The Annals of Statistics
2010, Vol. 38, No. 4, 2047–2091
DOI: 10.1214/09-AOS770

TESTING CONDITIONAL INDEPENDENCE USING MAXIMAL NONLINEAR CONDITIONAL CORRELATION[1,2,3]

BY TZEE-MING HUANG

National Chengchi University

In this paper, the maximal nonlinear conditional correlation of two random vectors X and Y given another random vector Z, denoted by $\rho_1(X, Y|Z)$, is defined as a measure of conditional association, which satisfies certain desirable properties. When Z is continuous, a test for testing the conditional independence of X and Y given Z is constructed based on the estimator of a weighted average of the form $\sum_{k=1}^{n_Z} f_Z(z_k)\rho_1^2(X, Y|Z = z_k)$, where f_Z is the probability density function of Z and the z_k's are some points in the range of Z. Under some conditions, it is shown that the test statistic is asymptotically normal under conditional independence, and the test is consistent.

1. Introduction. In this paper, the problem of interest is testing the conditional independence between two random vectors X and Y given a third random vector Z. The study of the problem of testing conditional independence has a long history. However, there are relatively few results on nonparametric tests when the vectors X, Y and Z are continuous. Some examples of such tests can be found in Su and White [12, 13], where they also proposed conditional independence tests based on a weighted Hellinger distance between the conditional densities or the difference between the conditional characteristic functions.

As mentioned in Daudin [2], X and Y are conditionally independent given Z means that for every $f(X, Z)$ and $g(Y, Z)$ such that $Ef^2(X, Z)$ and $Eg^2(Y, Z)$ are finite

$$E(f(X, Z)g(Y, Z)|Z) = E(f(X, Z)|Z)E(g(Y, Z)|Z).$$

Thus, the problem of testing conditional independence, as the problem of testing unconditional independence, is invariant when one-to-one transforms are applied to the marginals X and Y, respectively. Various authors have taken this invariant

Received September 2009; revised November 2009.

[1]A part of this work is done when the author was visiting the Institute of Statistical Science at Academia Sinica in Taiwan.

[2]Supported by the National Science Council of Taiwan through Grants NSC 95-2119-M-004-001- and NSC 97-2118-M-004-001-.

[3]Supported in part by Center for Service Innovation at National Chengchi University.

AMS 2000 subject classifications. Primary 62H20; secondary 62H15, 62G10.

Key words and phrases. Measure of association, measure of conditional association, conditional independence test.

property into consideration when constructing conditional or unconditional independence tests. For example, Su and White [13] used Hellinger distance in their test statistic for testing conditional independence, so that the test statistic is invariant. Dauxois and Nkiet [3] used measures of association to construct independence tests, and the measures are invariant under the above transforms. In this paper, to take invariance into account, the proposed test is based on the maximal nonlinear conditional correlation, which can be viewed as a measure of conditional association and satisfies the above invariance property.

To choose a reasonable measure of conditional association between X and Y, the following properties are considered.

(P1) The measure can be defined for all types of random vectors, including both discrete and continuous ones.
(P2) The measure is symmetric, that is, it remains the same when (X, Y) is replaced by (Y, X).
(P3) The measure is invariant when one-to-one transforms are applied to X and Y, respectively.
(P4) The measure is between 0 and 1.
(P5) The measure is 0 if and only if conditional independence holds.

The above properties are adapted from some of the conditions for a good measure of association proposed by Rényi [9]. In [9], the conditional independence in (P5) is replaced by the unconditional independence. Note that the symmetric property (P2) is not always required. For instance, Hsing et al. [6] proposed to use the coefficient of intrinsic dependence as a measure of dependence, which does not satisfy (P2). Here, (P2) is considered.

Many measures of conditional association satisfying (P1)–(P5) can be constructed. Dauxois and Nkiet [4] showed that a class of measures of association between two Hilbertian subspaces can be obtained by properly combining the canonical coefficients of the canonical analysis (CA) between the spaces. In particular, take the two subspaces to be $\tilde{H}_1 = \{f(X, Z) - E(f(X, Z)|Z) : Ef^2(X, Z) < \infty\}$ and $\tilde{H}_2 = \{g(Y, Z) - E(g(Y, Z)|Z) : Eg^2(Y, Z) < \infty\}$, then a class of measures of conditional association between X and Y given Z satisfying properties (P1)–(P5) can be obtained using the canonical coefficients. Denote the canonical coefficients (arranged in descending order) by $\tilde{\rho}_i(X, Y|Z) : i = 1, 2, \ldots$. When X and Y are not functions of Z, the largest canonical coefficient $\tilde{\rho}_1(X, Y|Z)$ is the maximal partial correlation defined by Romanovič [10], which is

$$\sup_{f,g} \operatorname{corr}\bigl(f(X, Z) - E(f(X, Z)|Z), g(Y, Z) - E(g(Y, Z)|Z)\bigr).$$

Another approach to construct measures of conditional association is to modify the CA between the spaces $H_1 = \{f(X) - Ef(X) : Ef^2(X) < \infty\}$ and $H_2 = \{g(Y) - Eg(Y) : Eg^2(Y) < \infty\}$ to obtain a conditional version of it. That is, to find pairs of functions $(f_i, g_i): i = 0, 1, \dots,$ such that for each i, (f_i, g_i) maximizes $E(f(X, Z)g(Y, Z)|Z)$ subject to

$$E(f^2(X, Z)|Z)I_{(0,\infty)}(E(f^2(X, Z)|Z)) = I_{(0,\infty)}(E(f^2(X, Z)|Z)), \tag{1.1}$$

$$E(g^2(Y, Z)|Z)I_{(0,\infty)}(E(g^2(Y, Z)|Z)) = I_{(0,\infty)}(E(g^2(Y, Z)|Z)) \tag{1.2}$$

and

$$E(f(X, Z)f_j(X, Z)|Z) = 0 = E(g(Y, Z)g_j(Y, Z)|Z) \qquad \text{for } 0 \le j < i.$$

Here, I_A denotes the indicator function on a set A, that is, $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$, otherwise. If the above (f_i, g_i)'s exist, then one can define $\rho_i(X, Y|Z) = E(f_i(X, Z)g_i(Y, Z)|Z)$ for each i and the $\rho_i(X, Y|Z)$'s can serve as a conditional version of canonical coefficients. A measure of conditional association satisfying (P1)–(P5) can be obtained by taking a proper combination of the $\rho_i(X, Y|Z)$'s, following the approach in [4]. Examples of such combinations include $\rho_1(X, Y|Z)$ and $1 - \exp(-\sum_i \rho_i^2(X, Y|Z))$. The measure of conditional association used in this paper is $\rho_1(X, Y|Z)$, which will be called the maximal nonlinear conditional correlation of two random vectors X and Y given Z from now on.

In the above definition of $\rho_i(X, Y|Z)$'s, it is assumed that the (f_i, g_i)'s exist. However, it is not clear what conditions can guarantee the existence of the (f_i, g_i)'s. To avoid the problem of finding such conditions, a more general definition for $\rho_1(X, Y|Z)$ is given in Section 2. To construct a test based on $\rho_1(X, Y|Z)$, it is assumed that Z has a Lebesgue probability density function f_Z. An estimator of $\sum_k f_Z(z_k)\rho_1^2(X, Y|Z = z_k)$ is then used as the test statistic, where the z_k's are some points in the range of Z. To study the asymptotic behavior of the test statistic under the hypothesis that X and Y are conditionally independent given Z, we follow the approach in [3] for finding the asymptotic distribution of a statistic for testing the independence between X and Y, which is based on estimators of the canonical coefficients from the CA of H_1 and H_2. To make the approach work for the conditional case, some strong approximation results for kernel estimators of certain conditional expectations are also established.

This paper is organized as follows. The new definition of $\rho_1(X, Y|Z)$ is given in Section 2. Section 3 deals with the estimation of $\rho_1(X, Y|Z = z)$ and test construction. An example is in Section 4 and proofs are given in Section 7.

2. Maximal nonlinear conditional correlation. In this section, a more general definition of the maximal nonlinear conditional correlation $\rho_1(X, Y|Z)$ will be given. Note that in the definition of $\rho_i(X, Y|Z)$'s in Section 1, one can take $f_0(X, Z) = 1 = g_0(Y, Z)$, which gives that $\rho_0(X, Y|Z) = 1$, and then $\rho_1(X, Y|Z)$ can be defined as $E(f_1(X, Z)g_1(Y, Z)|Z)$ if there exists $(f_1, g_1) \in S_0$ such that

$$E(f(X, Z)g(Y, Z)|Z) \le E(f_1(X, Z)g_1(Y, Z)|Z) \qquad \text{for every } (f, g) \in S_0,$$

where S_0 is the collection of pairs of functions (f, g)'s that satisfy (1.1), (1.2) and $E(f(X, Z)|Z) = 0 = E(g(Y, Z)|Z)$. Without assuming the existence of (f_1, g_1), it is reasonable to define $\rho_1(X, Y|Z)$ as

$$\sup_{(f,g)\in S_0} E(f(X, Z)g(Y, Z)|Z), \tag{2.1}$$

if the supremum can be defined.

The above approach can be considered as a "pointwise" approach. Indeed, when Z takes values in a countable set $\mathcal{Z}$, for each $z \in \mathcal{Z}$, one may define $\rho_1(X, Y|Z = z)$ as

$$\sup_{(f,g)\in S_0} E\big(f(X, z)g(Y, z)|Z = z\big), \tag{2.2}$$

then the $\rho_1(X, Y|Z)$ defined using (2.2) is a measurable function and can serve as the supremum in (2.1). However, if $\mathcal{Z}$ is uncountable, then it is not clear whether the $\rho_1(X, Y|Z)$ defined using (2.2) is measurable. Therefore, we use the following fact to define the supremum in (2.1) so that it is well defined and is a measurable function.

FACT 1. *There exists a sequence* $\{(\alpha_n, \beta_n)\}$ *in* S_0 *such that*:

(i) *The sequence* $\{E(\alpha_n(X, Z)\beta_n(Y, Z)|Z)\}$ *is nondecreasing, and*
(ii) *for every* $(f, g) \in S_0$,

$$E(f(X, Z)g(Y, Z)|Z) \le \lim_{n\to\infty} E(\alpha_n(X, Z)\beta_n(Y, Z)|Z).$$

Furthermore, if (i) *and* (ii) *hold for* $\{(\alpha_n, \beta_n)\} = \{(\alpha_{n,1}, \beta_{n,1})\}$ *or* $\{(\alpha_{n,2}, \beta_{n,2})\}$, *where* $\{(\alpha_{n,1}, \beta_{n,1})\}$ *and* $\{(\alpha_{n,2}, \beta_{n,2})\}$ *are sequences in* S_0, *then*

$$\lim_{n\to\infty} E(\alpha_{n,1}(X, Z)\beta_{n,1}(Y, Z)|Z) = \lim_{n\to\infty} E(\alpha_{n,2}(X, Z)\beta_{n,2}(Y, Z)|Z). \tag{2.3}$$

For the sake of brevity, from now on, some functions of (X, Z) or (Y, Z) may be expressed without the arguments (X, Z) or (Y, Z). For distinguishing purpose, functions of (X, Z) may have names starting with only α or f, and functions of (Y, Z) may have names starting with only β or g.

PROOF FOR FACT 1. We will first establish (2.3) if (i) and (ii) hold for $\{(\alpha_n, \beta_n)\} = \{(\alpha_{n,1}, \beta_{n,1})\}$ or $\{(\alpha_{n,2}, \beta_{n,2})\}$. Note that for each n, from (ii), we have that

$$E(\alpha_{n,2}\beta_{n,2}|Z) \le \lim_{n\to\infty} E(\alpha_{n,1}\beta_{n,1}|Z)$$

and

$$E(\alpha_{n,1}\beta_{n,1}|Z) \le \lim_{n\to\infty} E(\alpha_{n,2}\beta_{n,2}|Z).$$

Take the limits in these two inequalities as $n \to \infty$, and we have (2.3).

It remains to find a sequence $\{(\alpha_n, \beta_n)\}$ in S_0 that satisfies (i) and (ii). Let $\{(\alpha_{n,0}, \beta_{n,0})\}$ be a sequence in S_0 so that the sequence $\{E(\alpha_{n,0}\beta_{n,0})\}$ is nondecreasing and converges to $\sup_{(f,g)\in S_0} E(fg)$. We will construct $\{(\alpha_n, \beta_n)\}$ using $\{(\alpha_{n,0}, \beta_{n,0})\}$ as follows. For $n = 1$, define $(\alpha_1, \beta_1) = (\alpha_{1,0}, \beta_{1,0})$. For $n \geq 2$, define

$$\begin{aligned}&(\alpha_n(X,Z), \beta_n(Y,Z))\\ &\qquad = \begin{cases}(\alpha_{n,0}(X,Z), \beta_{n,0}(Y,Z)), & \text{if } E(\alpha_{n,0}\beta_{n,0}|Z) > E(\alpha_{n-1}\beta_{n-1}|Z);\\ (\alpha_{n-1}(X,Z), \beta_{n-1}(Y,Z)), & \text{otherwise.}\end{cases}\end{aligned}$$

Then $\{(\alpha_n, \beta_n)\}$ is a sequence in S_0 that satisfies (i), and the sequence $\{E\alpha_n\beta_n\}$ converges to $\sup_{(f,g)\in S_0} E(fg)$ since $E(\alpha_n\beta_n|Z) \geq E(\alpha_{n,0}\beta_{n,0}|Z)$. To see that $\{(\alpha_n, \beta_n)\}$ also satisfies (ii), for (α, β) in S_0, define

$$(\alpha_n^*, \beta_n^*) = \begin{cases}(\alpha, \beta), & \text{if } E(\alpha\beta|Z) > \lim_{n\to\infty} E(\alpha_n\beta_n|Z);\\ (\alpha_n, \beta_n), & \text{otherwise.}\end{cases}$$

Then $\{(\alpha_n^*, \beta_n^*)\}$ is a sequence in S_0 such that

$$\lim_{n\to\infty} E(\alpha_n^*\beta_n^*|Z) = \max\Big\{E(\alpha\beta|Z), \lim_{n\to\infty} E(\alpha_n\beta_n|Z)\Big\}. \tag{2.4}$$

From the monotone convergence theorem, we have

$$E \lim_{n\to\infty} E(\alpha_n^*\beta_n^*|Z) = \lim_{n\to\infty} E(\alpha_n^*\beta_n^*) \tag{2.5}$$

and

$$E \lim_{n\to\infty} E(\alpha_n\beta_n|Z) = \lim_{n\to\infty} E(\alpha_n\beta_n), \tag{2.6}$$

so (2.4) implies that

$$\sup_{(f,g)\in S_0} E(fg) \geq \lim_{n\to\infty} E(\alpha_n^*\beta_n^*) \geq \lim_{n\to\infty} E(\alpha_n\beta_n) = \sup_{(f,g)\in S_0} E(fg),$$

which gives

$$\lim_{n\to\infty} E(\alpha_n^*\beta_n^*) = \lim_{n\to\infty} E(\alpha_n\beta_n). \tag{2.7}$$

If $E(\alpha\beta|Z) > \lim_{n\to\infty} E(\alpha_n\beta_n|Z)$ with positive probability, then (2.4), (2.5) and (2.6) together implies that $\lim_{n\to\infty} E(\alpha_n^*\beta_n^*) > \lim_{n\to\infty} E(\alpha_n\beta_n)$, which contradicts (2.7). Thus, (ii) holds. The proof of Fact 1 is complete. $\square$

With Fact 1, the maximal nonlinear conditional correlation $\rho_1(X, Y|Z)$ can be redefined as follows.

DEFINITION 1. $\rho_1(X, Y|Z) = \sup_{(f,g)\in S_0} E(f(X,Z)g(Y,Z)|Z)$, which is defined as $\lim_{n\to\infty} E(\alpha_n(X,Z)\beta_n(Y,Z)|Z)$, where $\{(\alpha_n, \beta_n)\}$ is a sequence in S_0 that satisfies (i) and (ii) in Fact 1.

Below are some remarks for the $\rho_1(X, Y|Z)$.

1. If there exists (f_1, g_1) in S_0 such that $E(f_1 g_1|Z) \geq E(fg|Z)$ for all $(f, g) \in S_0$, then $\rho_1(X, Y|Z) = E(f_1 g_1|Z)$ using Definition 1. To see this, let $\{(\alpha_n, \beta_n)\}$ be a sequence in S_0 that satisfies (i) and (ii) in Fact 1. Then $\rho_1(X, Y|Z) = \lim_{n\to\infty} E(\alpha_n \beta_n|Z)$, so $E(f_1 g_1|Z) \leq \rho_1(X, Y|Z)$ by (ii). Also, $E(f_1 g_1|Z) \geq E(\alpha_n \beta_n|Z)$ for every n, so $E(f_1 g_1|Z) \geq \rho_1(X, Y|Z)$. Therefore, $\rho_1(X, Y|Z) = E(f_1 g_1|Z)$ and Definition 1 can be viewed as a generalized version of the definition of $\rho_1(X, Y|Z)$ given in Section 1.
2. $\rho_1(X, Y|Z)$ satisfies properties (P1)–(P5).
3. When X is a function of Y and Z or Y is a function of X and Z, it is not necessary that $\rho_1(X, Y|Z) = 1$. For instance, suppose that X and Z are independent standard normal random variables and $Y = X I_{(0,\infty)}(Z)$, then $\rho_1(X, Y|Z) = I_{(0,\infty)}(Z)$.
4. Let $\rho_1(X, Y)$ be the largest canonical coefficient from the CA between $H_1 = \{f(X) - Ef(X) : Ef^2(X) < \infty\}$ and $H_2 = \{g(Y) - Eg(Y) : Eg^2(Y) < \infty\}$. Then $\rho_1(X, Y|Z) = \rho_1(X, Y)$ if (X, Y) and Z are independent.
5. Let $\rho_1(X, Y)$ be as defined in item 4. It is stated in [3] that when the joint distribution of X and Y is bivariate normal
$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$
$\rho_1(X, Y) = |\rho|$. This result implies that, when the joint distribution for X, Y and Z is multivariate normal and X and Y are both univariate,
$$\begin{aligned}\rho_1(X, Y|Z) &= \left|\frac{E((X - E(X|Z))(Y - E(Y|Z))|Z)}{(E(X - E(X|Z))^2|Z)^{1/2}(E(Y - E(Y|Z))^2|Z)^{1/2}}\right| \\ &= \left|\frac{E(X - E(X|Z))(Y - E(Y|Z))}{(E(X - E(X|Z))^2)^{1/2}(E(Y - E(Y|Z))^2)^{1/2}}\right|,\end{aligned}$$
which also equals the absolute value of the usual partial correlation coefficient.

3. A test of conditional independence. Testing conditional independence is equivalent to testing $H_0 : \rho_1(X, Y|Z) = 0$, which involves testing $H_{0,z} : \rho_1(X, Y| Z = z) = 0$ for different z's in the range of Z. Let $\mathcal{Z}$ be the range of Z. In this section, an estimator $\hat{\rho}(z)$ is proposed for estimating $\rho_1(X, Y|Z = z)$ for each $z \in \mathcal{Z}$, and for distinct points $z_1, \dots, z_{n_Z}$ in $\mathcal{Z}$, the asymptotic joint distribution of $\hat{\rho}(z_1), \dots, \hat{\rho}(z_{n_Z})$ under H_0 is derived to construct a test for testing H_0.

3.1. *Estimation of* $\rho_1(X, Y|Z = z)$. To estimate
$$\rho_1(X, Y|Z) = \sup_{(f,g)\in S_0} E(fg|Z)$$
for $(f, g) \in S_0$, f and g are approximated using basis functions. Suppose that there exist Λ_1, Λ_2 and Λ_3: subsets of the set of all positive integers and three sets

of functions $\{\phi_{p,i}: 1 \le i \le p, p \in \Lambda_1\}$, $\{\psi_{q,j}: 1 \le j \le q, q \in \Lambda_2\}$ and $\{\theta_{r,k}: 1 \le k \le r, k \in \Lambda_3\}$ such that for $\alpha(X, Z)$ and $\beta(Y, Z)$ with finite second moments,

$$\lim_{p,r\to\infty} \inf_{a(i,k)} E\biggl(\alpha(X,Z) - \sum_{1\le i\le p, 1\le k\le r} a(i,k)\phi_{p,i}(X)\theta_{r,k}(Z)\biggr)^2 = 0 \tag{3.1}$$

and

$$\lim_{q,r\to\infty} \inf_{b(j,k)} E\biggl(\beta(Y,Z) - \sum_{1\le j\le q, 1\le k\le r} b(j,k)\psi_{q,j}(Y)\theta_{r,k}(Z)\biggr)^2 = 0. \tag{3.2}$$

Also, suppose that for each (p, q), there exist coefficients $a_{p,0,i}$'s and $b_{q,0,j}$'s such that

$$\sum_{1\le i\le p} a_{p,0,i}\phi_{p,i}(x) = 1 = \sum_{1\le j\le q} b_{q,0,j}\psi_{q,j}(y) \tag{3.3}$$

for every x in the range of X and every y in the range of Y.

Let S_1 be the collection of all (f, g)'s with finite second moments and let $S_{1,p,q}$ be the collection of all (f, g)'s in S_1 such that $f(X, Z) = \sum_{i=1}^p a_{p,i}(Z)\phi_{p,i}(X)$ for some $a_{p,i}(Z)$'s, and $g(Y, Z) = \sum_{j=1}^q b_{q,j}(Z)\psi_{q,j}(Y)$ for some $b_{q,j}(Z)$'s. Then (3.1) and (3.2) together imply that S_1 can be approximated by $S_{1,p,q}$ for large p and q. Since $S_0 \subset S_1$, S_0 can be approximated by $S_{1,p,q}$ as well. With the additional condition (3.3), S_0 can be easily approximated using the subspace $S_{0,p,q} = S_0 \cap S_{1,p,q}$. Note that (3.1), (3.2) and (3.3) hold for certain basis functions, for example, the tensor product splines in [11].

Assuming (3.1), (3.2) and (3.3), it is reasonable to define

$$\sup_{(f,g)\in S_{0,p,q}} E(fg|Z)$$

and use it to approximate $\rho_1(X, Y|Z)$. To define $\sup_{(f,g)\in S_{0,p,q}} E(fg|Z)$, one may follow the same approach for defining $\sup_{(f,g)\in S_0} E(fg|Z)$, or simply note that there exists $(f_1, g_1) \in S_{0,p,q}$ such that

$$E(f_1g_1|Z) \ge E(fg|Z) \qquad \text{for all } (f,g) \in S_{0,p,q} \tag{3.4}$$

and define $\sup_{(f,g)\in S_{0,p,q}} E(fg|Z) = E(f_1g_1|Z)$. The pair (f_1, g_1) can be obtained as follows. Let

$$\Sigma_{\phi,p}(Z) = \bigl(E(\phi_{p,i}(X)\phi_{p,j}(X)|Z) - E(\phi_{p,i}(X)|Z)E(\phi_{p,j}(X)|Z)\bigr)_{p\times p},$$

$$\Sigma_{\psi,q}(Z) = \bigl(E(\psi_{q,i}(Y)\psi_{q,j}(Y)|Z) - E(\psi_{q,i}(Y)|Z)E(\psi_{q,j}(Y)|Z)\bigr)_{q\times q}$$

and

$$\Sigma_{\phi,\psi,p,q}(Z) = \bigl(E(\phi_{p,i}(X)\psi_{q,j}(Y)|Z) - E(\phi_{p,i}(X)|Z)E(\psi_{q,j}(Y)|Z)\bigr)_{p\times q}.$$

Consider the following two cases:

(i) $\Sigma_{\phi,p}(Z)$ and $\Sigma_{\psi,q}(Z)$ are not zero matrices, and
(ii) at least one of $\Sigma_{\phi,p}(Z)$ and $\Sigma_{\psi,q}(Z)$ is a zero matrix.

In case (i), let $a_1 = (a_{1,1}(Z), \dots, a_{1,p}(Z))^T$ and $b_1 = (b_{1,1}(Z), \dots, b_{1,q}(Z))^T$ be such that (a_1, b_1) is the pair of (a, b) that maximizes

$$a^T \Sigma_{\phi,\psi,p,q}(Z) b$$

subject to

$$a^T \Sigma_{\phi,p}(Z) a = 1 = b^T \Sigma_{\psi,q}(Z) b,$$

and then take

$$f_1(X, Z) = \sum_{i=1}^{p} a_{1,i}(Z)\bigl(\phi_{p,i}(X) - E(\phi_{p,i}(X)|Z)\bigr)$$

and

$$g_1(Y, Z) = \sum_{j=1}^{q} b_{1,j}(Z)\bigl(\psi_{q,j}(Y) - E(\psi_{q,j}(Y)|Z)\bigr).$$

In case (ii), take $f_1(X, Z) = 0 = g_1(Y, Z)$. Then $(f_1, g_1) \in S_{0,p,q}$ and (3.4) holds. Denote $\sup_{(f,g)\in S_{0,p,q}} E(fg|Z)$ by $\rho_{p,q}(Z)$.

The following fact states that $\rho_1(X, Y|Z)$ can be reasonably approximated by $\rho_{p,q}(Z)$ if p and q are large.

FACT 2. *Suppose that (3.1), (3.2) and (3.3) hold and $\{p_n\}$ and $\{q_n\}$ are sequences of positive integers that tend to ∞ as $n \to \infty$. Then*

$$\lim_{n\to\infty} E\bigl(|\rho_1(X, Y|Z) - \rho_{p_n,q_n}(Z)|\bigr) = 0.$$

PROOF. Since $\rho_1(X, Y|Z) \ge \rho_{p_n,q_n}(Z)$ for every n, Fact 2 holds if for every $\varepsilon > 0$, there exists N_0 such that for $n \ge N_0$,

$$\rho_1(X, Y|Z) \le \rho_{p_n,q_n}(Z) + \Delta_1 \tag{3.5}$$

for some Δ_1 such that $E|\Delta_1| < \varepsilon$. To find such a Δ_1, we will first look for a pair $(f_m, g_m) \in S_0$ such that $E(f_m g_m|Z) \approx \rho_1(X, Y|Z)$, and then find $(f_n^*, g_n^*) \in S_{0,p_n,q_n}$ such that $(f_n^*, g_n^*) \approx (f_m, g_m)$. Take

$$\Delta_1 = E(f_m g_m|Z) - E(f_n^* g_n^*|Z) + \rho_1(X, Y|Z) - E(f_m g_m|Z), \tag{3.6}$$

then (3.5) holds and $E|\Delta_1|$ can be made small if m and n are large enough.

To find $(f_m, g_m) \in S_0$ such that $E(f_m g_m|Z) \approx \rho_1(X, Y|Z)$, let $\{(f_n, g_n)\}_{n=1}^{\infty}$ be a sequence in S_0 such that $\{E(f_n g_n|Z)\}$ is an increasing sequence and $\lim_{n\to\infty} E(f_n g_n|Z) = \rho_1(X, Y|Z)$. Let $\Delta_{2,n} = \rho_1(X, Y|Z) - E(f_n g_n|Z)$, then $\lim_{n\to\infty} E|\Delta_{2,n}| = 0$, which implies that for every $\delta > 0$, there exists m such that

$$E|\Delta_{2,m}| < \delta. \tag{3.7}$$

To find $(f_n^*, g_n^*) \in S_{0,p_n,q_n}$ such that $(f_n^*, g_n^*) \approx (f_m, g_m)$, note that it follows from (3.1) and (3.2) that for $n \geq N_0$, there exists some $(f_{n,1}, g_{n,1}) \in S_{1,p_n,q_n}$ such that

$$\sqrt{E(f_m - f_{n,1})^2} < \delta \quad \text{and} \quad \sqrt{E(g_m - g_{n,1})^2} < \delta. \tag{3.8}$$

Let $f_{n,2}(X,Z) = f_{n,1}(X,Z) - E(f_{n,1}|Z)$, $g_{n,2}(Y,Z) = g_{n,1}(Y,Z) - E(g_{n,1}|Z)$,

$$f_n^*(X,Z) = \frac{f_{n,2}(X,Z)}{\sqrt{E(f_{n,2}^2|Z)}} I_{(0,\infty)}(E(f_{n,2}^2|Z))$$

and

$$g_n^*(Y,Z) = \frac{g_{n,2}(Y,Z)}{\sqrt{E(g_{n,2}^2|Z)}} I_{(0,\infty)}(E(g_{n,2}^2|Z)),$$

then it follows from (3.3) that $(f_n^*, g_n^*) \in S_{0,p_n,q_n}$. To see that $(f_n^*, g_n^*) \approx (f_m, g_m)$, let $\Delta_3 = f_m - f_n^*$ and $\Delta_4 = g_m - g_n^*$, then it can be shown that

$$E\Delta_3^2 \leq 16\delta^2 + 8\delta \tag{3.9}$$

and

$$E\Delta_4^2 \leq 16\delta^2 + 8\delta. \tag{3.10}$$

Below we will verify (3.9) only since the verification for (3.10) is similar. Write $\Delta_3 = f_m - f_{n,2} + f_{n,2} - f_n^*$, then by (3.8),

$$E(f_m - f_{n,2})^2 \leq 4\delta^2 \tag{3.11}$$

since $E(f_m - f_{n,2})^2 \leq 2(E(f_m - f_{n,1})^2 + E(f_{n,1} - f_{n,2})^2)$ and $(f_{n,1} - f_{n,2})^2 = (E((f_m - f_{n,1})|Z))^2 \leq E((f_m - f_{n,1})^2|Z)$. Also,

$$\begin{aligned}
E\big((f_n^* - f_{n,2})^2|Z\big) &= \big(1 - \sqrt{E(f_{n,2}^2|Z)}\big)^2 I_{(0,\infty)}(E(f_{n,2}^2|Z)) \\
&\leq |1 - E(f_{n,2}^2|Z)| \\
&= \big|E\big((f_m - f_{n,2})^2|Z\big) - 2E\big(f_m(f_m - f_{n,2})|Z\big)\big| \\
&\leq E\big((f_m - f_{n,2})^2|Z\big) + 2\sqrt{E\big((f_m - f_{n,2})^2|Z\big)},
\end{aligned}$$

so

$$E(f_{n,2} - f_n^*)^2 \leq E(f_m - f_{n,2})^2 + 2\sqrt{E(f_m - f_{n,2})^2} \overset{(3.11)}{\leq} 4\delta^2 + 4\delta. \tag{3.12}$$

Therefore, (3.9) follows from (3.11), (3.12) and the inequality $E\Delta_3^2 \leq 2(E(f_m - f_{n,2})^2 + E(f_{n,2} - f_n^*)^2)$.

Finally, the Δ_1 in (3.6) is $E(f_n^*\Delta_4|Z) + E(g_n^*\Delta_3|Z) + E(\Delta_3\Delta_4|Z) + \Delta_{2,m}$, so it follows from (3.9), (3.10), (3.7) and the Cauchy inequality that

$$E|\Delta_1| \leq 3\sqrt{16\delta^2 + 8\delta} + \delta.$$

For $\varepsilon > 0$, one can choose δ so that $3\sqrt{16\delta^2 + 8\delta} + \delta < \varepsilon$, then $E|\Delta_1| < \varepsilon$ as required. The proof of Fact 2 is complete. $\square$

Based on Fact 2, it is reasonable to estimate $\rho_1(X, Y|Z)$ using an estimator for $\rho_{p,q}(Z)$, where p and q are large. To estimate $\rho_{p,q}(Z)$, the following assumption is made:

(A1) There exists a version of the conditional distribution of (X, Y) given Z such that for every bounded function $g(X, Y)$, $E(g(X, Y)|Z)$ calculated using that version is a continuous function of Z.

From now on, we will use the version of conditional distribution in (A1) to obtain $E(g(X, Y)|Z = z)$ for every bounded g and every z in the range of Z. It for each (p, q), $1 \le i \le p$, $1 \le j \le q$, $|\phi_{p,i}| \le 1$ and $|\psi_{q,j}| \le 1$, then each element in $\Sigma_{\phi,p}(z)$, $\Sigma_{\psi,q}(z)$ and $\Sigma_{\phi,\psi,p,q}(z)$ is a continuous function of z, and $\rho_{p,q}(z)$ is $\max_{a,b} a^T \Sigma_{\phi,\psi,p,q}(z) b$, where the maximum is taken over all vectors a and b such that

$$a^T \Sigma_{\phi,p}(z) a = 1 = b^T \Sigma_{\psi,q}(z) b.$$

To estimate $\rho_{p,q}(z)$, we consider the estimator

$$\hat{\rho}_{p,q}(z) = \max_{a,b} a^T \hat{\Sigma}_{\phi,\psi,p,q}(z) b,$$

where the maximum is taken over all vectors a and b such that

$$a^T \hat{\Sigma}_{\phi,p}(z) a = 1 = b^T \hat{\Sigma}_{\psi,q}(z) b,$$

and $\hat{\Sigma}_{\phi,p}(z)$, $\hat{\Sigma}_{\phi,\psi,p,q}(z)$ and $\hat{\Sigma}_{\psi,q}(z)$ are obtained by replacing the conditional expectations in $\Sigma_{\phi,p}(z)$, $\Sigma_{\phi,\psi,p,q}(z)$ and $\Sigma_{\psi,q}(z)$ by their kernel estimators. Specifically, each element in $\Sigma_{\phi,p}(z)$, $\Sigma_{\phi,\psi,p,q}(z)$ and $\Sigma_{\psi,q}(z)$ is of the form $E(UV|Z = z) - (E(U|Z = z))(E(V|Z = z))$, where U and V are functions of X or Y, so each of $E(UV|Z = z)$, $E(U|Z = z)$ and $E(V|Z = z)$ is of the form $E(g(X, Y)|Z = z)$, which is estimated by

$$\hat{E}\big(g(X, Y)|Z = z\big) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^n g(X_i, Y_i) k_h(z - Z_i)}{\sum_{i=1}^n k_h(z - Z_i)}, \tag{3.13}$$

where $k_h(z) = h^{-d} k_0(z/h)$ and k_0 is a kernel function on R^d satisfying certain conditions which will be specified later. For each $z \in \mathcal{Z}$, to make $\hat{\rho}_{p,q}(z)$ a reasonable estimator for $\rho_1(X, Y|Z = z)$, we will take $p = p_n$, $q = q_n$ and $h = h_n$, where $p_n \to \infty$, $q_n \to \infty$ and $h_n \to 0$ as $n \to \infty$. The estimator $\hat{\rho}_{p_n,q_n}(z)$ will be abbreviated as $\hat{\rho}(z)$ for each $z \in \mathcal{Z}$.

The estimator $\hat{\rho}(z)$ can be expressed in a different form that is easier to analyze. Let X_* and Y_* be random vectors of length p_n and q_n, respectively, such that given the data $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$,

$$(X_*^T, Y_*^T) = (\phi_{p_n,1}(X_\ell), \ldots, \phi_{p_n,p_n}(X_\ell), \psi_{q_n,1}(Y_\ell), \ldots, \psi_{q_n,q_n}(Y_\ell))$$

with probability $k_h(z - Z_\ell)/\sum_{i=1}^n k_h(z - Z_i)$ for $1 \le \ell \le n$. Then $\hat{\Sigma}_{\phi,\psi,p,q}(z) = EX_*Y_*^T - EX_*EY_*^T$, $\hat{\Sigma}_{\phi,p}(z) = EX_*X_*^T - EX_*EX_*^T$ and $\hat{\Sigma}_{\psi,q}(z) = EY_*Y_*^T - EY_*EY_*^T$, where the expectations are conditional expectations given the data. Therefore, the estimator $\hat{\rho}(z)$ is the largest canonical coefficient from the centered canonical analysis between X_* and Y_*. Note that it follows from (3.3) that

$$a_{n,*}^T X_* = 1 = b_{n,*}^T Y_*, \tag{3.14}$$

where

$$a_{n,*} = (a_{p_n,0,1}, \dots, a_{p_n,0,p_n})^T \quad \text{and} \quad b_{n,*} = (b_{q_n,0,1}, \dots, b_{q_n,0,q_n})^T,$$

so $\hat{\rho}(z)$ can also be obtained from the noncentered canonical analysis between X_* and Y_*. Let

$$V_{1,1}(z) = \big(E\big(\phi_{p_n,i}(X)\phi_{p_n,j}(X)|Z = z\big)\big)_{p_n \times p_n},$$

$$V_{1,2}(z) = \big(E\big(\phi_{p_n,i}(X)\psi_{q_n,j}(Y)|Z = z\big)\big)_{p_n \times q_n},$$

$$V_{2,2}(z) = \big(E\big(\psi_{q_n,i}(Y)\psi_{q_n,j}(Y)|Z = z\big)\big)_{q_n \times q_n} \quad \text{and} \quad V_{2,1}(z) = V_{1,2}(z)^T$$

for $1 \le i, j \le 2$, let $\hat{V}_{i,j}(z)$ be the estimator of $V_{i,j}(z)$ obtained by replacing the conditional expectations in $V_{i,j}(z)$ by their kernel estimators as in (3.13). Then $\hat{V}_{1,1}(z) = EX_*X_*^T$, $\hat{V}_{1,2}(z) = EX_*Y_*^T$, $\hat{V}_{2,2}(z) = EY_*Y_*^T$, so $\hat{\rho}(z)$ is the square root of the largest eigenvalue of the matrix

$$\hat{V}_{1,2}(z)\hat{V}_{2,2}^{-1}(z)\hat{V}_{2,1}(z)\hat{V}_{1,1}(z)^{-1} - \hat{V}_{1,1}(z)a_{n,*}a_{n,*}^T.$$

Also, $\rho_{p_n,q_n}(z)$ is the square root of the largest eigenvalue of the matrix

$$V_{1,2}(z)V_{2,2}^{-1}(z)V_{2,1}(z)V_{1,1}(z)^{-1} - V_{1,1}(z)a_{n,*}a_{n,*}^T.$$

To simplify the above matrix expressions, some notation is introduced as follows. For a $(p_n + q_n) \times (p_n + q_n)$ matrix U, express U as

$$\begin{pmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{pmatrix},$$

where the dimension of $U_{1,1}$ is $p_n \times p_n$. For $1 \le i, j \le 2$, let $g_{i,j}$ be the mapping that maps U to $U_{i,j}$. For a $p_n \times 1$ vector a and a $(p_n + q_n) \times (p_n + q_n)$ matrix U, define

$$g(U, a) = g_{1,2}(U)g_{2,2}(U)^{-1}g_{2,1}(U)g_{1,1}(U)^{-1} - g_{1,1}(U)aa^T,$$

if $g_{2,2}(U)$ and $g_{1,1}(U)$ are invertible. Let

$$V(z) = \begin{pmatrix} V_{1,1}(z) & V_{1,2}(z) \\ V_{2,1}(z) & V_{2,2}(z) \end{pmatrix}$$

and

$$\hat{V}(z) = \begin{pmatrix} \hat{V}_{1,1}(z) & \hat{V}_{1,2}(z) \\ \hat{V}_{2,1}(z) & \hat{V}_{2,2}(z) \end{pmatrix},$$

then $\hat{\rho}(z)$ is the square root of the largest eigenvalue of $g(\hat{V}(z), a_{n,*})$ and $\rho_{p_n,q_n}(z)$ is the square root of the largest eigenvalue of $g(V(z), a_{n,*})$.

The matrix $g(\hat{V}(z), a_{n,*})$ can be replaced by a different matrix if basis change is performed. That is, suppose that

$$\phi = (\phi_{p_n,1}, \ldots, \phi_{p_n,p_n})^T \quad \text{and} \quad \psi = (\psi_{q_n,1}, \ldots, \psi_{q_n,q_n})^T$$

are replaced by $\phi^* = P_1\phi$ and $\psi^* = Q_1\psi$, respectively, and $\hat{V}(z)$ becomes $\hat{V}^*(z)$ after such a change is made. Then $\hat{\rho}(z)$ is also the square root of the largest eigenvalue of the matrix $g(\hat{V}^*(z), \alpha^*)$, where $\alpha^* = (P_1^{-1})^T a_{n,*}$ is a vector such that $(\alpha^*)^T\phi^* = 1$. To make the expression for $g(V^*(z), \alpha^*)$ simple, the matrices P_1 and Q_1 are chosen so that

$$\phi_1^* = 1 = \psi_1^*, \tag{3.15}$$

$g_{1,1}(V^*(z))$ and $g_{2,2}(V^*(z))$ are identity matrices, and for $1 \le i \le p_n$ and $1 \le j \le q_n$,

$$E\big(\phi_i^*(X)\psi_j^*(Y)|Z = z\big) = \delta_{i,j}\sqrt{\lambda_i}, \tag{3.16}$$

where ϕ_i^* and ψ_j^* denote the ith element in ϕ^* and the jth element in ψ^*, respectively, $\delta_{i,j}$ denotes the Kronecker symbol and the λ_i's are the eigenvalues of $g(V^*(z), \alpha^*)$. Note that $(\alpha^*)^T = (1, 0, \ldots, 0)$ with the above choice of P_1 and Q_1.

3.2. *Asymptotic properties and a test of conditional independence.* In this section, we will give asymptotic properties of the estimators $\hat{\rho}(z_k) : 1 \le k \le n_Z$, where the z_k's are distinct points in $\mathcal{Z}$. First, we will establish the consistency of the estimators, which relies on the fact that for each k, the two matrices $g(\hat{V}^*(z_k), \alpha^*)$ and $g(V^*(z_k), \alpha^*)$ are close, and their largest eigenvalues are $\hat{\rho}^2(z_k)$ and $\rho^2_{p_n,q_n}(z_k)$. The difference between $g(\hat{V}^*(z_k), \alpha^*)$ and $g(V^*(z_k), \alpha^*)$ depends on the difference of $\hat{V}^*(z_k)$ and $V^*(z_k)$, and the difference between some conditional expectation $E(g(X, Y, Z)|Z = z)$ and its kernel estimator $\hat{E}(g(X, Y, Z)|Z = z) = \sum_{i=1}^n w_{0,i}(z)g(X_i, Y_i, z)/\sum_{i=1}^n w_{0,i}(z)$, where $w_{0,i}(z) = k_0(h_n^{-1}(z - Z_i))$. To make it easier to derive the asymptotic properties of $\hat{E}(g(X, Y, Z)|Z = z)$, some regularity conditions on the distribution of (X, Y, Z) are imposed as follows.

(R1) There exists a σ-finite measure μ such that for every $z \in \mathcal{Z}$, the conditional distribution of (X, Y) given $Z = z$ has a p.d.f. $f(\cdot|z)$ with respect to μ. Also, Z has a Lebesgue p.d.f. f_Z, and $f(x, y|z)$ and $f_Z(z)$ are twice differentiable with respect to z.

(R2) There exists a function h on $\mathcal{X} \times \mathcal{Y}$ such that

$$\sup_{z\in\mathcal{Z}} \max\left(|f(x, y|z)|, \max_{1\le i\le d}\left|\frac{\partial}{\partial z_i} f(x, y|z)\right|, \max_{1\le i,j\le d}\left|\frac{\partial^2}{\partial z_i\,\partial z_j} f(x, y|z)\right|\right) \le h(x, y)$$

and $\int h(x, y)\,d\mu(x, y) < \infty$.

(R3) There exist constants c_0 and c_1 such that

$$\sup_{z\in\mathcal{Z}} \max\left(|f_Z(z)|, \max_{1\le i\le d}\left|\frac{\partial}{\partial z_i} f_Z(z)\right|, \max_{1\le i,j\le d}\left|\frac{\partial^2}{\partial z_i\,\partial z_j} f_Z(z)\right|\right) \le c_0$$

and $1/f_Z(z) \le c_1$ for $z \in \mathcal{Z}$.

Note that (R2) implies condition (A1) in Section 3.1. For the kernel function k_0, conditions (K1) and (K2) are assumed. The notation $\|\cdot\|$ denotes the Euclidean norm for a vector or the Frobenius norm for a matrix.

(K1) $k_0 \ge 0$, $\sup_u k_0(u) < \infty$, $\int k_0(u)\,du = 1$, $\int u k_0(u)\,du = 0$ and $\sigma_0^2 = \int \|u\|^2 k_0(u)\,du < \infty$.

(K2) There exists positive constants γ_2 and γ_3 that does not depend on d such that

$$k_0(a) \le (\gamma_2)^d e^{-\gamma_3\|a\|^2} \qquad \text{for every } a \in R^d.$$

REMARK. If k_0 is a product kernel of the form $k_0(z_1,\dots,z_d) = k_{00}(z_1)\cdots k_{00}(z_d)$, and

$$k_{00}(x) \le \gamma_2 e^{-\gamma_3 x^2} \qquad \text{for every } x \in R,$$

then condition (K2) holds.

Assume the above conditions, then it is possible to control the difference between $\hat{V}^*(z_k)$ and $V^*(z_k)$ using the following result.

LEMMA 1. *Suppose that conditions* (R1)–(R3) *and* (K1)–(K2) *hold. Suppose that $f_{n,1},\dots,f_{n,k_n}$ are functions defined on $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$, where $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are the ranges of X, Y and Z, respectively. Let f_Z be the p.d.f. of Z, $\hat{f}_Z(z) = (nh_n^d)^{-1}\sum_{i=1}^n k_0(h_n^{-1}(z - Z_i))$ for $z \in \mathcal{Z}$ and $c_K = 1/\int k_0^2(s)\,ds$. For $z\in\mathcal{Z}$, let $w_i(z) = n^{-1}h_n^{-d} w_{0,i}(z)/\hat{f}_Z(z)$ for $1\le i\le n$ and*

$$W_{n,j}(z) = \sqrt{nh_n^d c_K f_Z(z)}\left(\left(\sum_{i=1}^n w_i(z) f_{n,j}(X_i,Y_i,z)\right) - E\bigl(f_{n,j}(X,Y,z)|Z=z\bigr)\right)$$

for $1\le j\le k_n$. Suppose that $\{h_n\}_{n=1}^\infty$ and $\{\varepsilon_n\}_{n=1}^\infty$ are sequences of positive numbers such that

$$c_{3,1}n^{-\alpha} \le h_n \le c_{3,2}n^{-\alpha}$$

for some positive constants $c_{3,1}$ and $c_{3,2}$ and $1/(d+4) < \alpha < 1/d$, and $h_n/\varepsilon_n = O(n^{-\beta})$ for some $\beta > 0$. Let

$$\mathcal{Z}(\varepsilon_n) = \{z \in \mathcal{Z} : \{z' \in R^d : \|z' - z\| < \varepsilon_n\} \subset \mathcal{Z}\} \tag{3.17}$$

and suppose that $z_1,\dots,z_{n_Z}$ are points in $\mathcal{Z}(\varepsilon_n)$ such that

$$\|z_k - z_{k^*}\| \ge h_n \qquad \text{for } 1\le k,k^*\le n_Z \text{ and } k\ne k^* \tag{3.18}$$

for large n and

$$\max_{1\le k\le n_Z} \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} |f_{n,j}(x,y,z_k)| \le C_n \qquad \text{for some } C_n \ge 1. \tag{3.19}$$

Suppose that $k_n n_Z C_n = O((\ln n)^{1/16})$. Then there exist $W_{n,1,j,k}$ and $W_{n,2,j,k} : 1 \le j \le k_n$, $1 \le k \le n_Z$ such that the joint distribution of $W_{n,1,j,k} + W_{n,2,j,k}$'s is the same as the joint distribution of $W_{n,j}(z_k)$'s, $\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z} W_{n,2,j,k}^2 = O_P(\exp(-(\ln n)^{1/9}))$, and $W_{n,1,j,k}$'s are jointly normal with $EW_{n,1,j,k} = 0$ and for $1 \le j, \ell \le k_n$ and $1 \le k, k^* \le n_Z$

$$\begin{aligned}&\operatorname{Cov}(W_{n,1,j,k}, W_{n,1,\ell,k^*})\\ &\quad = \begin{cases}\operatorname{Cov}\big(f_{n,j}(X,Y,z_k), f_{n,\ell}(X,Y,z_k)|Z=z_k\big), & \text{if } k=k^*;\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

The proof of Lemma 1 is given in Section 7.1.

The differences between $\hat{V}^*(z_k)$'s and $V^*(z_k)$'s can be controlled by applying Lemma 1 and taking the $f_{n,j}(X,Y,z)$'s to be the functions $\phi_\ell^*(X)\phi_{\ell'}^*(X)$, $\phi_\ell^*(X)\psi_m^*(Y)$ and $\psi_m^*(Y)\psi_{m'}^*(Y)$, where $1 \le \ell \le \ell' \le p_n$ and $1 \le m \le m' \le q_n$. In such case, (3.19) holds under the following conditions.

(B1) For each (p,q), $|\phi_{p,k}| \le 1$ and $|\psi_{q,\ell}| \le 1$ for $1 \le k \le p$ and $1 \le \ell \le q$.

(B2) There exists $\{\delta_n\}$: a sequence of positive numbers such that for $1 \le k \le n_Z$, the smallest eigenvalues of the matrices $V_{1,1}(z_k)$ and $V_{2,2}(z_k)$ are greater than or equal to δ_n.

Under the above conditions, the $\hat{\rho}(z_k)$'s are consistent, as stated in Theorem 3.1.

THEOREM 3.1. *Suppose that* (3.1), (3.2), (3.3), *conditions* (R1)–(R3), (K1)–(K2) *and* (B1)–(B2) *hold. Suppose that $\{h_n\}_{n=1}^\infty$ and $\{\varepsilon_n\}_{n=1}^\infty$ are sequences of positive numbers such that*

$$c_{3,1} n^{-\alpha} \le h_n \le c_{3,2} n^{-\alpha}$$

for some positive constants $c_{3,1}$ and $c_{3,2}$ and $1/(d+4) < \alpha < 1/d$, and $h_n/\varepsilon_n = O(n^{-\beta})$ for some $\beta > 0$. Suppose that $z_1, \dots, z_{n_Z}$ are points in $\mathcal{Z}(\varepsilon_n)$ [defined in (3.17)] *such that* (3.18) *holds and*

$$n_Z(p_n+q_n)^2 \max\{1, \delta_n^{-1}(p_n+q_n)\} = O((\ln n)^{1/16}). \tag{3.20}$$

Then

$$\sum_{k=1}^{n_Z} \big(\hat{\rho}^2(z_k) - \rho_{p_n,q_n}^2(z_k)\big)^2 = O_P((nh_n^d)^{-1}(\ln n)^{1/4}) \tag{3.21}$$

and

$$\left(\sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k) - \sum_{k=1}^{n_Z} f_Z(z_k)\rho_{p_n,q_n}^2(z_k)\right)^2 = O_P\left(\frac{(\ln n)^{5/16}}{nh_n^d}\right). \tag{3.22}$$

The proof of Theorem 3.1 is given in Section 7.2.

The next result deals with the asymptotic distribution of $\sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k)$ when X and Y are conditionally independent given Z.

THEOREM 3.2. *Suppose that the conditions in Theorem 3.1 hold and X and Y are conditionally independent given Z. Then there exist random variables $\tilde{f}_k$, $\tilde{\rho}^2(z_k)$ and $\lambda_k : 1 \le k \le n_Z$ such that $\sum_{k=1}^{n_Z} \tilde{f}_k \tilde{\rho}^2(z_k)$ has the same distribution as $\sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k)$ and*

$$nh_n^d c_K \sum_{k=1}^{n_Z} \tilde{f}_k \tilde{\rho}^2(z_k) - \sum_{k=1}^{n_Z} \lambda_k = O_P(\exp(-0.5(\ln n)^{1/9})(\ln n)^{3/32}),$$

where the λ_k's are independent and each λ_k has the same distribution as the largest eigenvalue of a matrix CC^T, where C is a $(p_n - 1) \times (q_n - 1)$ matrix whose elements are i.i.d. $N(0, 1)$.

The proof of Theorem 3.2 is given in Section 7.3. The result in Theorem 3.2 is similar to that in Lemma 7.2 in [3]. The difference is that the asymptotic result here is derived as the sample size n, p_n and q_n all tend to ∞, while in [3], the result is derived as n tends to ∞, but p_n and q_n are held fixed.

Theorem 3.2 suggests the test that rejects the conditional independence hypothesis at approximate level a if

$$nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k) > F_{n_Z,p,q}^{-1}(1-a), \tag{3.23}$$

where $F_{n_Z,p,q}$ is the cumulative distribution function of $\sum_{k=1}^{n_Z} \lambda_k$.

One can estimate $F_{n_Z,p,q}^{-1}(1-a)$ in (3.23) using simulated data, but it is also possible to use a normal approximation. Since the λ_k's are i.i.d., the central limit theorem suggests the asymptotic normality of $\sum_{k=1}^{n_Z} \lambda_k$ and $\sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k)$. The following corollary gives the conditions that guarantee the asymptotic normality of $\sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k)$.

COROLLARY 1. *Suppose that the conditions in Theorem 3.1 hold*

$$\lim_{n\to\infty} \frac{p_n^3 q_n^3}{\sqrt{n_Z}(\max(p_n, q_n))^{1/3}} = 0 \tag{3.24}$$

and (i) *or* (ii) *holds*:

(i) $q_n = h(p_n)$, *where h is an increasing function such that $\lim_{p\to\infty} h(p)/p$ exists and is greater than or equal to* 1.

(ii) $p_n = h(q_n)$, *where h is an increasing function such that $\lim_{q\to\infty} h(q)/q$ exists and is greater than or equal to* 1.

Let μ_{p_n,q_n} and $\sigma^2_{p_n,q_n}$ be the mean and variance of the largest eigenvalue of the matrix CC^T in Theorem 3.2, respectively, and let the λ_k's be as in Theorem 3.2, then

$$\frac{(\max(p_n, q_n))^{1/6}}{\sigma_{p_n,q_n}} = O(1) \tag{3.25}$$

and

$$\frac{\sum_{k=1}^{n_Z} \lambda_k - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}} \xrightarrow{\mathcal{D}} N(0, 1) \qquad \text{as } n \to \infty. \tag{3.26}$$

If X and Y are conditionally independent given Z, then

$$\frac{n h_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k) \hat{\rho}^2(z_k) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}} \xrightarrow{\mathcal{D}} N(0, 1) \qquad \text{as } n \to \infty. \tag{3.27}$$

The proof of Corollary 1 is given in Section 7.4. Corollary 1 gives the test that rejects the conditional independence hypothesis if

$$\frac{n h_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k) \hat{\rho}^2(z_k) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}} \geq \Phi^{-1}(1 - a), \tag{3.28}$$

where Φ is the cumulative distribution function for the standard normal distribution. Here, μ_{p_n,q_n} and $\sigma^2_{p_n,q_n}$ can be approximated by the sample mean and variance of a random sample from the distribution of the largest eigenvalue of the matrix CC^T.

To distinguish the two tests mentioned above, we will refer to the test with rejection region in (3.28) as test 1N and the test with rejection region in (3.23) as test 1. Note that under the conditions in Corollary 1, test 1 does not differ from test 1N much since the rejection region for test 1 can be written as

$$\frac{n h_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k) \hat{\rho}^2(z_k) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}} \geq I + \Phi^{-1}(1 - a),$$

where

$$I = \frac{F^{-1}_{n_Z,p,q}(1 - a) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}} - \Phi^{-1}(1 - a) = o(1) \tag{3.29}$$

by (3.26). Therefore, both tests 1 and 1N are of asymptotic significance level a. Below we will discuss the consistency and asymptotic power of test 1N only since the same properties of test 1 can be established similarly using (3.29).

Suppose all the conditions in Theorem 3.1 hold, then test that 1N is also consistent if the z_k's are chosen in a way such that there exist a constant $c_3 > 0$ and a sequence $\{\eta_{1,n}\}_{n=1}^{\infty}$ such that $\eta_{1,n} > 0$ for every n, $\lim_{n\to\infty} \eta_{1,n} = 0$ and

$$\frac{1}{n_Z}\sum_{k=1}^{n_Z} f_Z(z_k)\rho^2_{p_n,q_n}(z_k) - c_3 E\rho^2_{p_n,q_n}(Z) = o_P(\eta_{1,n}). \tag{3.30}$$

To see that test 1N is consistent, note that $0 \le \mu_{p_n,q_n} \le E\,\mathrm{tr}(CC^T)$ and $\sigma^2_{p_n,q_n} \le E(\mathrm{tr}(CC^T))^2$, where CC^T is as in Theorem 3.2. Therefore, $\mu_{p_n,q_n} = O(p_nq_n)$ and $\sigma^2_{p_n,q_n} = O(p_n^2q_n^2)$. Then it follows from (3.22), (3.30) and Fact 2 that $n_Z^{-1}\sum_{k=1}^{n_Z}\hat f_Z(z_k)\hat\rho^2(z_k) - c_3E\rho_1^2(X,Y|Z) = O_P((\ln n)^{5/32}/n_Z\sqrt{nh_n^d}) + o_P(\eta_{1,n}) + c_3E\rho^2_{p_n,q_n}(Z) - c_3E\rho_1^2(X,Y|Z) = o_P(1)$, so

$$\begin{aligned}
&\frac{nh_n^dc_K\sum_{k=1}^{n_Z}\hat f_Z(z_k)\hat\rho^2(z_k) - n_Z\mu_{p_n,q_n}}{\sqrt{n_Z\sigma^2_{p_n,q_n}}}\\
&\qquad \ge \frac{\sqrt{n_Z}(nh_n^dc_K(c_3E\rho_1^2(X,Y|Z) + o_P(1)) + O(p_nq_n))}{c_{2,1}p_nq_n},
\end{aligned}$$

where $c_{2,1} > 0$ is a constant. Thus, the left-hand side in (3.28) tends to ∞ as $n\to\infty$ when $E\rho_1^2(X,Y|Z) > 0$, which implies that the probability that (3.28) holds tends to 1 if X and Y are not conditionally independent given Z.

Test 1N can also reject an alternative where $E\rho^2_{p_n,q_n}(Z)$ is small under the conditions in Theorem 3.1. Indeed, for $\{\eta_{1,n}\}_{n=1}^{\infty}$ such that $\eta_{1,n} > 0$ for every n, $\lim_{n\to\infty}\eta_{1,n} = 0$ and (3.30) holds, if

$$\frac{\max(\eta_{1,n}, (\ln n)^{5/32}/(n_Z\sqrt{nh_n^d}))}{E\rho^2_{p_n,q_n}(Z)} = o(1), \tag{3.31}$$

then the probability that (3.28) holds tends to 1 since

$$\begin{aligned}
&\frac{nh_n^dc_K\sum_{k=1}^{n_Z}\hat f_Z(z_k)\hat\rho^2(z_k) - n_Z\mu_{p_n,q_n}}{\sqrt{n_Z\sigma^2_{p_n,q_n}}}\\
&\quad \ge \biggl(\sqrt{n_Z}\biggl(nh_n^dc_K\biggl(c_3E\rho^2_{p_n,q_n}(Z)\\
&\qquad\qquad + O_P\biggl(\frac{(\ln n)^{5/32}}{n_Z\sqrt{nh_n^d}}\biggr) + o_P(\eta_{1,n})\biggr) + O(p_nq_n)\biggr)\biggr)\\
&\qquad \times (c_{2,1}p_nq_n)^{-1},
\end{aligned}$$

where $p_nq_n/(nh_n^dE\rho^2_{p_n,q_n}(Z)) = O((\ln n)^{1/16}/(n_Znh_n^dE\rho^2_{p_n,q_n}(Z))) = o(1)$ by (3.20) and (3.31), and $p_nq_n/(\sqrt{n_Z}nh_n^dE\rho^2_{p_n,q_n}(Z)) = o(1)$. In summary, test 1N can reject an alternative where $E\rho^2_{p_n,q_n}(Z)$ tends to zero at a rate that is slower than $\max(\eta_{1,n}, (\ln n)^{5/32}/(n_Z\sqrt{nh_n^d}))$, where $\eta_{1,n}$ is determined by (3.30). An example that satisfies (3.30) and the conditions in Corollary 1 will be given in Section 4. In that example, $\eta_{1,n} = p_n^{11} n_Z^{-1/d}$.

4. An example. In this section, an example is given to illustrate the verification of the conditions in Corollary 1, assuming (R1)–(R3) and the condition that there exists a positive constant $c_{1,1}$ such that

$$f_{X|Z}(x|z) \geq c_{1,1} \quad \text{and} \quad f_{Y|Z}(y|z) \geq c_{1,1} \qquad \text{for all } (x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}, \tag{4.1}$$

where $f_{X|Z}(\cdot|z)$ and $f_{Y|Z}(\cdot|z)$ are conditional probability densities of X and Y, respectively, given $Z = z$, with respect to Lebesgue measures.

EXAMPLE 1. *Suppose that X, Y and Z are random vectors that take values in $[0, 1]^{d_x}$, $[0, 1]^{d_y}$ and $[0, 1]^d$, respectively. Suppose that* (R1)–(R3), *and* (4.1) *hold. Choose the basis functions as follows. Let Λ be the set of all positive integers and $\Lambda(k) = \{m^k : m \in \Lambda\}$ for $k \in \Lambda$. For $k, i_1, \ldots, i_k \in \Lambda$ and $h_0 > 0$, let*

$$h_{k,h_0,i_1,\ldots,i_k}(x_1, \ldots, x_k) = \prod_{j=1}^{k} I_{A_{i_j,h_0}}(x_j) \qquad \textit{for } (x_1, \ldots, x_k) \in [0, 1]^k,$$

where

$$A_{i_j,h_0} = \begin{cases} (h_0(i_j - 1), h_0 i_j], & \textit{if } i_j > 1; \\ [h_0(i_j - 1), h_0 i_j], & \textit{if } i_j = 1. \end{cases}$$

For $p, q, r \in \Lambda$, let

$$\{\phi_{p,i} : 1 \leq i \leq p\} = \{h_{d_x, p^{-1/d_x}, i_1,\ldots,i_{d_x}} : 1 \leq i_1, \ldots, i_{d_x} \leq p^{1/d_x}\},$$

$$\{\psi_{q,j} : 1 \leq j \leq q\} = \{h_{d_y, q^{-1/d_y}, i_1,\ldots,i_{d_y}} : 1 \leq i_1, \ldots, i_{d_y} \leq q^{1/d_y}\}$$

and

$$\{\theta_{r,k} : 1 \leq k \leq r\} = \{h_{d, r^{-1/d}, i_1,\ldots,i_d} : 1 \leq i_1, \ldots, i_d \leq r^{1/d}\}.$$

Take k_0 to be the product kernel function such that

$$k_0(z_1, \ldots, z_d) = k_{00}(z_1) \cdots k_{00}(z_d),$$

where k_{00} is the probability density function for the standard normal distribution. Let $h_n = n^{-a}$, where $1/(d + 4) < a < 1/d$. Let n_Z^* to be the largest number in $\Lambda(d)$ such that $n_Z^* \leq (\ln n)^{1/32}$, and let

$$\{z_k : 1 \leq k \leq n_Z\} = \left\{\left(\frac{i_1}{(n_Z^*)^{1/d}}, \ldots, \frac{i_d}{(n_Z^*)^{1/d}}\right) : 1 \leq i_1, \ldots, i_d < (n_Z^*)^{1/d}\right\},$$

so $n_Z = ((n_Z^*)^{1/d} - 1)^d$. *Suppose that* $\{p_n\}$ *is a sequence in* $\Lambda(d_x) \cap \Lambda(d_y)$ *such that* $\lim_{n\to\infty} p_n = \infty$ *and* $q_n = p_n$. *If*

$$p_n^{12} \le n_Z, \tag{4.2}$$

then all the conditions in Corollary 1 *hold. If*

$$p_n^{12} \le n_Z^{1/d}, \tag{4.3}$$

then (3.30) *holds with* $\eta_{1,n} = p_n^{11} n_Z^{-1/d}$.

PROOF. We will first show that all the conditions in Corollary 1 hold assuming (4.2). It is clear that (3.1), (3.2) and (3.3), and conditions (B1), (K1) and (K2) hold.

To find the δ_n in condition (B2), note that for $z \in \mathcal{Z}$, the smallest eigenvalue of $V_{1,1}(z)$ is the minimum of $\{E(\phi_{p_n,i}(X)|Z = z) : 1 \le i \le p_n\}$, which is the minimum of $\{E(h_{d_x,p_n^{-1/d_x},i_1,\dots,i_{d_x}}(X)|Z = z) : 1 \le i_1, \dots, i_{d_x} \le p_n^{1/d_x}\}$. Under (4.1), for $m \in \Lambda$ and $1 \le i_1, \dots, i_{d_x} \le m$,

$$\begin{aligned} &E\big(h_{d_x,1/m,i_1,\dots,i_{d_x}}(X)|Z = z\big) \\ &\quad = \int_{(i_1-1)/m}^{i_1/m} \cdots \int_{(i_{d_x}-1)/m}^{i_{d_x}/m} f_{X|Z}(x_1, \dots, x_{d_x}|z)\, dx_{d_x} \cdots dx_1 \ge \frac{c_{1,1}}{m^{d_x}}. \end{aligned}$$

Take $m = p_n^{1/d_x}$, and we have that the smallest eigenvalue of $V_{1,1}(z)$ is at least $c_{1,1}/p_n$. Similarly, $c_{1,1}/p_n$ is also a lower bound for the smallest eigenvalue of $V_{2,2}(z)$ and (B2) holds with $\delta_n = c_{1,1}/p_n$. Furthermore, (3.20) holds since

$$n_Z(p_n + q_n)^2 \max\{1, \delta_n^{-1}(p_n + q_n)\} = O(n_Z p_n^4) = O(n_Z^2).$$

Finally, the z_k's are in $\mathcal{Z}(\varepsilon_n)$ with $\varepsilon_n = (n_Z^*)^{-1/d}$ and $h_n/\varepsilon_n = O(n^{-\beta})$ for $0 < \beta < \alpha$. For $1 \le k, k^* \le n_Z$, and $k \ne k^*$, $\|z_k - z_{k^*}\| \ge (n_Z^*)^{-1/d} \ge n^{-a}$, so (3.18) holds. Also, (3.24) holds since

$$\frac{p_n^3 q_n^3}{\sqrt{n_Z}(\max(p_n, q_n))^{1/3}} = p_n^{-1/3}\sqrt{\frac{p_n^{12}}{n_Z}} = o(1).$$

Therefore, all the conditions in Corollary 1 hold for this example.

The verification of (3.30) is based on the fact that there exist positive constants $c_{4,1}$ and η_0 such that

$$|\rho_{p_n,q_n}^2(z) - \rho_{p_n,q_n}^2(z')| \le c_{4,1} p_n^{11}\|z - z'\| \qquad \text{if } p_n^3\|z - z'\| < \eta_0. \tag{4.4}$$

Below we will first check (3.30) assuming that (4.4) holds and then prove (4.4). Suppose that (4.3) holds. Let $g_n(z) = f_Z(z)\rho_{p_n,q_n}^2(z)$. Since f_Z is Lipschitz continuous, (4.4) implies that there exists a constant $c_{4,2} > 0$ such that

$$|g_n(z) - g_n(z')| \le c_{4,2} p_n^{11}\|z - z'\| \qquad \text{if } p_n^3\|z - z'\| < \eta_0.$$

Let $\{z_{1+n_Z}, \dots, z_{n_Z^*}\}$ be the set

$$\left\{\left(\frac{i_1}{(n_Z^*)^{1/d}}, \dots, \frac{i_d}{(n_Z^*)^{1/d}}\right) : 1 \le i_1, \dots, i_d \le (n_Z^*)^{1/d}\right\} \cap \{z_k : 1 \le k \le n_Z\}^c,$$

then

$$\left|\sum_{k=1}^{n_Z^*} g_n(z_k)\left(\frac{1}{(n_Z^*)^{1/d}}\right)^d - \int_{\mathcal{Z}} g_n(z)\,dz\right| \le 2c_{4,2} p_n^{11}\sqrt{d}\left(\frac{1}{n_Z^*}\right)^{1/d},$$

if $p_n^3(n_Z^*)^{-1/d} < \eta_0$. Since $|g_n(z)| \le c_0$ by (R3) and there exists a positive constant $c_{4,3}$ depending on d such that

$$n_Z^* - n_Z \begin{cases} \le c_{4,3}(n_Z^*)^{1/d}, & \text{if } d \ge 2; \\ = 1, & \text{if } d = 1, \end{cases}$$

we have

$$\begin{aligned}
&\left|n_Z^{-1}\sum_{k=1}^{n_Z} f_Z(z_k)\rho_{p_n,q_n}^2(z_k) - \frac{\int_{\mathcal{Z}} f_Z(z)\rho_{p_n,q_n}^2(z)\,dz}{\int_{\mathcal{Z}} 1\,dz}\right| \\
&\quad = \left|\frac{n_Z^*}{n_Z}\left(\frac{1}{n_Z^*}\sum_{k=1}^{n_Z^*} g_n(z_k) - \int_{\mathcal{Z}} g_n(z)\,dz\right)\right. \\
&\qquad \left. - \frac{\sum_{k=1+n_Z}^{n_Z^*} g_n(z_k)}{n_Z} + \left(\frac{n_Z^*}{n_Z} - 1\right)\int_{\mathcal{Z}} g_n(z)\,dz\right| \\
&\quad \le \frac{n_Z^*}{n_Z}\left|\frac{1}{n_Z^*}\sum_{k=1}^{n_Z^*} g_n(z_k) - \int_{\mathcal{Z}} g_n(z)\,dz\right| + c_0\left(1 + \int_{\mathcal{Z}} 1\,dz\right)\left(\frac{n_Z^* - n_Z}{n_Z}\right) \\
&\quad \le \frac{c_{4,4} p_n^{11}}{n_Z^{1/d}}
\end{aligned}$$

for some constant $c_{4,4} > 0$ if $p_n^3(n_Z^*)^{-1/d} < \eta_0$. Since $p_n^{12} \le n_Z^{1/d}$, $p_n^3 n_Z^{-1/d} = o(1)$, so

$$\left|n_Z^{-1}\sum_{k=1}^{n_Z} f_Z(z_k)\rho_{p_n,q_n}^2(z_k) - \frac{\int_{\mathcal{Z}} f_Z(z)\rho_{p_n,q_n}^2(z)\,dz}{\int_{\mathcal{Z}} 1\,dz}\right| = O_P\left(\frac{p_n^{11}}{n_Z^{1/d}}\right)$$

and $p_n^{11} n_Z^{-1/d} = o(1)$. Take $\eta_{1,n} = p_n^{11} n_Z^{-1/d}$ and $c_3 = (\int_{\mathcal{Z}} 1\,dz)^{-1} = 1$, then (3.30) holds.

It remains to prove (4.4). Recall that for $z \in \mathcal{Z}$, $\rho_{p_n,q_n}^2(z)$ is the largest eigenvalue of $g(V(z), a_{n,*})$, as mentioned in Section 3.1. Thus, $|\rho_{p_n,q_n}^2(z) - \rho_{p_n,q_n}^2(z')|$ is bounded by $\|g(V(z), a_{n,*}) - g(V(z'), a_{n,*})\|$. For $1 \le i, j \le 2$, let $g_{i,j}^*$ be as defined in (7.8) and let $\Delta_{i,j} = g^*_{i,j}(V(z')) - g^*_{i,j}(V(z))$ for $1 \le i, j \le 2$, then from the fact that $\|AB\| \le \|A\|\|B\|$ for two matrices A and B, we have

$$\begin{aligned}
&\|g(V(z), a_{n,*}) - g(V(z'), a_{n,*})\| \\
&\quad \le \prod_{i=1}^{2}\prod_{j=1}^{2}\bigl(\|g^*_{i,j}(V(z))\| + \|\Delta_{i,j}\|\bigr) - \prod_{i=1}^{2}\prod_{j=1}^{2}\|g^*_{i,j}(V(z))\| \\
&\qquad + \|g_{1,1}(V(z')) - g_{1,1}(V(z))\|\|a_{n,*}\|^2.
\end{aligned} \tag{4.5}$$

The bounds for the $\|g^*_{i,j}(V(z))\|$'s are derived as follows. Since the elements in $V(z)$ are bounded by 1 and the smallest eigenvalue of $g_{i,i}(V(z))$ is at least $c_{1,1}/p_n$ for $1 \le i \le 2$, we have

$$\begin{aligned}
\max(\|g^*_{1,2}(V(z))\|, \|g^*_{2,1}(V(z))\|) &\le p_n, \\
\|g^*_{1,1}(V(z))\|^2 &\le \frac{p_n^2}{(c_{1,1}/p_n)^2} = \frac{p_n^4}{c_{1,1}^2}
\end{aligned}$$

and

$$\|g^*_{2,2}(V(z))\| \le \frac{p_n^2}{c_{1,1}}.$$

To find bounds for $\|g_{1,1}(V(z')) - g_{1,1}(V(z))\|$ and $\|\Delta_{i,j}\|$'s, note that from (R3), each element in $g_{i,j}(V(z')) - g_{i,j}(V(z))$ is bounded by $\sqrt{d}\int h(x, y)\, d\mu(x, y)\|z - z'\|$, so

$$\begin{aligned}
&\max\bigl(\|\Delta_{1,2}\|, \|\Delta_{2,1}\|, \|g_{1,1}(V(z')) - g_{1,1}(V(z))\|\bigr) \\
&\quad \le p_n\sqrt{d}\int h(x, y)\, d\mu(x, y)\|z - z'\|.
\end{aligned}$$

For $1 \le i \le 2$, by Fact 4,

$$\|\Delta_{i,i}\| \le \frac{\|g^*_{i,i}(V(z))\|^2\|g_{i,i}(V(z')) - g_{i,i}(V(z))\|}{1 - \|g^*_{i,i}(V(z))\|\|g_{i,i}(V(z')) - g_{i,i}(V(z))\|},$$

if $\|g^*_{i,i}(V(z))\|\|g_{i,i}(V(z')) - g_{i,i}(V(z))\| < 1$, so

$$\|\Delta_{i,i}\| \le \frac{2\sqrt{d}\,p_n^5}{c_{1,1}^2}\int h(x, y)\, d\mu(x, y)\|z - z'\|,$$

if

$$\frac{\sqrt{d}\,p_n^3}{c_{1,1}}\int h(x, y)\, d\mu(x, y)\|z - z'\| < \frac{1}{2}. \tag{4.6}$$

To give a bound for $\|a_{n,*}\|$, note that the smallest eigenvalue of $g_{1,1}(V(z))$ is at least $c_{1,1}/p_n$ and at most

$$\frac{a_{n,*}^T g_{1,1}(V(z))a_{n,*}}{a_{n,*}^T a_{n,*}} = \frac{1}{\|a_{n,*}\|^2},$$

so

$$\|a_{n,*}\| \le \sqrt{\frac{p_n}{c_{1,1}}}.$$

From (4.5) and the above bounds for $\|a_{n,*}\|$, the $\|g_{i,j}^*(V(z))\|$'s and $\|\Delta_{i,j}\|$'s, we have

$$\|g(V(z), a_{n,*}) - g(V(z'), a_{n,*})\| \le c_{4,1} p_n^{11} \|z - z'\|$$

for some constant $c_{4,1}$ if (4.6) holds. Therefore, (4.4) holds and the proof for the results in Example 1 is complete. □

5. Simulation studies. In this section, results of several simulation experiments are presented. Those experiments are designed to demonstrate the performance of test 1 introduced in Section 3.2.

In Section 3.2, test 1N is also introduced, but no simulation studies are done for it in this section. The reason is as follows. Test 1N is constructed based on the normal approximation for $\sum_{k=1}^{n_Z} \lambda_k$. Using the parameter set-up in Table 2, the selected n_Z is only 4 or 5 and the normal approximation for $\sum_{k=1}^{n_Z} \lambda_k$ is not expected to work well.

For simplicity, in all the simulation experiments here, X, Y, Z are one dimensional and only the following distributions for (X, Y, Z) are considered.

(M1) $(X, Y) = (\Phi(Z\epsilon_1), \Phi(Z\epsilon_2))$, where ϵ_1, ϵ_2 and Z are independent, Z follows the uniform distribution on $[0, 1]$, and ϵ_i follows the standard normal distribution for $i = 1, 2$.

(M2) Z follows the standard normal distribution, and the conditional distribution of (X, Y) given $Z = z$ is bivariate normal with mean μ and covariance matrix Σ, where

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & \rho(z) \\ \rho(z) & 1 \end{pmatrix} \tag{5.1}$$

and the $\rho(z)$ in (5.1) is taken to be $a(|1 - 2\Phi(z)|)$ with $a \in \{0, 0.1, 0.3\}$.

(M3) $(X, Y, Z) = (\Phi(X_0), \Phi(Y_0), \Phi(Z_0))$, where Z_0 follows the t-distribution with degree of freedom 1, and the conditional distribution of (X_0, Y_0) given $\Phi(Z_0) = z$ is bivariate normal with mean μ and covariance matrix Σ, where μ and Σ are as in (5.1) and the $\rho(z)$ in (5.1) is taken to be $a(|1 - 2z|)$ with $a \in \{0, 0.1, 0.3\}$.

Here, (M1) is used for parameter selection and (M2) and (M3) are used for checking the power of test 1. In (M1), X and Y are conditionally independent given Z. In (M2) and (M3), $\rho_1(X, Y|Z = z) = \rho(z)$ and $E\rho_1(X, Y|Z)$ is proportional to a.

The details of parameter selection are given in Section 5.1 and the experimental results are given in Section 5.2.

5.1. *Parameter selection.* To apply test 1, certain parameters need to be chosen, including the kernel function k_0, the kernel bandwidth h_n, the basis functions $\phi_{p_n,i}$'s and $\psi_{q_n,j}$'s and the evaluation points z_k's, which are chosen as follows.

(S1) k_0 and the basis functions $\phi_{p,i}$'s and $\psi_{q,j}$'s are chosen as in Example 1 in Section 4 with $p_n = q_n = 2$. Since the basis functions are supported on [0, 1], if X, Y and Z do not take values in [0, 1] [such as in (M2)], then the data $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ will be transformed to $\{(\Phi(X_i), \Phi(Y_i), \Phi(Z_i))\}_{i=1}^n$ before applying test 1. The bandwidth h_n is chosen to be the h that minimizes

$$\int_{0.143h^{0.121}}^{1-0.143h^{0.121}} E\big(\hat{f}_Z(z) - 1\big)^2\, dz \tag{5.2}$$

over $(0, 0.5]$, where $\hat{f}_Z$ is the kernel density estimator based on a sample of size n from the uniform distribution on [0, 1] with kernel k_0 and bandwidth h. Below are the h_n's used for different n's.

The z_k's are points in $I_n = [0.143h_n^{0.121}, 1 - 0.143h_n^{0.121}]$ such that $z_k = 0.143h_n^{0.121} + (k-1)h_{0,n}$, where $h_{0,n}$ is a given positive number. Here, the ε_n is taken to be $0.143h_n^{0.121}$, so the z_k's are chosen so that they are $0.143h_n^{0.121}$ away from the boundary and the integral in (5.2) is over $[0.143h^{0.121}, 1 - 0.143h^{0.121}]$.

With the parameter set-up in (S1), it remains to choose $h_{0,n}$. The $h_{0,n}$ is chosen to be the smallest multiple of 0.01 such that the distribution for the test 1 statistic $nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_k \hat{\rho}^2(z_k)$ based on 1000 samples of size n from (M1) is similar to the distribution of $\sum_{k=1}^{n_Z} \lambda_k$ (χ^2 with n_Z degrees of freedom), as stated in Theorem 3.2. The one-sample Kolmogorov–Smirnov test is used to determine whether the two distributions are similar. Below are the $h_{0,n}$'s used for $n = 10{,}000$ and $n = 5000$.

For the above procedure for selecting $h_{0,n}$, when $n = 500$ or $n = 1000$, it seems that the distribution of $nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_k \hat{\rho}^2(z_k)$ cannot be approximated well by the distribution of $\sum_{k=1}^{n_Z} \lambda_k$, regardless what $h_{0,n}$ is used. To overcome this problem, one may use local bootstrap to determine the rejection region.

The idea of using local bootstrap is to draw samples $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ from the distribution of (X^*, Y^*, Z^*), where Z^*'s distribution is close to the distribution of Z and the conditional distributions of X^* given $Z^* = z$ and Y^* given $Z^* = z$ are close to the conditional distributions of X given $Z = z$ and Y given $Z = z$, yet X^* and Y^* are conditionally independent given Z^*. Therefore, if X and Y are conditionally independent given Z, then the local bootstrap resamples

TABLE 1
Selected h_n's for different n's

n	**10,000**	**5000**	**1000**	**500**
h_n	0.05935281	0.06525282	0.08533451	0.0983018

$\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ should behave like a random sample from (X, Y, Z). One can then compute the test 1 statistic $nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_k \hat{\rho}^2(z_k)$ for the original sample and for each local bootstrap resample. If the statistic computed based on the original sample is larger than $(1-a)\%$ of the statistics computed based on the local bootstrap resamples, then the conditional independence hypothesis is rejected at level a.

The local bootstrap procedure used here is the same as the one proposed by Paparoditis and Politis [8] except that here the Z_i's are not lagged variables. For a given sample $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, a local bootstrap resample $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ is generated as follows.

- Step 1. Draw a random sample $(Z_1^*, \dots, Z_n^*)$ from the empirical cumulative distribution function $\hat{F}_Z$, where
$$\hat{F}_Z(z) = \frac{1}{n}\sum_{i=1}^n I_{(-\infty, Z_i]}(z).$$
- Step 2. For $1 \le i \le n$, for each Z_i^* from Step 1, draw X_i^* and Y_i^* independently from the empirical conditional cumulative distribution functions $\hat{F}_{X|Z=Z_i^*}$ and $\hat{F}_{Y|Z=Z_i^*}$, respectively, where
$$\hat{F}_{X|Z=Z_i^*}(x) = \frac{\sum_{i=1}^n k_0((Z_i^* - Z_i)/b) I_{(-\infty, X_i]}(x)}{\sum_{i=1}^n k_0((Z_i^* - Z_i)/b)}$$
and
$$\hat{F}_{Y|Z=Z_i^*}(y) = \frac{\sum_{i=1}^n k_0((Z_i^* - Z_i)/b) I_{(-\infty, Y_i]}(y)}{\sum_{i=1}^n k_0((Z_i^* - Z_i)/b)}.$$
The parameters for test 1 with local bootstrap are chosen as follows. The bandwidth b is taken to be $h_n^{0.4}$, $p_n = q_n = 2$ and $h_{0,n} = 0.4$, where h_n is as in Table 1.

TABLE 2
$h_{0,n}$'s for different n's

n	**10,000**	**5000**
$h_{0,n}$	0.16	0.2

5.2. *Experiments*. The objective of the first experiment is to compare the power of test 1 with that of a Hellinger distance-based test proposed by Su and White [13]. The critical value for Su and White's test can be determined using the asymptotic distribution of the test statistic or using local bootstrap. To distinguish between the two cases, we use test 2A to denote the asymptotic distribution-based version of Su and White's test and test 2B to denote the local bootstrap version. While test 2B is recommended by Su and White [13], test 2A is used here to save time for computation.

In this experiment, both tests 1 and 2A are carried out for 1000 random samples of size $n = 10^4$, where the distribution of (X, Y, Z) is as in (M2) or (M3). Under (M2), test 1 is applied to transformed data, as mentioned in Section 5.1. Test 2A is applied to normalized data and the bandwidth parameter in the kernel estimators in the test statistic is taken to be $n^{-1/8.5}$, as in [13]. The power estimates based on data from (M2) and (M3) with $n = 10^4$ are given in Table 3. The asymptotic significance level is 0.05. It is shown in Table 3 that power estimates for test 1 when $a = 0$ and $a = 0.1$ are larger that those for test 2A.

To explore the power performance of test 2B without actually running the local bootstrap procedure, approximate critical values for test 2B under (M2) and (M3) are used. To obtain these approximate critical values, note that under (M2) or (M3), for large n, a local bootstrap resample for $a = 0.1$ or $a = 0.3$ is approximately distributed as a random sample for the $a = 0$ case, so the critical value for test 2B can be approximated by the 95% sample quantile of the 1000 test 2A statistics from the first experiment for the $a = 0$ case. Then the power estimates for test 2B can be approximated by the proportions of the 1000 test 2A statistics from the first experiment under different alternatives that exceed the approximate critical values. The approximate power estimates are given in Table 4. Note that the approximate power estimates for test 2B are often larger than the power estimates for test 2A in Table 3, which suggests that test 2B is more powerful than test 2A.

To investigate the performance of test 1 when the sample size is smaller, in the next experiment, power estimates for test 1 are computed based on 1000 random samples of size $n = 5000$ from (M2) and (M3). The results are given in Table 5. The results for $n = 10^4$ from the first experiment are also included for comparison. The asymptotic significance level is 0.05 as before. Table 5 shows that test 1 is more powerful when n is larger.

TABLE 3
Power comparison between tests 1 *and* 2A

	$a = 0$		$a = 0.1$		$a = 0.3$	
	Test 1	**Test 2A**	**Test 1**	**Test 2A**	**Test 1**	**Test 2A**
(M2)	0.049	0.028	0.65	0.076	1	0.95
(M3)	0.041	0.029	0.572	0.119	1	1

TABLE 4
Approximated power estimates for test 2B

	$a = 0.1$	$a = 0.3$
(M2)	0.128	0.971
(M3)	0.241	1

Finally, for smaller sample size such as $n = 500$ or $n = 1000$, since the approximation in Theorem 3.2 does not work well, the local bootstrap version of test 1 is considered. Here 1000 samples of size n from (M2) are used, and for each sample, 1000 local bootstrap resamples are used to determine the rejection region. The level is 0.05. The power estimates for the test are given in Table 6.

In the above results, the power estimates for test 1 are larger when a is larger. This is expected. Under (M2) or (M3), $E\rho^2_{p_n,q_n}(Z) = E\rho^2_{2,2}(Z)$ increases as a increases ($a \in [0, 1]$), so test 1 should be more powerful for larger a, if the approximation in (3.22) and (3.30) work. Table 7 gives the values of $E\rho^2_{p_n,q_n}(Z)$ for $a = 0.1$ and 0.3. For (M2), the calculation of $E\rho^2_{p_n,q_n}(Z)$ is done for the transformed (X, Y, Z), which is obtained by applying the function Φ to the original (X, Y, Z).

6. Concluding remarks. A test statistic for testing conditional independence based on maximal nonlinear conditional correlation is proposed. Two tests, tests 1 and 1N, are constructed using the test statistic. Both tests are consistent and have similar asymptotic properties, as discussed in Section 3.2. Some simulation experiments are carried out to check the performance of test 1. The simulation results show that when the sample size $n = 10^4$, the power of test 1 is comparable with that of test 2A. The simulation results also indicate that test 1 has better power when $E\rho^2_{p_n,q_n}(Z)$ is larger, as expected.

Below are a few remarks.

1. Equation (3.20) requires that p_n, q_n and n_Z grow slowly comparing to n. The parameter selection result in Table 2 in Section 5 seems to agree with such a

TABLE 5
Test 1 *power estimates for* $n = 5000$ *and* $n = 10^4$

	$a = 0$		$a = 0.1$		$a = 0.3$	
	(M2)	**(M3)**	**(M2)**	**(M3)**	**(M2)**	**(M3)**
$n = 5000$	0.052	0.039	0.373	0.321	0.998	1
$n = 10^4$	0.049	0.041	0.65	0.572	1	1

TABLE 6
Power estimates for test 1 *with local bootstrap*

	$\boldsymbol{a=0}$	$\boldsymbol{a=0.1}$	$\boldsymbol{a=0.3}$
$n=500$	0.041	0.071	0.309
$n=1000$	0.033	0.099	0.531

requirement. With $n = 10^4$, n_Z is only 5 and $p_n = q_n = 2$. When $p_n = q_n = 3$, even with $h_{0,n} = 0.4$ (this corresponds to the smallest n_Z for $n = 10^4$), the distribution of the test statistic cannot be approximated well by the distribution of $\sum_{k=1}^{n_Z} \lambda_k$.

2. The parameter selection criteria given in Section 5 needs to be studied to see whether the asymptotic properties of test 1 still hold using such a criteria.
3. When the distribution of the test statistic cannot be approximated well by the distribution of $\sum_{k=1}^{n_Z} \lambda_k$, it is possible to use local bootstrap version of test 1. However, it takes a lot of time to obtain the bootstrap resamples, so this approach is recommended when the sample size n is small.
4. In all theorems proved in this paper, it is assumed that the (X_i, Y_i, Z_i)'s are i.i.d. It is also expected that test 1 works for some stationary weakly dependent data such as the vector ARMA processes, where the central limit theorem for the i.i.d. case still applies. However, to carry out the details in the proofs, one needs the strong approximation result in Lemma 2, which is a stronger result than the central limit theorem and requires a version of Lemma 5 that works for dependent data.
5. Test 1 can be modified to work for discrete Z. Modification is necessary since the rate of convergence for each $\hat{\rho}(z_k)$ is faster in the discrete case.
6. In Lemma 1 and Theorems 3.1 and 3.2, the z_k's are chosen in $\mathcal{Z}(\varepsilon_n)$ so that they are ε_n-away from the boundary, and it is assumed that $h_n/\varepsilon_n = O(n^{-\beta})$ to ensure that certain error terms in the bias/variance calculation are negligible. For implementation, the condition $h_n/\varepsilon_n = O(n^{-\beta})$ still leaves some room for choosing ε_n. This problem can be eliminated by using a kernel function with compact support, as pointed out by a reviewer. In particular, if the kernel function k_0 is supported on $[-1, 1]^d$, then one can simply take $\varepsilon_n = h_n$. In such

TABLE 7
$E\rho^2_{p_n,q_n}(Z)$ *under* (M2) *and* (M3)

	$\boldsymbol{a=0.1}$	$\boldsymbol{a=0.3}$
(M2)	0.001345575	0.01908246
(M3)	0.002044604	0.01765322

case, even though the condition $h_n/\varepsilon_n = O(n^{-\beta})$ does not hold, the results in Lemma 1 and Theorems 3.1 and 3.2 remain valid.

7. Proofs.

7.1. *Proof of Lemma 1*. Recall that for $1 \le j \le k_n$,

$$W_{n,j}(z) = \sqrt{nh_n^d c_K f_Z(z)}\left(\left(\sum_{i=1}^n w_i(z) f_{n,j}(X_i, Y_i, z)\right) - E\big(f_{n,j}(X, Y, z)|Z = z\big)\right).$$

To prove the asymptotic normality of $W_{n,j}(z_k)$'s, we will approximate $W_{n,j}(z)$ using sums of i.i.d. random variables. For $1 \le i \le n$, let $w_{0,i}(z) = k_0(h_n^{-1}(z - Z_i))$and let $\hat f_Z(z) = n^{-1}h_n^{-d}\sum_{i=1}^n w_{0,i}(z)$. Then $w_i(z) = n^{-1}h_n^{-d} w_{0,i}(z)/\hat f_Z(z)$. For $1 \le j \le k_n$, let

$$\begin{aligned}\tilde W_{n,j}(z) = (nh_n^d f_Z(z))^{-1/2}(c_K)^{1/2}\sum_{i=1}^n\big(&w_{0,i}(z) f_{n,j}(X_i, Y_i, z)\\ &- E w_{0,i}(z) f_{n,j}(X_i, Y_i, z)\big)\end{aligned}$$

and $\tilde W_{n,k_n+1}(z) = \sqrt{nh_n^d c_K}(f_Z(z))^{-1/2}(\hat f_Z(z) - E\hat f_Z(z))$, then

$$\begin{aligned}W_{n,j}(z) &= \frac{f_Z(z)}{\hat f_Z(z)}\tilde W_{n,j}(z) + \sqrt{nh_n^d c_K f_Z(z)}E\big(f_{n,j}(X, Y, z)|Z = z\big)\left(\frac{f_Z(z)}{\hat f_Z(z)} - 1\right)\\ &\quad + \frac{\sqrt{nh_n^d c_K f_Z(z)}}{\hat f_Z(z)}\big(h_n^{-d} E(w_{0,1}(z) f_{n,j}(X_1, Y_1, z))\\ &\qquad - E\big(f_{n,j}(X, Y, z)|Z = z\big) f_Z(z)\big)\\ &= \hat W_{n,j}(z) + \sum_{\ell=1}^4 R_{\ell,n,j}(z),\end{aligned}$$

where $\hat W_{n,j}(z) = \tilde W_{n,j}(z) - \tilde W_{n,k_n+1}(z)E(f_{n,j}(X, Y, z)|Z = z)$,

$$R_{1,n,j}(z) = \left(\frac{f_Z(z)}{\hat f_Z(z)} - 1\right)\tilde W_{n,j}(z),$$

$$\begin{aligned}R_{2,n,j}(z) = \frac{\sqrt{nh_n^d c_K f_Z(z)}}{\hat f_Z(z)}\big(&h_n^{-d} E(w_{0,1}(z) f_{n,j}(X_1, Y_1, z))\\ &- E\big(f_{n,j}(X, Y, z)|Z = z\big) f_Z(z)\big),\end{aligned}$$

$$R_{3,n,j}(z) = \frac{\sqrt{nh_n^d c_K}E(f_{n,j}(X, Y, z)|Z = z)(f_Z(z) - \hat f_Z(z))^2}{\hat f_Z(z)\sqrt{f_Z(z)}}$$

and

$$R_{4,n,j}(z) = -\frac{\sqrt{nh_n^d c_K}}{\sqrt{f_Z(z)}} E\big(f_{n,j}(X,Y,z)|Z=z\big)\big(E\hat{f}_Z(z) - f_Z(z)\big).$$

We will complete the proof by showing that the following results hold for $T_n = \exp(-(\ln n)^{1/9})$.

(C1) $\sum_{j=1}^{k_n} \sum_{k=1}^{n_Z} (\sum_{\ell=1}^4 R_{\ell,n,j}(z_k))^2 = O_p(T_n)$.

(C2) There exist random variables $N_{1,j,k}$ and $\varepsilon_{1,j,k} : 1 \le j \le k_n,\ 1 \le k \le n_Z$ such that the joint distribution of $(N_{1,j,k} + \varepsilon_{1,j,k})_{j,k}$ is the same as that of $(\hat{W}_{n,j}(z_k))_{j,k}$, $N_{1,j,k}$'s are jointly normal with $EN_{1,j,k} = 0$ and $\mathrm{Cov}(N_{1,j,k}, N_{1,\ell,k^*}) = \mathrm{Cov}(\hat{W}_{n,j}(z_k), \hat{W}_{n,\ell}(z_{k^*}))$ and $\sum_{j=1}^{k_n} \sum_{k=1}^{n_Z} \varepsilon_{1,j,k}^2 = O_p(T_n)$.

(C3) There exist random variables $N_{2,j,k}$ and $\varepsilon_{2,j,k} : 1 \le j \le k_n,\ 1 \le k \le n_Z$ such that the joint distribution of $(N_{2,j,k} + \varepsilon_{2,j,k})_{j,k}$ is the same as that of $(N_{1,j,k})_{j,k}$, $N_{2,j,k}$'s are jointly normal with $EN_{2,j,k} = 0$ and

$$\mathrm{Cov}(N_{2,j,k}, N_{2,\ell,k^*}) = \begin{cases} \mathrm{Cov}\big(f_{n,j}(X,Y,z_k), f_{n,\ell}(X,Y,z_k)|Z=z_k\big), & \text{if } k = k^*; \\ 0, & \text{otherwise,} \end{cases}$$

and $\sum_{j=1}^{k_n} \sum_{k=1}^{n_Z} \varepsilon_{2,j,k}^2 = O_p(T_n)$.

Note that Lemma 1 follows from (C1)–(C3) since one can construct random variables $\tilde{N}_{2,j,k}$, $\tilde{\varepsilon}_{2,j,k}$, $\tilde{\varepsilon}_{1,j,k}$ and $R_{5,n,j,k} : 1 \le j \le k_n$, $1 \le k \le n_Z$ on the same probability space such that the joint distribution of $(\tilde{N}_{2,j,k}, \tilde{\varepsilon}_{2,j,k})_{j,k}$ is the same as that of $(N_{2,j,k}, \varepsilon_{2,j,k})_{j,k}$, the joint distribution of $(\tilde{\varepsilon}_{1,j,k}, \tilde{N}_{2,j,k} + \tilde{\varepsilon}_{2,j,k})_{j,k}$ is the same as that of $(\varepsilon_{1,j,k}, N_{1,j,k})_{j,k}$, and the joint distribution of $(R_{5,n,j,k}, \tilde{N}_{2,j,k} + \tilde{\varepsilon}_{2,j,k} + \tilde{\varepsilon}_{1,j,k})_{j,k}$ is the same as that of $(\sum_{\ell=1}^4 R_{\ell,n,j}(z_k), \hat{W}_{n,j}(z_k))_{j,k}$. Take $W_{n,1,j,k} = \tilde{N}_{2,j,k}$ and $W_{n,2,j,k} = \tilde{\varepsilon}_{2,j,k} + \tilde{\varepsilon}_{1,j,k} + R_{5,n,j,k}$, then we have Lemma 1.

To establish (C1)–(C3), we need certain expectations and covariances, which are computed below. Under (R1)–(R3) and the conditions that $\int u k_0(u)\,du = 0$ and $\sigma_0^2 = \int \|u\|^2 k_0(u)\,du < \infty$, for $z \in \mathcal{Z}(\varepsilon_n)$, we have

$$\begin{aligned} &(h_n^d)^{-1} E(w_{0,1}(z) f_{n,j}(X_1, Y_1, z)) \\ &\quad = E\big(f_{n,j}(X,Y,z)|Z=z\big) f_Z(z) + r_{n,j,1}(z) C_n h_n^2, \end{aligned} \tag{7.1}$$

where

$$\begin{aligned} r_{n,j,1}(z) = c_0 \int & h(x,y)\,d\mu(x,y) \\ & \times \big(2d\sigma_0^2 \theta_{n,j,1} + \theta_{n,j,2} h_n^{-2}(2+h_n)\gamma_4^d \exp(-\gamma_5 \varepsilon_n^2 h_n^{-2})\big), \end{aligned}$$

$|\theta_{n,j,1}|, |\theta_{n,j,2}| \le 1$, and γ_4 and γ_5 are positive constants that depend on γ_2 and γ_3 only. Also, for $k \ne k^*$, $z_k, z_k^* \in \mathcal{Z}(\varepsilon_n)$, we have

$$\begin{aligned}
&(h_n^d)^{-2}\operatorname{Cov}(w_{0,1}(z_k) f_{n,j}(X_1, Y_1, z_k), w_{0,1}(z_{k^*}) f_{n,\ell}(X_1, Y_1, z_{k^*}))\\
&\quad = \theta_{j,\ell,k,k^*}(h_n^d)^{-2}(\gamma_2)^{2d}\exp(-0.5\gamma_3 h_n^{-2}\|z_k - z_{k^*}\|^2)C_n^2\\
&\qquad - f_Z(z_k) f_Z(z_{k^*}) E\big(f_{n,j}(X, Y, z_k)|Z = z_k\big) E\big(f_{n,\ell}(X, Y, z_{k^*})|Z = z_{k^*}\big)\\
&\qquad - f_Z(z_k) E\big(f_{n,j}(X, Y, z_k)|Z = z_k\big) r_{n,\ell,1}(z_{k^*}) C_n h_n^2\\
&\qquad - f_Z(z_{k^*}) E\big(f_{n,\ell}(X, Y, z_{k^*})|Z = z_{k^*}\big) r_{n,j,1}(z_k) C_n h_n^2\\
&\qquad - r_{n,j,1}(z_k) r_{n,\ell,1}(z_{k^*}) C_n^2 h_n^4,
\end{aligned} \tag{7.2}$$

where $|\theta_{j,\ell,k,k^*}| \le 1$. Finally, for $z \in \mathcal{Z}(\varepsilon_n)$,

$$\begin{aligned}
&(h_n^d)^{-1}\operatorname{Cov}(w_{0,1}(z) f_{n,j}(X_1, Y_1, z), w_{0,1}(z) f_{n,\ell}(X_1, Y_1, z))\\
&\quad = f_Z(z) E\big(f_{n,j}(X, Y, z) f_{n,\ell}(X, Y, z)|Z = z\big)\int k_0^2(u)\,du + r_{n,j,\ell,2}(z) C_n^2 h_n\\
&\qquad - h_n^d f_Z^2(z) E\big(f_{n,j}(X, Y, z)|Z = z\big) E\big(f_{n,\ell}(X, Y, z)|Z = z\big)\\
&\qquad - h_n^{d+2} C_n r_{n,j,1}(z) f_Z(z) E\big(f_{n,\ell}(X, Y, z)|Z = z\big)\\
&\qquad - h_n^{d+2} C_n r_{n,\ell,1}(z) f_Z(z) E\big(f_{n,j}(X, Y, z)|Z = z\big)\\
&\qquad - h_n^{d+4} C_n^2 r_{n,j,1}(z) r_{n,\ell,1}(z)
\end{aligned} \tag{7.3}$$

and

$$h_n^{-d} E(w_{0,1}(z) f_{n,j}(X_1, Y_1, z))^3 \le C_n^3 c_0 \int k_0^3(u)\,du, \tag{7.4}$$

where

$$|r_{n,j,\ell,2}(z)| \le 2c_0 \int h(x, y)\,d\mu(x, y)\bigg(\sqrt{d}\int \|u\| k_0^2(u)\,du + h_n^{-1}\gamma_6^d e^{-\gamma_7\varepsilon_n^2/h_n^2}\bigg)$$

for some positive constants γ_6 and γ_7 that depend on γ_2 and γ_3 only. Below we will prove (C1)–(C3).

PROOF OF (C1). Let $S_n = \sum_{k=1}^{n_z}(\hat{f}_Z(z_k) - f_Z(z_k))^2$ and $A_n = \{\sqrt{S_n} < \min\{1, (2c_1)^{-1}\}\}$. From (7.1) and (7.3), $ES_n = O(n_Z(h_n^4 + (nh_n^d)^{-1})) = O(n_Z(n \times h_n^d)^{-1})$ and $1/f_Z(z_k) \le c_1$ for all k, $P(A_n^c) \to 0$ as $n \to \infty$. From (7.1), on A_n,

$$\begin{aligned}
&\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}\Bigg(\sum_{\ell=1}^{4}|R_{\ell,n,j}(z_k)|\Bigg)^2\\
&\qquad \le O(1)\Bigg(S_n\Bigg(\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}\tilde{W}_{n,j}^2(z_k)\Bigg) + k_n n_Z C_n^2(nh_n^{d+4}) + k_n C_n^2 n h_n^d S_n^2\Bigg),
\end{aligned}$$

and it follows from (7.3) that

$$E\left(\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}\tilde{W}_{n,j}^2(z_k)\right)=O(k_n n_Z C_n^2).$$

Take

$$T_{1,n}=\frac{k_n n_Z^2 C_n^2}{n h_n^d}+k_n n_Z C_n^2 n h_n^{d+4},$$

then (C1) holds with $T_n=\exp(-(\ln n)^{1/9})$ since $T_{1,n}=O(T_n)$. $\square$

The proof of (C2) is based on the following lemma, which deals with the normal approximation of sum of i.i.d. random vectors.

LEMMA 2. *Suppose that $X_1,\dots,X_n$ are i.i.d. random vectors in R^{d_1} with mean 0 and variance Σ. Suppose that there exist positive constants C, a_2 and a_3 such that $1\le a_2\le a_3\le C$, $\|X_1\|\le C$ and $E\|X_1\|^k\le a_k^k$ for $k=2,3$. Then for $T\ge 1$, there exist random vectors S and Y on the same probability space such that S is distributed as $(X_1+\cdots+X_n)/\sqrt{n}$, Y is multivariate normal with mean 0 and variance Σ and for $n\ge(25/(16a_2^2)+25d_1/12)C^2T^4\exp(3T^2/16)$,*

$$P(\|S-Y\|\ge\alpha)\le\alpha,$$

if

$$\alpha\ge\frac{33.75a_3^3}{\sqrt{n}}(12)^{d_1}e^{(d_1+3)T^2/8}+(48)^{d_1}e^{-3T^2/(32a_2^2)}.$$

The proof of Lemma 2 is given in Section 7.1.1. To prove (C2), note that $\tilde{W}_{n,j}(z_k)=\sum_{i=1}^n(g_{n,j,k}(X_i,Y_i,Z_i)-Eg_{n,j,k}(X_i,Y_i,Z_i))/\sqrt{n}$, where

$$\begin{aligned}&g_{n,j,k}(X_i,Y_i,Z_i)\\&\quad=\frac{\sqrt{c_K}}{\sqrt{f_Z(z_k)h_n^d}}k_0\left(\frac{z_k-Z_i}{h_n}\right)\\&\qquad\times\big(f_{n,j}(X_i,Y_i,z_k)-E\big(f_{n,j}(X,Y,z_k)|Z=z_k\big)\big).\end{aligned}$$

From (7.1)–(7.4), we have

$$\left(\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}\big(g_{n,j,k}(X_i,Y_i,Z_i)-Eg_{n,j,k}(X_i,Y_i,Z_i)\big)^2\right)^{1/2}\le\frac{O(1)C_n\sqrt{k_n n_Z}}{\sqrt{h_n^d}},$$

$$\left(\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}E\big(g_{n,j,k}(X_i,Y_i,Z_i)-Eg_{n,j,k}(X_i,Y_i,Z_i)\big)^2\right)^{1/2}\le O(1)C_n\sqrt{k_n n_Z}$$

and

$$\left(E\left(\sum_{j=1}^{k_n}\sum_{k=1}^{n_Z}\big(g_{n,j,k}(X_i,Y_i,Z_i)-Eg_{n,j,k}(X_i,Y_i,Z_i)\big)^2\right)^{3/2}\right)^{1/3}$$
$$\le C_n\sqrt{k_nn_Z}h_n^{-d/6}O(1).$$

Note that for every constant $M>0$, the condition

$$n\ge\left(\frac{25}{16}+\frac{25k_nn_Z}{12}\right)\left(\frac{MC_n\sqrt{k_nn_Z}}{\sqrt{h_n^d}}\right)^2T_{3,n}^4e^{3T_{3,n}^2/16}$$

holds for large n with $T_{3,n}=(\ln n)^{1/8}$, so Lemma 2 is applicable. From Lemma 2, (C2) holds with any T_n such that $T_{2,n}=O(T_n)$, where

$$T_{2,n}=\frac{(C_n\sqrt{k_nn_Z})^6 12^{2k_nn_Z}e^{(k_nn_Z+3)T_{3,n}^2/4}}{nh_n^d}+(48)^{2k_nn_Z}e^{-\gamma T_{3,n}^2/(C_n\sqrt{k_nn_Z})^2},$$

$\gamma>0$ is a constant. Since $T_{2,n}=O(\exp(-\gamma_1(\ln n)^{1/8}))$ for some constant $\gamma_1>0$, (C2) holds with $T_n=\exp(-(\ln n)^{1/9})$.

The proof of (C3) is based on the following result.

FACT 3. *Suppose that A and B are $d_1\times d_1$ nonnegative definite matrices. Then*

$$\|\sqrt{A}-\sqrt{B}\|\le d_1^{3/4}\sqrt{\|A-B\|}.$$

The proof of Fact 3 is given at the end of the proof of (C3). Note that Fact 3 implies the following: suppose that X_0 and Y_0 are two $d_1\times 1$ normal vectors of mean 0 and covariance matrices A and B, respectively. Let Z be a $d_1\times 1$ normal vector whose elements are i.i.d. $N(0,1)$. Then $\sqrt{A}Z$ is distributed as X_0 and $\sqrt{B}Z$ is distributed as Y_0 and

$$\|\sqrt{A}Z-\sqrt{B}Z\|^2\le\|\sqrt{A}-\sqrt{B}\|^2\|Z\|^2\le d_1^{3/2}\|A-B\|\|Z\|^2$$
$$=O_p(d_1^{5/2}\|A-B\|).$$

Therefore, (C3) holds if $\operatorname{Cov}(\hat{W}_{n,j}(z_k),\hat{W}_{n,\ell}(z_{k^*}))$ is close to

$$\operatorname{Cov}\big(f_{n,j}(X,Y,z_k),f_{n,\ell}(X,Y,z_k)|Z=z_k\big)\delta_{k,k^*},$$

where δ_{k,k^*} is 1 if $k=k^*$ and is 0 otherwise. From (7.1)–(7.4), we have

$$\sum_{j,\ell,k,k^*}\big(\operatorname{Cov}(\hat{W}_{n,j}(z_k),\hat{W}_{n,\ell}(z_{k^*}))$$
$$-\operatorname{Cov}\big(f_{n,j}(X,Y,z_k),f_{n,\ell}(X,Y,z_k)|Z=z_k\big)\delta_{k,k^*}\big)^2$$
$$=h_nC_n^2(k_nn_Z)^2O(1),$$

so (C3) holds with $T_n = \exp(-(\ln n)^{1/9})$ since $(k_n n_Z)^{5/2}\sqrt{h_n C_n^2 (k_n n_Z)^2} = O(\exp(-(\ln n)^{1/9}))$.

PROOF OF FACT 3. Consider first the case where A is diagonal. Let D be a diagonal matrix such that $B = Q^T D Q$ for some Q such that $QQ^T = I$. Let $D = \operatorname{diag}(\lambda_1, \dots, \lambda_{d_1})$, $A = \operatorname{diag}(\alpha_1, \dots, \alpha_{d_1})$, $Q = (q_{i,j})$ and $E = B - A = (e_{i,j})$. Let q_i be the ith column of Q, then $q_i^T D q_j = \alpha_i \delta_{i,j} + e_{i,j}$, where $\delta_{i,j} = 1$ for $i = j$ and $\delta_{i,j} = 0$, otherwise. Write $Dq_k = \sum_{j=1}^{d_1} (q_k^T D q_j) q_j$, then

$$
\begin{aligned}
\|\sqrt{D}q_k - \sqrt{\alpha_k}q_k\|^2 &= \sum_{j=1}^{d_1} \bigl(\sqrt{\lambda_j}q_{j,k} - \sqrt{\alpha_k}q_{j,k}\bigr)^2 \\
&= \sum_{j=1}^{d_1} \bigl(\sqrt{\lambda_j|q_{j,k}|} - \sqrt{\alpha_k|q_{j,k}|}\bigr)^2 |q_{j,k}| \\
&\le \sum_{j=1}^{d_1} \bigl|\lambda_j|q_{j,k}| - \alpha_k|q_{j,k}|\bigr| |q_{j,k}| \\
&\le \left(\sum_{j=1}^{d_1} (\lambda_j q_{j,k} - \alpha_k q_{j,k})^2\right)^{1/2} \left(\sum_{j=1}^{d_1} q_{j,k}^2\right)^{1/2} \\
&= \left(\sum_{j=1}^{d_1} e_{k,j}^2\right)^{1/2}
\end{aligned}
$$

and

$$
\begin{aligned}
\|\sqrt{Q^T D Q} - \sqrt{A}\|^2 &= \sum_{i=1}^{d_1}\sum_{j=1}^{d_1} \bigl(q_i^T \sqrt{D} q_j - q_i^T \sqrt{\alpha_j} q_j\bigr)^2 \\
&\le \sum_{i=1}^{d_1}\sum_{j=1}^{d_1} \|\sqrt{D}q_j - \sqrt{\alpha_j}q_j\|^2 \\
&\le d_1 \sum_{j=1}^{d_1} \left(\sum_{\ell=1}^{d_1} e_{j,\ell}^2\right)^{1/2} \\
&\le (d_1)^{3/2} \left(\sum_{j=1}^{d_1}\sum_{\ell=1}^{d_1} e_{j,\ell}^2\right)^{1/2},
\end{aligned}
$$

so the result in Fact 3 holds if A (or B) is diagonal. For general A and B, write $A = P^T A_0 P$ and $B = Q^T D Q$, where A_0 and D are diagonal and $P^T P = Q^T Q = I$.

Let $B_0 = PQ^T DQP^T$, then we have

$$\begin{aligned}\|\sqrt{A} - \sqrt{B}\| &= \|P^T\sqrt{A_0}P - Q^T\sqrt{D}Q\| \\ &= \|\sqrt{A_0} - PQ^T\sqrt{D}QP^T\| \le d_1^{3/4}\sqrt{\|A_0 - B_0\|} \\ &= d_1^{3/4}\sqrt{\|P^T A_0 P - P^T B_0 P\|} = d_1^{3/4}\sqrt{\|A - B\|}.\end{aligned}$$

The proofs of Fact 3 and Lemma 1 are complete. □

7.1.1. *Proof of Lemma* 2. The proof Lemma 2 is based on several facts, which are taken directly or adapted from some existing results and are stated/proved below in Lemmas 3–5.

In the statements of Lemmas 3 and 4, (S_0, d_0) is a metric space, $\mathcal{B}$ denotes the collection of Borel sets in (S_0, d_0), and for two measures μ_1 and μ_2 defined on $\mathcal{B}$, $\rho_0(\mu_1, \mu_2)$ denotes the Prohorov distance of μ_1 and μ_2, which is defined as

$$\rho_0(\mu_1, \mu_2) = \inf\{\epsilon > 0 : \mu_1(A) < \mu_2(A^\epsilon) + \epsilon, \text{ for all } A \in \mathcal{B}\},$$

where $A^\epsilon = \{x : d^*(x, A) < \epsilon\}$ and $d^*(x, A) = \inf\{d_0(x, y) : y \in A\}$. Here are Lemmas 3–5.

LEMMA 3 (Lemma 2.1 in Berkes and Philipp [1]). *Suppose that P_1 and P_2 are two measures defined on $\mathcal{B}$ and $\rho_0(P_1, P_2) < \alpha$. Then there exists a probability measure Q on the Borel sets of $S_0 \times S_0$ with marginals P_1 and P_2 such that*

$$Q\{(x, y) : d_0(x, y) > \alpha\} \le \alpha.$$

LEMMA 4 (Adapted from Lemma 2.2 in [1]). *Suppose that F and G are two distributions on R^{d_1} with characteristic functions f and g, respectively. Then for $\sigma \in (0, 1]$ and $T > 0$, the Prohorov distance $\rho_0(F, G) \le \alpha$, where*

$$\begin{aligned}\alpha = \sigma T + 3(2^{d_1})e^{-3T^2/32} + \left(\frac{T}{\pi}\right)^{d_1} \int |f(u) - g(u)| e^{-\sigma^2\|u\|^2/2}\, du \\ + F\left(\left\{x : \|x\| \ge \frac{T}{2}\right\}\right).\end{aligned}$$

PROOF. Let H be the $N(0, \sigma^2 I)$ distribution on R^{d_1}, where I is the identity matrix and $\sigma > 0$. Let F_1 be the convolution of F and H and G_1 be the convolution of G and H. Then

$$\rho_0(F, G) \le \rho_0(F_1, G_1) + 2\max\{r, H(\{x : \|x\| \ge r\})\} \qquad \text{for every } r > 0. \tag{7.5}$$

Let f_1, g_1 and h be the characteristic functions of F_1, G_1 and H, respectively, and let γ_F and γ_G be the densities of F_1 and G_1, respectively. Then

$$\begin{aligned}|\gamma_F(x) - \gamma_G(x)| &= (2\pi)^{-d_1}\left|\int e^{-iu^T x}(f_1(u) - g_1(u))\, du\right| \\ &\le (2\pi)^{-d_1}\int |f(u) - g(u)||h(u)|\, du,\end{aligned}$$

which implies that for every borel set B in R^{d_1},

$$\begin{aligned}
&F_1(B) - G_1(B) \\
&\quad \le F_1(B \cap \{x : \|x\| \le T\}) - G_1(B \cap \{x : \|x\| \le T\}) + F_1(\{x : \|x\| \ge T\}) \\
&\quad \le \int_{\{x : \|x\| \le T\}} |\gamma_F(x) - \gamma_G(x)|\, dx + F(\{x : \|x\| \ge T/2\}) \\
&\qquad + H(\{x : \|x\| \ge T/2\}) \\
&\quad \le \underbrace{\left(\frac{T}{\pi}\right)^{d_1} \int |f(u) - g(u)||h(u)|\, du + F(\{x : \|x\| \ge T/2\}) + H(\{x : \|x\| \ge T/2\})}_{II}.
\end{aligned}$$

Note that II is an upper bound for the Prohorov distance $\rho_0(F_1, G_1)$, so for $r \le T/2$, it follows from (7.5) that

$$\begin{aligned}
\rho_0(F, G) &\le II + 2r + 2H(\{x : \|x\| \ge r\}) \\
&\le \left(\frac{T}{\pi}\right)^{d_1} \int |f(u) - g(u)||h(u)|\, du + F(\{x : \|x\| \ge T/2\}) + 2r \\
&\quad + 3P\bigl(\chi^2(d_1) \ge (r/\sigma)^2\bigr).
\end{aligned}$$

Since $h(u) = e^{-\sigma^2\|u\|^2/2}$ and

$$\begin{aligned}
P\bigl(\chi^2(d_1) \ge A\bigr) &\le e^{-tA} E e^{t\chi^2(d_1)}|_{t=3/8} \\
&= e^{-3A/8}(2^{d_1}) \qquad \text{for every } A > 0.
\end{aligned} \tag{7.6}$$

Lemma 4 holds if $r = \sigma T/2$ and $\sigma \in (0, 1]$. □

LEMMA 5 (Adapted from Theorem 1(a) in pages 204–208 in Gnedenko and Kolmogorov [5]). *Suppose that $X_1, \dots, X_n$ are i.i.d. random vectors with mean* 0 *and variance Σ. Suppose that C and a are positive constants such that $\|X_1\| \le C$, $a \le C$ and $E\|X_1\|^k \le a^k$ for $k = 2, 3$. Let f_n be the characteristic function of $(X_1 + \cdots + X_n)/\sqrt{n}$. Then*

$$\left| f_n(u) - \exp\left(-\frac{1}{2}u^T \Sigma u\right)\right| \le \frac{0.25\|u\|^3 a^3}{\sqrt{n}},$$

if $\|u\| \le (0.4\sqrt{n})/C$.

PROOF. Consider first the case where X_1 is univariate. Let $U = f_1(u/\sqrt{n}) - 1$, then

$$U = \frac{\theta_1^* E X_1^2}{2}\left(\frac{u}{\sqrt{n}}\right)^2$$

and

$$U = \frac{EX_1^2}{2}\left(\frac{iu}{\sqrt{n}}\right)^2 + \frac{\theta_1 E|X_1|^3}{3!}\left(\frac{u}{\sqrt{n}}\right)^3,$$

where $|\theta_1^*| \le 1$ and $|\theta_1| \le 1$. Suppose that $|u| \le (0.4\sqrt{n})/C$, then $|U| < 0.1$ and

$$\log(1+U) = U + 0.62\theta_2 U^2,$$

where $|\theta_2| \le 1$. Let $V = \log f_n(u) + E(X_1^2)u^2/2 = E(X_1^2)u^2/2 + n\log(1+U)$, then

$$\begin{aligned}
V &= \frac{n\theta_1 E|X_1|^3 u^3}{3! n^{3/2}} + (0.62)n\theta_2\left(\frac{EX_1^2}{2}\left(\frac{iu}{\sqrt{n}}\right)^2 + \frac{\theta_1 E|X_1|^3}{3!}\left(\frac{u}{\sqrt{n}}\right)^3\right)^2 \\
&= \frac{\lambda_1 |u|^3 a^3}{6\sqrt{n}} + 0.62\left(\frac{\lambda_2 a^4 u^4}{4n} + \frac{\lambda_3 a^5 |u|^5}{6(\sqrt{n})^3} + \frac{\lambda_4 a^6 u^6}{36 n^2}\right) \\
&= \frac{|u|^3 a^3}{\sqrt{n}}\left(\frac{\lambda_1}{6} + 0.62\left(\frac{\lambda_2 a|u|}{4\sqrt{n}} + \frac{\lambda_3 a^2 u^2}{6n} + \frac{\lambda_4 a^3 |u|^3}{36(\sqrt{n})^3}\right)\right),
\end{aligned}$$

where $|\lambda_k| \le 1$ for $k = 1, 2, 3, 4$. Since $a|u|/\sqrt{n} \le 0.4$,

$$V = \frac{\theta_3 (0.25)|u|^3 a^3}{\sqrt{n}},$$

where $|\theta_3| \le 1$. Since $e^V = 1 + \theta_4 |V| e^{|V|}$, where $|\theta_4| \le 1$,

$$\begin{aligned}
f_n(u) &= \exp\left(-\frac{E(X_1^2)u^2}{2}\right)\left(1 + \theta_4 |V| e^{|V|}\right) \\
&= \exp\left(-\frac{E(X_1^2)u^2}{2}\right) + \theta_5\left(\frac{0.25|u|^3 a^3}{\sqrt{n}}\right) e^{|V| - E(X_1^2)u^2/2},
\end{aligned}$$

where $|\theta_5| \le 1$. To find an upper bound for $|V| - E(X_1^2)u^2/2$, note that

$$\left|nU + \frac{E(X_1^2)u^2}{2}\right| = \frac{|\theta_1| E|X_1|^3 |u|^3}{6\sqrt{n}} \le \frac{CEX_1^2|u|^3}{6\sqrt{n}} \le \frac{(0.4)u^2 E(X_1^2)}{6},$$

$n|U| = |\theta_1^*| u^2 E(X_1^2)/2 \le u^2 E(X_1^2)/2$ and

$$\left|n\left(\log(1+U) - U\right)\right| = 0.62 n |\theta_2 U^2| \le 0.62(0.1)\left(\frac{E(X_1^2)u^2}{2}\right)$$

since $|U| < 0.1$. Therefore,

$$\begin{aligned}
|V| - \frac{u^2 E(X_1^2)}{2} &= \left|\frac{E(X_1^2)u^2}{2} + nU + n\left(\log(1+U) - U\right)\right| - \frac{u^2 E(X_1^2)}{2} \\
&\le \frac{(0.4)u^2 E(X_1^2)}{6} + \frac{0.062 E(X_1^2)u^2}{2} - \frac{u^2 E(X_1^2)}{2} \le 0
\end{aligned}$$

and Lemma 5 holds for the univariate case. The result for the general case can be obtained by applying the univariate result with u and X_i replaced by $\|u\|$ and $Y_i = u^T X_i/\|u\|$. $\square$

Now we are ready to prove Lemma 2.

PROOF OF LEMMA 2. Let f_n be the characteristic function of $(X_1 + \cdots + X_n)/\sqrt{n}$ and g be the characteristic function of G, the $N(0, \Sigma)$ distribution. From Lemmas 3–5, there exist random vectors S and Y on the same probability space such that S is distributed as $(X_1 + \cdots + X_n)/\sqrt{n}$, Y is multivariate normal with mean 0 and variance Σ and

$$P(\|S - Y\| \geq \alpha_1) \leq \alpha_1,$$

where

$$\begin{aligned}\alpha_1 = {} & \sigma T + 3(2^{d_1})e^{-3T^2/32} + \frac{0.25a_3^3}{\sqrt{n}}\left(\frac{2}{\pi}\right)^{d_1/2}\frac{T^{d_1}}{\sigma^{d_1+3}}E(\chi^2(d_1))^{3/2} \\ & + 2\left(\frac{2}{\pi}\right)^{d_1/2}\frac{T^{d_1}}{\sigma^{d_1}}P\left(\chi^2(d_1) \geq \frac{0.16n\sigma^2}{C^2}\right) + P\big(\|N(0, \Sigma)\| \geq T/2\big).\end{aligned}$$

From the facts that $E(\chi^2(d_1))^{3/2} \leq (E(\chi^2(d_1))^2)^{3/4}$ and $P(\|N(0, \Sigma)\| \geq T/2) \leq P(\chi^2(d_1) \geq T^2/(4a_2^2))$, (7.6) and the condition $a_2 \geq 1$, we have

$$\begin{aligned}\alpha_1 \leq {} & \sigma T + 4(2^{d_1})e^{-3T^2/(32a_2^2)} + \frac{0.25a_3^3}{\sqrt{n}}\left(\frac{2}{\pi}\right)^{d_1/2}\frac{T^{d_1}}{\sigma^{d_1+3}}(2d_1 + d_1^2)^{3/4} \\ & + 2\left(\frac{2}{\pi}\right)^{d_1/2}\frac{T^{d_1}}{\sigma^{d_1}}(2^{d_1})e^{-0.06n\sigma^2/(C^2)}.\end{aligned}$$

Set $\sigma = T^{-1}e^{-3T^2/32}$, then $0 < \sigma \leq 1$, $T/\sigma < 12e^{T^2/8}$ and $1/\sigma < 3e^{T^2/8}$, which, together with the fact that $(2/\pi)^{d_1/2}(2d_1 + d_1^2)^{3/4} < 5$, gives that

$$\begin{aligned}\alpha_1 & \leq \big(1 + 4(2^{d_1})\big)e^{-3T^2/(32a_2^2)} + \frac{33.75a_3^3}{\sqrt{n}}(12)^{d_1}e^{(d_1+3)T^2/8} \\ & \quad + 2(19.15)^{d_1}e^{d_1T^2/8}e^{-0.06n\sigma^2/(C^2)} \\ & \leq \frac{33.75a_3^3}{\sqrt{n}}(12)^{d_1}e^{(d_1+3)T^2/8} + (48)^{d_1}e^{-3T^2/(32a_2^2)} \leq \alpha,\end{aligned}$$

if $0.06n\sigma^2/(C^2) \geq d_1T^2/8 + 3T^2/(32a_2^2)$, which corresponds to $n \geq (25/(16 \times a_2^2) + 25d_1/12)C^2T^4\exp(3T^2/16)$ and we have Lemma 2. $\square$

7.2. *Proof of Theorem 3.1.* To prove Theorem 3.1, we apply Lemma 1 by taking the $f_{n,j}(X,Y,z)$'s to be the functions $\phi_\ell^*(X)\phi_{\ell'}^*(X)$, $\phi_\ell^*(X)\psi_m^*(Y)$ and $\psi_m^*(Y)\psi_{m'}^*(Y)$, where $1 \le \ell \le \ell' \le p_n$ and $1 \le m \le m' \le q_n$. In such case, (3.19) holds under conditions (B1) and (B2). To see this, for each $1 \le k \le n_Z$ and $1 \le j \le p_n$, let $\phi_{n,j,k}^*$ be the jth component of ϕ^* when $z = z_k$. Then $\phi_{n,j,k}^*(x) = \sum_{i=1}^{p_n} a_{n,i,j,k}\phi_{n,i}(x)$ for some $a_{n,i,j,k}$'s and

$$\begin{aligned}1 &= E\big((\phi_{n,j,k}^*(X))^2 | Z = z_k\big)\\ &= E\left(\left(\sum_{i=1}^{p_n} a_{n,i,j,k}\phi_{n,i}(X)\right)^2 \middle| Z = z_k\right)\\ &\ge \delta_n \sum_{i=1}^{p_n} a_{n,i,j,k}^2,\end{aligned}$$

so $|\phi_{n,j,k}^*(x)| \le \sqrt{\sum_{i=1}^{p_n} a_{n,i,j,k}^2}\sqrt{\sum_{i=1}^{p_n} \phi_{n,i}^2(x)} \le \sqrt{p_n/\delta_n}$. Similarly, for each $1 \le k \le n_Z$ and $1 \le j \le q_n$, let $\psi_{n,j,k}^*$ be the jth component of ψ^* when $z = z_k$, then $|\psi_{n,j,k}^*(x)| \le \sqrt{q_n/\delta_n}$. Thus, (3.19) holds with $C_n = \max\{1, (p_n + q_n)/\delta_n\}$ and it follows from Lemma 1 that $\sum_{k=1}^{n_Z} \|\hat{V}^*(z_k) - V^*(z_k)\|^2$ has the same distribution as $\sum_{k=1}^{n_Z}(nh_n^d c_K f_Z(z_k))^{-1}\|W_{n,1,k} + W_{n,2,k}\|^2$, where the $W_{n,1,k}$'s and $W_{n,2,k}$'s are random matrices such that each element in $W_{n,1,k}$ is normal with mean zero and variance bounded by $C_n^2 = (\max\{1, (p_n + q_n)/\delta_n\})^2$, and $\sum_{k=1}^{n_Z} \|W_{n,2,k}\|^2 = O_P(\exp(-(\ln n)^{1/9}))$. Therefore,

$$\sum_{k=1}^{n_Z} \|\hat{V}^*(z_k) - V^*(z_k)\|^2 = O_P((nh_n^d)^{-1}(\ln n)^{1/8}). \tag{7.7}$$

To control the difference between $g(\hat{V}^*(z_k), \alpha^*)$ and $g(V^*(z_k), \alpha^*)$ for $1 \le k \le n_Z$, for a $(p_n + q_n) \times (p_n + q_n)$ matrix U, let

$$g_{i,j}^*(U) = \begin{cases} g_{i,j}(U), & \text{if } (i,j) = (1,2) \text{ or } (2,1);\\ g_{i,j}^{-1}(U), & \text{if } (i,j) = (1,1) \text{ or } (2,2).\end{cases} \tag{7.8}$$

For $1 \le k \le n_Z$, let $\Delta_{i,j,k} = g_{i,j}^*(\hat{V}^*(z_k)) - g_{i,j}^*(V^*(z_k))$ for $1 \le i, j \le 2$. Then from the fact that $\|AB\| \le \|A\|\|B\|$ for two matrices A and B, we have

$$\begin{aligned}&\|g(\hat{V}^*(z_k), \alpha^*) - g(V^*(z_k), \alpha^*)\|\\ &\quad \le \prod_{i=1}^{2}\prod_{j=1}^{2}\big(\|g_{i,j}^*(V^*(z_k))\| + \|\Delta_{i,j,k}\|\big) - \prod_{i=1}^{2}\prod_{j=1}^{2}\|g_{i,j}^*(V^*(z_k))\|\\ &\qquad + \|g_{1,1}(\hat{V}^*(z_k)) - g_{1,1}(V^*(z_k))\|\|\alpha^*(\alpha^*)^T\|.\end{aligned} \tag{7.9}$$

To control the $\Delta_{1,1,k}$ and $\Delta_{2,2,k}$ in (7.9), the following result is needed.

FACT 4. *Suppose that A is a $p \times p$ invertible matrix and $\Delta = A - I_p$. Then $\|A^{-1} - I_p + \Delta\| \le \|A^{-1} - I_p\|\|\Delta\|$ and*

$$\|A^{-1} - I_p\| \le \frac{\|\Delta\|}{1 - \|\Delta\|} \qquad \textit{if } \|\Delta\| < 1.$$

PROOF. Let $B = A^{-1} - I_p$. Then $B = -\Delta - B\Delta$, so $\|B + \Delta\| = \|B\Delta\| \le \|B\|\|\Delta\|$. Also,

$$\|B\| \le \|\Delta\|(1 + \|B\|). \tag{7.10}$$

Apply (7.10) and we have

$$\|B\| \le \frac{\|\Delta\|}{1 - \|\Delta\|} \qquad \text{if } \|\Delta\| < 1.$$

Since $\|\alpha^*\| = 1$ and for $1 \le k \le n_Z$, $g_{1,1}(V^*(z_k)) = I_{p_n}$, $g_{2,2}(V^*(z_k)) = I_{q_n}$ and $\|g_{1,2}(V^*(z_k))\|^2 = \|g_{2,1}(V^*(z_k))\|^2 \le (p_n + q_n)$, from (7.9) and Fact 4, we have

$$\begin{aligned}\sum_{k=1}^{n_Z} &\|g(\hat{V}^*(z_k), \alpha^*) - g(V^*(z_k), \alpha^*)\|^2 \\ &= O_P\big((nh_n^d)^{-1}(\ln n)^{1/8} n_Z^2 (p_n + q_n)^3\big) \\ &= O_P((nh_n^d)^{-1}(\ln n)^{1/4}),\end{aligned}$$

which gives (3.21) since $|\hat{\rho}^2(z_k) - \rho^2_{p_n,q_n}(z_k)| \le \|g(\hat{V}^*(z_k), \alpha^*) - g(V^*(z_k), \alpha^*)\|$ for $1 \le k \le n_Z$. (3.22) follows from (3.21) and the fact that $\sum_{k=1}^{n_Z}(\hat{f}_Z(z_k) - f_Z(z_k))^2$ is $O_P(n_Z(nh_n^d)^{-1})$. The proof of Theorem 3.1 is complete. $\square$

7.3. *Proof of Theorem* 3.2. From Lemma 1, the joint distribution of $\hat{V}^*(z_k): 1 \le k \le n_Z$ is the same as that of $V^*(z_k) + (nh_n^d c_K f_Z(z_k))^{-1/2}(W_{n,1,k} + W_{n,2,k}): 1 \le k \le n_Z$, where

$$\sum_{k=1}^{n_Z} \|W_{n,2,k}\|^2 = O_P(\exp(-(\ln n)^{1/9})) \tag{7.11}$$

and $W_{n,1,k}$'s are independent symmetric normal matrices of mean zero. To describe the covariance structure of each $W_{n,1,k}$, let $\phi^* = (\phi_1^*, \dots, \phi_{p_n}^*)^T$, $\psi^* = (\psi_1^*, \dots, \psi_{q_n}^*)^T$ and let V_0 be the $(p_n + q_n) \times (p_n + q_n)$ symmetric matrix such that $g_{1,1}(V_0) = \phi^*(X)\phi^*(X)^T$, $g_{1,2}(V_0) = \phi^*(X)\psi^*(Y)^T$ and $g_{2,2}(V_0) = \psi^*(Y)\psi^*(Y)^T$. For $1 \le k \le n_Z$ and $1 \le m, \ell \le p_n + q_n$, let $U_{k,m,\ell}$ and $V_{0,m,\ell}$ be the (m, ℓ)th elements of $W_{n,1,k}$ and V_0, respectively, then

$$\mathrm{Cov}(U_{k,m,\ell}, U_{k,m',\ell'}) = \mathrm{Cov}(V_{0,m,\ell}, V_{0,m',\ell'} | Z = z_k)$$

for (m,ℓ), $(m',\ell') \in \{(i,j): 1 \le i \le j \le (p_n+q_n)\}$. For $1 \le k \le n_Z$, let $\tilde{V}_k = V^*(z_k) + (nh_n^d c_K f_Z(z_k))^{-1/2}(W_{n,1,k} + W_{n,2,k})$ and

$$\begin{aligned} A_1(z_k) &= g(\tilde{V}_k, \alpha^*) g_{1,1}(\tilde{V}_k) \\ &= g_{1,2}(\tilde{V}_k)(g_{2,2}(\tilde{V}_k))^{-1} g_{2,1}(\tilde{V}_k) \\ &\quad - g_{1,1}(\tilde{V}_k)\alpha^*(\alpha^*)^T g_{1,1}(\tilde{V}_k), \end{aligned}$$

and let $\tilde{\rho}_0^2(z_k)$ be the largest eigenvalue of $A_1(z_k)(g_{1,1}(\tilde{V}_k))^{-1}$, then the joint distribution of $\hat{\rho}^2(z_k): 1 \le k \le n_Z$ is the same as that of $\tilde{\rho}_0^2(z_k): 1 \le k \le n_Z$. For $1 \le i, j \le 2$ and $1 \le k \le n_Z$, let $\Delta_{i,j,k} = g_{i,j}(\tilde{V}_k) - g_{i,j}(V^*(z_k))$, then from (7.7),

$$\sum_{k=1}^{n_Z}\sum_{i=1}^{2}\sum_{j=1}^{2} \|\Delta_{i,j,k}\|^2 = O_P((nh_n^d)^{-1}(\ln n)^{1/8}) \tag{7.12}$$

and

$$\begin{aligned} A_1(z_k) &= g_{1,2}(V^*(z_k))(g_{2,2}(\tilde{V}_k))^{-1} g_{2,1}(V^*(z_k)) \\ &\quad - g_{1,1}(\tilde{V}_k)\alpha^*(\alpha^*)^T g_{1,1}(\tilde{V}_k) + g_{1,2}(V^*(z_k))\Delta_{2,1,k} \\ &\quad + \Delta_{1,2,k} g_{2,1}(V^*(z_k)) + \Delta_{1,2,k}\Delta_{2,1,k} \\ &\quad - g_{1,2}(V^*(z_k))\Delta_{2,2,k}\Delta_{2,1,k} \\ &\quad - \Delta_{1,2,k}\Delta_{2,2,k} g_{2,1}(V^*(z_k)) + R_{1,n,k}, \end{aligned} \tag{7.13}$$

where

$$\begin{aligned} R_{1,n,k} &= \Delta_{1,2,k}\big(g_{2,2}(\tilde{V}_k)^{-1} - I_{q_n}\big)\Delta_{2,1,k} \\ &\quad + g_{1,2}(V^*(z_k))\big(g_{2,2}(\tilde{V}_k)^{-1} - I_{q_n} + \Delta_{2,2,k}\big)\Delta_{2,1,k} \\ &\quad + \Delta_{1,2,k}\big(g_{2,2}(\tilde{V}_k)^{-1} - I_{q_n} + \Delta_{2,2,k}\big) g_{2,1}(V^*(z_k)). \end{aligned}$$

To simplify the expression for $A_1(z_k)$ in (7.13), we will make use of the following properties.

(C4) The elements of the matrix $g_{1,2}(V^*(z_k))$ are zeros except that the $(1,1)$th element is 1.
(C5) For $(i,j) \in \{(1,2),(2,1)\}$, $g_{i,j}(V^*(z_k))$'s first row (or first column) is either the first row or the first column of $g_{i',j'}(V^*(z_k))$ for $(i',j') \ne (i,j)$.
(C6) The $(1,1)$th element in $g_{2,2}(\hat{V}^*(z_k))$ is 1.

Here (C4) follows from the conditional independence assumption and (3.16), and (C5) and (C6) follow from (3.15). From (C6), $g_{2,2}(\tilde{V}_k)$ can be expressed as

$$g_{2,2}(\tilde{V}_k) = \begin{pmatrix} 1 & B_k^T \\ B_k & D_k \end{pmatrix}$$

for some matrices B_k and D_k, so the $(1,1)$th element of $g_{2,2}(\tilde{V}_k)^{-1}$ is $(1 + B_k^T(D_k - B_k B_k^T)^{-1} B_k)$. Let $J = \alpha^*(\alpha^*)^T$, then by (C4) and (C5), we have

$$g_{1,2}(V^*(z_k))(g_{2,2}(\tilde{V}_k))^{-1} g_{2,1}(V^*(z_k)) = \bigl(1 + B_k^T(D_k - B_k B_k^T)^{-1} B_k\bigr) J,$$

$g_{1,2}(V^*(z_k))\Delta_{2,1,k} = J\Delta_{1,1,k}$ and $B_k^T B_k J = g_{1,2}(V^*(z_k))(\Delta_{2,2,k})^2 g_{2,1}(V^*(z_k))$, so the expression for $A_1(z_k)$ in (7.13) becomes

$$\begin{aligned} &B_k^T\bigl((D_k - B_k B_k^T)^{-1} - I_{q_n-1}\bigr)B_k J + g_{1,2}(V^*(z_k))(\Delta_{2,2,k})^2 g_{2,1}(V^*(z_k)) \\ &\quad - \Delta_{1,1,k} g_{1,2}(V^*(z_k)) g_{2,1}(V^*(z_k))\Delta_{1,1,k} + \Delta_{1,2,k}\Delta_{2,1,k} \\ &\quad - g_{1,2}(V^*(z_k))\Delta_{2,2,k}\Delta_{2,1,k} - \Delta_{1,2,k}\Delta_{2,2,k} g_{2,1}(V^*(z_k)) + R_{1,n,k}. \end{aligned}$$

Let

$$\begin{aligned} A_2(z_k) = {}& g_{1,2}(V^*(z_k))(g_{2,2}(W_{1,n,k}))^2 g_{2,1}(V^*(z_k)) \\ &- g_{1,1}(W_{1,n,k}) g_{1,2}(V^*(z_k)) g_{2,1}(V^*(z_k)) g_{1,1}(W_{1,n,k}) \\ &+ g_{1,2}(W_{1,n,k}) g_{2,1}(W_{1,n,k}) - g_{1,2}(V^*(z_k)) g_{2,2}(W_{1,n,k}) g_{2,1}(W_{1,n,k}) \\ &- g_{1,2}(W_{1,n,k}) g_{2,2}(W_{1,n,k}) g_{2,1}(V^*(z_k)) \end{aligned}$$

and

$$\begin{aligned} R_{2,n,k} = {}& B_k^T\bigl((D_k - B_k B_k^T)^{-1} - I_{q_n-1}\bigr)B_k J \\ &- (nh_n^d c_K f_Z(z_k))^{-1} A_2(z_k) + g_{1,2}(V^*(z_k))(\Delta_{2,2,k})^2 g_{2,1}(V^*(z_k)) \\ &- \Delta_{1,1,k} g_{1,2}(V^*(z_k)) g_{2,1}(V^*(z_k))\Delta_{1,1,k} + \Delta_{1,2,k}\Delta_{2,1,k} \\ &- g_{1,2}(V^*(z_k))\Delta_{2,2,k}\Delta_{2,1,k} - \Delta_{2,1,k}\Delta_{2,2,k} g_{2,1}(V^*(z_k)), \end{aligned}$$

then

$$A_1(z_k) = \frac{A_2(z_k)}{nh_n^d c_K f_Z(z_k)} + R_{1,n,k} + R_{2,n,k}, \tag{7.14}$$

where

$$\sum_{k=1}^{n_Z} (\|R_{1,n,k}\|^2 + \|R_{2,n,k}\|^2) = O_P\biggl(\frac{\exp(-(\ln n)^{1/9})(\ln n)^{1/8}}{(nh_n^d)^2}\biggr) \tag{7.15}$$

from Fact 4, (7.11) and (7.12), and a simple expression for $A_2(z_k)$ can be obtained as stated below in (C7), which follows from (C4) and (C5).

(C7) For $1 \le k \le n_Z$, $A_2(z_k) = C_k C_k^T$, where C_k is the $p_n \times q_n$ matrix obtained by replacing elements in the first row and first column of $g_{1,2}(W_{1,n,k})$ with zeros.

Note that from (C7), we have that

$$\sum_{k=1}^{n_Z} \|A_2(z_k)\|^2 = O_P\big(n_Z(p_n-1)^2(q_n-1)^2\big) = O_P((\ln n)^{1/8}),$$

which, together with (7.14) and (7.15), implies that

$$\sum_{k=1}^{n_Z} \|A_1(z_k)\|^2 = O_P((nh_n^d)^{-2}(\ln n)^{1/8}), \tag{7.16}$$

and then it follows from (7.16), Fact 4 and (7.12) that

$$\sum_{k=1}^{n_Z} \|A_1(z_k)(g_{1,1}(\tilde{V}_k))^{-1} - A_1(z_k)\|^2 = O_p((nh_n^d)^{-3}(\ln n)^{1/4}). \tag{7.17}$$

For $1 \le k \le n_Z$, let $\lambda_{0,k}$ be the largest eigenvalue of $A_2(z_k)$ and recall that $\tilde{\rho}_0^2(z_k)$ is the largest eigenvalue of $A_1(z_k)(g_{1,1}(\tilde{V}_k))^{-1}$. Then by (7.14), (7.15) and (7.17),

$$\sum_{k=1}^{n_Z} \big(nh_n^d c_K f_Z(z_k)\tilde{\rho}_0^2(z_k) - \lambda_{0,k}\big)^2 = O_P(\exp(-(\ln n)^{1/9})(\ln n)^{1/8}). \tag{7.18}$$

Let $\tilde{f}_k$, $\tilde{\rho}(z_k)$ and $\lambda_k : 1 \le k \le n_Z$ be random variables such that the joint distribution of $(\tilde{f}_k, \tilde{\rho}(z_k)) : 1 \le k \le n_Z$ is the same as that of $(\hat{f}_Z(z_k), \hat{\rho}(z_k)) : 1 \le k \le n_Z$, and the joint distribution of $(\tilde{\rho}(z_k), \lambda_k) : 1 \le k \le n_Z$ is the same as that of $(\tilde{\rho}_0(z_k), \lambda_{0,k}) : 1 \le k \le n_Z$. Note that from (7.18) and the fact that

$$\sum_{k=1}^{n_Z} \|A_2(z_k)\|^2 = O_P\big(n_Z(p_n-1)^2(q_n-1)^2\big),$$

we have that

$$\sum_{k=1}^{n_Z} nh_n^d c_K f_Z(z_k)\tilde{\rho}^2(z_k) = \sqrt{O_P\big(n_Z^2(p_n-1)^2(q_n-1)^2\big)} = O_P((\ln n)^{1/16}),$$

so $nh_n^d \sum_{k=1}^{n_Z} (\hat{\rho}(z_k))^2 = O_P((\ln n)^{1/16})$,

$$\begin{aligned}
&\left| nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k)(\hat{\rho}(z_k))^2 - nh_n^d c_K \sum_{k=1}^{n_Z} f_Z(z_k)(\hat{\rho}(z_k))^2 \right| \\
&\quad \le nh_n^d c_K \left(\sum_{k=1}^{n_Z} \big(\hat{f}_Z(z_k) - f_Z(z_k)\big)^2 \right)^{1/2} \sum_{k=1}^{n_Z} (\hat{\rho}(z_k))^2 \\
&\quad = O_P((\ln n)^{1/16})(O_P(n_Z(nh_n^d)^{-1}))^{1/2} \\
&\quad = O_P((nh_n^d)^{-1/2}(\ln n)^{3/32})
\end{aligned}$$

and

$$\left|nh_n^d c_K \sum_{k=1}^{n_Z} \tilde{f}_k(\tilde{\rho}(z_k))^2 - \sum_{k=1}^{n_Z} \lambda_k\right|$$
$$\le O_P((nh_n^d)^{-1/2}(\ln n)^{3/32}) + \left|nh_n^d c_K \sum_{k=1}^{n_Z} f_Z(z_k)(\tilde{\rho}(z_k))^2 - \sum_{k=1}^{n_Z} \lambda_k\right|$$
$$[\text{by (7.18)}] \le O_P((nh_n^d)^{-1/2}(\ln n)^{3/32}) + \sqrt{n_Z}(O_P(\exp(-(\ln n)^{1/9})(\ln n)^{1/8}))^{1/2}$$
$$= O_P(\exp(-0.5(\ln n)^{1/9})(\ln n)^{3/32}).$$

The proof of Theorem 3.2 is complete.

7.4. *Proof of Corollary* 1. To prove Corollary 1, it is sufficient to establish (3.25) and (3.26). To see this, let $\tilde{f}_k$, $\tilde{\rho}^2(z_k)$ and $\lambda_k : 1 \le k \le n_Z$ be as in Theorem 3.2, then

$$\frac{nh_n^d c_K \sum_{k=1}^{n_Z} \hat{f}_Z(z_k)\hat{\rho}^2(z_k) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}}$$

has the same distribution as

$$\frac{nh_n^d c_K \sum_{k=1}^{n_Z} \tilde{f}_k \tilde{\rho}^2(z_k) - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}}$$
$$= \underbrace{\frac{nh_n^d c_K \sum_{k=1}^{n_Z} \tilde{f}_k \tilde{\rho}^2(z_k) - \sum_{k=1}^{n_Z} \lambda_k}{\sqrt{n_Z \sigma^2_{p_n,q_n}}}}_{I} + \underbrace{\frac{\sum_{k=1}^{n_Z} \lambda_k - n_Z \mu_{p_n,q_n}}{\sqrt{n_Z \sigma^2_{p_n,q_n}}}}_{II}.$$

Suppose that (3.25) holds, then $I \to 0$ almost surely by (3.24) and Theorem 3.2. Also, (3.26) says that *II* converges to $N(0, 1)$ in distribution. Therefore, (3.27) holds if (3.25) and (3.26) hold.

To establish (3.26), we will verify the Lyapounov condition

$$\lim_{n\to\infty} \sum_{k=1}^{n_Z} \frac{E|\lambda_k - \mu_{p_n,q_n}|^3}{(n_Z \sigma^2_{p_n,q_n})^{3/2}} = 0, \tag{7.19}$$

and then apply Lindeberg's central limit theorem. Let λ be the largest eigenvalue of CC^T. Then $\lambda \le \operatorname{tr}(CC^T)$, where $\operatorname{tr}(CC^T)$ is the trace of CC^T, which follows the χ^2 distribution with degrees of freedom $m_{1,n} = (p_n - 1)(q_n - 1)$. Therefore,

$$E\lambda^3 \le E(\operatorname{tr}(CC^T))^3 = m_{1,n}(m_{1,n} + 2)(m_{1,n} + 4),$$

which implies that $E|\lambda_1 - \mu_{p_n,q_n}|^3 = O(p_n^3 q_n^3)$, so (7.19) follows from (3.25) and (3.26) holds.

It remains to prove (3.25). Consider first the case where (i) holds. By Theorem 1.1 in Johnstone [7],

$$\frac{\lambda_1 - \mu_n}{\sigma_n} \text{ converges in distribution} \qquad \text{as } n \to \infty, \tag{7.20}$$

where

$$\mu_n = (\sqrt{q_n - 2} + \sqrt{p_n - 1})^2$$

and

$$\sigma_n = (\sqrt{q_n - 2} + \sqrt{p_n - 1})\left(\frac{1}{q_n - 2} + \frac{1}{p_n - 1}\right)^{1/3}.$$

Here the limiting distribution is the Tracy–Widom law of order 1. Let F denote its cumulative distribution function. Suppose that ϵ, t_1 and t_2 are real numbers such that $t_1 < t_1 + \epsilon < t_2 - \epsilon$, which implies that $F(t_2) > F(t_2 - \epsilon)$ and $F(t_1 + \epsilon) > F(t_1)$. From (7.20),

$$P\big(\lambda_1 > \mu_n + (t_2 - \epsilon)\sigma_n\big) \geq 1 - F(t_2)$$

and

$$P\big(\lambda_1 < \mu_n + (t_1 + \epsilon)\sigma_n\big) \geq F(t_1),$$

if n is large enough. For such n, we have

$$\sigma^2_{p_n,q_n} \geq \frac{\min(F(t_1), 1 - F(t_2))(t_2 - t_1 - 2\epsilon)^2\sigma_n^2}{4},$$

which gives (3.25). The proof of (3.25) for the case where (ii) holds can be done by reversing the roles of p_n and q_n. The proof of Corollary 1 is complete.

Acknowledgments. The author thank Dr. Su-Yun Huang, Dr. I-Ping Tu and Dr. Hung Chen for helpful discussions, and thank the reviewers for constructive comments.

REFERENCES

[1] BERKES, I. and PHILIPP, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54. MR0515811

[2] DAUDIN, J.-J. (1980). Partial association measures and an application to qualitative regression. *Biometrika* **67** 581–590. MR0601095

[3] DAUXOIS, J. and NKIET, G. M. (1998). Nonlinear canonical analysis and independence tests. *Ann. Statist.* **26** 1254–1278. MR1647653

[4] DAUXOIS, J. and NKIET, G. M. (2002). Measures of association for Hilbertian subspaces and some applications. *J. Multivariate Anal.* **82** 263–298. MR1921388

[5] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA. MR0233400

[6] HSING, T., LIU, L.-Y., BRUN, M. and DOUGHERTY, E. R. (2005). The coefficient of intrinsic dependence. *Pattern Recognition* **38** 623–636.

[7] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[8] PAPARODITIS, E. and POLITIS, D. N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Statist. Math.* **52** 139–159. MR1771485

[9] RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10** 441–451. MR0115203

[10] ROMANOVIČ, V. A. (1975). The maximal partial correlation coefficient of two σ-algebras relative to a third σ-algebra. *Izv. Vysš. Učebn. Zaved. Matematika* **10** 94–96. MR0420757

[11] SCHUMAKER, L. L. (1981). *Spline Functions*. Wiley, New York. MR0606200

[12] SU, L. and WHITE, H. (2007). A consistent characteristic function-based test for conditional independence. *J. Econometrics* **141** 807–834. MR2413488

[13] SU, L. and WHITE, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* **24** 829–864. MR2428851

DEPARTMENT OF STATISTICS
NATIONAL CHENGCHI UNIVERSITY
NO. 64, SEC. 2, ZHINAN RD.
TAIPEI 11605, TAIWAN
REPUBLIC OF CHINA
E-MAIL: tmhuang@nccu.edu.tw