.889	0.559	0.992	0.850	1.000	0.980
0.6	0.722	0.375	0.907	0.545	0.986	0.798	0.999	0.944
0.7	0.785	0.371	0.896	0.495	0.976	0.711	0.999	0.906
0.8	0.856	0.447	0.925	0.487	0.977	0.649	0.998	0.802
0.9	0.953	0.624	0.968	0.548	0.983	0.557	0.997	0.672
0.95	0.982	0.798	0.995	0.731	0.999	0.658	0.999	0.590

Table 4. Simulated power (%) for T_n^u with a unimodal kernel and T_n^b with a bimodal kernel at size $\alpha = 0.05$ when $m(x) = 300x^3(1-x)^3 + 1$ for $x \in [0,1]$ and $h = n^{-1/5}$.

ϕ	T_{100}^u	T_{100}^b	T_{200}^u	T_{200}^b	T_{400}^u	T_{400}^b	T_{800}^u	T_{800}^b
-0.95	0.995	0.994	1.000	1.000	1.000	1.000	1.000	1.000
-0.9	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
-0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.4	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.2	1.000	0.994	1.000	1.000	1.000	1.000	1.000	1.000
-0.1	0.999	0.992	1.000	1.000	1.000	1.000	1.000	1.000
0.0	0.999	0.980	1.000	1.000	1.000	1.000	1.000	1.000
0.1	0.998	0.972	1.000	1.000	1.000	1.000	1.000	1.000
0.2	0.995	0.938	1.000	0.999	1.000	1.000	1.000	1.000
0.3	0.986	0.895	1.000	0.997	1.000	1.000	1.000	1.000
0.4	0.986	0.864	1.000	0.992	1.000	1.000	1.000	1.000
0.5	0.979	0.802	1.000	0.977	1.000	1.000	1.000	1.000
0.6	0.976	0.780	1.000	0.951	1.000	0.997	1.000	1.000
0.7	0.972	0.708	0.997	0.893	1.000	0.991	1.000	0.999
0.8	0.962	0.699	0.995	0.841	1.000	0.962	1.000	0.998
0.9	0.984	0.764	0.997	0.829	0.999	0.854	1.000	0.967
0.95	0.990	0.884	0.998	0.868	0.998	0.870	1.000	0.899

Table 5. Simulated mean and standard deviation for $R_n = \max_{j=1,\ldots,n} \left|(nh\hat{\sigma}_r^2)^{-1/2}\sum_{i=1}^n K((x_j - x_i)/h)\hat{\epsilon}_i\right|$ with a unimodal kernel when $m(x) = 300x^3(1-x)^3$ for $x \in [0,1]$ and $h = n^{-1/5}$.

ϕ	$n = 100$		$n = 200$		$n = 400$		$n = 800$	
-0.95	40.90	18.81	41.77	17.24	43.30	17.93	46.06	17.55
-0.9	14.43	6.32	14.83	6.54	15.74	6.22	16.82	6.37
-0.8	5.22	2.31	5.60	2.33	5.93	2.44	6.31	2.51
-0.7	3.07	1.39	3.22	1.39	3.59	1.42	3.74	1.43
-0.6	2.15	0.98	2.29	1.02	2.53	1.05	2.67	1.05
-0.5	1.71	0.77	1.83	0.79	1.90	0.78	2.10	0.81
-0.4	1.44	0.68	1.57	0.67	1.69	0.69	1.74	0.69
-0.3	1.27	0.59	1.37	0.59	1.45	0.61	1.58	0.62
-0.2	1.17	0.53	1.28	0.56	1.35	0.55	1.45	0.56
-0.1	1.11	0.51	1.18	0.53	1.31	0.55	1.39	0.54
0.0	1.10	0.51	1.19	0.52	1.29	0.54	1.36	0.52
0.1	1.09	0.53	1.23	0.53	1.30	0.54	1.35	0.52
0.2	1.18	0.54	1.25	0.56	1.34	0.56	1.47	0.59
0.3	1.27	0.60	1.33	0.58	1.50	0.62	1.55	0.61
0.4	1.40	0.63	1.54	0.66	1.65	0.69	1.79	0.67
0.5	1.71	0.79	1.77	0.80	1.94	0.81	2.07	0.80
0.6	2.13	0.99	2.31	0.98	2.43	1.00	2.68	1.05
0.7	2.94	1.32	3.19	1.41	3.45	1.40	3.70	1.46
0.8	4.78	2.26	5.33	2.33	5.88	2.39	6.25	2.44
0.9	11.02	5.33	13.17	5.93	15.03	6.42	16.13	6.40
0.95	23.64	11.81	31.43	15.23	38.30	17.12	43.67	17.27

$m(x) = 300x^3(1-x)^3 + 1$ as H_1. From Table 4, the overall adjustments made by T_n^b tend to disappear as n increases. From Tables 3 and 4, one can infer that, as we have more distant m and large n, T_n^u and T_n^b achieve similar powers whether the errors are correlated or not. Conclusively, our simulation recommends the use of T_n^b because it corrects the size distortion due to correlated error reasonably well, and performs similarly to T_n^u for distant H_1 and large n, irrespective of error correlatedness.

Table 5 presents the mean and standard deviation of

$$R_n = \max_{j=1,\ldots,n} \left| (nh\hat{\sigma}_r^2)^{-1/2} \sum_{i=1}^{n} K\Big(\frac{x_j - x_i}{h}\Big)\hat{\epsilon}_i \right|$$

for various values of ϕ and $n = 100, 200, 400$, and 800 when H_0 is true or $m(x) = 300x^3(1-x)^3$. There, as $|\phi|$ increases from 0 to 0.95 (or dependence gets strong), both mean and standard deviation of R_n clearly increase. This shows strong effect of dependency on R_n and hence likely size distortion, as discussed in Remark 4. Also as n increases, the mean and standard deviation of R_n tend to grow.

Acknowledgement

TY Kim's research was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2011-314-C00046) and by the 2012 Bisa research grant of Keimyung University. J Ha's research has been supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning(2011-0014741). S.Y. Hwang's work was supported by a grant from the National Research Foundation of Korea (NRF - 2012012872).

Appendix

Verification of (2.4). Assume that H_0 is true and let

$$T_{1n} = (n^2h)^{-1} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_{ij}\epsilon_i\epsilon_j,$$

where $i, j = 1, \ldots, n, K_{ij} = K((i-j)/nh)$; then, from the proof of Theorem 1 below, it suffices to check that (2.4) holds for T_{1n}. By using (C1)$-$(C3) and Lemma A.1 below

$$\begin{aligned} T_{1n} &= (n^2h)^{-1}\Big\{ \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_{ij}E\epsilon_i\epsilon_j \Big\} \\ &= (n^2h)^{-1}\Big\{ \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + K(0) \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} E\epsilon_i\epsilon_j \end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} [K_{ij} - K(0)] E\epsilon_i\epsilon_j \Big\} \\
&= (n^2h)^{-1} \Big\{ \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + K(0) \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} E\epsilon_i\epsilon_j \Big\} \\
&\quad + O\Big((n^3h^2)^{-1} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} |i-j| E\epsilon_i\epsilon_j \Big) \\
&= (n^2h)^{-1} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + 2(nh)^{-1} K(0) \sum_{i=1}^{n-1} E\epsilon_0\epsilon_i + O((n^2h^2)^{-1}) \\
&= (n^2h)^{-1} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + 2(nh)^{-1} K(0) \sum_{i=1}^{n-1} E\epsilon_0\epsilon_i + o((nh^{1/2})^{-1}) \\
&= U_{1n} + 2(nh)^{-1} K(0) \sum_{i=1}^{n-1} E\epsilon_0\epsilon_i + o((nh^{1/2})^{-1}).
\end{aligned}
$$

Since U_{n1} can be shown to converges to Z in distribution (see the proof of Theorem 1), we have

$$
P\Big[\frac{nh^{1/2}T_n}{\hat{\sigma}_0} > z_\alpha \Big] = P\Big(Z \geq z_\alpha - 2h^{-1/2}K(0) \sum_{i=1}^{n-1} \frac{E\epsilon_0\epsilon_i}{\hat{\sigma}_0} \Big) + o(1).
$$

<u>Verification of Remark 3</u>. If $\delta_n = (n^{-1/2}h^{-1/4})^{\epsilon}$ for some ϵ, then by (i) of Theorem 3 we have

$$
\begin{aligned}
p_{n0} &= P\Big[\frac{nh^{1/2}T_n}{\hat{\sigma}_0} > z_\alpha \Big] = P\Big(Z \geq z_\alpha - \frac{nh^{1/2}\delta_n^2\mu_1}{\hat{\sigma}_0} \Big) \\
&= P\Big(Z \geq z_\alpha - \frac{(nh^{1/2})^{(1-\epsilon)}\mu_1}{\hat{\sigma}_0} \Big) \to \begin{cases} 1, & \epsilon_0 < \epsilon < 1, \\ P\left(Z \geq z_\alpha - \frac{\mu_1}{\sigma_0} \right), & \epsilon = 1, \\ P\left(Z \geq z_\alpha \right), & \epsilon > 1, \end{cases} \qquad \text{(A.1)}
\end{aligned}
$$

where $0 < \epsilon_0 = (1-\eta)/(1-\eta/2)$ if $h = n^{-\eta}$ for some $0 < \eta < 1$. Note that $nh\delta_n^2 \to \lambda > 0$ when $\epsilon = \epsilon_0$ and that $nh\delta_n^2 \to \infty$ when $0 < \epsilon < \epsilon_0$. Thus, (A.1) indicates that if the local alternative is of a rate slower than $n^{-1/2}h^{-1/4}$ (or $\epsilon_0 < \epsilon < 1$), the test or p_{n0} has asymptotic power 1 with rate $s_n = (nh^{1/2})^{1-\epsilon}$. If the local alternative is of a rate faster than $n^{-1/2}h^{-1/4}$ (or $1 < \epsilon$), the test has trivial power.

In order to prove that the power rate s_n is a continuous, but a nonmonotonic function of δ_n, we observe that if $0 < \epsilon < \epsilon_0$, then by (ii) of Theorem 3, we have

$$
p_{n0} = P\Big[\frac{nh^{1/2}T_n}{\hat{\sigma}_0} > z_\alpha \Big] = P\Big(Z \geq (2^{1/2}\sigma_2)^{-1}(z_\alpha\hat{\sigma}_0(n^{1/2}h^{1/2}\delta_n)^{-1} - n^{1/2}\delta_n\mu_1) \Big)
$$

$$= P\Big(Z \geq (2^{1/2}\sigma_2)^{-1}(z_\alpha \hat{\sigma}_0 (n^{1/2}h^{1/4})^{(1-\epsilon)} h^{1/4} - n^{(1-\epsilon)/2} h^{-\epsilon/4}\mu_1)\Big) \to 1. \quad \text{(A.2)}$$

If $\epsilon = \epsilon_0$, then by (iii) of Theorem 3, we have

$$p_{n0} = P\Big[\frac{nh^{1/2}T_n}{\hat{\sigma}_0} > z_\alpha\Big] = P\Big(Z \geq \Big(\frac{2\sigma_2^2}{\lambda} + \sigma_0^2\Big)^{-1/2}(z_\alpha \hat{\sigma}_0 - \lambda h^{-1/2}\mu_1)\Big) \to 1. \quad \text{(A.3)}$$

Here p_{n0} has power rate $s_n = h^{-1/2}$ under the local alternative rate of δ_n with $\epsilon = \epsilon_0$ and, in addition, it has power rate $s_n = n^{1/2}(nh)^{-\epsilon/2}$ under the local alternative rate δ_n with $0 < \epsilon < 1$. Thus, (A.1)−(A.3) may be summarized in terms of power rate s_n as follows;

$$s_n = \begin{cases} 0 & \epsilon \geq 1; \\ (nh^{1/2})^{(1-\epsilon)} = n^{(1-\eta/2)(1-\epsilon)}, & \epsilon_0 = \frac{1-\eta}{1-\eta/2} < \epsilon < 1; \\ h^{-1/2} = n^{-\eta/2}, & \epsilon = \epsilon_0; \\ n^{(1-\epsilon)/2}h^{-\epsilon/4} = n^{1/2-(1-\eta/2)\epsilon/2} & 0 < \epsilon < \epsilon_0; \\ n^{1/2} & \epsilon = 0. \end{cases} \quad \text{(A.4)}$$

From (A.4), as ϵ decreases to ϵ_0 (or the rate of the local alternative slows down), the power rate s_n slows down to $h^{-1/2}$ (or $\epsilon = \epsilon_0$) and then increases to $n^{1/2}$.

Proof of Theorem 1. Under H_0, $\hat{\epsilon}_i = \epsilon_i - [g(x_i, \hat{\gamma}) - g(x_i, \gamma_0)]$, and we can rewrite T_n as

$$\begin{aligned} T_n &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \epsilon_i \epsilon_j K_{ij} - \frac{2}{n^2 h}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} [g(x_i, \hat{\gamma}) - g(x_i, \gamma_0)]\epsilon_j K_{ij} \\ &\quad + \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} [g(x_i, \hat{\gamma}) - g(x_i, \gamma_0)][g(x_j, \hat{\gamma}) - g(x_j, \gamma_0)]K_{ij} \\ &= T_{1n} - 2T_{2n} + T_{3n}, \end{aligned} \quad \text{(A.5)}$$

where $K_{ij} = K(\frac{i-j}{nh})$. We prove Theorem by showing that (i) $nh^{1/2}T_{1n} \to N(0, \sigma_0^2)$ in distribution; (ii) $T_{2n} = o_p((nh^{1/2})^{-1})$; (iii) $T_{3n} = O_p(n^{-1})$.

Proof of (i). First note that

$$\begin{aligned} T_{1n} &= (n^2h)^{-1}\Big\{\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} K_{ij}E\epsilon_i\epsilon_j\Big\} \\ &= (n^2h)^{-1}\Big\{\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} [K_{ij} - K(0)]E\epsilon_i\epsilon_j\Big\} \\ &= (n^2h)^{-1}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + O((n^3h^2)^{-1}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} |i-j|E\epsilon_i\epsilon_j) \end{aligned}$$

$$= (n^2h)^{-1}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + O((n^2h^2)^{-1})$$
$$= (n^2h)^{-1}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}[\epsilon_i\epsilon_j - E\epsilon_i\epsilon_j] + o((nh^{1/2})^{-1}). \qquad \text{(A.6)}$$

We have used $K(0) = 0$, (A.1)−(A.3), and Lemma A.1. Let

$$U_{1n} = (n^2h)^{-1}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}\epsilon_i\epsilon_j.$$

Then the proof of (i) is complete if one shows

$$(n^2h)^{1/2}[U_{1n} - EU_{1n}] \to N(0, \sigma_0^2). \qquad \text{(A.7)}$$

We sketch the proof because its details are established by following the proof of Theorem 2 of Kim et al. (2014) which is a CLT for reduced U statistics under dependence. Here U_{1n} is basically a reduced degenerate U statistics in its structure because K is compactly supported by (A.1) and $K(0) = 0$. The reduced U statistics is defined as

$$U_{nr} = \frac{\sum_{1\le|i-j|\le\kappa_n}\Psi(Z_i, Z_j)}{N(\kappa_n)}, \qquad \text{(A.8)}$$

where Ψ is a symmetric function, $N(\kappa_n)$ is the number of distinct pairs satisfying $1 \le |i-j| \le \kappa_n$ and $1 \le \kappa_n \le n$. Following the verification of (15) of Lemma 1 of Kim et al. (2014), we can obtain the variance of n^2hU_{1n} as follows.

$$Var(\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}\epsilon_i\epsilon_j)$$
$$= \sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}^2E(\epsilon_0^2)E(\epsilon_1^2) + \sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{k=1,k\neq i,j}^n K_{ij}K_{ik}E(\epsilon_0^2)E(\epsilon_j\epsilon_k)$$
$$+ \sum_{\text{all different indices},i,j,k,l} K_{ij}K_{kl}E(\epsilon_i\epsilon_k)E(\epsilon_j\epsilon_l).$$

Refer to σ_3^2 of (2) of Kim et al. (2014). Using

$$\sum_{i=1}^n\sum_{j=1,j\neq i}^n K_{ij}^2E(\epsilon_0^2)E(\epsilon_1^2) = 2n^2h\int K^2[E(\epsilon_0^2)]^2 + o(n^2h),$$
$$\sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{k=1,k\neq i,j}^n K_{ij}K_{ik}E(\epsilon_0^2)E(\epsilon_j\epsilon_k)$$

$$= 4n^2h \int K^2 E(\epsilon_0^2) \sum_j E(\epsilon_0\epsilon_j) + o(n^2h),$$

$$\sum_{all\ different\ indices, i,j,k,l} K_{ij}K_{kl}E(\epsilon_i\epsilon_k)E(\epsilon_j\epsilon_l)$$

$$= 8n^2h \int K^2 \sum_i E(\epsilon_0\epsilon_i) \sum_j E(\epsilon_0\epsilon_j) + o(n^2h),$$

we have

$$Var\Big(\sum_{i=1}^n \sum_{j=1, j\neq i}^n K_{ij}\epsilon_i\epsilon_j\Big) = n^2h \int K^2(u)du \Big[(E(\epsilon)^2)^2 + (\sum_{j=-\infty}^{\infty} E(\epsilon_0\epsilon_j))^2 + o(n^2h)\Big].$$

For establishing a CLT for $nh^{1/2}T_{1n}$, one can follow the proof of Theorem 2 of Kim et al. (2014). Indeed, since

$$E(\sum_{i=1}^n \sum_{j=1, j\neq i}^n K_{ij}\epsilon_i\epsilon_j)^r = \sum_{i_1}\sum_{i_2}\cdots\sum_{i_{2r}} K_{i_1i_2}\cdots K_{i_{2r-1}i_{2r}} E(\epsilon_{i_1}\epsilon_{i_2}\cdots\epsilon_{i_{2r-1}}\epsilon_{i_{2r}}),$$

the proof of Theorem 2 of Kim et al. (2014) applies in a straightforward fashion, and the proof of (i) is complete.

Proof of (ii). Using $g(x_t,\hat{\gamma}) - g(x_t,\gamma_0) = g^{(1)}(x_t,\gamma_0)(\hat{\gamma}-\gamma_0) + 1/2g^{(2)}(x_t,\tilde{\gamma})(\hat{\gamma}-\gamma_0)^2$, where $\tilde{\gamma}$ is between $\hat{\gamma}$ and γ_0, we get

$$T_{2n} = \frac{1}{n^2h}\Big((\hat{\gamma}-\gamma_0)\sum\sum_{t\neq s} \epsilon_s g^{(1)}(x_t,\gamma_0)K_{ts} + (\hat{\gamma}-\gamma_0)^2 \sum\sum_{t\neq s} \epsilon_s g^{(2)}(x_t,\tilde{\gamma})\frac{K_{ts}}{2}\Big).$$

Then it is not hard to check that, under the conditions of the Theorem,

$$\sum\sum_{t\neq s} \epsilon_s g^{(1)}(x_t,\gamma_0)K_{ts} = O_p(n^{3/2}h) \text{ and } \sum\sum_{t\neq s} \epsilon_s g^{(2)}(x_t,\tilde{\gamma})K_{ts} = O_p(n^{3/2}h). \tag{A.9}$$

Also refer to Theorem 2 of Kim, Luo, and Ha (2012). Thus

$$\frac{1}{n^2h}(\hat{\gamma}-\gamma_0)\sum\sum_{t\neq s} \epsilon_s g^{(1)}(x_t,\gamma_0)K_{ts} = O_p(n^{-1}) = o_p((nh^{1/2})^{-1})$$

$$\frac{1}{n^2h}(\hat{\gamma}-\gamma_0)^2\sum\sum_{t\neq s} \epsilon_s g^{(2)}(x_t,\tilde{\gamma})K_{ts} = O_p(n^{-1}) = o_p((nh^{1/2})^{-1}).$$

We have also used $(\hat{\gamma}-\gamma_0) = O_p(n^{-1/2})$ and H_0. This completes the proof of (ii).

Proof of (iii). This follows from the Mean Value Theorem, (A4)(i), and the fact that $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$. It is trivial to check that $\sum\sum_{s\neq t} M_g(x_t)M_g(x_s)K_{st} = O(n^2h)$. This completes the proof of (iii).

Proof of Theorem 2. Since, under H_1, $\hat{\epsilon}_i = \theta_i - [g(x_i,\hat{\gamma}) - g(x_i,\gamma_*)]$ where $\theta_i = Y_i - g(x_i,\gamma_*) = \eta_i + m(x_i) - g(x_i,\gamma_*)$, we can rewrite T_n as

$$\begin{aligned} T_n &= \frac{1}{n^2h}\sum\sum_{i\neq j}\theta_i\theta_j K_{ij} - \frac{2}{n^2h}\sum\sum_{i\neq j}[g(x_i,\hat{\gamma}) - g(x_i,\gamma_*)]\theta_i K_{ij} \\ &\quad + \frac{1}{n^2h}\sum\sum_{i\neq j}[g(x_i,\hat{\gamma}) - g(x_i,\gamma_*)][g(x_j,\hat{\gamma}) - g(x_j,\gamma_*)]K_{ij} \\ &= T_{1n} - 2T_{2n} + T_{3n}. \end{aligned} \tag{A.10}$$

Now, T_{1n} may be rewritten as $T_{1n} = T_{11n} + 2T_{12n} + T_{13n}$ where

$$\begin{aligned} T_{11n} &= \frac{1}{n^2h}\sum\sum_{i\neq j}K_{ij}(\theta_i - E\theta_i)(\theta_j - E\theta_j), \\ T_{12n} &= \frac{1}{n^2h}\sum\sum_{i\neq j}K_{ij}(\theta_i - E\theta_i)E\theta_j, \\ T_{13n} &= \frac{1}{n^2h}\sum\sum_{i\neq j}K_{ij}E\theta_i E\theta_j, \end{aligned}$$

and $E\theta_i = m(x_i) - g(x_i,\gamma_*)$. Now one can show that under the conditions of the Theorem

$$T_{11n} = O_p(n^{-1}h^{-1/2}) = o_p(n^{-1/2}), \tag{A.11}$$

$$T_{12n} = \frac{1}{n}\sum_i \eta_i(m(x_i) - g(x_i,\gamma_*)) + o_p(n^{-1/2}), \tag{A.12}$$

$$T_{13n} = \int [m(x) - g(x,\gamma_*)]^2 dx + o(1), \tag{A.13}$$

by using Theorem 2 of Kim, Luo, and Ha (2012). Application of the CLT for triangular array of random variables (see Lemma A.2 of Kim et al. (2014)) yields

$$n^{1/2}(\sigma_1)^{-1}T_{12n} \to N(0,1) \tag{A.14}$$

in distribution where

$$\sigma_1^2 = \sum_{i=-\infty}^{\infty} Cov[(Y_0 - m(x_0))(g(x_0,\tilde{\gamma}) - m(x_0)), (Y_i - m(x_i))(g(x_i,\tilde{\gamma}) - m(x_i))].$$

From the above results we have $n^{1/2}(2\sigma_1)^{-1}T_{1n} \to N(\mu,1)$ where $\mu = \int[m(x) - g(x,\gamma_*)]^2dx$. Using $\hat{\gamma} - \gamma_* = o_p(1)$ under H_1, it is easy to see that $T_{2n} =$

$o_p(n^{-1/2})$ and $T_{3n} = o_p(n^{-1/2})$, which proves $n^{1/2}(2\sigma_1)^{-1}T_n \to N(\mu, 1)$ where $\mu = \int[m(x) - g(x, \gamma_*)]^2 dx$.

Proof of Theorem 3. (i) If $nh\delta_n^2 \to 0$, then $nh^{1/2}T_{12n} = O_p(n^{1/2}h^{1/2}\delta_n) = o_p(1)$ and

$$nh^{1/2}T_{13n} = nh^{1/2}\int[m(x) - g(x, \gamma_*)]^2 dx + o(1) = nh^{1/2}\delta_n^2\mu_1 + o(1).$$

Application of Theorem 2 of Kim, Luo, and Ha (2012) yields $nh^{1/2}T_{11n} \to N(0, \sigma_0^2)$. Thus under H_{1n} we have $nh^{1/2}(T_n - \delta_n^2\mu_1) \to N(0, \sigma_0^2)$ in distribution.

(ii) If $nh\delta_n^2 \to \infty$, then $n^{1/2}\delta_n^{-1}T_{11n} = o_p(1)$ since $T_{11n} = O_p(n^{-1}h^{-1/2})$. Application of CLT for triangular array of random variables (see Lemma A.2 of Kim et al. (2014)) to T_{12n} yields $n^{1/2}\delta_n^{-1}T_{12n} \to N(0, \sigma_2^2)$ where

$$\sigma_2^2 = \sum_{i=-\infty}^{\infty} Cov[(Y_0 - m(x_0))l(x_0), (Y_i - m(x_i))l(x_i)].$$

Then since

$$n^{1/2}\delta_n^{-1}T_{13n} = n^{1/2}\delta_n^{-1}[\int[m(x) - g(x, \gamma_*)]^2 dx + o(1)],$$

under H_{1n} we have $n^{1/2}\delta_n^{-1}(T_n - \delta_n^2\mu_1) \to N(0, 2\sigma_2^2)$ in distribution.

(iii) If $nh\delta_n^2 \to \lambda > 0$, then $nh^{1/2}T_{11n} \to N(0, \sigma_0^2)$ and $nh^{1/2}T_{12n} \to N(0, \sigma_2^2/\lambda)$. It is then easy to check that, under the conditions the of Theorem, $nh^{1/2}T_{13n} = \lambda h^{-1/2}[\mu_1 + o(1)]$. Thus under H_{1n} we have $nh^{1/2}(T_n - \lambda(nh)^{-1}\mu_1) \to N(0, \sigma_0^2 + 2\sigma_2^2/\lambda)$ in distribution. This can be proved by Cramer-Wold device, as in Lemma 6 of Kim, Luo, and Kim (2011).

Lemma A.1. Let $\zeta_i \in \mathcal{M}_{s_i}^{t_i}$ be α-mixing random variables, where $s_1 < t_1 < s_2 < t_2 < \cdots < t_m$ and $s_{i+1} - t_i \geq \tau$ for all i. Assume that, for a positive integer ℓ, $\| \zeta_i \|_{p_i} = (E|\eta_i|^{p_i})^{1/p_i} < \infty$, for some $p_i > 1$ with $q = \sum_{i=1}^m p_i^{-1} < 1$. Then

$$\left|E\prod_{i=1}^m \zeta_i - \prod_{i=1}^m E\,\zeta_i\right| \leq 10(m-1)\alpha(\tau)^{1-q}\prod_{i=1}^m \| \zeta_i \|_{p_i}.$$

For complex valued random variables, it holds that

$$\left|E\prod_{i=1}^m \zeta_i - \prod_{i=1}^m E\,\zeta_i\right| \leq 40(m-1)\alpha(\tau)^{1-q}\prod_{i=1}^m \| \zeta_i \|_{p_i}.$$

Proof of Lemma A.1. The proof can be found at Theorem 7.4 of Roussas and Ioannides (1987).

References

Altman, N. S. (1994). Krige, smooth, both or neither ? Technical Report. Biometrics Unit, Cornell Univ.

Eubank, R. and Spiegelman, S. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques, *J. Ameri. Statist. Assoc.* **85**, 387-392.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonpara. Statist.* **10**, 245-271.

Glesser, L. J. and Moore, D. S. (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *Ann. Statist.* **11**, 1100-1108.

Hart, J. D. (1997). Nonparametric Smoothing and Lack-of-Fit Tests. *Springer.*

Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995-1034.

Kim, T. Y., Ha, J., Hwang, S. Y., Park, C. and Luo, Z. (2014). Central limit theorems for reduced U-statistics under dependence and their usefulness. *Austral. N. Z. J. Statist.* **55**, 387-399.

Kim, T. Y., Luo, Z. and Ha, J. (2012). Moment bounds for quadratic forms useful in nonparametric functional estimation under dependency. *Quantitative Bioscience* **31**, 1-5.

Kim, T. Y., Luo, Z. and Kim, C. (2011). Central limit theorem for degenerate variable U-statistics under dependence. *J. Nonpara. Statist.* **23**, 683-699.

Kim, T. Y., Park, B. U., Moon, M. S. and Kim, C. (2009). Using bimodal kernel for inference in nonparametric regression with correlated errors. *J. Multivariate Anal.* **100**, 1487-1497.

Laslett, G. (1994). Kriging and splines: An empirical comparison of their predictive performance in some applications. *J. Ameri. Statist. Assoc.* **89**, 392-409.

Luo, Z., Kim, T. Y. and Song, G. M. (2011). Central limit theorem for quadratic errors of Nadaraya-Waston regression estimator under dependence. *J. Korean Stat. Soc.* **40**, 425-435.

Opsomer, J., Wang, Y. and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* **16**, 134-153.

Roussas, G. G. and Ioannides, D. (1987). Moment inequalities for mixing sequences of random variables. *Stochas. Anal. and Appl.* **5**, 61-120.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263-289.

Department of Statistics, Keimyung University, Daegu 704-701, Korea.

E-mail: cypark1@kmu.ac.kr

Department of Statistics, Keimyung University, Daegu 704-701, Korea.

E-mail: tykim@kmu.ac.kr

Department of Statistics, Keimyung University, Daegu 704-701, Korea.

E-mail: jeicy@kmu.ac.kr

Department of Statistics, Keimyung University, Daegu 704-701, Korea.

E-mail: zhimingluo@kmu.ac.kr

Department of Statistics, Sookmyung Womens University, Seoul 140-742, Korea.

E-mail: shwang@sookmyung.ac.kr

(Received January 2014; accepted July 2014)