Ann. Inst. Statist. Math.
Vol. 57, No. 1, 105–127 (2005)

A TEST FOR INDEPENDENCE OF TWO STATIONARY INFINITE ORDER AUTOREGRESSIVE PROCESSES

EUNHEE KIM AND SANGYEOL LEE

Department of Statistics, Seoul National University, Seoul 151-742, Korea

(Received June 6, 2002; revised January 16, 2004)

Abstract. This paper considers the independence test for two stationary infinite order autoregressive processes. For a test, we follow the empirical process method and construct the Cramér-von Mises type test statistics based on the least squares residuals. It is shown that the proposed test statistics behave asymptotically the same as those based on true errors. Simulation results are provided for illustration.

Key words and phrases: Independence test, infinite order autoregressive processes, the Cramér-von Mises test, residual empirical process, weak convergence.

1. Introduction

In this paper, we consider the problem of testing the independence of two stationary time series. For the past two decades, the issue has drawn much attention from many researchers. For instance, Haugh (1976) proposed an independence test for the errors in ARMA models based on the sum of squares of residual cross correlations. Later, adopting his idea, Pierce (1977), Geweke (1981) and Hong (1996) studied the independence test for two stationary time series. In fact, the method using the cross correlations has been much popular in the time series context since it is a crucial task to figure out the dependence structure of given time series in a correct manner, and any model selection procedures require a step for diagnostics to set up a true model. However, the cross correlation method merely guarantees the uncorrelatedness of observations and does not ensure the independence. Moreover, the cross correlation just checks the linear relationship but cannot find a nonlinear dependence. Therefore, instead of it, here we employ the empirical process method devised by Hoeffding (1948) and Blum *et al.* (1961). Their test statistics essentially fall into the category of Cramér-von Mises (CV) statistics, and have been applied under a variety of circumstances. Recently, their method has been adopted by Skaug and Tjøstheim (1993), Delgado (1996), Hong (1998) and Delgado and Mora (2000) aimed at developing a serial independence test.

In this paper, we focus on the independence test for two stationary infinite order autoregressive processes. We adopted autoregressive processes since they include the most popular ARMA processes in time series analysis, and a method based on residuals usually discards correlation effects. In fact, the infinite order autoregressive process has been in a central position in the study of the asymptotic efficiency of model selection criteria (see Shibata (1980), Lee and Karagrigoriou (2001) and the papers therein). For a test, we construct the CV statistic based on residuals. It will be seen that the limiting distribution of the residual based CV statistic is the same as the CV statistic based on the true errors.

In fact, the CV statistic is designed for testing the independence for a specific lag k. However, in real situations one should check the independence for several lags, say, $|k| \leq K$, where K is a positive integer larger than 1, since the test for only one lag is not sufficient to ensure the independence between the two times series. To this end, we consider three types of test statistics. First, we consider the summation of the CV statistics based on residuals, which Skaug and Tjøstheim (1993) suggested as a test statistic for testing serial independence. Second, we consider the weighted summation of the CV statistics which was proposed in Hong (1998) for testing serial independence, since the summation type test statistic may suffer from severe size distortions as K increases. Finally, we propose as a test statistic the maximum of the CV test statistics. We consider this because it is less affected by the CV statistics with large values and will have more stability compared to other tests. Our simulation study shows that the method based on the CV statistic with residuals turns out to be suitable for the independence test of two stationary time series.

The rest of the paper is organized as follows. In Section 2, we present the procedure for the independence test of two stationary infinite order autoregresive processes. In particular, we derive the asymptotic distribution of the proposed test statistics. This task requires extending the result of Lee and Wei (1999) to the residual empirical process with bivariate time parameters, which may be of independent interest in its own sake. In Section 3, we report the result of our simulation study. Finally, in Section 4 we provide the proofs of the theorems presented in Section 2.

2. Main results

Suppose that $\{X_t\}$ and $\{Y_t\}$ satisfy the following difference equations:

$$X_t - \mu - \sum_{j=1}^{\infty} \phi_j (X_{t-j} - \mu) = \epsilon_t, \qquad t = 1, \ldots, n$$

and

$$Y_t - \nu - \sum_{j=1}^{\infty} \theta_j (Y_{t-j} - \nu) = \eta_t, \qquad t = 1, \ldots, n,$$

where (ϵ_t, η_t) are random vectors with common distribution F, ϵ_t and η_t are iid r.v.'s with marginal distributions F_1 and F_2, respectively, and $E\epsilon_t^4 + E\eta_t^4 < \infty$. Furthermore, both $A_1(z) := 1 - \sum_{j=1}^{\infty} \phi_j z^j$ and $A_2(z) := 1 - \sum_{j=1}^{\infty} \theta_j z^j$ are assumed to be analytic on an open neighborhood of the closed unit disk D in the complex plane and have no zeroes on D. It can be easily seen that the last condition implies

$$|\phi_j| + |\theta_j| \leq C\rho^j, \qquad C > 0, \; 0 < \rho < 1,$$

(cf. Lee and Wei (1999)). It is well-known that the AR(∞) process covers a broad class of stationary processes including invertible ARMA processes (cf. Brockwell and Davis (1990)).

Suppose that one wishes to test the hypotheses

$$H_0 : \{X_t\} \text{ and } \{Y_t\} \text{ are independent.} \quad \text{vs.} \quad H_1 : \text{not } H_0.$$

The above is equivalent to testing

$$H_0^{'} : \{\epsilon_t\} \text{ and } \{\eta_t\} \text{ are independent.} \quad \text{vs.} \quad H_1^{'} : \text{not } H_0^{'}$$

since the independence between the processes themselves implies that of the error processes and the converse is also true (cf. Lee and Wei (1999), p. 247). For testing H_0' vs. H_1', we consider employing the CV test statistic (cf. Hoeffding (1948)),

$$B_{nk} = \begin{cases} (n-|k|)^{-1} \sum_{i=|k|+1}^{n} S_{nk}^2(\epsilon_{i-k}, \eta_i), & k \geq 0, \\ (n-|k|)^{-1} \sum_{i=|k|+1}^{n} S_{nk}^2(\epsilon_i, \eta_{i-|k|}), & k < 0, \end{cases}$$

with

$$S_{nk}(x,y) = \begin{cases} (n-|k|)^{-1} \sum_{t=|k|+1}^{n} I(\epsilon_{t-k} \leq x) I(\eta_t \leq y) \\ \quad - (n-|k|)^{-2} \sum_{t=|k|+1}^{n} I(\epsilon_{t-k} \leq x) \sum_{t=|k|+1}^{n} I(\eta_t \leq y), & k \geq 0 \\ (n-|k|)^{-1} \sum_{t=|k|+1}^{n} I(\epsilon_t \leq x) I(\eta_{t-|k|} \leq y) \\ \quad - (n-|k|)^{-2} \sum_{t=|k|+1}^{n} I(\epsilon_t \leq x) \sum_{t=|k|+1}^{n} I(\eta_{t-|k|} \leq y), & k < 0. \end{cases} \tag{2.1}$$

Note that since verifying the same lag independence itself is not enough to ensure the independence between two time series, we consider the test statistic based on the empirical distribution of (ϵ_{t-k}, η_t). Here, utilizing the similar method of Skaug and Tjøstheim (1993) and the result of Carlstein (1988), one can show that under H_0 for each k,

$$(n-k)B_{nk} \xrightarrow{d} \mathcal{W}_k := \sum_{i,j=1}^{\infty} \lambda_{ij} W_{ijk}^2, \tag{2.2}$$

where W_{ijk}, $i,j = 1,2,\ldots$ are iid $N(0,1)$ r.v.'s and λ_{ij} are the numbers in Theorem 2 of Skaug and Tjøstheim (1993). It is well known that $\lambda_{ij} = (ij\pi^2)^{-2}$ for continuous type r.v.'s.

In order to test H_0 vs. H_1, however, true errors should be replaced by residuals since they are unknown in practice. For this task, we fit finite order autoregressive models to the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. Let $p = p_n$ and $q = q_n$ be certain sequences of positive integers that diverge to ∞ and satisfy $p^3/n \to 0$ and $q^3/n \to 0$ as $n \to \infty$. Writing

$$X_t - \mu = \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu) + r_{1t} + \epsilon_t$$

and

$$Y_t - \nu = \sum_{j=1}^{q} \theta_j (Y_{t-j} - \nu) + r_{2t} + \eta_t,$$

where

$$r_{1t} = \sum_{j=p+1}^{\infty} \phi_j (X_{t-j} - \mu) \quad \text{and} \quad r_{2t} = \sum_{j=q+1}^{\infty} \theta_j (Y_{t-j} - \nu),$$

we estimate $\boldsymbol{\phi}_n = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\theta}_n = (\theta_1, \ldots, \theta_q)'$ by the least squares estimates $\tilde{\boldsymbol{\phi}}_n = (\tilde{\phi}_1, \ldots, \tilde{\phi}_p)'$ and $\tilde{\boldsymbol{\theta}}_n = (\tilde{\theta}_1, \ldots, \tilde{\theta}_q)'$, i.e.,

$$\tilde{\boldsymbol{\phi}}_n = \left\{ \sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)' \right\}^{-1} \sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(X_t - \hat{\mu}_n) \tag{2.3}$$

and

$$\tilde{\boldsymbol{\theta}}_n = \left\{ \sum_{t=q+1}^{n} (\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)' \right\}^{-1} \sum_{t=q+1}^{n} (\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)(Y_t - \hat{\nu}_n), \tag{2.4}$$

where $\boldsymbol{X}_t = (X_t, \ldots, X_{t-p+1})'$, $\boldsymbol{Y}_t = (Y_t, \ldots, Y_{t-q+1})'$, $\mathbf{1}_r$, $r \geq 1$, denotes the vector in R^r whose components are all equal to one, and $\hat{\mu}_n$ and $\hat{\nu}_n$ are suitable estimates of μ and ν. Then calculating the residuals

$$\tilde{\epsilon}_t = X_t - \hat{\mu}_n - \tilde{\boldsymbol{\phi}}_n'(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)$$

and

$$\tilde{\eta}_t = Y_t - \hat{\nu}_n - \tilde{\boldsymbol{\theta}}_n'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q),$$

we define

$$\tilde{B}_{nk} = \begin{cases} (n - |k|)^{-1} \sum\limits_{i=|k|+1}^{n} \tilde{S}_{nk}^2(\tilde{\epsilon}_{i-k}, \tilde{\eta}_i), & k \geq 0, \\ (n - |k|)^{-1} \sum\limits_{i=|k|+1}^{n} \tilde{S}_{nk}^2(\tilde{\epsilon}_i, \tilde{\eta}_{i-|k|}), & k < 0, \end{cases}$$

where $\tilde{S}_{nk}(x, y)$ is defined in the same way as $S_{nk}(x, y)$ in (2.1) with ϵ_t and η_t replaced by $\tilde{\epsilon}_t$ and $\tilde{\eta}_t$. Throughout this paper, we assume that

(C1) $n^{-1}(p^5 + q^5)(\log n)^2 \to 0$ and $n^2(p\rho^p + q\rho^q) \to 0$ for all $\rho \in (0, 1)$ as $n \to \infty$;

(C2) $\sup_z |\frac{\partial F_i(z)}{\partial z}| < \infty$, and $\sup_z |\frac{\partial^2 F_i(z)}{\partial z^2}| < \infty$, $i = 1, 2$;

(C3) $n^{1/2}(\hat{\mu}_n - \mu) = O_p(1)$ and $n^{1/2}(\hat{\nu}_n - \nu) = O_p(1)$.

The first condition in (C1) implies that the rate of p and q is not so fast; otherwise, we are in a situation that there are too many parameters to be estimated, while the second condition requires those to be large enough for a good approximation. A typical example of p and q is $p = q = [C(\log n)^2]$ for some $C > 0$. Then it can be shown that under H_0, for each nonnegative integer k,

$$\tilde{B}_{nk} - B_{nk} = o_P(n^{-1}), \tag{2.5}$$

proof of which result is provided in Section 4.

However, with $\tilde{B}_{nk}$, it is just tested that (ϵ_{t-k}, η_t) are dependent for given k. In order to testing the independence for k's, one should consider the test statistic based on more than one $\tilde{B}_{nk}$'s. If the true errors were known, typically one could consider the summation type test statistic $G_{nK} := n \sum_{k=-K}^{K} B_{nk}$ as Skaug and Tjøstheim (1993) examined this for testing the serial independence of random observation. In fact, similarly to Serfling ((1980), pp. 194–199), it can be shown that under H_0,

$$G_{nK} \xrightarrow{d} \sum_{i,j=1}^{\infty} \lambda_{ij} C_{ij}(2K + 1), \tag{2.6}$$

where $C_{ij}(K)$, $i,j = 1,2,\ldots$ are independent chi-square r.v.'s with K degrees of freedom. Combining (2.6) and (2.5), we obtain the following.

THEOREM 2.1. *Assume that* (C1)–(C3) *hold, and let* $\tilde{G}_{nK} = n\sum_{k=-K}^{K}\tilde{B}_{nk}$, *where* K *is a nonnegative integer. Then under* H_0,

$$\tilde{G}_{nK} \xrightarrow{d} \mathcal{G}_K := \sum_{i,j=1}^{\infty}(ij\pi^2)^{-2}C_{ij}(2K+1) \qquad as \quad n \to \infty,$$

where $C_{ij}(K)$ *are independent chi-square r.v.'s with* K *degrees of freedom.*

Remark. Under the assumption of Theorem 2.1, we have that $\tilde{V}_{nK} := \sum_{k=-K}^{K}(n-k)\tilde{B}_{nk}$ has the same limiting distribution as $\tilde{G}_{nK}$.

Although the idea of using the summation type statistic sounds quite natural, Hong (1998) pointed out that it suffers from severe size distortions as K increases and suggested a weight sum of the CV test statistic for testing the serial independence. Similarly, in our set-up, we can also employ weighted sum of $\tilde{B}_{nk}$'s:

$$\tilde{H}_{nK_n} := \left\{2\tilde{V}_0\sum_{k=2-n}^{n-2}g^4(k/K_n)\right\}^{-1/2}\sum_{k=1-n}^{n-1}g^2(k/K_n)\{(n-k)\tilde{B}_{nk} - \tilde{M}_{0n}\},$$

where g is a kernel function, $\{K_n\}$ is a sequence of positive real numbers,

$$\tilde{M}_{0n} = \frac{1}{n}\sum_{t=1}^{n}\tilde{F}_{1n}(\tilde{\epsilon}_t)\{1-\tilde{F}_{1n}(\tilde{\epsilon}_t)\}\frac{1}{n}\sum_{t=1}^{n}\tilde{F}_{2n}(\tilde{\eta}_t)\{1-\tilde{F}_{2n}(\tilde{\eta}_s)\}$$

and

$$\tilde{V}_{0n} = \frac{1}{n^2}\sum_{s,t=1}^{n}\{\tilde{F}_{1n}(\tilde{\epsilon}_s \wedge \tilde{\epsilon}_t) - \tilde{F}_{1n}(\tilde{\epsilon}_s)\tilde{F}_{1n}(\tilde{\epsilon}_t)\}^2$$
$$\times \frac{1}{n^2}\sum_{s,t=1}^{n}\{\tilde{F}_{2n}(\tilde{\eta}_s \wedge \tilde{\eta}_t) - \tilde{F}_{2n}(\tilde{\eta}_s)\tilde{F}_{2n}(\tilde{\eta}_t)\}^2$$

with $\tilde{F}_{1n}(u) = \frac{1}{n}\sum_{t=1}^{n}I(\tilde{\epsilon}_t \le u)$ and $\tilde{F}_{2n}(u) = \frac{1}{n}\sum_{t=1}^{n}I(\tilde{\eta}_t \le u)$. Then we have the following result, of which proof is provided in Section 4.

THEOREM 2.2. *Assume that* (C1)–(C3) *hold and the function* $g : R \to [-1,1]$ *is symmetric, continuous at* 0 *and all except a finite number of points, with* $g(0) = 1$, $\int_{-\infty}^{\infty}g(z)^2dz < \infty$ *and* $|g(z)| \le C|z|^{-b}$ *as* $z \to \infty$ *for some* $b > \frac{1}{2}$ *and* $0 < C < \infty$, *and* $K_n = cn^{\nu}$ *for some* $0 < \nu < 1$ *and* $0 < c < \infty$. *Then under* H_0, *we have*

$$\tilde{H}_{nK_n} \xrightarrow{d} N(0,1).$$

Notice that the truncated ($g(z) = I(|z| < 1)$), Bartlett ($g(z) = (1-|z|)I(|z| < 1)$) and Daniell ($g(z) = \sin(\pi z)/(\pi z)$) kernels satisfy the above conditions (cf. Hong

(1998) and Priestley (1981)). As it will be seen in our simulation study, $\tilde{H}_{nK_n}$ cures the drawback of $\tilde{G}_{nK}$.

Now, we propose another test statistic, which is obtained as the maximum of $\tilde{B}_{nk}$'s, namely,

$$\tilde{M}_{nK} = n \max_{|k| \le K} \tilde{B}_{nk}.$$

It will be seen in our simulation study that the maximum type statistic is the most stable among the test statistics considered here. The proof of the following theorem is given in Section 4.

THEOREM 2.3. *Assume that* (C1)–(C3) *hold. Then under* H_0,

$$\tilde{M}_{nK} \xrightarrow{d} \mathcal{M}_K := \max_{|k| \le K} \mathcal{W}_k \quad as \quad n \to \infty,$$

where $\mathcal{W}_k$ *is as defined in* (2.2) *with* $\lambda_{ij} = (ij\pi^2)^{-2}$.

Remark. The results of Theorems 2.1–2.3 are applicable to the serial independent test for the autoregressive models. In practice, some may prefer to use ARMA models instead of infinite order autoregressive models despite the residuals in ARMA models are harder to deal with in deriving the limiting distribution of the residual empirical process (cf. Bai (1994)). Since this is not our primary concern, we do not intend to pursue it here. However, in light of the result of Bai (1994), one can easily guess that the same results hold in ARMA models.

One may argue that using the autoregressive model approach may cause a model bias, and the observations themselves must be used in constructing test statistics. But it is well known that the empirical process for dependent observations has a limiting distribution depending upon their correlation structure. Therefore, the limiting distributions of the test statistics as in Theorems 2.1–2.3 will depend upon the correlation structure as well. This will certainly cause a serious trouble when performing the test in real situations. It may be worthwhile from a theoretical viewpoint to investigate the limiting distribution of the test statistics, for instance, for strictly stationary strong mixing processes. However, this issue is somewhat beyond the scope of the present paper. So, we leave it as a task of future study.

Recently, there exists a tendency to release the iid assumption on the true errors in time series models. To our knowledge, there are few literatures pertaining to the residual empirical process based on a sequence of martingale differences. Conventionally, the functional central theorem for martingales requires higher moment condition. Therefore, one may expect that Theorems 2.1–2.3 will hold for strictly stationary martingale differences under fairly mild conditions. But, a caution should be taken since serial dependence in high order moments, for instance, induced by volatility clustering, may affect the asymptotic variance (and so the form of the tests statistics) of the tests in Theorems 2.1 and 2.3. This, however, is not necessarily true for the test in Theorem 2.2.

In the next section, we will study the empirical sizes and powers of the test statistic based on our asymptotic results.

3. Simulation results and discussion

In this section we evaluate the performance of the test statistics introduced in Section 2 through a simulation study. First, we calculate the asymptotic critical values of the test statistics under H_0. The figures in Table 1 denote the numbers $g_{K,\alpha}$ and $m_{K,\alpha}$, such that for $\alpha = 0.01, 0.05, 0.1$, and $K = 5, 10, 15$, $P(\mathcal{G}_K > g_{K,\alpha}) = \alpha$, and $P(\mathcal{M}_K > m_{K,\alpha}) = \alpha$, respectively.

In calculating $g_{K,\alpha}$ and $m_{K,\alpha}$, we followed the same method as Skaug and Tjøstheim (1993): we calculated the empirical quantiles for each α and K from the 10000 numbers generated from the truncated r.v.'s. $\sum_{ij=1}^{200}(ij\pi^2)^{-2}C_{ij}(2K+1)$ and $\max_{|k|\le K}\sum_{ij=1}^{200}(ij\pi^2)^{-2}W_{ijk}^2$, and repeated this procedure 1000 times.

Here, as in Hong (1998), we also examine the performance of the leave-one-out test statistics. Towards this end, we define

$$\tilde{B}^*_{nk} = \begin{cases} (n-k-1)^{-1}\sum\limits_{i=k+1}^{n}\tilde{S}^{2*}_{nk}(\tilde{\epsilon}_{i-k},\tilde{\eta}_i), & k \ge 0,\\ (n-|k|-1)^{-1}\sum\limits_{i=|k|+1}^{n}\tilde{S}^{2*}_{nk}(\tilde{\epsilon}_i,\tilde{\eta}_{i-|k|}), & k < 0, \end{cases}$$

where

$$\tilde{S}^*_{nk}(\tilde{\epsilon}_{i-k},\tilde{\eta}_i) = \begin{cases} (n-k-1)^{-1}\sum\limits_{t=k+1,t\ne i}^{n} I(\tilde{\epsilon}_{t-k}\le\tilde{\epsilon}_{i-k})I(\tilde{\eta}_t\le\tilde{\eta}_i) \\ \quad -(n-k-1)^{-2}\sum\limits_{t=k+1,t\ne i}^{n} I(\tilde{\epsilon}_{t-k}\le\tilde{\epsilon}_{i-k}) \\ \quad \times\sum\limits_{t=k+1,t\ne i}^{n} I(\tilde{\eta}_t\le\tilde{\eta}_i), & k\ge 0,\\ (n-|k|-1)^{-1}\sum\limits_{t=|k|+1,t\ne i}^{n} I(\tilde{\epsilon}_t\le\tilde{\epsilon}_i)I(\tilde{\eta}_{t-|k|}\le\tilde{\eta}_{i-|k|}) \\ \quad -(n-|k|-1)^{-2}\sum\limits_{t=|k|+1,t\ne i}^{n} I(\tilde{\epsilon}_t\le\tilde{\epsilon}_i) \\ \quad \times\sum\limits_{t=|k|+1,t\ne i}^{n} I(\tilde{\eta}_{t-|k|}\le\tilde{\eta}_{i-|k|}), & k<0. \end{cases}$$

We denote by $\tilde{G}^*_{nK}$, $\tilde{V}^*_{nK}$, $\tilde{H}^*_{nK_n}$ and $\tilde{M}^*_{nK}$ be the leave-one-out test statistics corresponding to $\tilde{G}_{nK}$, $\tilde{V}_{nK}$, $\tilde{H}_{nK_n}$ and $\tilde{M}_{nK}$, respectively. In this simulation, for all test statistics considered in Section 2 and their associated leave-one-out test statistics, we calculate the empirical sizes and powers at a nominal level 0.05. In each simulation, 200 initial observations are discarded to remove initialization effects. For calculating the empirical

Table 1. Asymptotic critical values of $\mathcal{G}_K$ and $\mathcal{M}_K$.

		$\mathcal{G}_K$			$\mathcal{M}_K$		
	α	0.1	0.05	0.01	0.1	0.05	0.1
	5	0.3731	0.3982	0.4511	0.0892	0.1018	0.1323
K	1	0.6747	0.7070	0.7736	0.1074	0.1205	0.1531
	15	0.9707	1.0087	1.0873	0.1187	0.1323	0.1669

size and power, sets of 100 and 300 observations are generated from the following models:

$$X_t = 0.5X_{t-1} + \epsilon_t \quad \text{and} \quad Y_t = 0.5Y_{t-1} + \eta_t,$$

and

$$X_t = 0.5X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1} \quad \text{and} \quad Y_t = 0.5Y_{t-1} + \eta_t + 0.5\eta_{t-1}.$$

Under H_0, both ϵ_t and η_t are assumed to be iid standard normal r.v.'s. Meanwhile, under H_1, we assume that

$$\rho := \mathrm{Corr}(X_{t-2}, Y_t) = 0.2 \text{ and } 0.5.$$

Here we use $p = q = [0.1(\log n)^2]$ for the AR approximation.

Table 2. Empirical sizes in AR(1) model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.112	.202	.352	.066	.070	.108
$\tilde{V}_{nK}$	.088	.100	.126	.062	.060	.058
$\tilde{H}^T_{nK_n}$	.104	.122	.134	.056	.066	.062
$\tilde{H}^B_{nK_n}$	.090	.100	.106	.048	.046	.054
$\tilde{H}^D_{nK_n}$	.098	.126	.152	.044	.052	.066
$\tilde{M}_{nK}$	.054	.036	.030	.030	.026	.026
$\tilde{G}^*_{nK}$	.120	.192	.334	.074	.076	.112
$\tilde{V}^*_{nK}$	.086	.102	.130	.060	.058	.060
$\tilde{H}^{*T}_{nK_n}$	.108	.142	.130	.052	.072	.064
$\tilde{H}^{*B}_{nK_n}$	.108	.114	.114	.050	.050	.056
$\tilde{H}^{*D}_{nK_n}$	.108	.120	.142	.052	.052	.060
$\tilde{M}_{nK}$	.062	.024	.036	.036	.032	.020

Table 3. Empirical powers with $\rho = 0.2$ in AR(1) model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.226	.278	.428	.514	.442	.410
$\tilde{V}_{nK}$	.174	.146	.170	.502	.386	.348
$\tilde{H}^T_{nK_n}$	.202	.176	.172	.546	.416	.330
$\tilde{H}^B_{nK_n}$	.336	.298	.266	.700	.664	.628
$\tilde{H}^D_{nK_n}$	.316	.278	.272	.714	.628	.588
$\tilde{M}_{nK}$	.158	.094	.078	.602	.480	.392
$\tilde{G}^*_{nK}$	.176	.234	.360	.456	.388	.378
$\tilde{V}^*_{nK}$	.132	.108	.114	.440	.344	.304
$\tilde{H}^{*T}_{nK_n}$	.152	.118	.138	.506	.382	.302
$\tilde{H}^{*B}_{nK_n}$	.282	.242	.214	.652	.616	.584
$\tilde{H}^{*D}_{nK_n}$	.260	.220	.190	.682	.590	.536
$\tilde{M}_{nK}$	.140	.076	.064	.560	.414	.356

The figures in Tables 2–7 indicate the percentages of rejections of the null hypothesis. Tables 2–7 exhibit the empirical sizes and powers of $\tilde{G}_{nK}$, $\tilde{V}_{nK}$, $\tilde{H}^T_{nK_n}$, $\tilde{H}^B_{nK_n}$, $\tilde{H}^D_{nK_n}$, $\tilde{M}_{nK}$ and their leave-one-out test statistics, where $\tilde{H}^T_{nK_n}$, $\tilde{H}^B_{nK_n}$, $\tilde{H}^D_{nK_n}$ denote the test statistic using the truncated, Bartlett and Daniell kernels, respectively.

From Tables 2 and 5, we can observe that the sizes of all test statistic except $\tilde{M}_{nK}$ are larger than the nominal size and increases as K increases. However, when n is 300, all test statistics except $\tilde{G}_{nK}$ achieved the sizes close to the nominal level.

Tables 3 and 6 and Tables 4 and 7 show the powers of the test statistics under the alternative with $\rho = 0.2$ and $\rho = 0.5$, respectively. We can see that all test statistics produce good powers as either ρ gets close to 1 or n increases. In some cases, such as $n = 100$ and $\rho = 0.2$, $\tilde{M}_{nK}$ produced lower powers than the others, which is, however,

Table 4. Empirical powers with $\rho = 0.5$ in AR(1) model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.848	.824	.856	1	1	.998
$\tilde{V}_{nK}$	.816	.704	.638	1	.998	.996
$\tilde{H}^T_{nK_n}$	.860	.734	.652	1	1	.996
$\tilde{H}^B_{nK_n}$	.976	.962	.934	1	1	1
$\tilde{H}^D_{nK_n}$	.964	.938	.900	1	1	1
$\tilde{M}_{nK}$	.920	.856	.798	1	1	1
$\tilde{G}^*_{nK}$	.786	.770	.798	1	.998	.996
$\tilde{V}^*_{nK}$	.746	.636	.562	1	.998	.996
$\tilde{H}^{*T}_{nK_n}$	.790	.656	.586	1	1	.996
$\tilde{H}^{*B}_{nK_n}$	.962	.932	.890	1	1	1
$\tilde{H}^{*D}_{nK_n}$	.944	.890	.828	1	1	1
$\tilde{M}_{nK}$	.884	.784	.744	1	1	1

Table 5. Empirical sizes in ARMA(1, 1) model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.132	.193	.349	.057	.078	.091
$\tilde{V}_{nK}$	.111	.102	.114	.050	.061	.049
$\tilde{H}^T_{nK_n}$	.126	.119	.120	.082	.070	.064
$\tilde{H}^B_{nK_n}$	.099	.116	.135	.066	.073	.073
$\tilde{H}^D_{nK_n}$	.114	.151	.165	.070	.075	.073
$\tilde{M}_{nK}$	.069	.051	.040	.044	.037	.033
$\tilde{G}^*_{nK}$	.123	.193	.327	.057	.075	.081
$\tilde{V}^*_{nK}$	.099	.097	.101	.053	.056	.046
$\tilde{H}^{*T}_{nK_n}$	.119	.121	.111	.078	.063	.061
$\tilde{H}^{*B}_{nK_n}$	.113	.119	.121	.062	.080	.075
$\tilde{H}^{*D}_{nK_n}$	.118	.142	.139	.060	.075	.075
$\tilde{M}_{nK}$	.066	.053	.046	.041	.037	.029

due to size distortions. In the case that $n = 300$ and $\rho = 0.2$, we can see that $\tilde{M}_{nK}$ produces better powers than $\tilde{G}_{nK}$, $\tilde{V}_{nK}$ and $\tilde{H}^{T}_{nK_n}$, and slightly less powers than $\tilde{H}^{B}_{nK_n}$ and $\tilde{H}^{D}_{nK_n}$.

In fact, in all tests one can observe power losses when K becomes large. In particular, the powers of $\tilde{V}_{nK}$ and $\tilde{M}_{nK}$ are remarkably reduced, but both $\tilde{H}^{B}_{nK_n}$ and $\tilde{H}^{D}_{nK_n}$ produced less power losses. Despite we do not report here, we simulated the powers for different alternatives, for instance, H_1 : Y_t is correlated with X_{t-k}: the correlation is 0.2, where k is taken to be 5, 10 and 15. From this study, we could see that when n is small, e.g., $n = 100$, and the correlation is low, e.g., $\rho = 0.2$, $\tilde{M}_{nK}$ has the best powers among all test statistics when Y_t is correlated with X_{t-K}, but its power is diminished significantly compared to the others when Y_t is correlated with X_{t-l}, $l < K$. From this result, we

Table 6. Empirical powers with $\rho = 0.2$ in ARMA$(1, 1)$ model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.252	.358	.480	.538	.463	.442
$\tilde{V}_{nK}$	.210	.199	.204	.517	.415	.363
$\tilde{H}^{T}_{nK_n}$	.253	.219	.229	.557	.432	.370
$\tilde{H}^{B}_{nK_n}$	.416	.341	.324	.682	.659	.630
$\tilde{H}^{D}_{nK_n}$	.365	.330	.330	.726	.648	.594
$\tilde{M}_{nK}$	.203	.135	.099	.617	.500	.441
$\tilde{G}^{*}_{nK}$	.208	.304	.420	.489	.421	.405
$\tilde{V}^{*}_{nK}$	.172	.160	.174	.474	.373	.325
$\tilde{H}^{*T}_{nK_n}$	.203	.190	.203	.523	.396	.339
$\tilde{H}^{*B}_{nK_n}$	.339	.286	.267	.646	.630	.597
$\tilde{H}^{*D}_{nK_n}$	.291	.274	.246	.689	.618	.546
$\tilde{M}_{nK}$	.160	.105	.083	.582	.465	.380

Table 7. Empirical powers with $\rho = 0.5$ in ARMA$(1, 1)$ model.

n	100			300		
K	5	10	15	5	10	15
$\tilde{G}_{nK}$	.835	.821	.861	1	1	1
$\tilde{V}_{nK}$	.793	.696	.633	1	1	1
$\tilde{H}^{T}_{nK_n}$	.830	.729	.672	.998	.998	.998
$\tilde{H}^{B}_{nK_n}$	.970	.946	.920	1	1	1
$\tilde{H}^{D}_{nK_n}$	.949	.923	.890	1	1	1
$\tilde{M}_{nK}$	.900	.828	.779	1	1	1
$\tilde{G}^{*}_{nK}$	.782	.745	.806	1	1	1
$\tilde{V}^{*}_{nK}$	.752	.623	.569	1	.998	.998
$\tilde{H}^{*T}_{nK_n}$	.790	.653	.610	.998	.998	.998
$\tilde{H}^{*B}_{nK_n}$	.952	.921	883	1	1	1
$\tilde{H}^{*D}_{nK_n}$	.931	.878	.844	1	1	1
$\tilde{M}_{nK}$	.858	.782	.717	1	1	1

could conclude that in small samples, $\tilde{M}_{nK}$ is somewhat sensitive to a choice of K, and might produce poor powers with high chance if a very large K is selected. But in practice, there are no universal ways to choose the best K. Meanwhile, one can notice that using the leave-one-out test statistics alleviates the size distortions in some cases, but it also drops powers. Overall, we did not find a solid proof that the leave-one-out method outperforms the original one in a significant manner.

So far, we have seen the performance of several residual based CV tests. As anticipated, $\tilde{G}_{nK}$ has severe size distortions and the others cure this drawback to a great degree. In particular, we could see that $\tilde{M}_{nK}$ is the most stable among all test statistics considered here. From this simulation study, we figured out that for fairly large samples, the three tests: $\tilde{M}_{nK}$, $\tilde{H}^B_{nK_n}$ and $\tilde{H}^D_{nK_n}$ are all recommendable. However, when the sample size is small, one should keep in mind that the first test has a power loss problem as mentioned earlier depending upon situations, and the other two tests have the size distortion problem. In conclusion, when one prefers a conservative test in small samples, we recommend to use $\tilde{M}_{nK}$. Otherwise, we recommend $\tilde{H}^B_{nK_n}$ and $\tilde{H}^D_{nK_n}$.

4. Proofs

In this section, we provide the proofs of the results presented in Section 2. For brevity, we will consider the nonnegative k case. Before we proceed, we introduce some notation. Note that the least squares estimates of $\boldsymbol{\phi}_n$ and $\boldsymbol{\theta}_n$ in (2.3) and (2.4) can be rewritten as $\tilde{\boldsymbol{\phi}}_n = \boldsymbol{\phi}_n + \boldsymbol{\gamma}_{1n} + \boldsymbol{\delta}_{1n} + \boldsymbol{\zeta}_{1n}$ and $\tilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n + \boldsymbol{\gamma}_{2n} + \boldsymbol{\delta}_{2n} + \boldsymbol{\zeta}_{2n}$, where

$$\boldsymbol{\gamma}_{1n} = \left\{\sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)'\right\}^{-1} \sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p) r_{1t},$$

$$\boldsymbol{\delta}_{1n} = \left\{\sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)'\right\}^{-1} \sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p) \epsilon_t,$$

$$\boldsymbol{\zeta}_{1n} = \left\{\sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)'\right\}^{-1} \sum_{t=p+1}^{n} (\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)(\mu_n - \mu) \times \left(\sum_{j=1}^{p} \phi_j - 1\right),$$

and $\boldsymbol{\gamma}_{2n}$, $\boldsymbol{\delta}_{2n}$ and $\boldsymbol{\zeta}_{2n}$ are similarly defined. Note that if we set $\hat{\boldsymbol{\phi}}_n = \boldsymbol{\phi}_n + \boldsymbol{\delta}_{1n} + \boldsymbol{\zeta}_{1n}$ and $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n + \boldsymbol{\delta}_{2n} + \boldsymbol{\zeta}_{2n}$, $\hat{\boldsymbol{\phi}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are the least square estimates of $\boldsymbol{\phi}_n$ and $\boldsymbol{\theta}_n$ when $r_{jt} = 0$, $j = 1, 2$. Imitating the proof of Lemma 3 of Berk (1974), we can easily show that under (C1),

$$\|\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\|^2 = O_P(n^{-1}p) \quad \text{and} \quad \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\|^2 = O_P(n^{-1}q). \tag{4.1}$$

(cf. Lee and Wei (1999), Lemma 3.3). If we put

$$\hat{\epsilon}_t := X_t - \hat{\mu}_n - \hat{\boldsymbol{\phi}}_n'(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p) \quad \text{and} \quad \hat{\eta}_t := Y_t - \hat{\nu}_n - \hat{\boldsymbol{\theta}}_n'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q),$$

we can write

$$\tilde{\epsilon}_t = \hat{\epsilon}_t - \boldsymbol{\gamma}_{1n}'(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p) \quad \text{and} \quad \tilde{\eta}_t = \hat{\eta}_t - \boldsymbol{\gamma}_{2n}'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q).$$

The following lemma is proven in a similar fashion to Lemma 2.2 of Lee and Wei (1999)(cf. Lemmas 5.1 and 5.2 of Neuhaus (1971)), and the proof is omitted for brevity.

LEMMA 4.1. *Suppose that k is a nonnegative integer, (ϵ_t^*, η_t^*) are random vectors satisfying*

$$\begin{aligned}\alpha_n := \sup_{x,y}\Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\epsilon_{t-k}^* \le x) I(\eta_t^* \le y) - F(x + \epsilon_{t-k} - \epsilon_{t-k}^*, y + \eta_t - \eta_t^*) \\ + F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y)]\Bigg| \\ = o_P(1),\end{aligned}$$

and β_{jn}, $j = 1, 2$, are r.v.'s with $\beta_{jn} = o_P(n^{-1/2})$. Then if (C2) *holds,*

$$\begin{aligned}\sup_{x,y}\Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\epsilon_{t-k}^* \le x + \beta_{1t}) I(\eta_t^* \le y + \beta_{2t}) \\ - F(x + \epsilon_{t-k} - \epsilon_t^* + \beta_{1t}, y + \eta_t - \eta_t^* + \beta_{2t}) \\ + F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y)]\Bigg| \\ = o_P(1).\end{aligned}$$

LEMMA 4.2. *Let k be a nonnegative integer. Suppose that* (C1)–(C3) *hold. In addition, assume that*

(a) $$\begin{aligned}\sup_{x,y}\Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\hat\epsilon_{t-k} \le x) I(\hat\eta_t \le y) - F(x + \epsilon_{t-k} - \hat\epsilon_{t-k}, y + \eta_t - \hat\eta_t) \\ + F(x,y) - I(\epsilon_{t-k} \le x)(\eta_t \le y)]\Bigg| = o_P(1);\end{aligned}$$

(b) $$\sup_{x}\Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\hat\epsilon_{t-k} \le x) - F_1(x + \epsilon_{t-k} - \hat\epsilon_{t-k}) + F_1(x) - I(\epsilon_{t-k} \le x)]\Bigg| = o_P(1);$$

(c) $$\sup_{y}\Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\hat\eta_{t-k} \le y) - F_2(y + \eta_{t-k} - \hat\eta_{t-k}) + F_2(y) - I(\eta_{t-k} \le y)]\Bigg| = o_P(1).$$

Then, under H_0, we have

$$\sup_{x,y} |\tilde S_{nk}(x,y) - \hat S_{nk}(x,y)| = o_P(n^{-1/2}), \tag{4.2}$$

where $\hat S_{nk}(\cdot)$ is the same as $S_{nk}(\cdot)$ in (2.1) with ϵ_t and η_t replaced by $\hat\epsilon_t$ and $\hat\eta_t$.

PROOF. Let

$$\hat{\mathcal{E}}_{nk}(x,y) = (n-k)^{-1/2} \sum_{t=k+1}^{n} [I(\hat{\epsilon}_{t-k} \le x) I(\hat{\eta}_t \le y) - F(x,y)], \tag{4.3}$$

$$\hat{\mathcal{E}}_{1nk}(x) = (n-k)^{-1/2} \sum_{t=k+1}^{n} [I(\hat{\epsilon}_{t-k} \le x) - F_1(x)] \tag{4.4}$$

and

$$\hat{\mathcal{E}}_{2nk}(y) = (n-k)^{-1/2} \sum_{t=k+1}^{n} [I(\hat{\eta}_{t-k} \le y) - F_2(y)]. \tag{4.5}$$

Define $\tilde{\mathcal{E}}_{nk}(x,y)$, $\tilde{\mathcal{E}}_{1nk}(x)$ and $\tilde{\mathcal{E}}_{2nk}(y)$ similarly to $\hat{\mathcal{E}}_{nk}(x,y)$, $\hat{\mathcal{E}}_{1nk}(x)$ and $\hat{\mathcal{E}}_{2nk}(y)$ by replacing $\hat{\epsilon}_t$ and $\hat{\eta}_t$ by $\tilde{\epsilon}_t$ and $\tilde{\eta}_t$. Using this notation, we can write that

$$\begin{aligned} &n^{1/2}\{\tilde{S}_{nk}(x,y) - \hat{S}_{nk}(x,y)\} \\ &\quad = \{n(n-k)^{-1}\}^{1/2}[\{\tilde{\mathcal{E}}_{nk}(x,y) - \hat{\mathcal{E}}_{nk}(x,y)\} - F_2(y)\{\tilde{\mathcal{E}}_{1nk}(x) - \hat{\mathcal{E}}_{1nk}(x)\} \\ &\qquad - F_1(x)\{\tilde{\mathcal{E}}_{2n0}(y) - \hat{\mathcal{E}}_{2n0}(y)\} \\ &\qquad - n^{-1/2}\{\tilde{\mathcal{E}}_{1nk}(x) - \hat{\mathcal{E}}_{1nk}(x)\}\{\tilde{\mathcal{E}}_{2n0}(y) - \hat{\mathcal{E}}_{2n0}(y)\} \\ &\qquad - n^{-1/2}\{\tilde{\mathcal{E}}_{1nk}(x) - \hat{\mathcal{E}}_{1nk}(x)\}\hat{\mathcal{E}}_{2n0}(y) \\ &\qquad - n^{-1/2}\{\tilde{\mathcal{E}}_{2n0}(y) - \hat{\mathcal{E}}_{2n0}(y)\}\hat{\mathcal{E}}_{1nk}(x)]. \end{aligned} \tag{4.6}$$

Since $\tilde{\epsilon}_t = \hat{\epsilon}_t - \boldsymbol{\gamma}'_{1n}(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)$ and $\tilde{\eta}_t = \hat{\eta}_t - \boldsymbol{\gamma}'_{2n}(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)$, the first term in the right hand side of (4.6), viz., $\sup_{x,y} |\tilde{\mathcal{E}}_{nk}(x,y) - \hat{\mathcal{E}}_{nk}(x,y)|$ is bounded by $I_{1nk} + I_{2nk} + I_{3nk}$, where

$$\begin{aligned} I_{1nk} = \sup_{x,y} \Bigg| (n-k)^{-1/2} \sum_{t=k+1}^{n} &[I(\hat{\epsilon}_{t-k} \le x) I(\hat{\eta}_t \le y) \\ &- F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k}, y + \eta_t - \hat{\eta}_t) \\ &+ F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y)] \Bigg|, \end{aligned}$$

$$\begin{aligned} I_{2nk} = \sup_{x,y} \Bigg| (n-k)^{-1/2} \sum_{t=k+1}^{n} &[I(\hat{\epsilon}_{t-k} \le x + \boldsymbol{\gamma}'_{1n}(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p)) I(\hat{\eta}_t \le y \\ &+ \boldsymbol{\gamma}'_{2n}(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) - F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k} \\ &+ \boldsymbol{\gamma}'_{1n}(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), y + \eta_t - \hat{\eta}_t \\ &+ \boldsymbol{\gamma}'_{2n}(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) + F(x,y) \\ &- I(\epsilon_{t-k} \le x) I(\eta_t \le y)] \Bigg|, \end{aligned}$$

$$\begin{aligned} I_{3nk} = \sup_{x,y} \Bigg| (n-k)^{-1/2} \sum_{t=k+1}^{n} &[F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k} + \boldsymbol{\gamma}'_{1n}(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), y + \eta_t - \hat{\eta}_t \\ &+ \boldsymbol{\gamma}'_{2n}(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) - F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k}, y + \eta_t - \hat{\eta}_t)] \Bigg|. \end{aligned}$$

First, note that $I_{1nk} = o_P(1)$ due to (a). Next, using (C1) and the arguments used to prove Lemmas 2.1 and 3.2 of Lee and Wei (1999), we can show that

$$\max_{1\le t\le n} |\gamma_{1n}'(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p)| = o_P(n^{-1/2}) \tag{4.7}$$

and

$$\max_{1\le t\le n} |\gamma_{2n}'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)| = o_P(n^{-1/2}), \tag{4.8}$$

whose detailed proof is omitted for brevity. The arguments (4.7) and (4.8) with Lemma 4.1 and (a) imply $I_{2nk} = o_P(1)$. Finally, I_{3nk} is $o_P(1)$ from (4.7), (4.8), (C2) and Taylor's series expansion. Combining all these results, we have

$$\sup_{x,y} |\tilde{\mathcal{E}}_{nk}(x,y) - \hat{\mathcal{E}}_{nk}(x,y)| = o_P(1). \tag{4.9}$$

Then (4.2) is yielded by (4.6), (4.9) and the fact that:

$$\sup_{x} |\tilde{\mathcal{E}}_{1nk}(x) - \hat{\mathcal{E}}_{1nk}(x)| = o_P(1),$$
$$\sup_{y} |\tilde{\mathcal{E}}_{2nk}(y) - \hat{\mathcal{E}}_{2nk}(y)| = o_P(1),$$

and $\sup_x |\hat{\mathcal{E}}_{ink}(x)| = O_P(1)$ (cf. Lee and Wei (1999)). □

Notice that the arguments in (b) and (c) hold due to Corollary 2.2 of Lee and Wei (1999) under (C1)–(C3). The following lemma is concerned with (a).

LEMMA 4.3. *Let k be a nonnegative integer, and assume that* (C1)–(C3) *hold. Then, under H_0,*

$$\begin{aligned} C_{nk} := \sup_{x,y} \Bigg| n^{-1/2} \sum_{t=k+1}^{n} & [I(\hat{\epsilon}_{t-k} \le x) I(\hat{\eta}_t \le y) - F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k}, y + \eta_t - \hat{\eta}_t) \\ & + F(x,y) - I(\epsilon_{t-k} \le x)(\eta_t \le y)] \Bigg| \\ = o_P(1). \end{aligned}$$

PROOF. Letting

$$\epsilon_t^* = \epsilon_t - (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p), \qquad \eta_t^* = \eta_t - (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q),$$
$$\mu_n^* = (\hat{\mu}_n - \mu)\left(\sum_{j=1}^{p} \phi_j - 1\right), \qquad \nu_n^* = (\hat{\nu}_n - \nu)\left(\sum_{j=1}^{q} \theta_j - 1\right),$$
$$x^* = x - \mu_n^*, \qquad y^* = y - \nu_n^*,$$

we can write $C_{nk} \le I_{1nk} + I_{2nk}$, where

$$I_{1nk} = \sup_{x,y} \Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\epsilon_{t-k}^* \le x^* - r_{1,t-k}) I(\eta_t^* \le y^* - r_{2t}) - F(x^* + \epsilon_{t-k}$$

$$- \epsilon_{t-k}^* - r_{1,t-k}, y^* + \eta_t - \eta_t^* - r_{2t})$$
$$+ F(x^*, y^*) - I(\epsilon_{t-k} \le x^*) I(\eta_t \le y^*)]\Big|,$$

and

$$I_{2nk} = \sup_{x,y} \Big| n^{-1/2} \sum_{t=k+1}^{n} [I(\epsilon_{t-k} \le x) I(\eta_t \le y) - F(x,y)$$
$$- I(\epsilon_{t-k} \le x^*) I(\eta_t \le y^*) + F(x^*, y^*)]\Big|.$$

First, we can see that $I_{2nk} = o_P(1)$ in view of Lemmas 5.1 and 5.2 of Neuhaus (1971). On the other hand, due to (C1), we have that

$$\max_{p+1 \le t \le n} |r_{1t}| = o_P(n^{-1/2}) \quad \text{and} \quad \max_{q+1 \le t \le n} |r_{2t}| = o_P(n^{-1/2})$$

(cf. Lee and Wei (1999), Lemma 3.2). Therefore, in view of Lemma 4.1, it suffices to show that

$$\begin{aligned}\sup_{x,y} |C_{nk}^{\circ}(x,y)| := \sup_{x,y} \Big| n^{-1/2} \sum_{t=k+1}^{n} & [I(\epsilon_{t-k} \le x + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p)) \\ & \times I(\eta_t \le y + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & - F(x + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), \\ & y + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & + F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y)]\Big| = o_P(1).\end{aligned}$$

Let x_i and y_j, $i,j = 1,\ldots,n$ be such that $-\infty = x_0 < \cdots < x_n = \infty$, $-\infty = y_0 < \cdots < y_n = \infty$, $F_1(x_i) = i/n$, and $F_2(y_j) = j/n$, where F_1 and F_2 denote the marginal distribution of ϵ_1 and η_1, respectively. Observe that for any $x \in (x_{nr}, x_{n,r+1}]$ and $y \in (y_{ns}, y_{n,s+1}]$, $C_{nk}^{\circ}(x,y)$ is bounded by $II_{1nk} + II_{2nk} + II_{3nk}$, where

$$\begin{aligned}II_{1nk} = \max_{i=r,r+1,j=s,s+1} \Big| n^{-1/2} \sum_{t=k+1}^{n} & [I(\epsilon_{t-k} \le x_i + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p)) \\ & \times I(\eta_t \le y_j + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & - F(x_i + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), \\ & y_j + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & + F(x_i, y_j) - I(\epsilon_{t-k} \le x_i) I(\eta_t \le y_j)]\Big|,\end{aligned}$$

$$\begin{aligned}II_{2nk} = \max_{i=r,r+1,j=s,s+1} \Big| n^{-1/2} \sum_{t=k+1}^{n} & [F(x_i + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), y_j \\ & + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q))\end{aligned}$$

$$- F(x + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p),$$
$$y + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q))]\Bigg|,$$

$$II_{3nk} = \max_{i=r,r+1,j=s,s+1} \Bigg| n^{-1/2} \sum_{t=k+1}^{n} [I(\epsilon_{t-k} \le x_i) I(\eta_t \le y_j) - F(x_i, y_j) + F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y)] \Bigg|.$$

Using Taylor's series expansion and the proposition in the Appendix of Lee and Wei (1999), we can readily show that $\sup_{x,y} |II_{2nk}| = o_P(1)$. On the other hand, it follows from Lemmas 5.1 and 5.2 of Neuhaus (1971) that $\sup_{x,y} |II_{3nk}| = o_P(1)$. Therefore, it suffices to prove that

$$\begin{aligned} C^*_{nk} := \max_{1 \le r,s \le n} \Bigg| n^{-1/2} \sum_{t=k+1}^{n} & [I(\epsilon_{t-k} \le x_r + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p)) \\ & \times I(\eta_t \le y_s + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & - F(x_r + (\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n)'(\boldsymbol{X}_{t-k-1} - \hat{\mu}_n \mathbf{1}_p), \\ & y_s + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)'(\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q)) \\ & + F(x_r, y_s) - I(\epsilon_{t-k} \le x_r) I(\eta_t \le y_s)] \Bigg| = o_P(1). \end{aligned} \tag{4.10}$$

Let γ be any positive real number. In view of (4.1), we can choose a positive real number K, such that $P(\cup_{i=1}^3 S_i^c) \le \gamma$ for all sufficiently large n, where

$$\begin{aligned} S_1 &= \{\|\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\| \le M(n^{-1}p)^{1/2}, \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\| \le M(n^{-1}q)^{1/2}\}, \\ S_2 &= \left\{ \sum_{t=1}^{n} \|\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p\| \le Mnp^{1/2}, \sum_{t=1}^{n} \|\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q\| \le Mnq^{1/2} \right\}, \\ S_3 &= \left\{ \max_{1 \le t \le n} \|\boldsymbol{X}_{t-1} - \hat{\mu}_n \mathbf{1}_p\| \le M(np)^{1/2}, \max_{1 \le t \le n} \|\boldsymbol{Y}_{t-1} - \hat{\nu}_n \mathbf{1}_q\| \le M(nq)^{1/2} \right\}. \end{aligned}$$

Then for $\lambda > 0$,

$$\begin{aligned} P(C^*_{nk} > \lambda) &\le P(C^*_{nk} > \lambda, \cap_{i=1}^3 S_i) + \gamma \\ &\le P\left(\max_{0 \le r,s \le n, z_j \in \mathcal{Z}_j, j=1,2} \Bigg| n^{-1/2} \sum_{t=k+1}^{n} d_{tk}((x_r, y_s), \boldsymbol{z}) \Bigg| > \lambda, S_2 \cap S_3 \right) + \gamma, \end{aligned}$$

where $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2)', \boldsymbol{z}_j \in \mathcal{Z}_j$, $j = 1,2$, $\mathcal{Z}_1 = \{\boldsymbol{z}_1 \in R^p; \|\boldsymbol{z}_1\| \le K\}, \mathcal{Z}_2 = \{\boldsymbol{z}_2 \in R^q; \|\boldsymbol{z}_2\| \le K\}$, and

$$\begin{aligned} d_{tk}((x,y), \boldsymbol{z}) = & \; I(\epsilon_{t-k} \le x + \boldsymbol{z}_1' \boldsymbol{X}^*_{t-k-1}) I(\eta_t \le y + \boldsymbol{z}_2' \boldsymbol{Y}^*_{t-1}) \\ & - F(x + \boldsymbol{z}_1' \boldsymbol{X}^*_{t-k-1}, y + \boldsymbol{z}_2' \boldsymbol{Y}^*_{t-1}) + F(x,y) - I(\epsilon_{t-k} \le x) I(\eta_t \le y) \end{aligned}$$

with $\boldsymbol{X}^*_t = (n^{-1}p)^{1/2}(\boldsymbol{X}_t - \hat{\mu}_n \mathbf{1}_p)$ and $\boldsymbol{Y}^*_t = (n^{-1}q)^{1/2}(\boldsymbol{Y}_t - \hat{\nu}_n \mathbf{1}_q)$.

In order to verify (4.10), it suffices to show that

$$\sup_{0\le r,s\le n, z_j\in\mathcal{Z}_j, j=1,2}\left|n^{-1/2}\sum_{t=k+1}^{n} d_{tk}((x_r,y_s),\boldsymbol{z})\right| I(S_2\cap S_3)=o_P(1)$$

since γ can be chosen arbitrarily small.

To this end, we partition the rectangle $[-M,M]^p$ in R^p by subrectangles generated by the vertices $V_p=\{(z_{11j},\ldots,z_{1pj}); j=1,\ldots,n\}$, where $z_{1ij}=-M+2Mj/n$, $i=1,\ldots,p$, $j=0,\ldots,n$. Similarly, we partition $[-M,M]^q$ in R^q by $V_q=\{(z_{21j},\ldots,z_{2qj}); j=1,\ldots,n\}$, where $z_{2ij}=-M+2Mj/n$, $i=1,\ldots,q$, $j=0,\ldots,n$. Let $\mathcal{B}_j$ be the class of all subrectangles C_j such that $C_j\cap\mathcal{Z}_j\ne\emptyset$, and denote it by $\mathcal{B}_j=\{\mathcal{B}_{l_j}; l_j=1,\ldots,k_{jn}\}$, $j=1,2$. Here, m_{1n} and m_{2n} are at most n^p and n^q, respectively. For $\boldsymbol{z}_1\in\mathcal{B}_{l_1}$, we define $v^+_{tk,l_1}=\sup_{z_1\in\mathcal{B}_{l_1}}\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1}$ and $v^-_{tk,l_1}=\inf_{z_1\in\mathcal{B}_{l_1}}\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1}$. Similarly, for $\boldsymbol{z}_2\in\mathcal{B}_{l_2}$, we define $v^+_{tk,l_2}=\sup_{\boldsymbol{z}_2\in\mathcal{B}_{l_2}}\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-k-1}$ and $v^-_{tk,l_2}=\inf_{z_2\in\mathcal{B}_{l_2}}\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-k-1}$. Note that $(v^+_{tk,l_1},v^+_{t,0,l_2})$ and $(v^-_{tk,l_1},v^-_{t,0,l_2})$ are $\mathcal{F}_{t-k-1}$-measurable, where $\mathcal{F}_{t-k}=\sigma((\epsilon_{s-k},\eta_s); s\le t)$.

Now for $\boldsymbol{z}_j\in\mathcal{B}_{l_j}$, we have that $L_{tk}((x,y),\boldsymbol{z})\le d_{tk}((x,y),\boldsymbol{z})\le U_{tk}((x,y),\boldsymbol{z})$, where

$$\begin{aligned}U_{tk}((x,y),\boldsymbol{z}) &= I(\epsilon_{t-k}\le x+v^+_{tk,l_1})I(\eta_t\le y+v^+_{t,0,l_2})-F(x+v^+_{tk,l_1},y+v^+_{t-1,l_2})\\ &\quad+F(x,y)-I(\epsilon_{t-k}\le x)I(\eta_t\le y)\\ &\quad+F(x+v^+_{tk,l_1},y+v^+_{t,0,l_2})-F(x+\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1},y+\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-1}),\end{aligned}$$

and $L_{tk}((x,y),\boldsymbol{z})$ is the same as $U_{tk}((x,y),\boldsymbol{z})$ with v^+_{tk,l_j} replaced by v^-_{tk,l_j}. On $S_2\cap S_3$, we have that

$$\begin{aligned}&|F(x+v^+_{tk,l_1},y+v^+_{t,0,l_2})-F(x+\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1},y+\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-1})|\\ &\quad\le\sup_{x,y}\left|\frac{\partial F(x,y)}{\partial x}\right||v^+_{tk,l_1}-\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1}|+\sup_{x,y}\left|\frac{\partial F(x,y)}{\partial y}\right||v^+_{t,0,l_2}-\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-1}|\\ &\quad=O(p^{1/2}n^{-1})+O(q^{1/2}n^{-1})=o(n^{-1/2}),\end{aligned}$$

and similarly,

$$|F(x+v^-_{tk,l_1},y+v^-_{t,0,l_2})-F(x+\boldsymbol{z}_1'\boldsymbol{X}^*_{t-k-1},y+\boldsymbol{z}_2'\boldsymbol{Y}^*_{t-1})|=o(n^{-1/2}).$$

Therefore, we can write that

$$\sup_{0\le r,s\le n,z_j\in\mathcal{Z}_j,j=1,2}\left|n^{-1/2}\sum_{t=k+1}^{n}d_{tk}((x_r,y_s),\boldsymbol{z})\right|I(S_2\cap S_3)\le III_{1nk}+III_{2nk}+o(1),$$

where

$$\begin{aligned}III_{1nk}&=\max_{1\le l_j\le m_{jn},j=1,2}\max_{0\le r,s\le n}\left|n^{-1/2}\sum_{t=k+1}^{n}e_{tk}((x_r,y_s),(v^+_{tk,l_1},v^+_{t,0,l_2}))\right|I(S_2\cap S_3),\\ III_{2nk}&=\max_{1\le l_j\le m_{jn},j=1,2}\max_{0\le r,s\le n}\left|n^{-1/2}\sum_{t=k+1}^{n}e_{tk}((x_r,y_s),(v^-_{tk,l_1},v^-_{t,0,l_2}))\right|I(S_2\cap S_3),\end{aligned}$$

and

$$e_{tk}((x_1,y_1),(x_2,y_2)) = I(\epsilon_{t-k} \le x_1 + x_2)I(\eta_t \le y_1 + y_2) - F(x_1 + x_2, y_1 + y_2) + F(x_1,y_1) - I(\epsilon_{t-k} \le x_1)I(\eta_t \le y_1).$$

Here, we only prove that $III_{1nk} = o_P(1)$ since the negligibility of III_{2nk} is similarly proven.

Define

$$I\tilde{I}I_{1nk} = \max_{1\le l_j \le m_{jn}, j=1,2} \max_{0\le r,s\le n} \left| n^{-1/2} \sum_{t=k+1}^{n} \tilde{e}_{tk} \right|,$$

where

$$\tilde{e}_{tk} := \tilde{e}_{tk}((x_r,y_s),(v^+_{tk,l_1}, v^+_{t,0,l_2})) = e_{tk}((x_r,y_s),(v^+_{tk,l_1}, v^+_{t,0,l_2})) \times I\left(\sum_{i=p+k+1}^{t} \|\boldsymbol{X}_{i-k-1} - \hat{\mu}_n \mathbf{1}_p\| \le Mnp^{1/2}, \sum_{i=q+1}^{t} \|\boldsymbol{Y}_{i-1} - \hat{\nu}_n \mathbf{1}_q\| \le Mnq^{1/2} \right).$$

Note that $\{\tilde{e}_{tk}, \mathcal{F}_{t-k}\}$ forms a sequence of martingale differences with $|\tilde{e}_{tk}| \le 1$ a.s. for all t, and

$$\sum_{t=k+1}^{n} E(\tilde{e}^2_{tk} \mid \mathcal{F}_{t-k-1}) \le \sup_{x,y} \left| \frac{\partial F(x,y)}{\partial x} \right| p^{3/2} M^2 n^{1/2} + \sup_{x,y} \left| \frac{\partial F(x,y)}{\partial y} \right| q^{3/2} K^2 n^{1/2} \le B(p^{3/2} + q^{3/2}) n^{1/2}, \qquad B > 0.$$

Then, using Bernstein's inequality for martingales (cf. Shorack and Wellner (1986), p. 855) we have that for any $\lambda > 0$,

$$P(I\tilde{I}I_{1nk} > \lambda) \le Dn^{(p+q+2)} \exp\left(-\frac{n\lambda^2/2}{B(p^{3/2}+q^{3/2})n^{1/2} + n^{1/2}\lambda/3} \right), \qquad D > 0.$$

Since

$$P(\tilde{e}_{tk} \ne e_{tk}((x_r,y_s),(v^+_{tk,l_1}, v^+_{t,0,l_2})) \text{ for some } t \le n \text{ on } S_2 \cap S_3) = 0,$$

we have $P(III_{1nk} > \lambda) = P(I\tilde{I}I_{1nk} > \lambda) = o(1)$. This completes the proof. □

PROOF OF (2.5). Write that $n\tilde{B}_{nk} = nB_{nk} + I_{1nk} + I_{2nk}$, where

$$I_{1nk} = n \int \{\tilde{S}^2_{nk}(x,y) - S^2_{nk}(x,y)\} d\tilde{F}_{nk}(x,y),$$
$$I_{2nk} = n \int S^2_{nk}(x,y) d\{\tilde{F}_{nk}(x,y) - F_{nk}(x,y)\},$$

$F_{nk}(x,y) = (n-k)^{-1} \sum_{t=k+1}^{n} I(\epsilon_{t-k} \le x) I(\eta_t \le y)$, and $\tilde{F}_{nk}(x,y)$ is the same as $F_{nk}(x,y)$ with ϵ_t and η_t replaced by $\tilde{\epsilon}_t$ and $\tilde{\eta}_t$. Since

$$I_{1nk} \le n \sup_{x,y} |\tilde{S}^2_{nk}(x,y) - S^2_{nk}(x,y)| \le 2n^{1/2} \sup_{x,y} |\tilde{S}_{nk}(x,y) - S_{nk}(x,y)|,$$

it suffices to show that

$$\sup_{x,y} |\tilde{S}_{nk}(x,y) - S_{nk}(x,y)| = o_P(n^{-1/2}), \tag{4.11}$$

and

$$I_{2nk} = o_P(1). \tag{4.12}$$

First, we prove (4.11). Since $\sup_{x,y} |\tilde{S}_{nk}(x,y) - \hat{S}_{nk}(x,y)| = o_P(n^{-1/2})$ by Lemma 4.2, it suffices to show that

$$\sup_{x,y} |\hat{S}_{nk}(x,y) - S_{nk}(x,y)| = o_P(n^{-1/2}). \tag{4.13}$$

Write that

$$\begin{aligned} n^{1/2}\{\hat{S}_{nk}(x,y) - S_{nk}(x,y)\} \\ = \{n(n-k)^{-1}\}^{1/2}[\{\hat{\mathcal{E}}_{nk}(x,y) - \mathcal{E}_{nk}(x,y)\} - F_2(y)\{\hat{\mathcal{E}}_{1nk}(x) - \mathcal{E}_{1nk}(x)\} \\ - F_1(x)\{\hat{\mathcal{E}}_{2n0}(y) - \mathcal{E}_{2n0}(y)\} \\ - n^{-1/2}\{\hat{\mathcal{E}}_{1nk}(x) - \mathcal{E}_{1nk}(x)\}\{\hat{\mathcal{E}}_{2n0}(y) - \mathcal{E}_{2n0}(y)\} \\ - n^{-1/2}\{\hat{\mathcal{E}}_{1nk}(x) - \mathcal{E}_{1nk}(x)\}\mathcal{E}_{2n0}(y) \\ - n^{-1/2}\{\hat{\mathcal{E}}_{2n0}(y) - \mathcal{E}_{2n0}(y)\}\mathcal{E}_{1nk}(x)], \end{aligned} \tag{4.14}$$

where $\mathcal{E}_{nk}(x,y)$, $\mathcal{E}_{1nk}(x)$ and $\mathcal{E}_{2nk}(y)$ are the same as $\hat{\mathcal{E}}_{nk}(x,y)$, $\hat{\mathcal{E}}_{1nk}(x)$ and $\hat{\mathcal{E}}_{2nk}(y)$ in (4.3)–(4.5) with $\hat{\epsilon}_t$ and $\hat{\eta}_t$ replaced by ϵ_t and η_t. Here, split $\hat{\mathcal{E}}_{nk}(x,y) - \mathcal{E}_{nk}(x,y)$ into $II_{1nk}(x,y) + II_{2nk}(x,y)$, where

$$\begin{aligned} II_{1nk}(x,y) = (n-k)^{-1/2} \sum_{t=k+1}^{n} [I(\hat{\epsilon}_{t-k} \le x)I(\hat{\eta}_t \le y) - F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k}, y + \eta_t - \hat{\eta}_t) \\ + F(x,y) - I(\epsilon_{t-k} \le x)I(\eta_t \le y)] \end{aligned}$$

and

$$II_{2nk}(x,y) = (n-k)^{-1/2} \sum_{t=k+1}^{n} [F(x + \epsilon_{t-k} - \hat{\epsilon}_{t-k}, y + \eta_t - \hat{\eta}_t) - F(x,y)].$$

First, note that $\sup_{x,y} |II_{1nk}(x,y)| = o_P(1)$ due to (C1)–(C3) and Lemma 4.3. Next, using a Taylor's series expansion we have that

$$II_{2nk}(x,y) = -\frac{\partial F(x,y)}{\partial x} n^{1/2}\mu_n^* - \frac{\partial F(x,y)}{\partial y} n^{1/2}\nu_n^* + \xi_n(x,y)$$

with $\sup_{x,y} |\xi_n(x,y)| = o_P(1)$. Hence,

$$\hat{\mathcal{E}}_{nk}(x,y) - \mathcal{E}_{nk}(x,y) = -\frac{\partial F(x,y)}{\partial x} n^{1/2}\mu_n^* - \frac{\partial F(x,y)}{\partial y} n^{1/2}\nu_n^* + \xi_n^*(x,y) \tag{4.15}$$

with $\sup_{x,y} |\xi_n^*(x,y)| = o_P(1)$. Similarly, we can show that

$$\hat{\mathcal{E}}_{1nk}(x) - \mathcal{E}_{1nk}(x) = -\frac{\partial F_1(x)}{\partial x} n^{1/2}\mu_n^* + \xi_{1n}(x) \tag{4.16}$$

and

$$\hat{\mathcal{E}}_{2nk}(y) - \mathcal{E}_{2nk}(y) = -\frac{\partial F_2(y)}{\partial y} n^{1/2} \nu_n^* + \xi_{2n}(y) \tag{4.17}$$

with $\sup_x |\xi_{in}(x)| = o_P(1)$, $i = 1, 2$. Then (4.13) follows from (4.14)–(4.17) and the fact that $\sup_x |\mathcal{E}_{jnk}(x)| = O_P(1)$ (cf. Billingsley (1968), pp. 103–108).

Meanwhile, (4.12) is a direct result of the equality

$$I_{2nk} = n^{-1/2} \int \{n^{1/2} S_{nk}(x, y)\}^2 d\{\tilde{\mathcal{E}}_{nk}(x, y) - \mathcal{E}_{nk}(x, y)\},$$

(4.9), (4.15) and the fact that $\sup_{x,y} |n^{1/2} S_{nk}(x, y)| = O_P(1)$. This establishes the argument in (2.5). □

PROOF OF THEOREM 2.2. Define

$$H_{nK_n} := \left\{ 2V_0 \sum_{k=2-n}^{n-2} g^4(k/K_n) \right\}^{-1/2} \sum_{k=1-n}^{n-1} g^2(k/K_n)\{(n-k)B_{nk} - M_0\}$$

with

$$M_0 = \prod_{j=1}^{2} \left[\int F_j(u_j)\{1 - F_j(u_j)\} dF_j(u_j) dF_j(u_j) \right]$$

and

$$V_0 = \prod_{j=1}^{2} \left[\int \{F_j(u_j \wedge u_j') - F_j(u_j)F_j(u_j')\}^2 dF_j(u_j) dF_j(u_j') \right].$$

Put

$$\tilde{H}_{nK_n}^{\circ} := \left\{ 2V_0 \sum_{k=2-n}^{n-2} g^4(k/K_n) \right\}^{-1/2} \sum_{k=1-n}^{n-1} g^2(k/K_n)\{(n-|k|)\tilde{B}_{nk} - \tilde{M}_{0n}\}.$$

Then, provided

$$\tilde{M}_{0n} - M_0 = O_P(n^{-1/2}), \tag{4.18}$$

in view of (2.5) we have

$$\tilde{H}_{nK_n}^{\circ} - H_{nK_n} = o_P(1). \tag{4.19}$$

Note that under the conditions in Theorem 2.2, we have $H_{nK_n} \xrightarrow{d} N(0, 1)$ (cf. Hong (1998)), and then by (4.19),

$$\tilde{H}_{nK_n}^{\circ} \xrightarrow{d} N(0, 1). \tag{4.20}$$

Furthermore, if it holds that

$$\tilde{V}_{0n} - V_0 = o_P(1), \tag{4.21}$$

then from (4.20), (4.21) and Slusky's theorem, we have

$$\tilde{H}_{nK_n} \xrightarrow{d} N(0,1).$$

Hence, to establish the theorem, it suffices to show (4.18) and (4.21).

Put

$$\tilde{A}_n(u_j, u_j') := \tilde{F}_{jn}(u_j \wedge u_j') - \tilde{F}_{jn}(u_j)\tilde{F}_{jn}(u_j')$$

and

$$A(u_j, u_j') := F_j(u_j \wedge u_j') - F_j(u_j)F_j(u_j').$$

Then it is obvious that

$$\begin{aligned} \tilde{M}_{0n} - M_{0n} &= \int \left\{ \prod_{j=1}^{2} \tilde{A}_n(u_j, u_j) - \prod_{j=1}^{2} A_n(u_j, u_j) \right\} \prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) \\ &\quad + \int \prod_{j=1}^{2} A_n(u_j, u_j) \left\{ \prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) - \prod_{j=1}^{2} dF_j(u_j) \right\}. \end{aligned}$$

From (4.10), (4.11), (4.16) and (4.17), we have that

$$\prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) - \prod_{j=1}^{2} dF_j(u_j) = O_P(n^{-1/2}),$$

and

$$\prod_{j=1}^{2} \tilde{A}_n(u_j, u_j) - \prod_{j=1}^{2} A_n(u_j, u_j) = O_P(n^{-1/2}).$$

Therefore, since $\prod_{j=1}^{2} A(u_j, u_j')$ is bounded, we obtain (4.18).

Meanwhile, note that

$$\begin{aligned} \tilde{V}_{0n} - V_0 &= \int \prod_{j=1}^{2} \tilde{A}_n^2(u_j, u_j') - \prod_{j=1}^{2} A^2(u_j, u_j') \prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) d\tilde{F}_{jn}(u_j') \\ &\quad + \int \prod_{j=1}^{2} A^2(u_j, u_j') \left\{ \prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) d\tilde{F}_{jn}(u_j') - \prod_{j=1}^{2} dF_j(u_j) dF_j(u_j') \right\}. \end{aligned}$$

Since $\prod_{j=1}^{2} A^2(u_j, u_j')$ is bounded, it can be shown that

$$\prod_{j=1}^{2} d\tilde{F}_{jn}(u_j) d\tilde{F}_{jn}(u_j') - \prod_{j=1}^{2} dF_j(u_j) dF_j(u_j') = O_P(n^{-1/2}),$$

and

$$\prod_{j=1}^{2} \tilde{A}_n^2(u_j, u_j') - \prod_{j=1}^{2} A^2(u_j, u_j') = O_P(n^{-1/2}).$$

This establishes (4.21). □

Proof of Theorem 2.3. As in Skaug (1993), we can write that for $\boldsymbol{Z}_s = (\boldsymbol{X}_s, \boldsymbol{Y}_s)$,

$$B_{nk} = (n-k)^{-2} \sum_{s,t=k+1}^{n} h_k(\boldsymbol{Z}_s, \boldsymbol{Z}_t) + O_P(n^{-3/2}), \qquad |k| \leq K,$$

where

$$h_k(\boldsymbol{z}_s, \boldsymbol{z}_t) = \int q_k(\boldsymbol{u}, \boldsymbol{z}_s) q_k(\boldsymbol{u}, \boldsymbol{z}_t) dF(\boldsymbol{u}),$$

and

$$q_k(\boldsymbol{u}, \boldsymbol{z}_s) = \begin{cases} \{I(x_{s-k} \leq u_1) - F_1(u_1)\}\{I(y_s \leq u_2) - F_2(u_2)\}, & k \geq 0, \\ \{I(x_s \leq u_1) - F_1(u_1)\}\{I(y_{s-|k|} \leq u_2) - F_2(u_2)\}, & k < 0. \end{cases}$$

Also, we can write that

$$h_k(\boldsymbol{z}_s, \boldsymbol{z}_t) = \sum_{ij=1}^{\infty} \lambda_{ijk} \phi_{ijk}(\boldsymbol{z}_s) \phi_{ijk}(\boldsymbol{z}_t),$$

where $\{\lambda_{ijk}, \phi_{ijk}\}$ are an orthogonal set of eigenvalues and eigenfunctions of h_k (cf. Serfling (1980) and Dunford and Schwartz (1963)). Now following essentially the same arguments as in Serfling ((1980), pp. 194–199), one can readily show that

$$(nB_{n,-K}, \ldots, nB_{nK}) \xrightarrow{d} (\mathcal{W}_{-K}, \ldots, \mathcal{W}_K), \tag{4.22}$$

of which detailed proof is omitted for brevity since the proof is rather standard. Therefore, we have

$$\max_{|k| \leq K} n\tilde{B}_{nk} \xrightarrow{d} \max_{|k| \leq K} \mathcal{W}_k \quad \text{as} \quad n \to \infty$$

by continuous mapping theorem, (4.22) and (2.5). This completes the proof. □

Acknowledgements

We are grateful to Dr. Skaug and Professor Carlstein for showing us their manuscripts. We wish to thank the two referees for their valuable comments to improve the quality of this paper greatly. This work was supported by Korea Research Foundation Grant 2000-015-DP0052.

References

Bai, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models, *Annals of Statistics*, **22**, 2051–2061.

Berk, K. (1974). Consistent autoregressive spectral estimates, *Annals of Statistiscs*, **2**, 489–502.

Billingsley, P. (1968). *Convergence of Probability Measure*, Wiley, New York.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function, *Annals of Mathematical Statistics*, **32**, 485–498.

Brockwell, P. J. and Davis, R. A. (1990). *Time Series: Theory and Models*, 2nd ed., Springer-Verlag, New York.

Carlstein, E. (1988). Degenerate U-statistic based on non-independent observations, *Calcutta Statistical Association Bulletin*, **37**, 55–65.

Delgado, M. A. (1996). Testing serial independence using the sample distribution function, *Journal of Time Series Analysis*, **17**, 271–286.

Delgado, M. A. and Mora, J. (2000). A nonparametric test for serial independence of regression errors, *Biometrika*, **87**, 228–234.

Dunford, N. and Schwartz. J. T. (1963). *Linear Operators*, Wiley, New York.

Geweke, J. (1981). A comparison of tests of the independence of two covariance-stationary time series, *Journal of the American Statistical Association*, **76**, 363–373.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach, *Journal of the American Statistical Association*, **71**, 378–385.

Hoeffding, W. (1948). A nonparametric test of independence, *Annals of Mathematical Statistics*, **26**, 189–211.

Hong, Y. (1996). Testing for independence between two covariance stationary time series, *Biometrika*, **83**, 615–625.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function, *Journal of the Royal Statistical Society, Series B*, **60**, 429–453.

Lee, S. and Karagrigoriou, A. (2001). An asymptotically optimal selection of the order of a linear processes, *Sankhyā A*, **63**, 93–106.

Lee, S. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series, *Annals of Statistics*, **27**, 237–261.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter, *Annals of Mathematical Statistics*, **42**, 1285–1295.

Pierce, D. A. (1977). Relationships-and the lack thereof-between economic time series, with special reference to money and interest rates, *Journal of the American Statistical Association*, **72**, 11–22.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. **1**, Academic Press, London.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear processes, *Annals of Statisitics*, **8**, 147–164.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Skaug, H. J. (1993). The limit distribution of the Hoeffding statistic for test of serial independence (unpublished).

Skaug, H. J. and Tjøstheim, D. (1993). A nonparametic test of serial; independence based on the empirical distribution function, *Biometrika*, **80**, 591–602.