Les Cahiers du GERAD ISSN: 0711–2440

Rank-Based Extensions of the BDS Test for Serial Dependence

C. Genest, K. Ghoudi,
B. Rémillard

G–2006–23

April 2006

Rank-Based Extensions of the BDS Test for Serial Dependence

Christian Genest
Département de mathématiques et de statistique
Université Laval
Québec (Québec) Canada G1K 7P4
genest@mat.ulaval.ca

Kilani Ghoudi
United Arab Emirates University
P.O. Box 17555
Al Ain, United Arab Emirates
kghoudi@uaeu.ac.ae

Bruno Rémillard
GERAD and
Service de l'enseignement des méthodes quantitatives de gestion
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7
bruno.remillard@hec.ca

April 2006

Les Cahiers du GERAD

G–2006–23

Abstract

This paper proposes new tests of randomness for innovations of a large class of time series models. These tests are based on functionals of empirical processes constructed either from the model residuals or from their associated ranks. The asymptotic behavior of these empirical processes is determined under the null hypothesis of randomness. The limiting distributions are seen to be independent of estimation errors when appropriate regularity conditions hold. Several test statistics are derived from these processes; the classical BDS statistic and a rank-based analogue thereof are included as special cases. Since the limiting distributions of the rank-based test statistics are margin-free, their finite-sample P-values can easily be calculated by simulation. Monte Carlo experiments show that these statistics are quite powerful against several alternatives.

Key Words: Copulas; Empirical processes; BDS statistic; Pseudo-observations; Randomness; Ranks; Time series.

Résumé

Dans cet article, on propose de nouveaux tests d'indépendance pour les erreurs de modèles de séries chronologiques. Les tests sont basés sur des fonctionellles de processus empiriques construits à partir de résidus ou des rangs des résidus. Le comportement asymptotique des processus empiriques est déterminé sous l'hypothèse nulle d'indépendance, et l'on montre que sous certaines conditions, la loi limite ne dépend pas des estimations des paramètres du modèle. Plusieurs statistiques sont déduites de ces processus, incluant la statistique BDS, ainsi que son analogue basé sur des rangs. Comme la loi limite des statistiques de rangs ne dépend pas des marges, les valeurs critiques peuvent être estimées par simulation. Des expériences Monte Carlo sont aussi utilisées pour montrer que ces nouveaux tests sont très puissants par rapport à plusieurs hypothèses alternatives de dépendance.

Acknowledgments: Partial funding in support of this work was provided by the Natural Sciences and Engineering Research Council of Canada, by the Fonds québécois de la recherche sur la nature et les technologies, as well as by the Institut de finance mathématique de Montréal.

1 Introduction

Time series models typically involve error terms called innovations that are assumed to be mutually independent with common distribution function F. An important step in validating such models is to check this so-called "white noise assumption" for the sequence (ε_i) of innovations.

When the parameters of the model are known, the innovations are observable and a wide variety of tools is available for testing the null hypothesis H_0 of randomness. Common tests are based on autocorrelations (Moran 1948; Ljung and Box 1978; Dufour and Roy 1985; Hong 2000), entropy measures (Robinson 1991; Hong and White 2005), rank-based dependence measures (Hallin et al. 1985, 1987; Hallin and Puri 1992; Ferguson et al. 2000), empirical distribution functions (Skaug and Tjøstheim 1993; Delgado 1996; Ghoudi et al. 2001), empirical characteristic functions (Hong 1999; Bilodeau and Lafaye de Micheaux 2005), and empirical copulas (Genest and Rémillard 2004).

In practice, however, the model parameters are usually unknown, so that the ε_i are unobservable. In that case, a test for randomness must be based on "residuals" e_i. The latter are typically computed by plugging in the estimated parameter values in an equation defining the relation between ε_i and the observed data y_i at time i. This equation may also depend on finitely many previous values $y_{i-1}, \ldots, y_{i-p}$ and $\varepsilon_{i-1}, \ldots, \varepsilon_{i-q}$. A major stumbling block associated with such a procedure is that the limiting distribution of a test statistic will then generally depend both on the unknown parameter values and possibly on the (infinite-dimensional) nuisance parameter F. This issue, which is generally ignored in practice, is highlighted, e.g., in the work of Ghoudi and Rémillard (1998, 2004).

One ingenious way around this problem is provided by the so-called BDS statistic of Brock, Dechert, and Scheinkman, which was shown by Brock et al. (1996) to have the same limiting behavior under H_0, whether the model parameters are known or estimated. Inspired by the work of Grassberger and Procaccia (1983) on detecting chaotic behavior, the BDS statistic S_n is based on a comparison of the observed and expected numbers of pairs of vectors

$$w_i = (w_{i1}, \ldots, w_{im}) = (e_i, \ldots, e_{i+m-1}), \quad 1 \leq i \leq n \tag{1}$$

such that

$$\|w_i - w_j\| = \max_{1 \leq k \leq m} |w_{ik} - w_{jk}| \leq \delta$$

for some arbitrary constant $\delta > 0$. This statistic leads to the rejection of H_0 when $|S_n|$ is unduly large, by comparison with its distribution under the null.

More specifically, the BDS statistic is defined as

$$S_n = S_n\left(e_1, \ldots, e_{n+m-1}\right) = \frac{V_m - V^m}{s_n/\sqrt{n}} \tag{2}$$

in terms of

$$V_m = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \mathbf{1}(\|w_i - w_j\| \le \delta) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \prod_{k=1}^{m} \mathbf{1}(|w_{ik} - w_{jk}| \le \delta)$$

and

$$V = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \mathbf{1}(|e_j - e_i| \le \delta).$$

Here, s_n is an estimate of the standard deviation of $\sqrt{n}(V_m - V^m)$. The specific estimate proposed by Brock et al. (1996) does not require any knowledge of F. It is defined by

$$s_n^2 = 4\left(\gamma^m - V^{2m}\right) - 4m^2 V^{2m-2}\left(\gamma - V^2\right) + 8 \sum_{k=1}^{m-1} V^{2k}\left(\gamma^{m-k} - V^{2m-2k}\right),$$

where

$$\gamma = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{1}(|e_j - e_i| \le \delta) \mathbf{1}(|e_j - e_k| \le \delta).$$

It was shown by these authors that for a wide class of time series models, S_n has an asymptotic standard normal distribution under the null hypothesis of randomness, the same that one would obtain if the model parameters were known and the statistic were calculated using the (then observable) ε_i.

Nevertheless, the BDS procedure suffers from three major weaknesses. An obvious one is the arbitrariness in the choice of δ, which may affect both the power and the size of the test. In practice, Brock et al. (1996) recommend the use of $\delta \in [s/2, 3s/2]$, where s is the standard deviation of the pseudo-sample $e_1, \ldots, e_{n+m-1}$. A second limitation is the fact that the test is inconsistent, i.e., the probability of rejection of the null hypothesis does not necessarily approach 1 as $n \to \infty$ even when H_0 is false; indeed, alternatives may be found under which the expected value of the test statistic is equal to zero for at least some choice of δ.

However, the third and most critical difficulty associated with the BDS test is that although the statistic converges to a standard normal distribution under the null hypothesis, this convergence is often so slow that even for sample sizes as large as 1000, one is still far from the limit. This is most inconvenient from a practical point of view, because neither the level nor the power of the test can then be determined with any precision, unless F is known. While in the latter case, an appropriate table of critical values can then be constructed, it is widely acknowledged that the rate of convergence varies considerably from one choice of F to another; see, e.g., Brock et al. (1996).

In this paper, extensions of the BDS statistic are considered which have the following properties:

a) they are free of the arbitrary parameter δ;

b) their limiting distribution continues to be independent of the model parameters;
c) their finite-sample distribution is more tractable than that of S_n and well approximated by Monte Carlo methods.

In particular, note that the speed at which a test statistic converges in law is irrelevant when property c) holds.

A first alternative test statistic considered in Section 2 is a rank-based equivalent of the original BDS statistic. It still depends on δ but its asymptotic distribution, which is totally independent of F, is identified in Proposition 1, and a simple algorithm for simulating its finite-sample distribution is provided. Then, in Section 3, functional extensions of S_n are given in the form of empirical processes, one of which is rank-based, and their asymptotic behavior is studied. Calling on these empirical processes, alternative statistics are proposed in Section 4 and algorithms for computing associated P-values and quantiles are stated.

The finite-sample performance of the proposed statistics is considered in Section 5, where their power is estimated through Monte Carlo simulations for a wide range of serial dependence alternatives, including those studied in Hong and White (2005). This is followed by a discussion of the relative merits of these statistics. A small illustration of the methodology is treated in Section 6, and Section 7 contains examples of time series models that satisfy the conditions under which the asymptotic results are stated. The proofs of all results are relegated to Appendices A and B. As for Appendix C, it details technical conditions under which the central Assumption II is verified for common models.

2 A rank-based version of the BDS statistic

Given residuals $e_1, \dots, e_n$ from a time series model, let

$$\tilde{e}_i = \begin{cases} \text{rank}(e_i)/(n+1) & \text{for } i \in \{1, \dots, n\}; \\ \tilde{e}_{i-n} & \text{for } i \in \{n+1, \dots, n+m-1\}. \end{cases} \tag{3}$$

A natural nonparametric (circular) analogue of the BDS statistic S_n is then given by

$$\tilde{S}_n = S_n\left(\tilde{e}_1, \dots, \tilde{e}_{n+m-1}\right).$$

The asymptotic normality of this statistic (and of all other statistics to be introduced herein) depends critically on the Assumptions I and II stated below. It will be seen in Section 7 that the latter requirement is met by several well-known models including, e.g., linear and non-linear AR(p) as well as standard ARMA(p, q) models.

Assumption I: *F admits a continuous derivative that is square integrable.*

Assumption II: *For all $x = (x_1, \dots, x_m) \in \mathbb{R}^m$, let*

$$K_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(w_i \le x) = \frac{1}{n} \sum_{i=1}^{n} \prod_{k=1}^{m} \mathbf{1}(w_{ik} \le x_k)$$

be the empirical distribution function associated with the w_i defined in (1), and denote its theoretical counterpart by $K(x) = F(x_1) \times \cdots \times F(x_m)$. Write $\mathbb{K}_n = \sqrt{n}\,(K_n - K)$ and let also

$$\alpha_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \prod_{k=1}^{m} \mathbf{1}(\varepsilon_{i+k-1} \le x_k) - K(x) \right\}.$$

Then there exist processes $\alpha, \beta_1, \ldots, \beta_m$ in the Skorohod space $\mathcal{D}([-\infty,\infty]^m)$ of càdlàg processes with the property that $\beta_k(x)$ does not depend on x_k and such that as $n \to \infty$, $\alpha_n \rightsquigarrow \alpha$ and $\mathbb{K}_n \rightsquigarrow \mathbb{K}$ in $\mathcal{D}([-\infty,\infty]^m)$, where

$$\mathbb{K}(x) = \alpha(x) - \sum_{k=1}^{m} f(x_k)\beta_k(x), \quad x \in [-\infty,\infty]^m. \tag{4}$$

Proposition 1 *Suppose that F is symmetric and that Assumptions I and II hold under H_0. Then $\tilde{S}_n \rightsquigarrow \mathcal{N}(0,1)$ as $n \to \infty$.*

The rank-based statistic $\tilde{S}_n$ shares with S_n its dependence on an arbitrary parameter δ but, more importantly, the crucial property that its asymptotic distribution does not depend either on the estimated parameters nor on F. The advantage that $\tilde{S}_n$ has over S_n, however, is that its finite-sample distribution can be easily approximated, even when F is unknown. By resorting to Algorithm 1 below to construct tables for $\tilde{S}_n$, a user may then trust the nominal level of the statistic, whereas this could not be accomplished for S_n, unless F were known. In addition, the actual rate of convergence of $\tilde{S}_n$ to its normal limit is irrelevant.

Algorithm 1 (Critical values for $|\tilde{S}_n|$) *Repeat the following steps for each $\ell \in \{1, \ldots, L\}$ for some suitably large L.*

1. *Generate $U_{1,\ell}, \ldots, U_{n,\ell}$ mutually independent uniform random variates on $(0,1)$.*
2. *Define $\tilde{e}_i = \mathrm{rank}(U_{i,\ell})/(n+1)$ for $i \in \{1,\ldots,n\}$ and $\tilde{e}_i = \tilde{e}_{i-n}$ for $i \in \{n+1,\ldots,n+m-1\}$.*
3. *Compute $\tilde{S}_{n,\ell} = S_n(\tilde{e}_1,\ldots,\tilde{e}_{n+m-1})$.*

The $100\times\alpha\%$ critical value for the statistic $|\tilde{S}_n|$ is then approximated by the corresponding quantile in the set $|\tilde{S}_{n,1}|,\ldots,|\tilde{S}_{n,L}|$. Similarly, the P-value associated with an observed value $\tilde{S}_{n,0}$ can be estimated by

$$\frac{1}{L}\sum_{\ell=1}^{L} \mathbf{1}\left(|\tilde{S}_{n,\ell}| > |\tilde{S}_{n,0}|\right).$$

3 Empirical processes extending S_n and $\tilde{S}_n$

Although the rank-based statistic $\tilde{S}_n$ does not suffer from the slow rate of convergence associated with the original BDS statistic S_n, it still shares with it a dependence on the arbitrary parameter $\delta > 0$. This section describes empirical processes that will be exploited in Section 4 to get rid of this arbitrariness.

3.1 An empirical process extending S_n

Consider the process $B_n(t)$ defined for each $t = (t_1, \ldots, t_m) \in [0, \infty]^m$ by

$$B_n(t) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \prod_{k=1}^{m} \mathbf{1}(|w_{jk} - w_{ik}| \le t_k).$$

For all $s \in [0, \infty]$, let also

$$G_n(s) = B_n(s, \infty, \ldots, \infty) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \mathbf{1}(|e_j - e_i| \le s).$$

A test of randomness based on

$$\mathbb{D}_n(t) = \sqrt{n} \left\{ B_n(t) - \prod_{k=1}^{m} G_n(t_k) \right\},$$

would then represent an extension of S_n, since $V_m = B_n(\delta, \ldots, \delta)$ and $V = G_n(\delta)$, so that $S_n = \mathbb{D}_n(\delta, \ldots, \delta)/s_n$.

The limit of the general process $\mathbb{D}_n$ is characterized in Theorem 1 below, along with that of two related processes, namely

$$\mathbb{B}_n(t) = \sqrt{n} \left\{ B_n(t) - \prod_{k=1}^{m} G(t_k) \right\} \quad \text{and} \quad \mathbb{B}_n^\star(t) = \sqrt{n} \left\{ B_n^\star(t) - 2 \prod_{k=1}^{m} G(t_k) \right\},$$

where

$$B_n^\star(t) = \frac{2}{n} \sum_{i=1}^{n} \prod_{k=1}^{m} \left\{ F(w_{ik} + t_k) - F(w_{ik} - t_k) \right\}, \tag{5}$$

and for all $u \in \mathbb{R}_+$,

$$G(u) = \mathrm{P}(|\varepsilon_2 - \varepsilon_1| \le u) = \int_{-\infty}^{\infty} \{F(x + u) - F(x - u)\} dF(x). \tag{6}$$

In the special case where F is known (but only then), the processes $\mathbb{B}_n$ and $\mathbb{B}_n^\star$ could be used instead of $\mathbb{D}_n$ to construct tests of randomness.

Theorem 1 *Suppose that Assumptions I–II hold under H_0. Then as $n \to \infty$, $(\mathbb{B}_n, \mathbb{B}_n^\star, \mathbb{D}_n) \rightsquigarrow (\mathbb{B}, \mathbb{B}^\star, \mathbb{D})$ in $\mathcal{D}\left([0,\infty]^m\right)^{\otimes 3}$, where $\mathbb{B} = \mathbb{B}^\star$ and $\mathbb{D}$ are continuous centered Gaussian processes with covariance functions $\Gamma_{\mathbb{B}}$ and $\Gamma_{\mathbb{D}}$ defined respectively by*

$$\begin{aligned}
\Gamma_{\mathbb{B}}(s,t) &= 4\left\{\prod_{k=1}^m \gamma(s_k,t_k) - \prod_{k=1}^m G(s_k)G(t_k)\right\} \\
&\quad + 4\sum_{j=2}^m \left\{\prod_{k=1}^{j-1} G(s_k)G(t_{m+1-k})\right\}\left\{\prod_{k=j}^m \gamma(s_k, t_{k+1-j}) - \prod_{k=j}^m G(s_k)G(t_{k+1-j})\right\} \\
&\qquad + 4\sum_{j=2}^m \left\{\prod_{k=1}^{j-1} G(s_{m+1-k})G(t_k)\right\}\left\{\prod_{k=j}^m \gamma(s_{k+1-j}, t_k) - \prod_{k=j}^m G(s_{k+1-j})G(t_k)\right\}
\end{aligned}$$

and

$$\Gamma_{\mathbb{D}}(s,t) = \Gamma_{\mathbb{B}}(s,t) - 4\sum_{j=1}^m \sum_{k=1}^m \{\gamma(s_j,t_k) - G(s_j)G(t_k)\}\left\{\prod_{\ell\neq j} G(s_\ell)\right\}\left\{\prod_{\ell\neq k} G(t_\ell)\right\}.$$

As with the classical BDS statistic, it can be seen from Theorem 1 that the asymptotic covariances of the processes $\mathbb{B}_n$, $\mathbb{B}_n^\star$ and $\mathbb{D}_n$ do not depend on the model parameters or their estimates. However, they do depend on F through G and

$$\gamma(u,v) = \mathrm{P}(|\varepsilon_2 - \varepsilon_1| \le u, |\varepsilon_3 - \varepsilon_1| \le v), \quad u,v \in [0,\infty).$$

Nevertheless, consistent estimators of $G(u)$ and $\gamma(u,v)$ are respectively given by $G_n(u)$ and by

$$\gamma_n(u,v) = \frac{1}{n^3}\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \mathbf{1}(|e_i - e_k| \le u)\mathbf{1}(|e_j - e_k| \le v).$$

Remark 1 *When F is known, it would be tempting to work with the pseudo-observations $F(e_i)$. However, it turns out that this is not a good idea, because unless F is symmetric, the limiting distribution of the processes $\mathbb{B}_n$ and $\mathbb{B}_n^\star$ could depend on the estimated parameters. See the proof of Corollary 2 for more details.*

3.2 An empirical process extending $\tilde{S}_n$

Parallel results to those of Section 3.1 are presented here for the case where the e_i are replaced by the $\tilde{e}_i$ from (3). First introduce

$$\tilde{w}_i = (\tilde{w}_{i1}, \ldots, \tilde{w}_{im}) = (\tilde{e}_i, \ldots, \tilde{e}_{i+m-1}), \quad 1 \le i \le n.$$

For each $s \in [0,\infty]$ and $t = (t_1,\ldots,t_m) \in [0,\infty]^m$, let

$$\tilde{B}_n(t) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} \prod_{k=1}^{m} \mathbf{1}(|\tilde{w}_{jk} - \tilde{w}_{ik}| \le t_k),$$

$$\tilde{G}_n(s) = \tilde{B}_n(s,\infty,\ldots,\infty) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} \mathbf{1}(|\tilde{e}_j - \tilde{e}_i| \le s),$$

and

$$\tilde{B}^\star_n(t) = \frac{2}{n} \sum_{i=1}^{n} \prod_{k=1}^{m} \left\{ \tilde{F}(\tilde{w}_{ik} + t_k) - \tilde{F}(\tilde{w}_{ik} - t_k) \right\}. \tag{7}$$

Here, $\tilde{F}(s) = 0 \vee (s \wedge 1)$ is the distribution function of a uniform random variable on the interval $(0,1)$.

Next, the analogues of $\mathbb{D}_n$, $\mathbb{B}_n$ and $\mathbb{B}^\star_n$ are defined respectively by

$$\tilde{\mathbb{D}}_n(t) = \sqrt{n} \left\{ \tilde{B}_n(t) - \prod_{k=1}^{m} \tilde{G}_n(t_k) \right\},$$

$$\tilde{\mathbb{B}}_n(t) = \sqrt{n} \left\{ \tilde{B}_n(t) - \prod_{k=1}^{m} \tilde{G}(t_k) \right\} \quad \text{and} \quad \tilde{\mathbb{B}}^\star_n(t) = \sqrt{n} \left\{ \tilde{B}^\star_n(t) - 2\prod_{k=1}^{m} \tilde{G}(t_k) \right\},$$

where

$$\tilde{G}(u) = \int_{-\infty}^{\infty} \{\tilde{F}(x+u) - \tilde{F}(x-u) d\tilde{F}(x) = 0 \vee \{(2u-u^2) \wedge 1\}.$$

In practice, of course, there is no incentive to use $\tilde{\mathbb{D}}_n$ over $\tilde{\mathbb{B}}_n$, since $\tilde{G}_n$ is deterministic and

$$\sup_{t\in\mathbb{R}_+} \left|\tilde{G}_n(t) - \tilde{G}(t)\right| = O\left(\frac{1}{n}\right) \quad \Rightarrow \quad \sup_{t\in\mathbb{R}_+} \left|\tilde{\mathbb{D}}_n(t) - \tilde{\mathbb{B}}(t)\right| = O\left(\frac{1}{n}\right).$$

Theorem 2 *Suppose that Assumptions I–II hold under H_0. Then as $n \to \infty$, $(\tilde{\mathbb{B}}_n, \tilde{\mathbb{B}}^\star_n, \tilde{\mathbb{D}}_n) \rightsquigarrow (\tilde{\mathbb{B}}, \tilde{\mathbb{B}}, \tilde{\mathbb{B}})$ in $\mathcal{D}([0,1]^m)$, where $\tilde{\mathbb{B}}$ is a continuous centered Gaussian process with covariance function $\Gamma_{\tilde{\mathbb{B}}}(s,t)$ given by the same expression as $\Gamma_{\mathbb{D}}(s,t)$, but with G and γ respectively replaced by $\tilde{G}$ and*

$$\tilde{\gamma}(u,v) = \begin{cases} -u^2v - 2uv^2 + 4uv - \dfrac{u^3}{3} & \text{if } u \le \min(v, 1-v); \\[2ex] v - u^2 - uv^2 + \dfrac{v^3 - 1}{3} - v^2 + u + 2uv & \text{if } \max(u, 1-u) \le v; \\[2ex] -v^2u - 2vu^2 + 4uv - \dfrac{v^3}{3} & \text{if } v \le \min(u, 1-u); \\[2ex] u - v^2 - vu^2 + \dfrac{u^3 - 1}{3} - u^2 + v + 2uv & \text{if } \max(v, 1-v) \le u. \end{cases}$$

Note that under H_0 and for any $\delta \in (0,1)$, it follows from Theorem 2 that

$$\tilde{S}_n = \tilde{\mathbb{D}}_n(\delta, \ldots, \delta)/\tilde{s}_n \rightsquigarrow \mathcal{N}(0,1),$$

where $\tilde{s}_n^2 = s_n^2(\tilde{e}_1, \ldots, \tilde{e}_n) \to \Gamma_{\tilde{\mathbb{B}}}(\delta, \ldots, \delta)$ as $n \to \infty$.

4 Statistics based on functional extensions

In the light of Theorems 1 and 2, obvious extensions of statistics S_n and $\tilde{S}_n$ could be based on quadratic forms involving either $\mathbb{D}_n(t)$ or $\tilde{\mathbb{D}}_n(t)$ for finitely many, arbitrarily selected values of $t \in \mathbb{R}_+^m$. Once properly normalized, these quadratic forms would then be asymptotically distributed as chi-square random variables. Although sophisticated, this approach would provide no real relief. For, the quadratic form would now depend on several arbitrary choices of $t \in \mathbb{R}_+^m$ rather than on $t = (\delta, \ldots, \delta)$. In addition, the issue related to the rate of convergence would remain for the test statistic based on $\mathbb{D}_n$ or $\tilde{\mathbb{D}}_n$.

One obvious way around the arbitrariness of quadratic forms based on a finite number of evaluations of $\mathbb{D}_n$ or $\tilde{\mathbb{D}}_n$ is to resort to continuous functionals of these empirical processes that take into account their value over an infinite number of points. This section considers several statistics of this type, both in the cases where F is known or unknown.

4.1 The case where F is known

In the spirit of freeing the BDS statistic from the arbitrary parameter δ, an option would be to integrate S_n over all possible values of this parameter. This idea leads rather naturally to

$$\int_0^\infty \mathbb{D}_n(s, \ldots, s) dG_n(s).$$

Since F is known, however, it seems more convenient to base a test of randomness either on $\mathbb{B}_n$ or $\mathbb{B}_n^\star$ rather than on $\mathbb{D}_n$ itself. This yields

$$\begin{aligned} I_n &= \int_0^\infty \mathbb{B}_n(t,\ldots,t)dG(t) \\ &= \sqrt{n}\left\{\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}\min_{1\le k\le m}\bar{G}\left(|w_{ik}-w_{jk}|\right)-\frac{1}{m+1}\right\}, \end{aligned}$$

and

$$I_n^\star = \int_0^\infty \mathbb{B}_n^\star(t,\ldots,t)dG(t) = \frac{2}{\sqrt{n}}\sum_{i=1}^n\left\{W_m(w_i)-\frac{1}{m+1}\right\}.$$

Here, $\bar{G}=1-G$ and for arbitrary integer $p\le m$ and $t_1,\ldots,t_p\in[0,\infty)$,

$$W_p(t_1,\ldots,t_p) = \int_0^\infty \prod_{k=1}^p \left\{F(t_k+s)-F(t_k-s)\right\}dG(s).$$

By Theorem 1, one may conclude that both I_n and $I_n^\star$ converge in law to the centered Gaussian random variable

$$I = \int_0^\infty \mathbb{B}(s,\ldots,s)dG(s)$$

with variance

$$\int_0^\infty\int_0^\infty \Gamma_{\mathbb{B}}(s,\ldots,s,t,\ldots,t)dG(s)dG(t).$$

Other natural extensions of S_n based on the empirical processes $\mathbb{B}_n$ and $\mathbb{B}_n^\star$ could be constructed as follows from the Kolmogorov–Smirnov functional:

$$M_n = \sup_{s\in[0,\infty)}\left|\mathbb{B}_n(s,\ldots,s)\right| \quad\text{and}\quad M_n^\star = \sup_{s\in[0,\infty)}\left|\mathbb{B}_n^\star(s,\ldots,s)\right|.$$

Under the conditions stated in Theorem 1, M_n and $M_n^\star$ converge weakly to $\sup\{|\mathbb{B}(s,\ldots,s)| : s\in[0,\infty)\}$ as $n\to\infty$.

The Cramér–von Mises functional is yet another option, which has the advantage of leading to statistics that can be computed more or less explicitly. Indeed,

$$\begin{aligned} T_n &= \int \mathbb{B}_n^2(t_1,\ldots,t_m)dG(t_1)\times\cdots\times dG(t_m) \\ &= \frac{4}{n(n-1)^2}\sum_{1\le i_1<j_1\le n}\sum_{1\le i_2<j_2\le n}\prod_{k=1}^m \bar{G}\left(|w_{i_1k}-w_{j_1k}|\vee|w_{i_2k}-w_{j_2k}|\right) \\ &\quad -\frac{4}{n-1}\sum_{1\le i<j\le n}\prod_{k=1}^m\left\{\frac{1}{2}-\frac{1}{2}G^2\left(|w_{ik}-w_{jk}|\right)\right\}+\frac{n}{3^m}. \end{aligned}$$

Furthermore,

$$\begin{aligned} T_n^\star &= \int \mathbb{B}_n^{\star 2}(t_1,\ldots,t_m)dG(t_1)\times\cdots\times dG(t_m) \\ &= \frac{4n}{3^m}+\frac{4}{n}\sum_{i=1}^n\sum_{j=1}^n\prod_{k=1}^m W_2(w_{ik},w_{jk}) - 8\sum_{j=1}^n\prod_{k=1}^m N(w_{jk}), \end{aligned}$$

where for all $u, v \in \mathbb{R}^+$,

$$N(u)=\int_0^\infty \{F(u+s)-F(u-s)\}\, G(s)dG(s).$$

It follows from Theorem 1 that the asymptotic distribution of T_n and $T_n^\star$ is $\int \mathbb{B}^2(t_1, \ldots, t_m)dG(t_1)\times\cdots\times dG(t_m)$, which is an infinite sum of weighted chi-squares.

In view of the slow speed of convergence of the statistics S_n, I_n, $I_n^\star$, M_n, $M_n^\star$, T_n and $T_n^\star$ to limits that involve F in an intricate way, it seems wiser to rely on their finite-sample distribution for testing purposes. A Monte Carlo algorithm is provided below in the case of T_n. Its validity stems from Theorem 1. The modifications needed for other functionals of $\mathbb{B}_n$ or $\mathbb{B}_n^\star$ are obvious.

Algorithm 2 (Critical values for T_n) *Repeat the following steps for each $\ell \in \{1, \ldots, L\}$ for some suitably large L.*

1. *Generate a random sample $\epsilon_{1,\ell},\ldots,\epsilon_{n+m-1,\ell}$ from distribution F.*
2. *Set $e_i=\epsilon_{i,\ell}$ for all $i\in\{1,\ldots,n+m-1\}$.*
3. *Use the e_i to construct the w_i.*
4. *Compute the value $T_{n,\ell}$ of the Cramér–von Mises statistic T_n.*

The $100\times\alpha\%$ critical value for the statistic T_n is then approximated by the corresponding quantile in the set $T_{n,1},\ldots,T_{n,L}$. Similarly, the P-value associated with an observed value $T_{n,0}$ can be estimated by

$$\frac{1}{L}\sum_{\ell=1}^L \mathbf{1}\left(T_{n,\ell}>T_{n,0}\right).$$

4.2 The case where F is unknown

When F is unknown but symmetric, rank-based analogues of I_n and $I_n^\star$ are given by

$$\begin{aligned} \tilde{I}_n &= \int_0^1 \tilde{\mathbb{B}}_n(s,\ldots,s)d\tilde{G}(s) \\ &= \sqrt{n}\left\{\frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}\min_{1\le k\le m}\left(1-|\tilde{w}_{ik}-\tilde{w}_{jk}|\right)^2-\frac{1}{m+1}\right\}, \end{aligned}$$

and

$$\tilde{I}_n^\star = \int_0^1 \tilde{\mathbb{B}}_n^\star(s,\ldots,s)d\tilde{G}(s) = \frac{2}{\sqrt{n}}\sum_{i=1}^n \left\{\tilde{W}_m(\tilde{w}_i) - \frac{1}{m+1}\right\},$$

where $\tilde{W}_p(u) = \tilde{W}_p(u')$, with $u'_k = u_k \wedge (1-u_k)$, $k \in \{1,\ldots,p\}$, $\tilde{W}_p$ is symmetric in its arguments and if $0 = u_0 \le u_1 \le \cdots \le u_p \le u_{p+1} = 1/2$, then

$$\tilde{W}_p(u_1,\ldots,u_p) = 2\sum_{k=0}^p \int_{u_k}^{u_{k+1}} \left\{(1-s)(2s)^{p-k}\prod_{j=1}^k (u_j+s) + s\prod_{j=1}^k (u_j+1-s)\right\} ds.$$

Again, using Theorem 2, one may conclude that both $\tilde{I}_n$ and $\tilde{I}_n^\star$ converge in law to the centered Gaussian random variable

$$\tilde{I} = \int_0^1 \tilde{\mathbb{B}}(s,\ldots,s)d\tilde{G}(s)$$

with variance

$$\int_0^1\int_0^1 \Gamma_{\tilde{\mathbb{B}}}(s,\ldots,s,t,\ldots,t)d\tilde{G}(s)d\tilde{G}(t).$$

Rank-based analogues of M_n, and $M_n^\star$ are respectively given by

$$\tilde{M}_n = \max_{1\le i\le n}\left|\tilde{\mathbb{B}}_n\left(\frac{i}{n+1},\ldots,\frac{i}{n+1}\right)\right|$$

and

$$\tilde{M}_n^\star = \max_{1\le i\le n}\left|\tilde{\mathbb{B}}_n^\star\left(\frac{i}{n+1},\ldots,\frac{i}{n+1}\right)\right|.$$

Under the conditions stated in Theorem 2, $\tilde{M}_n$ and $\tilde{M}_n^\star$ converge weakly, as $n\to\infty$, to $\sup\{|\tilde{\mathbb{B}}(s,\ldots,s)| : s\in[0,1]\}$.

Finally, explicit expressions for the rank-based analogues of T_n and $T_n^\star$ are found to be

$$\begin{aligned}\tilde{T}_n &= \int \tilde{\mathbb{B}}_n^2(t_1,\ldots,t_m)d\tilde{G}(t_1)\times\cdots\times d\tilde{G}(t_m)\\ &= \frac{4}{n(n-1)^2}\sum_{1\le i_1<j_1\le n}\ \sum_{1\le i_2<j_2\le n}\ \prod_{k=1}^m \left\{1-\tilde{G}\left(|w_{i_1k}-w_{j_1k}|\vee|w_{i_2k}-w_{j_2k}|\right)\right\}\\ &\qquad -\frac{4}{n-1}\sum_{1\le i<j\le n}\ \prod_{k=1}^m\left\{\frac{1}{2}-\frac{1}{2}\,\tilde{G}^2\left(|w_{ik}-w_{jk}|\right)\right\}+\frac{n}{3^m},\end{aligned}$$

and

$$\begin{aligned}\tilde{T}_n &= \int \left| \tilde{\mathbb{B}}^\star_n(t_1,\ldots,t_m) \right|^2 d\tilde{G}(t_1)\times\cdots\times d\tilde{G}(t_m)\\ &= \frac{4n}{3^m} + \frac{4}{n}\sum_{i=1}^n\sum_{j=1}^n\prod_{k=1}^m \tilde{W}_2(\tilde{w}_{ik},\tilde{w}_{jk}) - 8\sum_{j=1}^n\prod_{k=1}^m \tilde{N}(\tilde{w}_{jk}).\end{aligned}$$

In these formulas,

$$\begin{aligned}\tilde{W}_2(u,v) &= \tilde{W}_2(u',v')\\ &= \int_0^1 \left\{\tilde{F}(u+s)-\tilde{F}(u-s)\right\}\left\{\tilde{F}(v+s)-\tilde{F}(v-s)\right\} d\tilde{G}(s)\\ &= \frac{1}{6} + u'v'(1-u'\vee v') + \frac{1}{3}u'\{1+(u')^3\} + \frac{1}{3}v'\{1+(v')^3\}\\ &\qquad -(u'\wedge v')^3 - \frac{2}{3}(u'\vee v')^3,\end{aligned}$$

with $u' = u\wedge(1-u)$, $v' = v\wedge(1-v)$, and

$$\begin{aligned}\tilde{N}(u) = \tilde{N}(u') &= \int_0^1 \left\{\tilde{F}(u+s)-\tilde{F}(u-s)\right\}\tilde{G}(s)d\tilde{G}(s)\\ &= \frac{187}{480} - \frac{3}{4}\left(u-\frac{1}{2}\right)^2 + \frac{1}{2}\left(u-\frac{1}{2}\right)^4\\ &= \frac{7}{30} + \frac{1}{2}u - u^3 + \frac{1}{2}u^4.\end{aligned}$$

A Monte Carlo algorithm is provided below for the determination of the distribution of $\tilde{T}_n$ under the null hypothesis of randomness. Its validity stems from Theorem 2. The modifications needed for $\tilde{M}_n$, $\tilde{M}^\star_n$, $\tilde{T}^\star_n$ or other functionals of $\tilde{\mathbb{B}}_n$ or $\tilde{\mathbb{B}}^\star_n$ are obvious.

Algorithm 3 (Critical values for $\tilde{T}_n$) *Repeat the following steps for each $\ell \in \{1, \ldots, L\}$ for some suitably large L.*

1. *Generate a random sample $U_{1,\ell},\ldots,U_{n+m-1,\ell}$ from $\mathcal{U}(0,1)$.*
2. *Call R_i the rank of $U_{i,\ell}$ among $U_{1,\ell},\ldots,U_{n+m-1,\ell}$.*
3. *Using $R_1,\ldots,R_{n+p-1}$, compute the $\tilde{e}_i$ as per Equation (3).*
4. *Use the $\tilde{e}_i$ to construct the $\tilde{w}_i$.*
5. *Compute the value $\tilde{T}_{n,\ell}$ of the rank-based Cramér–von Mises statistic $\tilde{T}_n$.*

The $100\times\alpha\%$ critical value for the statistic $\tilde{T}_n$ is then approximated by the corresponding quantile in the set $\tilde{T}_{n,1},\ldots,\tilde{T}_{n,L}$. Similarly, the P-value associated with an observed value $\tilde{T}_{n,0}$ can be estimated by

$$\frac{1}{L}\sum_{\ell=1}^L \mathbf{1}\left(\tilde{T}_{n,\ell} > \tilde{T}_{n,0}\right).$$

A final statistic that was included in the simulation study presented below is

$$\tilde{S}_n^\star = \tilde{\mathbb{B}}_n(\delta,\ldots,\delta)/\tilde{s}_n,$$

which converges to a standard normal random variable by Theorem 2.

Remark 2 *As illustrated in Table 1, note that in spite of the fact that $\tilde{M}_n$ and $\tilde{M}_n^\star$ converge in law to the same limit, under the null hypothesis of randomness, their respective speed of convergence could not be the same. In particular, for small sample sizes, quantiles might be different. It is therefore recommended to calculate P-values and quantiles for each statistic separately. The same advice applies to the pairs $\tilde{S}_n$, $\tilde{S}_n^\star$, $\tilde{T}_n$, $\tilde{T}_n^\star$, and $\tilde{I}_n$, $\tilde{I}_n^\star$.*

5 Finite-sample performance

This section carries out two sets of Monte Carlo experiments to compare the performance, for a sample size $n = 100$, under various alternatives and dimensions m, of the statistics $\tilde{S}_n$, $\tilde{S}_n^\star$, $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$. For $\tilde{S}_n$ and $\tilde{S}_n^\star$, $\delta = 0.3$ was used throughout all simulations. To estimate the power under a fixed alternative, 10,000 samples were

Table 1: 95% quantiles for statistics $\tilde{S}_n$, $\tilde{S}_n^\star$ with $\delta = 0.3$, $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$ for sample sizes $n = 20, 50, 100$ and for $m \in \{2, 3, 4, 5, 6\}$, based on 10,000 replicates.

		m				
Statistic	n	2	3	4	5	6
$\tilde{S}_n$	20	9.012032	9.134711	9.869765	10.590494	13.172255
	50	5.940659	5.730771	6.202165	6.629940	7.677685
	100	4.540863	4.401059	4.692272	5.100884	5.816306
$\tilde{S}_n^\star$	20	5.814012	4.296890	3.629241	3.201116	2.881487
	50	4.423979	3.423614	2.967845	2.712009	2.556186
	100	3.677803	2.945949	2.639806	2.458066	2.327052
$\tilde{I}_n$	20	0.075363	0.107253	0.121851	0.127242	0.127455
	50	0.052585	0.078264	0.090282	0.095680	0.097110
	100	0.043092	0.065551	0.076175	0.081853	0.083969
$\tilde{I}_n^\star$	20	0.105261	0.119070	0.124014	0.123530	0.120635
	50	0.078158	0.092435	0.099896	0.102787	0.103476
	100	0.063384	0.078290	0.085704	0.088411	0.090208
$\tilde{M}_n$	20	0.375645	0.507397	0.603171	0.688483	0.759761
	50	0.260807	0.360727	0.442342	0.508878	0.564885
	100	0.202400	0.287081	0.354657	0.409974	0.456697
$\tilde{M}_n^\star$	20	0.227611	0.314323	0.378650	0.431029	0.471951
	50	0.169251	0.240221	0.295861	0.342389	0.382811
	100	0.141433	0.209432	0.259929	0.300341	0.335972
$\tilde{T}_n$	20	0.017241	0.012620	0.008172	0.004973	0.002819
	50	0.007356	0.005982	0.004142	0.002596	0.001482
	100	0.004166	0.003801	0.002786	0.001800	0.001027
$\tilde{T}_n^\star$	20	0.011396	0.009728	0.006244	0.003404	0.001700
	50	0.006216	0.005944	0.004056	0.002420	0.001260
	100	0.004048	0.004104	0.002972	0.001784	0.000956

generated for each statistic, and the percentage of rejected samples was recorded. In order to speed up calculations, the 95% quantiles of Table 1 were used, instead of P-values.

In the first series of experiments, the alternatives are the time series models used in Hong and White (2005), restricted to Gaussian innovations. In that paper, the authors introduced a new statistic measuring entropy with respect to independence at various lags. In the second set of experiments, other models of alternatives are proposed, with variable degrees of dependence depending on the value of a parameter. A discussion of the results in presented afterwards.

5.1 First experiment

For the first set of comparisons, nine time series models exhibiting various forms of dependence were used, as in Hong and White (2005). In all these models, listed in Table 2, the (independent) innovations are Gaussian. Note that Hong and White (2005) also considered log-normal innovations.

In order to be able to make comparisons with the results of Hong and White (2005), the same procedure was followed to obtain nearly stationary time series: for each repetition and for each alternative, a time series of length 200 was generated, and only the last 100 observations were used. The results of these comparisons are given in two tables.

Table 3 contains the estimated power, under the nine alternatives, for the statistics $\tilde{S}_n$, $\tilde{S}_n^\star$, $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$ calculated with $m = 2$, and the statistic $\mathcal{T}_n(1)$ proposed by Hong and White (2005). All statistics are comparable since they are all based on the pairs (x_t, x_{t+1}), $t = 101, \ldots, 200$. No simulations needed to be done for $\mathcal{T}_n(1)$, as the results were kindly provided by the authors. In Table 4, similar results are presented for dimensions $m \in \{2, \ldots, 6\}$, for all statistics but $\mathcal{T}_n(1)$.

Table 2: List of models with Gaussian innovations used by Hong and White (2005)

Time Series Model	Equation
I.I.D.	$X_t = \varepsilon_t$
AR(1)	$X_t = 0.3X_{t-1} + \varepsilon_t$
ARCH(1)	$X_t = h_t^{1/2}\varepsilon_t$, $h_t = 1 + 0.8X_{t-1}^2$
Threshold GARCH(1, 1)	$X_t = h_t^{1/2}\varepsilon_t$, with $h_t^2 = 0.25 + 0.6h_{t-1}^2$ $+0.5X_{t-1}^2\mathbf{1}(\varepsilon_{t-1} < 0) + 0.2X_{t-1}^2\mathbf{1}(\varepsilon_{t-1} \geq 0)$
Bilinear AR(1)	$X_t = 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$
Nonlinear MA(1)	$X_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$
Threshold AR(1)	$X_t = 0.4X_{t-1}\mathbf{1}(X_{t-1} > 1) + \varepsilon_t$ $-0.5X_{t-1}\mathbf{1}(X_{t-1} \leq 1)$
Fractional AR(1)	$X_t = 0.8\lvert X_{t-1}\rvert^{1/2} + \varepsilon_t$
Sign AR(1)	$X_t = \mathrm{sign}(X_{t-1}) + 0.43\varepsilon_t$

Table 3: Percentage of rejection, at level $\alpha = .05$, of series of length $n = 100$ of the first set of alternatives, for tests based on the statistic $\mathcal{T}_n(1)$ of Hong and White (2005) and the statistics $\tilde{S}_n$, $\tilde{S}_n^\star$ (both with $\delta = 0.3$), $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$. For all statistics but $\mathcal{T}_n(1)$, the percentage of rejection was estimated with 10,000 replicates.

Model	$\tilde{S}_n^\star$	$\tilde{I}_n^\star$	$\tilde{M}_n^\star$	$\tilde{T}_n^\star$	$\tilde{S}_n$	$\tilde{I}_n$	$\tilde{M}_n$	$\tilde{T}_n$	$\mathcal{T}_n(1)$
I.I.D.	5.20	5.25	4.49	5.67	5.08	4.65	5.36	4.93	6.5
AR(1)	14.01	14.65	11.53	14.78	48.61	54.92	51.46	52.29	14.0
ARCH(1)	95.47	95.80	93.00	95.98	78.90	90.40	90.48	90.76	37.6
Threshold GARCH(1, 1)	72.14	72.71	66.22	72.90	48.99	61.85	62.22	61.75	20.6
Bilinear AR(1)	94.89	94.54	88.66	94.69	98.28	99.44	96.78	98.78	69.6
Nonlinear MA(1)	50.64	49.64	38.45	50.55	71.10	73.62	55.55	67.52	34.0
Threshold AR(1)	8.68	9.24	6.97	9.51	54.77	48.24	34.36	44.52	25.6
Fractional AR(1)	6.62	8.07	7.39	7.85	43.63	44.30	37.92	40.19	17.0
Sign AR(1)	32.06	33.92	32.03	34.03	57.59	59.15	58.42	59.57	60.8

5.2 Second experiment

A second set of comparisons was made using a set of alternatives allowing for various degrees of dependence reflected through a parameter θ. These models, in which $\theta = 0$ corresponds to independence, are listed in Table 5.

Remark 3 *The Threshold AR(1) model was proposed by Tong and Lim (1980). For the randomized tent map due to Genest et al. (2002), the choice $\theta = 1/4$ corresponds to the deterministic tent map, described by Chatterjee and Yilmaz (1992) as a prime example of a chaotic time series. Note that for the tent map, traditional measures of dependence for the pairs $(X_t, X_{t+\ell})$, like autocorrelations, Kendall's tau or Spearman rho, all have theoretical value 0. See, e.g., Genest et al. (2002).*

Table 6 gives the percentages of rejection of samples of size $n = 100$ for the test based on the statistic $\tilde{S}_n$ with $\delta = 0.3$, for $m \in \{2, \ldots, 6\}$ and for the 45 alternatives models. Finally, Table 7 provides a comparison of the performance of the test statistics $\tilde{S}_n$, $\tilde{S}_n^\star$, with $\delta = 0.3$, $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$, for dimensions $m \in \{2, \ldots, 6\}$ and the 9 alternatives corresponding to $\theta = 1/4$. In order to achieve stationarity for a given times series model, for each replicate, 120 observations were generated, and only the last 100 were considered.

5.3 Discussion

First, as shown by the results in Table 3, tests based on statistics $\tilde{S}_n$, $\tilde{I}_n$, $\tilde{M}_n$, and $\tilde{T}_n$ have better power than the test based on $\mathcal{T}_n(1)$ for all models of Table 2 with the exception of the sign AR(1), where they come very close. As for the tests based on statistics $\tilde{S}_n^\star$, $\tilde{I}_n^\star$, $\tilde{M}_n^\star$ and $\tilde{T}_n^\star$, they clearly dominate $\mathcal{T}_n(1)$ for the ARCH(1), Threshold GARCH(1, 1), Bilinear AR(1) and Nonlinear MA(1) models. They are also comparable to $\mathcal{T}_n(1)$ for the AR(1) model. They are, however, outperformed by $\mathcal{T}_n(1)$ for the remaining three models.

Table 4: Percentage of rejection, at level $\alpha = .05$, of the first set of alternatives, for the tests based on the statistics $\tilde{S}_n$, $\tilde{S}_n^\star$ (both with $\delta = 0.3$), $\tilde{I}_n$, $\tilde{I}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{T}_n$ and $\tilde{T}_n^\star$, with $m \in \{2, 2, 3, 4, 5, 6\}$, as estimated with 10,000 replicates of series of length $n = 100$.

Model	m	$\tilde{S}_n^\star$	$\tilde{I}_n^\star$	$\tilde{M}_n^\star$	$\tilde{T}_n^\star$	$\tilde{S}_n$	$\tilde{I}_n$	$\tilde{M}_n$	$\tilde{T}_n$
I.I.D.	2	5.20	5.25	4.49	5.67	5.08	4.65	5.36	4.93
I.I.D.	3	5.16	5.22	4.63	5.33	5.27	4.90	5.16	4.71
I.I.D.	4	5.15	5.24	4.83	5.17	5.04	4.96	5.21	4.95
I.I.D.	5	5.22	5.53	4.90	5.43	5.25	4.83	4.98	4.87
I.I.D.	6	5.40	5.24	4.93	5.28	4.71	4.95	4.96	5.14
AR(1)	2	14.01	14.65	11.53	14.78	48.61	54.92	51.46	52.29
AR(1)	3	13.10	12.20	10.96	13.68	42.73	49.97	45.54	49.95
AR(1)	4	12.31	11.47	10.82	12.55	34.81	44.94	39.52	45.61
AR(1)	5	11.53	10.95	10.59	11.85	29.29	40.38	35.24	41.64
AR(1)	6	11.33	10.05	10.04	11.43	23.22	36.60	31.74	38.94
ARCH(1)	2	95.47	95.80	93.00	95.98	78.90	90.40	90.48	90.76
ARCH(1)	3	94.18	93.77	91.25	94.34	76.21	89.96	90.05	88.98
ARCH(1)	4	91.53	91.06	89.35	91.45	67.91	87.76	88.21	85.87
ARCH(1)	5	88.32	88.09	87.28	88.60	60.50	85.45	86.16	81.86
ARCH(1)	6	85.09	84.39	84.60	85.71	50.29	82.85	84.15	78.36
Threshold GARCH(1, 1)	2	72.14	72.71	66.22	72.90	48.99	61.85	62.22	61.75
Threshold GARCH(1, 1)	3	80.84	79.31	75.76	80.68	57.79	73.10	73.60	71.89
Threshold GARCH(1, 1)	4	84.00	82.29	81.02	83.70	57.94	78.35	78.75	76.12
Threshold GARCH(1, 1)	5	85.53	83.48	83.62	85.45	58.24	80.62	81.22	77.94
Threshold GARCH(1, 1)	6	86.13	83.84	84.58	86.06	54.80	81.87	82.41	78.90
Bilinear AR(1)	2	94.89	94.54	88.66	94.69	98.28	99.44	96.78	98.78
Bilinear AR(1)	3	95.63	94.08	90.95	95.35	98.26	98.64	91.56	99.00
Bilinear AR(1)	4	93.59	91.08	89.85	93.23	95.47	95.41	83.93	98.26
Bilinear AR(1)	5	90.81	88.12	87.60	90.39	91.04	90.79	78.28	96.57
Bilinear AR(1)	6	87.78	83.84	85.00	87.37	83.03	85.32	73.90	94.03
Nonlinear MA(1)	2	50.64	49.64	38.45	50.55	71.10	73.62	55.55	67.52
Nonlinear MA(1)	3	42.61	39.07	32.38	41.46	66.55	57.57	34.66	63.43
Nonlinear MA(1)	4	35.89	32.45	29.45	34.76	54.88	41.19	23.50	54.72
Nonlinear MA(1)	5	31.48	28.93	27.58	30.56	45.49	31.08	19.01	45.87
Nonlinear MA(1)	6	28.25	25.77	25.35	27.71	34.80	24.93	16.84	40.42
Threshold AR(1)	2	8.68	9.24	6.97	9.51	54.77	48.24	34.36	44.52
Threshold AR(1)	3	9.42	8.63	7.00	9.41	48.39	30.04	16.84	38.82
Threshold AR(1)	4	9.20	8.33	7.79	9.02	38.27	20.10	9.81	32.32
Threshold AR(1)	5	8.81	8.30	8.13	8.93	31.49	14.57	7.49	26.57
Threshold AR(1)	6	8.76	7.86	8.08	8.86	23.39	11.46	6.42	23.70
Fractional AR(1)	2	6.62	8.07	7.39	7.85	43.63	44.30	37.92	40.19
Fractional AR(1)	3	7.09	7.38	6.86	7.78	37.13	36.49	29.32	37.64
Fractional AR(1)	4	7.23	7.14	7.35	7.57	29.64	30.13	22.59	33.97
Fractional AR(1)	5	7.02	7.17	7.07	7.43	25.25	24.91	18.37	30.86
Fractional AR(1)	6	6.84	6.80	6.91	7.43	19.75	21.24	15.30	28.33
Sign AR(1)	2	32.06	33.92	32.03	34.03	57.59	59.15	58.42	59.57
Sign AR(1)	3	35.43	38.03	37.19	37.75	58.22	59.63	59.18	60.38
Sign AR(1)	4	37.11	39.61	39.79	39.75	57.65	59.92	59.71	60.38
Sign AR(1)	5	37.75	40.90	40.91	40.79	57.60	59.93	60.17	60.09
Sign AR(1)	6	38.13	40.92	40.74	41.43	56.63	59.70	60.29	60.22

Second, based on the results of Tables 3–7, observe that among the statistics $\tilde{S}_n$, $\tilde{I}_n$, $\tilde{M}_n$, and $\tilde{T}_n$, the test based on $\tilde{I}_n$ is most often the best (or close to best) choice. While $\tilde{T}_n$ performs quite well also, its computational complexity makes it much less attractive.

Table 5: List of models used for the second experiment

Model	Equation
AR(1) Gaussian	$X_t = \theta X_{t-1} + \varepsilon_t$, with Gaussian innovations ε_t
AR(1) Laplace	$X_t = \theta X_{t-1} + \varepsilon_t$, with Laplace innovations ε_t
AR(1) Cauchy	$X_t = \theta X_{t-1} + \varepsilon_t$, with Cauchy innovations ε_t
MA(1)	$X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, with Gaussian innovations ε_t
GARCH(1, 1)	$X_t = h_t^{1/2}\varepsilon_t$, with $h_t^2 = 1 + \theta h_{t-1}^2 + 2\theta X_{t-1}^2$
ARCH(1)	$X_t = h_t^{1/2}\varepsilon_t$, with $h_t = 1 + \theta X_{t-1}^2$
Threshold AR(1)	$X_t = -\theta X_{t-1}\ \mathrm{sign}(X_{t-1} - 0.5) + \varepsilon_t$, with innovations $\varepsilon_t \sim \mathcal{U}(0,1)$
(Randomized) Tent Map	$X_t = (1-\eta_t)\varepsilon_t + \eta_t(1 - \lvert 2X_{t-1} - 1\rvert)$, with i.i.d. $\varepsilon_t \sim \mathcal{U}(0,1)$, independent of the i.i.d. Bernoulli $\eta_t \sim \mathcal{B}(4\theta)$
Clayton copula	X_t is Markovian, with $(X_{t-1}, X_t) \sim C_\theta$, with the Clayton copula, defined for $u, v \in (0,1)$ by $C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$.

Among the other group of statistics based on $\mathbb{B}_n^\star$, $\tilde{I}_n^\star$ and $\tilde{T}_n^\star$ are the top choices in terms of performance, with $\tilde{S}_n^\star$ not that far behind.

Last but not least, the simulation results suggest that there is a large difference between the performance of test statistics based on $\tilde{\mathbb{B}}_n$ versus $\tilde{\mathbb{B}}_n^\star$, depending on the type of alternatives. In fact:

a) For most alternatives with constant conditional variance given the past, e.g., AR(1), MA(1), Fractional AR(1), Threshold AR(1), the tests based on statistics $\tilde{S}_n$, $\tilde{I}_n$, $\tilde{M}_n$, and $\tilde{T}_n$ perform much better in general, than those based on $\tilde{S}_n^\star$, $\tilde{I}_n^\star$, $\tilde{M}_n^\star$ and $\tilde{T}_n^\star$.

b) For alternatives with non-constant conditional variance given the past, e.g., ARCH(1), GARCH(1, 1), and Threshold GARCH(1, 1), the tests based on statistics $\tilde{S}_n^\star$, $\tilde{I}_n^\star$, $\tilde{M}_n^\star$ and $\tilde{T}_n^\star$, are much more powerful in general, than those based on $\tilde{S}_n$, $\tilde{I}_n$, $\tilde{M}_n$, and $\tilde{T}_n$.

In practice, of course, the nature of the alternative being faced is usually unknown. For the statistic $\mathcal{T}_n$ of Hong and White (2005), this is not a concern since the simulations indicate that it performs equally well whether the conditional variance is constant or not. For the statistics proposed herein, however, this might be problematic. Luckily, the following general strategy can be used to circumvent the problem.

Consider a statistic $D_n = \phi(\tilde{\mathbb{B}}_n)$, calculated from a continuous functional ϕ of $\tilde{\mathbb{B}}_n$, and its parent statistic $D_n^\star = \phi(\tilde{\mathbb{B}}_n^\star)$. Let P_n and $P_n^\star$ represent respectively the (approximate) P-values of D_n and $D_n^\star$, as calculated using a method analogous to the one described in Algorithm 3 for $\tilde{T}_n$. A combined test of approximate level α is then obtained by the following rule:

$$\text{Reject } H_0 \quad \Leftrightarrow \quad \min(P_n, P_n^\star) < \alpha.$$

Table 6: Percentage of rejection, at level $\alpha = 5\%$, of the second set of alternatives with $\theta \in \{0, 1/32, 1/16, 1/8, 1/4\}$, for the test based on statistic $\tilde{S}_n$ with $\delta = 0.3$, as estimated with $10,000$ replicates of series of length $n = 100$.

Model	m	$\theta = 0$	$\theta = 1/32$	$\theta = 1/16$	$\theta = 1/8$	$\theta = 1/4$
AR(1) Gaussian	2	4.71	4.96	5.66	9.52	34.73
AR(1) Gaussian	3	4.83	4.86	5.97	8.95	30.11
AR(1) Gaussian	4	4.59	4.35	5.30	7.52	23.83
AR(1) Gaussian	5	4.82	4.73	5.46	7.17	20.41
AR(1) Gaussian	6	4.19	4.35	4.71	6.03	15.78
AR(1) Laplace	2	5.17	5.01	6.77	14.49	53.46
AR(1) Laplace	3	5.29	5.31	6.41	11.94	46.97
AR(1) Laplace	4	4.90	5.08	5.52	9.40	37.80
AR(1) Laplace	5	5.08	5.33	6.03	8.85	32.21
AR(1) Laplace	6	4.37	4.93	5.27	7.05	25.16
AR(1) Cauchy	2	4.90	10.01	20.83	54.47	96.34
AR(1) Cauchy	3	5.07	9.56	18.56	47.86	93.60
AR(1) Cauchy	4	4.60	7.77	13.99	39.34	88.35
AR(1) Cauchy	5	5.05	7.45	12.42	32.82	82.23
AR(1) Cauchy	6	4.65	6.23	9.88	25.15	73.26
MA(1) Gaussian	2	4.69	5.87	7.35	13.54	38.09
MA(1) Gaussian	3	4.73	5.89	7.15	12.03	33.08
MA(1) Gaussian	4	4.60	5.34	6.20	9.13	25.58
MA(1) Gaussian	5	4.96	5.63	6.05	8.50	21.09
MA(1) Gaussian	6	4.43	4.83	5.31	7.26	16.26
GARCH(1, 1)	2	4.60	7.17	10.49	21.37	55.38
GARCH(1, 1)	3	4.96	7.13	10.38	21.34	58.95
GARCH(1, 1)	4	4.67	6.18	8.72	17.51	54.38
GARCH(1, 1)	5	4.87	6.50	8.65	15.55	49.93
GARCH(1, 1)	6	4.51	5.57	6.95	11.98	42.31
ARCH(1)	2	5.01	6.00	6.96	10.44	21.30
ARCH(1)	3	5.30	6.25	6.58	9.71	19.76
ARCH(1)	4	4.53	5.37	5.65	8.23	15.40
ARCH(1)	5	4.81	5.77	5.90	7.71	13.02
ARCH(1)	6	4.26	5.01	5.34	6.40	10.52
Tent map	2	4.60	11.23	32.22	90.41	100.00
Tent map	3	4.81	10.07	26.43	85.13	100.00
Tent map	4	4.54	7.78	19.28	76.01	100.00
Tent map	5	4.77	7.27	16.06	66.32	100.00
Tent map	6	4.23	6.28	12.06	53.71	100.00
Threshold AR(1)	2	5.10	5.85	8.38	19.27	59.87
Threshold AR(1)	3	5.12	5.66	8.24	16.72	54.09
Threshold AR(1)	4	4.55	5.09	6.37	12.94	44.39
Threshold AR(1)	5	5.07	5.24	6.18	11.93	37.09
Threshold AR(1)	6	4.60	4.60	5.18	9.47	28.98
Clayton	2	4.83	5.11	5.67	8.94	21.89
Clayton	3	4.47	5.21	5.65	8.47	19.61
Clayton	4	4.30	4.59	4.85	7.06	15.50
Clayton	5	4.52	5.07	5.21	6.89	13.29
Clayton	6	4.07	4.19	4.71	5.85	10.50

Table 7: Percentage of rejection, at level $\alpha = 5\%$, of the second set of alternatives with $\theta = 1/4$, for test based on statistics $\tilde{S}_n$, $\tilde{S}_n^\star$, (both with $\delta = 0.3$), $\tilde{T}_n$, $\tilde{T}_n^\star$, $\tilde{M}_n$, $\tilde{M}_n^\star$, $\tilde{I}_n^\star$ and $\tilde{I}_n$, as estimated with 10,000 replicates of series of length $n = 100$.

Model	m	$\tilde{S}_n^\star$	$\tilde{I}_n^\star$	$\tilde{M}_n^\star$	$\tilde{T}_n^\star$	$\tilde{S}_n$	$\tilde{I}_n$	$\tilde{M}_n$	$\tilde{T}_n$
AR(1) Gaussian	2	10.15	10.75	9.46	10.63	34.73	38.38	36.18	36.07
AR(1) Gaussian	3	10.13	9.72	8.68	9.91	30.11	33.94	30.65	34.07
AR(1) Gaussian	4	9.35	8.89	8.79	9.38	23.83	29.59	26.10	31.06
AR(1) Gaussian	5	9.07	8.85	8.55	9.07	20.41	26.19	23.15	28.54
AR(1) Gaussian	6	8.77	8.28	8.40	8.85	15.78	23.95	20.97	26.80
AR(1) Laplace	2	27.98	25.59	18.27	25.99	53.46	63.34	59.40	61.35
AR(1) Laplace	3	25.26	21.20	17.16	22.50	46.97	56.69	54.28	58.50
AR(1) Laplace	4	22.50	18.13	16.75	19.58	37.80	51.24	49.22	53.83
AR(1) Laplace	5	20.17	17.17	16.61	17.73	32.21	46.46	45.37	49.55
AR(1) Laplace	6	18.91	15.15	16.00	16.83	25.16	43.04	41.71	46.18
AR(1) Cauchy	2	84.34	84.96	79.33	85.69	96.34	98.35	96.80	97.90
AR(1) Cauchy	3	79.27	77.79	73.72	79.51	93.60	97.12	94.52	97.23
AR(1) Cauchy	4	72.75	70.46	68.67	72.40	88.35	95.03	91.45	95.64
AR(1) Cauchy	5	66.60	64.48	63.88	66.12	82.23	92.24	88.09	93.56
AR(1) Cauchy	6	61.17	57.96	58.38	61.11	73.26	89.41	83.95	91.89
MA(1) Gaussian	2	10.74	11.38	9.49	11.06	38.09	42.27	38.03	38.98
MA(1) Gaussian	3	10.52	10.42	8.95	9.85	33.08	37.11	31.82	36.96
MA(1) Gaussian	4	10.04	9.76	9.05	9.15	25.58	32.38	26.81	33.38
MA(1) Gaussian	5	9.83	9.70	9.02	8.98	21.09	27.72	23.06	29.67
MA(1) Gaussian	6	9.63	9.03	8.91	8.90	16.26	24.67	20.92	26.99
GARCH(1,1)	2	81.25	81.83	75.49	82.47	55.38	69.18	70.05	70.36
GARCH(1,1)	3	84.92	83.95	80.10	84.79	58.95	76.59	77.37	74.80
GARCH(1,1)	4	84.18	82.79	81.18	84.03	54.38	78.21	78.28	74.67
GARCH(1,1)	5	82.04	81.46	80.66	82.51	49.93	76.94	77.90	73.07
GARCH(1,1)	6	80.09	78.84	79.26	80.56	42.31	75.69	76.83	70.52
ARCH(1)	2	41.89	43.06	35.18	43.42	21.30	29.12	29.70	29.56
ARCH(1)	3	39.26	37.39	32.48	39.47	19.76	29.82	29.59	27.18
ARCH(1)	4	35.23	32.95	30.77	34.95	15.40	29.07	28.25	24.36
ARCH(1)	5	31.62	30.27	28.96	31.72	13.02	26.63	26.56	21.78
ARCH(1)	6	29.04	26.38	26.82	29.31	10.52	25.08	25.06	20.63
Tent map	2	100.00	89.08	15.59	99.69	100.00	100.00	100.00	100.00
Tent map	3	99.85	55.94	76.78	96.45	100.00	100.00	100.00	100.00
Tent map	4	96.28	41.42	80.29	83.00	100.00	100.00	100.00	100.00
Tent map	5	88.15	37.58	79.17	70.42	100.00	100.00	83.34	100.00
Tent map	6	78.93	33.83	75.71	60.88	100.00	86.42	18.84	100.00
Threshold AR(1)	2	5.17	5.82	4.83	5.51	59.87	63.87	48.56	58.47
Threshold AR(1)	3	5.04	5.26	4.50	5.28	54.09	50.12	36.59	56.79
Threshold AR(1)	4	5.17	5.26	4.73	5.05	44.39	40.49	28.02	51.47
Threshold AR(1)	5	5.06	5.46	4.98	5.19	37.09	32.17	22.78	46.96
Threshold AR(1)	6	5.00	5.19	4.82	5.16	28.98	26.86	19.50	43.24
Clayton	2	11.39	11.56	9.57	11.72	21.89	24.64	22.61	21.56
Clayton	3	11.22	10.64	8.92	11.27	19.61	22.18	19.65	20.35
Clayton	4	10.48	10.29	9.15	10.81	15.50	19.79	17.05	18.83
Clayton	5	10.07	9.85	9.37	10.54	13.29	17.70	15.22	17.18
Clayton	6	9.86	8.96	9.23	10.38	10.50	16.70	14.56	16.32

To see that the limiting level of this decision rule is α, note that under H_0, both statistics converge in law to the same random variable $\mathbb{D} = \phi(\mathbb{B})$. Accordingly, both P_n and $P_n^\star$ converge to the same random variable $U \sim \mathcal{U}(0,1)$. Hence

$$\lim_{n\to\infty} P\left\{\min\left(P_n, P_n^\star\right) < \alpha\right\} = P(U \le \alpha) = \alpha.$$

5.4 Comparison between statistics based on $\tilde{\mathbb{B}}_n$ and $\tilde{\mathbb{B}}_n^\star$

As remarked before, there are really two different groups of statistics: those based either on $\tilde{\mathbb{B}}_n$ or on $\tilde{\mathbb{B}}_n^\star$. According to Theorem 2, their asymptotic behavior is the same under the null hypothesis of randomness. However, under an alternative making ε_t dependent but stationary and ergodic with common continuous distribution F, their power should depend respectively on $\sqrt{n}\,\mu(t)$ and $\sqrt{n}\,\mu^\star(t)$, μ and $\mu^\star$ being given respectively by

$$\mu(t) = \tilde{B}(t) - \prod_{k=1}^{m} G(t_k) \quad \text{and} \quad \mu^\star(t) = B^\star(t) - 2\prod_{k=1}^{m} G(t_k),$$

where

$$\tilde{B}(t) = P\left(\bigcap_{k=1}^{m}\{|U_k - V_k| \le t_k\}\right)$$

and

$$\tilde{B}^\star(t) = 2P\left(\bigcap_{k=1}^{m}\{|U_k - W_k| \le t_k\}\right)$$

are defined in terms of three independent random vectors U, V, and W with uniform marginals. Here, U and V are from the copula C associated with $(\varepsilon_1,\ldots,\varepsilon_m)$, i.e., C is the joint distribution function of $F(\varepsilon_1),\ldots,F(\varepsilon_m)$. As for W, its elements are taken to be mutually independent.

Under the assumptions of stationarity and ergodicity, $\tilde{B}_n$ and $\tilde{B}_n^\star$ are convergent estimators of $\tilde{B}$ and $\tilde{B}^\star$. Therefore, one could use Monte Carlo simulations to find out the value of δ for which the maximum distance between $\tilde{B}(\delta,\ldots,\delta)$ and $(2\delta-\delta^2)^m$ is achieved. Having done such simulations, it turns out that for many models of alternatives of the second list, $\delta = 0.3$ is almost always close to the optimum value.

6 An illustrative example

To illustrate the use of the proposed statistics, consider "Series G" of Box et al. (1994), which consists of 144 monthly totals x_t of thousands of international airline passengers. The series extends from January, 1949, to December, 1960. As a variance-stabilizing transformation, these authors consider the series $z_i = \log(x_i)$, for which they arrive at the model $\nabla\nabla_{12} z_i = (1-\theta B)(1-\Theta B^{12})\varepsilon_i$, written explicitly as

$$z_i - z_{i-1} - z_{i-12} - z_{i-13} = \varepsilon_i - \theta\varepsilon_{i-1} - \Theta\varepsilon_{i-12} + \theta\Theta\varepsilon_{i-13},$$

with $\hat{\theta} = 0.402$, $\hat{\Theta} = 0.557$, and $\hat{\sigma}_\varepsilon^2 = 1.34 \times 10^{-3}$. These estimates were obtained by the method of maximum likelihood, assuming that the ε_i are normally distributed. Based on the Ljung–Box statistic, Box et al. (1994) arrive at the conclusion that "the check does not provide any evidence of inadequacy in the model." The same conclusion is reached by Brockwell and Davis (1991).

The above model for $Y_i = \nabla\nabla_{12} z_i$ is an MA(1) $\times$ MA$_{12}$(1). Assumptions I and II can easily be seen to hold in this case, provided that second-order moments of ε_i exist and that their density is continuous, bounded and symmetric (as would be the case, e.g., under normality). Under these conditions, the statistics $\tilde{S}_n, \tilde{I}_n, \tilde{M}_n$ and $\tilde{T}_n$ and $\tilde{S}_n^\star, \tilde{I}_n^\star, \tilde{M}_n^\star$ and $\tilde{T}_n^\star$ provide alternative checks for the model.

Table 8 provides estimated P-values for the new statistics, based on 10,000 replicates. As can be seen from it, the null hypothesis of independence for m consecutive innovations of the fitted model is readily rejected at the 1% level for $p \geq 6$, for all statistics but $\tilde{S}_n$. Curiously, however, the rank-based version $\tilde{S}_n$ of the BDS statistic does not lead to the rejection of H_0 for any $2 \leq m \leq 10$.

Table 8: P-values (%) for the model proposed by Box et al. (1994), using 10,000 replicates

	Order m								
Statistic	2	3	4	5	6	7	8	9	10
$\tilde{S}_n^\star$	1.54	1.98	1.96	1.91	0.95	0.73	0.57	0.48	0.39
$\tilde{I}_n^\star$	2.23	2.98	2.60	1.86	0.98	0.57	0.47	0.40	0.41
$\tilde{M}_n^\star$	3.25	3.76	2.84	1.88	1.16	0.98	1.06	0.80	0.65
$\tilde{T}_n^\star$	1.80	2.12	2.05	1.63	0.81	0.50	0.44	0.36	0.29
$\tilde{S}_n$	22.21	57.13	75.73	46.91	27.07	30.18	37.01	26.23	12.78
$\tilde{I}_n$	6.12	7.42	4.96	1.76	0.81	0.62	0.58	0.48	0.38
$\tilde{M}_n$	5.74	3.87	2.78	1.27	0.92	1.21	0.95	0.66	0.49
$\tilde{T}_n$	3.80	6.10	5.60	2.40	1.30	0.80	0.70	0.80	0.80

7 Models satisfying Assumption II

Univariate stationary time series models can be divided into two major classes, according as the conditional variance given the past is constant or not. Many time series models (Y_i) from the first group can be represented in the form

$$Y_i = \phi(Z_{i-1}, \theta) + \varepsilon_i, \tag{8}$$

in terms of (possibly exogenous) random vectors (Z_i) and innovations (ε_i). Here, it is assumed that for $j > i$ the innovation ε_j is independent of Z_i, that $(Z_i)_{i\geq 1}$ is a stationary and ergodic series, and that the parameter space $\mathcal{O} \subset \mathbb{R}^d$ is open. For example, AR(p) and threshold AR(p) models are of this form. One could also enlarge that family and consider "recursive" models of the form

$$Y_i = \phi(Z_{i-1}, \varepsilon_{i-1}, , \ldots, \varepsilon_{i-q}, \theta) + \varepsilon_i, \tag{9}$$

whereof the standard ARMA(p, q) models are well-known representatives. The popular econometric model described below provides another example.

Example 1 *Consider the ARCH(p) model*

$$X_i = \left(\omega + \sum_{j=1}^{p} a_j X_{i-j}^2\right)^{1/2} \epsilon_i,$$

in which the innovations ϵ_i *are* $\mathcal{N}(0,1)$ *and the components of the parameter* $\theta = (\omega, a_1, \ldots, a_p)$ *satisfy* $\omega > 0$, $a_1 \geq 0, \ldots, a_p \geq 0$ *together with the second order stationarity condition*

$$\sum_{j=1}^{p} a_j < 1.$$

Note that the latter condition is also sufficient (but not necessary) to ensure strong stationarity. Setting $Y_i = \log\left(X_i^2\right)$, $Z_{i-1} = \left(X_{i-1}^2, \ldots, X_{i-p}^2\right)$ *and* $\varepsilon_i = \log\left(\epsilon_i^2\right)$, *it follows that*

$$Y_i = \phi(Z_{i-1}, \theta) + \varepsilon_i, \quad i \geq 1$$

with $\phi(z, \theta) = \phi(z, \omega, a) = \log\left(\omega + a^\top z\right)$.

Note that the density f *of* ε_i *is given by* $f(x) = \exp(-e^x/2)e^{x/2}/\sqrt{2\pi}$. *It it clearly continuous and its square integrates to* $1/(2\pi)$. *However,* f *is not symmetric. Nevertheless, since* F *is known and the estimation of* θ *behaves well. Furthermore, Assumption 2 holds true, so Theorem 2 can be applied to test the independence in the series* $|\varepsilon_i|$.

Incidentally, one can also check that $G(x) = (2/\pi)\arctan\left(e^{x/2}\right)$, *since* $\varepsilon_1 - \varepsilon_2 = 2\log\left(|\epsilon_1/\epsilon_2|\right)$ *and* ϵ_1/ϵ_2 *has standard Cauchy distribution.*

It is shown in Lemma 2 below that, under weak regularity conditions including Assumption I, time series models defined by (8) satisfy Assumption II. Moreover, building on the work of Bai (1994), one can show that Assumption II holds true for ARMA(p, q) models. The exact hypotheses are stated in Lemma 3, which in proven in Ghoudi and Rémillard (2006). It is likely that Assumption II also holds for time series models satisfying (9); this problem is currently under investigation.

Concerning time series models with non-constant conditional variance, the situation is quite different. In fact, unless these models can be transformed into models of the form (8) or (9), as done in Example 1 above for the ARCH(p) model, there is no hope that Assumption II could be satisfied. See, e.g., Berkes and Horváth (2003) for results on GARCH models restricted to $m = 1$.

Example 2 *Consider the ARCH(p) model defined by*

$$y_i = \mu + \sqrt{a + \sum_{j=1}^{p} b_j (y_{i-j} - \mu)^2}\, \epsilon_i = \mu + \sigma_i \epsilon_i.$$

For this model, it can be shown, e.g., using the techniques developed by Ghoudi and Rémillard (2004), that if

$$e_{i,n} = (y_i - \hat{\mu}_n) \Big/ \left\{ \hat{a}_n + \sum_{j=1}^{p} \hat{b}_{jn}(y_{i-j} - \hat{\mu}_n)^2 \right\}^{1/2}$$

and

$$\Theta_n = \sqrt{n}\,(\theta_n - \theta) = \sqrt{n}\,\left(\hat{\mu}_n - \mu, \hat{a}_n - a, \hat{b}_n - b\right) \rightsquigarrow \Theta = (\mathcal{M}, \mathcal{A}, \mathcal{B}),$$

then $\mathbb{K}_n \rightsquigarrow \mathbb{K}$ *in* $\mathcal{D}([-\infty,\infty]^m)$*, where for any* $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$,

$$\mathbb{K}(t) = \alpha(t) + \sum_{j=1}^{m} F'(t_j)\beta_j(t)$$

with

$$\begin{aligned} \beta_j(t) \quad = \quad & \left\{ \prod_{\ell > j} F(t_l) \right\} E\left[\left\{ \frac{\mathcal{M}}{\sigma_j} + \frac{t_j}{\sigma_j^2}\mathcal{A} + \frac{t_j}{\sigma_j^2} \sum_{\ell=1}^{p} \mathcal{B}_\ell (Y_{j-\ell} - \mu)^2 \right.\right. \\ & \qquad \left.\left. -2\frac{t_j}{\sigma_j^2} \sum_{\ell=1}^{p} \mathcal{B}_\ell (Y_{j-\ell} - \mu) \right\} \prod_{\ell < j} \mathbf{1}(\epsilon_\ell \le t_\ell) \right]. \end{aligned}$$

It is clear that β_j *depends on* t_j*, even if* $\mu = 0$*, so Assumption II is not met. As seen before, however, Assumption II holds true when* $\mu = 0$ *and* $\varepsilon_i = \log(|\epsilon_i|)$ *is considered instead of* ϵ_i.

Appendices

A Auxiliary results

This appendix contains a series of lemmas that will be used to prove the main convergence results. Suppose $w_1, \ldots, w_n$ are random vectors in $\mathbb{R}^m$, and for any $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, set

$$K_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(w_i \le x) = \frac{1}{n} \sum_{i=1}^{n} \prod_{k=1}^{m} \mathbf{1}(w_{ik} \le x_k)$$

and

$$F_{k,n}(x_k) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(w_{ik} \le x_k), \quad k = 1, \ldots, m.$$

Assume that K_n is an estimation of an arbitrary distribution function K with continuous margins $F_1, \ldots, F_m$. Then there exists a unique copula C such that for all $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$,

$$K(x_1, \ldots, x_m) = C\left\{F_1(x_1), \ldots, F_m(x_m)\right\}.$$

Accordingly, the so-called empirical copula

$$C_n(u_1, \ldots, u_m) = K_n\left\{F_{1,n}^{-1}(u_1), \ldots, F_{m,n}^{-1}(u_m)\right\}$$

is an estimation of $C(u_1, \ldots, u_m)$ for every $u = (u_1, \ldots, u_m) \in [0,1]^m$.

Further set, for all $x \in \mathbb{R}^m$,

$$K_\star(x) = \prod_{k=1}^m F_k(x_k). \tag{10}$$

Finally, assume that for each $k = 1, \ldots, m$, $w_{1k}, \ldots, w_{nk}$ are mutually distinct with probability one. It is then a simple exercise to show that

$$\sup_{0<u<1}\left|F_k \circ F_{k,n}^{-1}(u) - u\right| = \sup_{0<u<1}\left|F_{k,n} \circ F_k^{-1}(u) - u\right|. \tag{11}$$

This fact is instrumental in establishing the weak convergence of the processes $\mathbb{F}_{k,n} = \sqrt{n}\,(F_{k,n} - F_k)$ and $\mathbb{C}_n = \sqrt{n}\,(C_n - C)$, which is stated next.

Lemma 1 *Suppose that* $\mathbb{K}_n = \sqrt{n}(K_n - K) \rightsquigarrow \mathbb{K}$ *in* $\mathcal{D}([-\infty,\infty]^m)$ *as* $n \to \infty$*. Then also* $\mathbb{F}_{k,n} \rightsquigarrow \mathbb{F}_k$ *in* $\mathcal{D}([-\infty,\infty])$*, where*

$$\mathbb{F}_k(x_k) = \mathbb{K}(\infty, \ldots, \infty, x_k, \infty, \ldots, \infty).$$

Moreover, if C *has continuous derivatives of order one on* $[0,1]^m$*, then* $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ *in* $\mathcal{D}\left([0,1]^m\right)$ *as* $n \to \infty$*, where*

$$\mathbb{C}(u) = \mathbb{K}\left\{F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\right\} - \sum_{k=1}^m \mathbb{F}_k \circ F_k^{-1}(u_k)\frac{\partial}{\partial u_k}C(u), \tag{12}$$

for any $u = (u_1, \ldots, u_m) \in [0,1]^m$*.*

Proof: First, the convergence of $\mathbb{F}_{k,n}$ follows from the convergence of $\mathbb{K}_n$. Next, using (11) and the convergence of $\mathbb{F}_{k,n}$, one can see that for any $k \in \{1, \ldots, m\}$,

$$\sup_{0<u<1}\left|F_k \circ F_{k,n}^{-1}(u) - u\right| \xrightarrow{P} 0 \text{ as } n \to \infty.$$

Consequently,

$$\mathbb{K}_n\left\{F_{1,n}^{-1}(u_1), \ldots, F_{m,n}^{-1}(u_m)\right\} \rightsquigarrow \mathbb{K}\left\{F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\right\} \text{ in } \mathcal{D}\left([0,1]^m\right).$$

Next, note that for any $u = (u_1, \ldots, u_m) \in [0,1]^m$,

$$\begin{aligned}\mathbb{C}_n(u) &= \mathbb{K}_n\left\{F_{1,n}^{-1}(u_1), \ldots, F_{m,n}^{-1}(u_m)\right\} \\ &\quad + \sqrt{n}\,\left[C\left\{F_1 \circ F_{1,n}^{-1}(u_1), \ldots, F_m \circ F_{m,n}^{-1}(u_m)\right\} - C(u)\right].\end{aligned}$$

Also, the same kind of arguments used to show (11) yields the tightness of $\mathbb{Q}_{k,n}(u_k) = \sqrt{n}\,\{F_k \circ F_{k,n}^{-1}(u_k) - u_k\}$ for any $k = 1, \ldots, m$. Furthermore, it is easy to check that the finite-dimensional distributions of $\mathbb{Q}_{k,n}$ converge to those of $-\mathbb{F}_k \circ F_k^{-1}$. Hence, one may conclude that

$$\sqrt{n}\,\left[C\left\{F_1 \circ F_{1,n}^{-1}(u_1), \ldots, F_m \circ F_{m,n}^{-1}(u_m)\right\} - C(u)\right] \rightsquigarrow -\sum_{k=1}^{m} \mathbb{F}_k \circ F_k^{-1}(u_k) \frac{\partial}{\partial u_k} C(u),$$

which completes the proof. ∎

Remark 4 *The conclusion of Lemma 1 is well known to hold in the special case where the observations $w_1, \ldots, w_n$ on which K_n is based form a random sample from an $m-$variate distribution. See, e.g., Gänßler and Stute (1987) or Fermanian et al. (2004). The extension provided here, however, shows that the result remains valid in the more frequent contexts where the w_i are not identically distributed or even serially dependent.*

Let $\mathcal{S}_m$ be the set of all subsets of $\{1, \ldots, m\}$. For $A \in \mathcal{S}_m$, let $t_A \in \mathbb{R}^m$ be such that

$$(t_A)_k = \begin{cases} -t_k & \text{if } k \in A; \\ t_k & \text{if } k \notin A. \end{cases}$$

For any $t \in [0, \infty]^m$, set

$$B_n(t) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{k=1}^{m} \mathbf{1}(|w_{jk} - w_{ik}| \le t_k).$$

Next, for any $h \in \mathcal{D}([-\infty, \infty]^m)$, define the mappings $\psi(h)$ and $\psi^\star(h)$, by

$$\psi(h)(t) = \sum_{A \in S_m} (-1)^{|A|} \int h(x + t_A) dK(x), \tag{13}$$

$$\psi^\star(h)(t) = \sum_{A \in S_m} (-1)^{|A|} \int h(x + t_A) dK_\star(x). \tag{14}$$

It is easy to check that both ψ and $\psi^\star$ are continuous linear mappings from $\mathcal{D}([-\infty,\infty]^m)$ to $\mathcal{D}\left([0,\infty]^m\right)$. Further set $\mathbb{B}_n = \sqrt{n}\,(B_n - B)$, where

$$B(t) = \psi(K)(t) = \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int K(x+t_A)dK(x), \quad t\in[0,\infty]^m.$$

Next, set $\mathbb{B}_n^\star = 2\psi^\star(\mathbb{K}_n)$. Then, using the multinomial formula

$$\prod_{k=1}^m (x_k+y_k) = \sum_{A\subset\mathcal{S}_m} \left(\prod_{k\in A} x_k\right)\times\left(\prod_{j\in\mathcal{S}_m\setminus A} y_j\right),$$

it is easy to check that

$$\mathbb{B}_n^\star(t) = \frac{2}{\sqrt{n}}\sum_{i=1}^n \left[\prod_{k=1}^m \{F_k(w_{ik}+t_k) - F_k(w_{ik}-t_k)\} - \prod_{k=1}^m G_k(t_k)\right], \tag{15}$$

where

$$G_k(t_k) = \int \{F_k(x_k+t_k) - F_k(x_k-t_k)\}\, dF_k(x_k), \quad k=1,\ldots,m.$$

The asymptotic behaviors of $\mathbb{B}_n$ and $\mathbb{B}_n^\star$ are given next.

Proposition 2 *Suppose that $\mathbb{K}_n \rightsquigarrow \mathbb{K}$ in $\mathcal{D}([-\infty,\infty]^m)$ as $n\to\infty$. Then*

$$\sup_{t\in\mathbb{R}^m} |\mathbb{B}_n(t) - 2\psi(\mathbb{K}_n)(t)| \xrightarrow{P} 0 \quad \text{as } n\to\infty.$$

Moreover, $\mathbb{B}_n \rightsquigarrow \mathbb{B} = 2\psi(\mathbb{K})$ and $\mathbb{B}_n^\star \rightsquigarrow \mathbb{B}^\star = 2\psi^\star(\mathbb{K})$ in $\mathcal{D}\left([0,\infty]^m\right)$.

Proof: First, using the weak convergence of $\mathbb{K}_n$ to $\mathbb{K}$, it follows that

$$\sup_{t\in[0,\infty]^m}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{K}_n(w_i+t) - \int \mathbb{K}_n(x+t)dK(x)\right| \xrightarrow{P} 0 \quad \text{as} \quad n\to\infty.$$

Next, an application of the multinomial formula yields

$$\begin{aligned}
B_n(t) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\prod_{k=1}^m \mathbf{1}(|w_{jk}-w_{ik}|\le t_k) \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\prod_{k=1}^m \{\mathbf{1}(w_{ik}\le w_{jk}+t_k) - \mathbf{1}(w_{ik}< w_{jk}-t_k)\} \\
&= \sum_{A\in\mathcal{S}_m}\frac{(-1)^{|A|}}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left\{\prod_{k\in A}^m \mathbf{1}(w_{ik}< w_{jk}-t_k)\right\}
\end{aligned}$$

$$
\begin{aligned}
& \times \prod_{k\in S_m\setminus A}^{m} \mathbf{1}(w_{ik} \le w_{jk} + t_k) \\
= & \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \left\{ \prod_{k\in A}^{m} \mathbf{1}(w_{ik} \le w_{jk} - t_k) \right\} \\
& \times \prod_{k\in\mathcal{S}_m\setminus A}^{m} \mathbf{1}(w_{ik} \le w_{jk} + t_k) + o_P(1) \\
= & \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}(w_i \le w_j + t_A) + o_P(1) \\
= & \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \sum_{j=1}^{n} K_n(w_j + t_A) + o_P(1) \\
= & \frac{1}{\sqrt{n}} \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \sum_{j=1}^{n} \mathbb{K}_n(w_j + t_A) \\
& + \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \sum_{j=1}^{n} K(w_j + t_A) + o_P(1).
\end{aligned}
$$

Next, one has

$$
\sum_{A\in S_m} (-1)^{|A|}\mathbf{1}(x \le w_j + t_A) = \mathbf{1}\{|x - w_j| \le t\} = \sum_{A\in S_m} (-1)^{|A|}\mathbf{1}(w_j \le x + t_A),
$$

almost surely, so that

$$
\begin{aligned}
\sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \sum_{j=1}^{n} K(w_j + t_A) &= \sum_{j=1}^{n} \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \int \mathbf{1}(x \le w_j + t_A) dK(x) \\
&= \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{n} \sum_{j=1}^{n} \int \mathbf{1}(w_j \le x + t_A) dK(x) \\
&= \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int K_n(x + t_A) dK(x) \\
&= \frac{1}{\sqrt{n}} \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int \mathbb{K}_n(x + t_A) dK(x) \\
&\quad + \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int K(x + t_A) dK(x) \\
&= \sum_{A\in\mathcal{S}_m} \frac{(-1)^{|A|}}{\sqrt{n}} \int \mathbb{K}_n(x + t_A) dK(x) + B(t).
\end{aligned}
$$

Hence the following chain of identities holds uniformly in $t \in [0,\infty]^m$:

$$\begin{aligned}
\mathbb{B}_n(t) &= \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \frac{1}{n}\sum_{j=1}^n \mathbb{K}_n(w_j + t_A) \\
&\qquad + \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int \mathbb{K}_n(x+t_A)dK(x) + o_P(1) \\
&= 2\sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int \mathbb{K}_n(x+t_A)dK(x) + o_P(1) \\
&= 2\psi(\mathbb{K}_n)(t) + o_P(1).
\end{aligned}$$

To complete the proof, one may then invoke the continuous mapping theorem to conclude that $\psi(\mathbb{K}_n) \rightsquigarrow \psi(\mathbb{K})$ and $\psi^\star(\mathbb{K}_n) \rightsquigarrow \psi^\star(\mathbb{K})$ in $\mathcal{D}([0,\infty]^m)$. Hence the result. ■

The main result of this Appendix can now be stated.

Theorem 3 *Suppose that the margins $F_1,\ldots,F_m$ of K admit continuous and square integrable derivatives $F_1',\ldots,F_m'$, respectively. Further assume that the copula associated with K is the independence copula and that there exist processes $\alpha,\beta_1,\ldots,\beta_m \in \mathcal{D}([-\infty,\infty]^m)$ with the property that $\beta_k(x)$ does not depend on x_k and such that $\mathbb{K}_n \rightsquigarrow \mathbb{K}$ in $\mathcal{D}([-\infty,\infty]^m)$ as $n\to\infty$, where*

$$\mathbb{K}(x) = \alpha(x) - \sum_{k=1}^m F_k'(x_k)\beta_k(x), \quad x \in [-\infty,\infty]^m.$$

Then

$$\sup_{t\in\mathbb{R}^m} |\mathbb{B}_n(t) - \mathbb{B}_n^\star(t)| \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and $\mathbb{B}_n \rightsquigarrow \mathbb{B} = 2\psi(\alpha)$ in $\mathcal{D}([0,\infty]^m)$.

Proof: First, note that because of the independence hypothesis, $\psi^\star = \psi$. Next, it follows from Proposition 2 that $\sup_{t\in\mathbb{R}^m} |\mathbb{B}_n(t) - \mathbb{B}_n^\star(t)|$ converges in probability to zero and

$$\mathbb{B}_n \rightsquigarrow \mathbb{B} = 2\psi(\mathbb{K}) = 2\psi(\alpha) - 2\sum_{k=1}^m \psi(F_k'\beta_k).$$

Thus to complete the proof, it suffices to show that for any $k=1,\ldots,m$, $\psi(F_k'\beta_k) \equiv 0$. As the argument is the same for all $k\in\{1,\ldots,m\}$, one takes $k=m$ for sake of simplicity. Then for any $t\in[0,\infty]^m$,

$$\psi(F_m'\beta_m)(t) = \sum_{A\in\mathcal{S}_m} (-1)^{|A|} \int F_m'\{x_m + (t_A)_m\}\beta_m(x+t_A)dK(x)$$

$$= \sum_{A \in \mathcal{S}_{m-1}} (-1)^{|A|} \left\{ \int \{F_m'(x_m + t_m) - F_m'(x_m - t_m)\} dF_m(x_m) \right\} \times \int \beta_m(x + t_A) dF_1(x_1) \cdots dF_{m-1}(x_{m-1}),$$

since $\beta_m(x)$ does not depend on x_m. That the whole expression vanishes then follows from the fact that

$$\int \{F_m'(x_m + t_m) - F_m'(x_m - t_m)\} dF_m(x_m) = 0, \tag{16}$$

which is a simple consequence of the square integrability of F_m'. ■

Now for every $k \in \{1, \ldots, m\}$ and $t = (t_1, \ldots, t_m) \in [0, \infty]^m$, set

$$G_{k,n}(t_k) = B_n(\infty, \ldots, \infty, t_k, \infty, \ldots, \infty) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}(|w_{jk} - w_{ik}| \le t_k)$$

and define

$$\mathbb{D}_n(t) = \sqrt{n} \left\{ B_n(t) - \prod_{k=1}^{m} G_{k,n}(t_k) \right\}.$$

As shown below, the weak convergence of the process $\mathbb{D}_n$ is then a consequence of the previous result. Before stating this fact precisely, define for any $k \in \{1, \ldots, m\}$ the mapping ψ_k by

$$\psi_k(h)(t_k) = \int \{h(x_k + t_k) - h(x_k - t_k)\} dF_k(x_k), \quad t_k \in [0, \infty], \quad h \in \mathcal{D}([-\infty, \infty]).$$

Corollary 1 *Suppose that the conditions of Theorem 3 hold true. For every $k \in \{1, \ldots, m\}$ and $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, set*

$$\alpha_k(x_k) = \alpha(\infty, \ldots, \infty, x_k, \infty, \ldots, \infty).$$

Then $\mathbb{D}_n \rightsquigarrow \mathbb{D}$ in $\mathcal{D}([0, \infty]^m)$ as $n \to \infty$, and for all $k \in \{1, \ldots, m\}$, $\mathbb{G}_{k,n} \rightsquigarrow \mathbb{G}_k = 2\psi_k(\alpha_k)$ in $\mathcal{D}([0, \infty])$, where

$$\mathbb{D} = 2\psi(\alpha) - 2 \sum_{k=1}^{m} \psi_k(\alpha_k)(t_k) \prod_{j \neq k} G_j(t_j), \quad t = (t_1, \ldots, t_m) \in [0, \infty]^m.$$

Proof: First note that for every $k \in \{1, \ldots, m\}$ and $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$,

$$\mathbb{F}_{k,n} \rightsquigarrow \mathbb{F}_k(x_k) = \alpha_k(x_k) - F_k'(x_k) \beta_k(\infty, \ldots, \infty)$$

as $n \to \infty$. It follows from Proposition 2 that

$$\mathbb{G}_{k,n}(t_k) = 2\psi_k(\mathbb{F}_{k,n})(t_k) + o_P(1),$$

so

$$\mathbb{G}_{k,n} \rightsquigarrow 2\psi_k(\mathbb{F}_k) = 2\psi_k(\alpha_k) \text{ in } \mathcal{D}([0,\infty]),$$

in view of (16).

To complete the proof, note that, uniformly in $t \in \mathbb{R}_+^m$,

$$\begin{aligned}
\mathbb{D}_n(t) &= \mathbb{B}_n(t) + \sqrt{n}\left\{\prod_{k=1}^m G_{k,n}(t_k) - B(t)\right\} \\
&= \mathbb{B}_n(t) + \sum_{k=1}^m \mathbb{G}_{k,n}(t_k)\left\{\prod_{j\neq k} G_j(t_j)\right\} + o_P(1).
\end{aligned}$$

■

The next result pertains to the weak convergence of rank-based analogues of $\mathbb{D}_n$ and $\mathbb{B}_n$. To this end, define for every $i \in \{1,\ldots,n\}$,

$$\tilde{w}_i = (\tilde{w}_{i1},\ldots,\tilde{w}_{im}) = (F_{1,n}(w_{i1}),\ldots,F_{m,n}(w_{im})).$$

Note that for every $t \in [-\infty,\infty]^m$, one has

$$C_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(\tilde{w}_i \leq t).$$

Suppose that C is the independence copula. For $u \in [0,1]^m$, define

$$\tilde{B}_n(u) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbf{1}\left(||\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_j|| \leq u\right)$$

and define the mappings $\tilde{\psi} : h \in \mathcal{D}([-\infty,\infty]^m) \mapsto \tilde{\psi}(h) \in \mathcal{D}\left([0,1]^m\right)$ and $\tilde{\psi}_1 : h_1 \in \mathcal{D}([-\infty,\infty]) \mapsto \tilde{\psi}_1(h_1) \in \mathcal{D}\left([0,1]\right)$, by

$$\tilde{\psi}(h)(u) = \sum_{A\in\mathcal{S}_m}(-1)^{|A|}\int_{[0,1]^m} h(x+u_A)dx$$

and

$$\tilde{\psi}_1(h_1)(u_1) = \int_0^1 \left\{h_1(s+u_1) - h_1(s-u_1)\right\} ds.$$

Next, set $\tilde{\mathbb{B}}_n^\star = 2\tilde{\psi}(\mathbb{C}_n)$. It follows that for every $u \in [0,1]^m$, one has

$$\begin{aligned}\tilde{\mathbb{B}}_n^\star(u) &= 2\tilde{\psi}(\mathbb{C}_n)(u) = 2\sum_{A\in\mathcal{S}_m}(-1)^{|A|}\int_{[0,1]^m}\mathbb{C}_n(x+u_A)dx \\ &= \frac{2}{\sqrt{n}}\sum_{i=1}^n\left[\prod_{k=1}^m\{\tilde{F}(\tilde{w}_{ik}+u_k)-\tilde{F}(\tilde{w}_{ik}-u_k)\}-\prod_{k=1}^m\tilde{G}(u_k)\right],\end{aligned}$$

with $\tilde{F}(x) = \max\{0, \min(x,1)\} = P(U_1 \le x)$ for every $x \in \mathbb{R}$ and

$$\tilde{G}(s) = P(|U_2 - U_1| \le s) = \psi_1(\tilde{F})(s) = \int_0^1\{\tilde{F}(u+s)-\tilde{F}(u-s)\}du = 2s - s^2,$$

for all $s \in [0,1]$, where $U_1, U_2 \sim \mathcal{U}(0,1)$ are independent. Finally, set $\tilde{\mathbb{B}}_n = \sqrt{n}\,(\tilde{B}_n - C)$.

The stage is now set for the final result of this Appendix.

Corollary 2 *Assume that the conditions of Theorem 3 hold true. For every $k \in \{1, \ldots, m\}$, let α_k be defined as in Corollary 1. For all $u = (u_1, \ldots, u_m) \in [0,1]^m$, set*

$$\tilde{\alpha}_k(u_k) = \alpha_k \circ F_k^{-1}(u_k)$$

and define

$$\tilde{\alpha}(u) = \alpha\left\{F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\right\}.$$

Then

$$\sup_{u\in[0,1]^m}\left|\tilde{\mathbb{B}}_n(u) - \tilde{\mathbb{B}}_n^\star(u)\right| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

and $\tilde{\mathbb{B}}_n \rightsquigarrow \tilde{\mathbb{B}} = 2\tilde{\psi}(\mathbb{C})$ in $\mathcal{D}([0,1]^m)$. Moreover, if $F_1, \ldots, F_m$ are symmetric (not necessarily with respect to the origin), then for all $u = (u_1, \ldots, u_m) \in [0,1]^m$,

$$\tilde{\mathbb{B}}(u) = 2\tilde{\psi}(\tilde{\alpha})(u) - 2\sum_{k=1}^m\tilde{\psi}_1(\tilde{\alpha}_k)(u_k)\prod_{j\neq k}\tilde{G}_j(u_j).$$

Proof: Let $u = (u_1, \ldots, u_m) \in [0,1]^m$ be given. First, note that from Lemma 1, $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ as $n \to \infty$, and using (12), one has

$$\begin{aligned}\mathbb{C}(u) &= \mathbb{K}\left\{F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\right\} - \sum_{k=1}^m\mathbb{F}_k\circ F_k^{-1}(u_k)\frac{\partial}{\partial u_k}C(u) \\ &= \tilde{\alpha}(u) - \sum_{k=1}^m\tilde{\alpha}_k(u_k)\prod_{j\neq k}u_j\end{aligned}$$

$$-\sum_{k=1}^{m} F_k' \circ F_k^{-1}(u_k) \left\{ \tilde\beta_k(u) - \tilde\beta_k(1,\ldots,1) \prod_{j\neq k} u_j \right\},$$

with $\tilde\beta_k(u) = \beta_k\left\{F_1^{-1}(u_1),\ldots,F_m^{-1}(u_m)\right\}$. It then follows from Proposition 2 that

$$\sup_{u\in[0,1]^m} \left|\tilde{\mathbb{B}}_n(u) - \tilde{\mathbb{B}}_n^\star(u)\right| \xrightarrow{P} 0 \quad \text{as } n\to\infty$$

and that $\mathbb{B}_n \rightsquigarrow \mathbb{B} = 2\tilde\psi(\mathbb{C})$ in $\mathcal{D}\left([0,1]^m\right)$.

Since $\tilde\beta_k(u)$ does not depend on u_k, the proof will be complete if one can show that $\tilde\psi_1(F_k' \circ F_k^{-1}) \equiv 0$ on $[0,1]$ for all $k\in\{1,\ldots,m\}$. To this end, note that for all $x\in(0,1)$, one has

$$\begin{aligned}
\tilde\psi_1(F_k' \circ F_k^{-1})(x) &= \int_0^1 \left\{F_k' \circ F_k^{-1}(s+x) - F_k' \circ F_k^{-1}(s-x)\right\} ds \\
&= \int_x^1 F_k' \circ F_k^{-1}(s)ds - \int_0^{1-x} F_k' \circ F_k^{-1}(s)ds \\
&= \int_{F_k^{-1}(x)}^{\infty} \left\{F_k'(y)\right\}^2 dy - \int_{-\infty}^{F_k^{-1}(1-x)} \left\{F_k'(y)\right\}^2 dy.
\end{aligned}$$

Set $Q_k(s) = F_k^{-1}(s)$, $s\in(0,1)$. Then $Q_k'(s) = 1/F_k' \circ F_k^{-1}(s)$, so that $\tilde\psi_1(F_k' \circ F_k^{-1}) \equiv 0$ is satisfied if and only if $Q_k'(s) = Q_k'(1-s)$ for all $s\in(0,1)$, meaning that F_k is symmetric. Hence the result. ■

B Proof of the main results

Since Proposition 1 is a particular case of Theorem 1, only the latter is proved. Further note that although it is convenient to define the last $m-1$ values of $\tilde e_i$ in a circular way as in (3), this does not affect the limiting distribution of any statistic based on $\tilde e_1,\ldots,\tilde e_{n+m-1}$. Therefore, arguments in this section are presented as if a sample of size $n+m-1$ (rather than n) had been collected.

B.1 Proof of Theorem 1

First, set

$$\overset{\circ}{B}_n(t) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{k=1}^{m} \mathbf{1}(|\varepsilon_{j+k-1} - \varepsilon_{i+k-1}| \le t_k), \qquad t\in[0,\infty]^m,$$

and define $\overset{\circ\star}{\mathbb{B}}_n = 2\psi(\alpha_n)$. Using the multinomial formula, one can write

$$\overset{\circ\star}{\mathbb{B}}_n(t) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[\prod_{k=1}^{m}\{F(\varepsilon_{i+k-1}+t_k) - F(\varepsilon_{i+k-1}-t_k)\} - \prod_{k=1}^{m} G(t_k)\right].$$

Next, by hypothesis, $\alpha_n \rightsquigarrow \alpha$ in $\mathcal{D}([-\infty,\infty]^m)$ as $n \to \infty$, so $\overset{\circ}{\mathbb{B}}{}_n^\star = 2\psi(\alpha_n) \rightsquigarrow 2\psi(\alpha) = \mathbb{B}^\star$ and $\overset{\circ}{\mathbb{B}}_n \rightsquigarrow \mathbb{B}^\star$ in $\mathcal{D}([0,\infty]^m)$ as a consequence of Proposition 2. The weak convergence of $\mathbb{B}_n$ to $\mathbb{B} = 2\psi(\alpha)$ then follows directly from Theorem 3.

The formula for the covariance of $\mathbb{B}$ can thus be recovered from the relation $\mathbb{B} = 2\psi(\alpha)$, together with the fact that the limiting covariance of $\overset{\circ}{\mathbb{B}}{}_n^\star$ is the covariance of $\mathbb{B}$. The validity of Algorithm 2 is also a consequence of the weak convergence of $\overset{\circ}{\mathbb{B}}_n$ to $\mathbb{B}$.

Finally, the weak convergence of $\mathbb{D}_n$ follows from Corollary 1. Moreover, one can write

$$\mathbb{D}(t) = 2\psi(\alpha) - 2\sum_{k=1}^m \psi_1(\alpha_1)(t_k) \prod_{j\neq k} G(t_j)$$

for all $t \in [0,\infty]^m$. This is because for every $s \in [0,\infty]$, one has

$$\alpha_1(s) = \alpha(s,\infty,\ldots,\infty) = \alpha_2(s) = \cdots = \alpha_m(s).$$

To complete the proof, one can easily check that for all $s \in [0,\infty]$ and $t \in [0,\infty]^m$,

$$\text{cov}\{\psi_1(\alpha_1)(s), \psi(\alpha)(t)\} = \text{cov}\{\psi_1(\alpha_1)(s), \psi_1(\alpha_1)(t_1)\} = \gamma(s,t_1) - G(s)G(t_1).$$

■

B.2 Proof of Theorem 2

The weak convergence of $\tilde{\mathbb{D}}_n$ and $\tilde{\mathbb{B}}_n$ follows from Corollary 1. The validity of Algorithm 3 is a consequence of the weak convergence of $\tilde{\alpha}_n$ to $\tilde{\alpha}$, together with the fact that $\tilde{\mathbb{B}}$ depends only on $\tilde{\alpha}$. Moreover, since $\tilde{\mathbb{B}}$ has the same form as $\mathbb{D}$ when F and G are replaced by $\tilde{F}$ and $\tilde{G}$, the formula for the covariance given in Theorem 1 remains valid. ■

C Conditions for the validity of Assumption II

This appendix gives precise conditions under which Assumption II holds for models (8) or ARMA models. Consider first the time series model (8). One needs some regularity conditions on F and ϕ. Suppose that the density F' of ε_i is uniformly continuous and that in addition, $\phi(z,\theta)$ is continuously differentiable with respect to θ and $\dot{\phi}(z,\theta) = \nabla_\theta \phi(z,\theta)$ is the d-dimensional row vector.

Assume that for any fixed $\theta \in \mathcal{O} \subset \mathbb{R}^d$,

$$E\left\{\left\|\dot{\phi}(Z_1,\theta)\right\|^2\right\} < \infty$$

and that

$$\lim_{\delta\to 0} E\left\{\sup_{|\theta'-\theta|<\delta} \left\|\dot{\phi}(Z_0,\theta') - \dot{\phi}(Z_0,\theta)\right\|^2\right\} = 0. \tag{17}$$

Suppose that the estimator θ_n of θ is such that $\Theta_n = \sqrt{n}\,(\theta_n - \theta)$ converge in law to some variable Θ. Next, set $e_{i,n} = Y_i - \phi(Z_{i-1}, \theta_n)$, i.e., the $e_{i,n}$ are the residuals. Recall that for all $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$,

$$K_n(t_1, \ldots, t_m) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(e_{i,n} \le t_1, \ldots, e_{i+m-1,n} \le t_m)$$

and

$$\alpha_n(t) = \sqrt{n}\,\left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\,(\varepsilon_i \le t_1, \ldots, \varepsilon_{i+m-1} \le t_m) - K(t) \right\},$$

where $K(t) = \prod_{j=1}^m F(t_j)$. The following lemma is proved at the end of the section.

Lemma 2 *Suppose that* $(\alpha_n, \Theta_n) \rightsquigarrow (\alpha, \Theta)$ *in* $\mathcal{D}([-\infty,\infty]^m) \times \mathbb{R}^d$. *Under the above assumptions,* $\mathbb{K}_n \rightsquigarrow \mathbb{K}$, *where*

$$\mathbb{K}(t) = \alpha(t) + \sum_{j=1}^{d} F'(t_j) \left\{ \prod_{\ell > j} F(t_\ell) \right\} E\left\{ \dot{\phi}(Z_{j-1}, \theta) \prod_{\ell < j} \mathbf{1}(\varepsilon_\ell \le t_\ell) \right\} \Theta.$$

It is then clear from this lemma that Assumption II is satisfied since, for any $j \in \{1, \ldots, m\}$,

$$\beta_j(t) = \left\{ \prod_{\ell > j} F(t_\ell) \right\} E\left\{ \dot{\phi}(Z_{j-1}, \theta) \prod_{\ell < j} \mathbf{1}(\varepsilon_\ell \le t_\ell) \right\} \Theta$$

does not depend on t_j.

Next, consider ARMA(p,q) models of the form

$$Y_i - \mu - \sum_{k=1}^{p} \phi_k (Y_{i-k} - \mu) = \varepsilon_t - \sum_{j=1}^{q} \varphi_j \varepsilon_{i-j}, \quad i \ge 1 \tag{18}$$

where the innovations (ε_i) have mean zero and finite variance σ_ε^2, and the coefficients $\phi = (\phi_1, \ldots, \phi_p)^\top$ and $\varphi = (\varphi_1, \ldots, \varphi_q)^\top$ satisfy the usual conditions, i.e., the (complex) roots of the polynomials $1 - \sum_{k=1}^p \phi_k z^k$ and $1 - \sum_{k=1}^q \varphi_k z^k$ all lies outside the unit circle.

Let $\theta_n = \left(\hat{\mu}_n, \hat{\phi}_n, \hat{\varphi}_n\right)$ denote an estimation of (μ, ϕ, φ), and set $\Theta_n = \sqrt{n}\,(\hat{\theta}_n - \theta)$.

Lemma 3 *Suppose that for the ARMA model* (18),

$$(\alpha_n, \Theta_n) \rightsquigarrow (\alpha, \Theta) \quad in \quad \mathcal{D}([-\infty,\infty]^m) \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q, \tag{19}$$

where (α, Θ) *is a centered Gaussian process. If in addition the density* F' *of* ε_i *is continuous and bounded, then Assumption II is satisfied, i.e.,* $\mathbb{K}_n \rightsquigarrow \mathbb{K}$ *as* $n \to \infty$ *with* $\mathbb{K}$ *having representation* (4).

Lemma 3 follows easily from results in Ghoudi and Rémillard (2006). The conclusion is not surprising anyway, since an ARMA(p, q) model can be seen as an AR(∞) model, and $AR(p)$ models are covered by Lemma 2.

Note that the convergence condition (19) holds true for OLS estimators or MLE estimators provided the density F' is sufficiently smooth.

C.1 Proof of Lemma 2

First note that proving the weak convergence of $\mathbb{K}_n$ on $\mathcal{D}([-\infty,\infty]^m)$ is equivalent to show that $\mathbb{E}_n = \mathbb{K}_n\left(F^{-1},\ldots,F^{-1}\right) \rightsquigarrow \mathbb{E} = \mathbb{K}\left(F^{-1},\ldots,F^{-1}\right)$ on $\mathcal{D}\left([0,1]^m\right)$. For each $i \in \{1,\ldots,n\}$, introduce $U_i = F(\varepsilon_i)$ and $u_{i,n} = F(e_{i,n}) = F\left\{Y_i - \phi(Z_{i-1},\theta_n)\right\}$.

With these new definitions, one has $\mathbb{E}_n = \sqrt{n}(E_n - E)$, where

$$E_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(u_{i,n} \le u_1,\ldots,u_{i+m-1,n} \le u_m) \quad \text{and} \quad E(u) = \prod_{k=1}^{m} u_k$$

for all $u = (u_1,\ldots,u_m) \in [0,1]^m$. Furthermore,

$$\tilde{\alpha}_n(u) = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(U_i \le u) - \prod_{j=1}^{m} u_j\right\}$$

and

$$\mathbb{E}(u) = \tilde{\alpha}(u) + \sum_{j=1}^{d} F' \circ F^{-1}(u_j)\left(\prod_{\ell>j} u_\ell\right) E\left\{\dot{\phi}(Z_{j-1},\theta)\prod_{\ell<j}\mathbf{1}(U_\ell \le u_\ell)\right\}\Theta, \tag{20}$$

where $\tilde{\alpha} = \alpha\left(F^{-1},\ldots,F^{-1}\right)$ and $\dot{\phi}(z,\theta) = \nabla_\theta \phi(z,\theta)$ is a d–dimensional row vector.

The proof uses the asymptotic theory of empirical processes based on pseudo-observations developed by Ghoudi and Rémillard (2004). In particular, the convergence of $\mathbb{E}_n$ will follow from their Theorem 2.4, once its assumptions have been checked. To simplify this operation, it will be convenient to cast the problem in their notation. To do so, set

$$X_i = (Y_i, Z_{i-1}, Y_{i+1}, Z_i, \ldots, Y_{i+m-1}, Z_{i+m-2}) \in \mathfrak{X} = \left([-\infty,\infty]^{1+p}\right)^{\otimes m},$$

and write $\epsilon_i = (U_i,\ldots,U_{i+m-1})$ for every integer $i \ge 1$. Further set $X = X_1$ and $\epsilon = \epsilon_1$. Next, for all $x = (y_1, z_0, \ldots, y_m, z_{m-1}) \in \mathfrak{X}$, define

$$\mathbb{H}_n(x) = (\mathbb{H}_n^{(1)}(x),\ldots,\mathbb{H}_n^{(m)}(x)),$$

where for any $j \in \{1,\ldots,m\}$, $\mathbb{H}_n^{(j)}(x) = \sqrt{n}\left\{H_n^{(j)}(x) - H^{(j)}(x)\right\}$ with

$$H^{(j)}(x) = F\left\{y_j - \phi(z_{j-1},\theta)\right\} \quad \text{and} \quad H_n^{(j)}(x) = F\left\{y_j - \phi(z_{j-1},\theta_n)\right\}.$$

Now, set $r = \left(r^{(1)}, \ldots, r^{(m)}\right)$ where for $j \in \{1, \ldots, m\}$, and any $x \in \mathfrak{X}$,

$$r^{(j)}(x) = 1 + \|\dot{\phi}(z_{j-1}, \theta)\| + \varphi(z_{j-1}, \delta_0),$$

where δ_0 is such that $\{\theta' : \|\theta' - \theta\| < \delta_0\} \subset \mathcal{O}$, and

$$\varphi(x, \delta) = \sup_{|\theta' - \theta| < \delta} \left\| \dot{\phi}(Z_0, \theta') - \dot{\phi}(Z_0, \theta) \right\|, \quad (x, \delta) \in \mathfrak{X} \times [0, \delta_0].$$

Finally, let $\mathcal{C}_r$ be the set of all $\mathbb{R}^m$-valued functions on $\mathfrak{X} \times \mathbb{R}^d$ such that

$$f^{(j)}(x, a) = -F' \left\{y_j - \phi(z_{j-1}, \theta)\right\} \dot{\phi}(z_{j-1}, \theta) a, \quad 1 \le j \le m.$$

Observe that since F' is uniformly continuous, there exists a non-decreasing bounded function c on $[0, \infty)$ such that $c(0) = 0$ and such that

$$\left|F(x) - F(y) - (x - y)F'(y)\right| \le |x - y| \; c(|x - y|).$$

Next,

$$\sqrt{n} \; |\phi(z, \theta_n) - \phi(z, \theta) - \nabla_\theta(z, \theta)\Theta_n| \le \|\Theta_n\| \varphi\left(z, \|\theta_n - \theta\|\right).$$

Setting $d_n^{(j)}(x) = \phi(z_{j-1}, \theta_n) - \phi(z_{j-1}, \theta)$, for any $j \in \{1, \ldots, m\}$, it follows that

$$\begin{aligned} \left|\mathbb{H}_n^{(j)}(x) - f^{(j)}(x, \Theta_n)\right| \quad \le \quad & \sqrt{n} \, |d_n^{(j)}(x)| \; c\left\{|d_n^{(j)}(x)|\right\} \\ & + \|\Theta_n\| \varphi\left(z_{j-1}, \|\theta_n - \theta\|\right), \end{aligned}$$

so, an $n \to \infty$,

$$\sup_x \left|\mathbb{H}_n^{(j)}(x) - f^{(j)}(x, \Theta_n)\right| \Big/ r^{(j)}(x) \xrightarrow{P} 0,$$

using the weak convergence of Θ_n and condition (17).

Therefore, if $(\alpha_n, \Theta_n) \rightsquigarrow (\alpha, \Theta)$ in $\mathcal{D}([-\infty, \infty]^m) \times \mathbb{R}^d$, and given that condition (17) is satisfied, then $(\tilde{\alpha}_n, \mathbb{H}_n)$ converges in $\mathcal{D}([0, 1]^m) \times \mathcal{D}([-\infty, \infty]^p)$ to $(\tilde{\alpha}, \mathbb{H})$, where

$$\mathbb{H}^{(j)}(x) = -f^{(j)}(x, \Theta), \quad j \in \{1, \ldots, m\}.$$

In view of the above, and given that $E\left\{\|r(X)\|^2\right\} < \infty$ by hypothesis, Lemma 7.2 of Ghoudi and Rémillard (2004) now implies that Hypothesis II of their paper is verified.

Next, for $f \in \mathcal{C}_r$, it follows that for $j \in \{1, \ldots, m\}$, $\mu_j\left\{u, f^{(j)}(\cdot, a)\right\}$ is given by

$$\mu_j\left\{u, f^{(j)}(\cdot, a)\right\} = F' \circ F^{-1}(u_j) \left\{\prod_{\ell > j} u_\ell\right\} E\left\{\dot{\phi}(Z_{j-1}, \theta) \prod_{\ell < j} \mathbf{1}(U_\ell \le u_\ell)\right\} a.$$

Hence Hypothesis I is also verified and because the density of U_i is uniform on $[0, 1]$, Hypothesis III is not needed. Thus one may conclude that $\mathbb{E}_n \rightsquigarrow \tilde{\mathbb{E}}$ in $\mathcal{D}([0, 1]^m)$, where $\mathbb{E}$ has representation (20). ■

References

J. Bai. Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.*, 22:2051–2061, 1994.

István Berkes and Lajos Horváth. Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Process. Appl.*, 105:271–298, 2003.

M. Bilodeau and P. Lafaye de Micheaux. A multivariate empirical characteristic function test of independence with normal marginals. *J. Multivariate Anal.*, 95:345–369, 2005.

G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis*. Prentice Hall Inc., Englewood Cliffs, NJ, third edition, 1994. Forecasting and control.

W. A. Brock, W. D. Dechert, B. LeBaron, and J. A. Scheinkman. A test for independence based on the correlation dimension. *Econom. Rev.*, 15:197–235, 1996.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

Sangit Chatterjee and Mustafa R. Yilmaz. Chaos, fractals and statistics. *Statist. Sci.*, 7: 49–68, 1992.

M. A. Delgado. Testing serial independence using the sample distribution function. *J. Time Series Anal.*, 17:271–285, 1996.

J.-M. Dufour and R. Roy. Some robust exact results on sample autocorrelations and tests of randomness. *J. Econom.*, 29:257–273, 1985.

T. S. Ferguson, C. Genest, and M. Hallin. Kendall's tau for serial dependence. *Canad. J. Statist.*, 28:587–604, 2000.

Jean-David Fermanian, Dragan Radulović, and Marten J. Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860, 2004.

P. Gänßler and W. Stute. *Seminar on Empirical Processes*, volume 9 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1987.

Christian Genest and Bruno Rémillard. Tests of independence or randomness based on the empirical copula process. *Test*, 13:335–369, 2004.

Christian Genest, Jean-François Quessy, and Bruno Rémillard. Tests of serial independence based on Kendall's process. *Canad. J. Statist.*, 30:441–461, 2002.

Kilani Ghoudi and Bruno Rémillard. Empirical processes based on pseudo-observations. In *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)*, pages 171–197. North-Holland, Amsterdam, 1998.

Kilani Ghoudi and Bruno Rémillard. Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic Methods in Stochastics*, volume 44 of *Fields Inst. Commun.*, pages 381–406. Amer. Math. Soc., Providence, RI, 2004.

Kilani Ghoudi and Bruno Rémillard. Empirical distribution functions and copulas based on residuals of ARMA models. *Les Cahiers du GERAD* G–2006–29, HEC Montréal, 2006.

Kilani Ghoudi, Reg J. Kulperger, and Bruno Rémillard. A nonparametric test of serial independence for time series and residuals. *J. Multivariate Anal.*, 79:191–218, 2001.

Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Phys. D*, 9:189–208, 1983.

M. Hallin and M. L. Puri. Rank tests for time series analysis: A survey. In D. R. Brillinger, E. Parzen, and M. Rosenblatt, editors, *New Directions in Time Series Analysis, Part I*, pages 111–153. Springer, New York, 1992.

M. Hallin, J.-F. Ingenbleek, and M. L. Puri. Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.*, 13:1156–1181, 1985.

M. Hallin, J.-F. Ingenbleek, and M. L. Puri. Linear and quadratic serial rank tests for randomness against serial dependence. *J. Time Series Anal.*, 8:409–424, 1987.

Y. Hong. Generalized spectral tests for serial dependence. *J. Royal Statist. Soc., Series B*, 62:557–574, 2000.

Y. Hong and H. White. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica*, 73:837–901, 2005.

Yongmiao Hong. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J. Amer. Statist. Assoc.*, 94:1201–1220, 1999.

G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:297–303, 1978.

P. A. P. Moran. Some theorems on time series II: The significance of the serial correlation coefficient. *Biometrika*, 35:255–260, 1948.

P. M. Robinson. Consistent nonparametric testing entropy-based testing. *Review of Economic Studies*, 58:437–453, 1991.

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

H. Tong and K. S. Lim. Threshold autoregression,limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B.*, 42:245–292, 1980.