MEASURING NONLINEAR DEPENDENCE IN TIME SERIES, A DISTANCE CORRELATION APPROACH

By Zhou Zhou[1]

University of Toronto

September 15, 2014

Abstract

We extend the concept of distance correlation of Szekely, Rizzo and Bakirov (The Annals of Statistics, 2007) and propose the auto distance correlation function (ADCF) to measure the temporal dependence structure of time series. Unlike the classic measures of correlations such as the autocorrelation function, the proposed measure is zero if and only if the measured time series components are independent. In this paper, we propose and theoretically verify a subsampling methodology for the inference of sample ADCF for dependent data. Our methodology provides a useful tool for exploring nonlinear dependence structures in time series.

1 Introduction

Measuring the temporal dependence structure is of fundamental importance in time series analysis. The autocorrelation function (ACF) (Box and Jenkins, 1970), which measures the strength of *linear* dependencies in time series, has been one of the primary tools for exploring and testing time series dependence for many decades. For a univariate time series, the ACF measures the Pearson correlation between observations of the series. In particular, the ACF equals 0 when the measured observations are not at all linearly related and it achieves its maximum absolute value 1 if measured observations are perfectly linearly related.

[1]Corresponding author. Department of Statistics, 100 St. George St., Toronto, Ontario, M5S 3G3 Canada

E-mail: zhou@utstat.toronto.edu

Key words and phrases. Auto distance correlation function, nonlinear time series, nonlinear dependence, subsampling.

Nevertheless, the traditional linear time series models are inadequate in explaining many phenomenons observed in contemporary time series, such as volatility clustering, asymmetric cycles, extreme value dependence, time irreversibility, bimodality and mean reverting, among others. This leads to a recent surge in nonlinear time series research. Prominent examples include Brillinger (1977), Engle (1982), Ashley et. al (1986), Priestley (1988), Tong (1990), Chen and Tsay (1993), Franses and van Dijk (2000), Fan and Yao (2003), Kantz and Schreiber (2000), Tjøstheim and Auestad (1994a, b) and Brockwell (2007), among others.

Despite the fast advancement in nonlinear time series model building, on the other hand, there have been few works on how to explore and measure the complex dependence structures in nonlinear time series. Exceptions include, among others, Bagnato et. al (2011) who considered a graphical device of displaying general dependence structure in time series using the χ^2 tests of independence in contingency tables. It seems that nowadays the Pearson correlation related quantities, which measure the amount of linear dependence, are frequently used to explore the temporal dependence structures in nonlinear time series.

Recently Szekely, Rizzo and Bakirov (2007) (SRB hereafter) proposed the distance correlation to measure and test linear and nonlinear dependence between two samples each composed of iid observations; see also Szekely and Rizzo (2009). The purpose of this paper is to extend the concept of distance correlation into temporally dependent data and analyze a corresponding measure of time series dependence, the auto distance correlation function (ADCF), to explore and test nonlinear dependence structures in time series.

The distinct feature of ADCF is that the theoretical ADCF equals 0 if and only if the measured time series components are independent. The latter feature implies that theoretical ADCF is capable of measuring all forms of departures from independence. In particular, the ADCF is capable of digging out complex nonlinear dependence structures which are buried under the Pearson correlation related measures. For instance, it is well known that many financial return time series show no signs of Pearson correlation but the squared returns exhibits strong dependence. Hence the latter financial returns are strongly nonlinearly related. The ACF plot of the original return series would show no interesting signal at all. However, the ADCF of the original series usually shows a strong signal of dependence, which correctly reflects the underlying nonlinear dynamics of the return series. See also Section 6.1 for a detailed discussion of the SP500 monthly excess return data.

Time series with heavy tailed marginal distributions are frequently observed in areas such as finance, hydrology and telecommunications. For instance, Rachev and Mittnik (2000) did an extensive empirical study and found that the tail indices of many high-frequency financial return data are between 1 and 2. It is difficult and sometimes misleading to apply Pearson correlation related measures to such time series; see for instance Figure 1 in Section 3 and Davis and Mikosch (1998). On the other hand, however, the ADCF is defined for multivariate time series with finite first moment. In Section 4, we prove that the sample ADCF is consistent as long as the time series has $(1+r)$th moment for some $r > 0$. Therefore one extra nice feature of the ADCF is that it can be used to measure the strength of dependence for many heavy tailed multivariate time series.

Just as the sample ACF lets the data 'speak for themselves' and provides a first step in the analysis of linear time series, the sample ADCF is a nonparametric measure which allows one to visually explore the pattern of the strength of nonlinear dependence in stationary series and provides guidance for subsequent parametric or semi-parametric analysis. Specifically, three major uses of the ADCF are as follows. First, the sample ADCF displays the existence/nonexistence of temporal dependence of any kind which can be used to select the time lag of nonlinear time series models and check the adequacy of certain parametric nonlinear models by investigating the independence of its residuals. Second, the pattern of sample ADCF, such as speed of decay or periodicity, can be used to do a preliminary selection of nonlinear time series models: a model whose ADCF does not follow the latter pattern should be removed from consideration. Third, cut-off of ADCF at a small lag directly suggests a nonlinear m-dependent model; see Section 3.2 for more details.

One of the most important tasks in the study of ADCF is to perform finite sample inference. The presence of temporal dependence invalidates the permutation test in SRB for the latter purpose. In this paper, we propose a subsampling methodology for the inference of the ADCF under dependence. Utilizing the contemporary martingale theory and empirical process theory, we are able to control the uniform oscillation rate of the empirical characteristic function of dependent data and consequently theoretically verify the latter subsampling procedure for weakly dependent data. In fact, our theoretical results can be generalized to facilitate the asymptotic theory of a class of V-statistics under dependence. Simulation studies in Section 5.2 shows satisfactory performance of the subsampling procedure for low dimensional time series.

The rest of the paper is structured as follows. Section 2 defines the ADCF function and its sample version. Section 3 discusses the properties of the ADCF for multivariate linear time series models, multivariate nonlinear moving average processes and multivariate nonlinear auto regressive processes. Comparisons between the ADCF and the ACF for the latter processes will also be performed in Section 3. In Section 4, we will establish the consistency of the sample ADCF and obtain the asymptotic null distribution of the statistic for testing ADCF$=0$. In Section 5, a subsampling method is proposed for finite sample inference of the ADCF. A small simulation study on the accuracy of the finite sample ADCF test is conducted in Section 5.2. The SP500 data and the Canadian lynx data are analyzed in Section 6. Some previously unfound dependence signals are uncovered from our data analysis. Finally the asymptotic results are proved in Section 7.

2 Definitions of ADCF

Throughout this paper the Euclidean norm of $x \in \mathbb{R}^p$ is denoted by $|x|$. The inner product on $\mathbb{R}^p$ is denoted $\langle \cdot, \cdot \rangle$. For a complex valued function $f(\cdot)$, the complex conjugate of f is denoted by $\bar{f}$ and $|f|^2 = f\bar{f}$. Denote by f_X the characteristic function of a random vector X and the joint characteristic function of random vectors X and Y is denoted $f_{X,Y}$. The symbol i denotes the imaginary unit with $i^2 = -1$.

Motivated by the notion of distance correlation in SRB, for a strictly stationary time series $\{X_j\}$, the auto distance correlation function (ADCF) at order k, $k \geq 0$ is defined as the distance correlation between X_j and X_{j+k}. For the completeness of the paper, we briefly introduce the definitions and notation in this section.

Let $\{X_j\}_{j=-\infty}^{\infty}$ be a strictly stationary multivariate time series of dimension p and we observe $\{X_j\}_{j=1}^{n}$. Throughout this paper we assume that $p \ll n$ and therefore it is reasonable to assume that p is fixed and do not vary with the series length n. It is well known that X_j and X_{j+k} are independent if and only if $f_{X_j,X_{j+k}}(t,s) = f_{X_j}(t) f_{X_{j+k}}(s)$ for almost all t and s in $\mathbb{R}^p$. The latter property naturally leads to the following definition of auto distance covariance between X_j and X_{j+k}.

Definition 1 (Auto Distance Covariance). *For $k \geq 0$, the auto distance covariance and* *sample auto distance covariance between X_j and X_{j+k} are defined as*

$$\mathcal{V}_X(k) = \frac{1}{c_p^2} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \frac{|f_{X_j, X_{j+k}}(t,s) - f_{X_j}(t) f_{X_{j+k}}(s)|^2}{|t|^{p+1}|s|^{p+1}} dtds \text{ and} \tag{1}$$

$$\mathcal{V}_X^n(k) = \frac{1}{c_p^2} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \frac{|f_k^n(t,s) - f^n(t) f^{n,k}(s)|^2}{|t|^{p+1}|s|^{p+1}} dtds \tag{2}$$

respectively, where $c_p = \pi^{(1+p)/2}/\Gamma((1+p)/2)$,

$$f_k^n(t,s) = \frac{1}{n-k} \sum_{j=1}^{n-k} \exp\{i\langle t, X_j\rangle + i\langle s, X_{j+k}\rangle\}$$

is the empirical characteristic function of $\{(X_i, X_{i+k})\}$ and

$$f^n(t) = \sum_{j=1}^{n-k} \exp\{i\langle t, X_j\rangle\}/(n-k), \quad f^{n,k}(t) = \sum_{j=1}^{n-k} \exp\{i\langle t, X_{j+k}\rangle\}/(n-k)$$

are the marginal empirical characteristic functions of $\{X_j\}_{j=1}^{n-k}$ and $\{X_j\}_{j=k+1}^{n}$, respectively. We call $\mathcal{V}_X(0)$ and $\mathcal{V}_X^n(0)$ the distance variance and sample distance variance of (X_j), respectively.

It is clear from Definition 1 that $\mathcal{V}_X(k)$ equals 0 if and only if X_j and X_{j+k} are independent. The quantity $\mathcal{V}_X(k)$ equals the distance covariance between X_j and X_{j+k} in SRB.

Definition 2 (Auto Distance Correlation)**.** *For $k \geq 1$,define the auto distance correlation between X_j and X_{j+k}*

$$\mathcal{R}_X(k) = \sqrt{\frac{\mathcal{V}_X(k)}{\mathcal{V}_X(0)}}, \quad \text{if} \quad \mathcal{V}_X(0) \neq 0; \tag{3}$$

Otherwise let $\mathcal{R}_X(k) = 0$. The sample auto distance correlation between X_j and X_{j+k} is the nonnegative number $\mathcal{R}_X^n(k)$ defined by

$$[\mathcal{R}_X^n(k)]^2 = \frac{\mathcal{V}_X^n(k)}{\sqrt{\mathcal{V}_Y^n(0)\mathcal{V}_{X^*}^n(0)}}, \quad \text{if} \quad \mathcal{V}_Y^n(0)\mathcal{V}_{X^*}^n(0) \neq 0, \tag{4}$$

where $\mathcal{V}_Y^n(0)$ is the sample distance variance of $\{Y_j\}_{j=1}^{n-k}$, $Y_j = X_{j+k}$ and $\mathcal{V}_{X^*}^n(0)$ is the sample distance variance of $\{X_j\}_{j=1}^{n-k}$. Otherwise let $\mathcal{R}_X^n(k) = 0$.

It can be shown that the auto distance covariance, the auto distance correlation are well defined as long as X_j has finite first moment. Additionally, the theoretical ADCF achieves its minimum 0 if and only if X_j and X_{j+k} are independent. And the theoretical ADCF achieves its maximum 1 if X_j and X_{j+k} are perfectly linearly related by an orthogonal matrix. We refer to SRB and Szekely and Rizzo (2009) for more details on the properties of the distance correlation.

Recall that $Y_j = X_{j+k}$ for $j = 1, 2, \cdots, n-k$. Let $a_{rl} = |X_r - X_l|$ and $b_{rl} = |Y_r - Y_l|$. Define

$$\bar{a}_{r\cdot} = \frac{\sum_{l=1}^{n-k} a_{rl}}{n-k}, \quad \bar{a}_{\cdot l} = \frac{\sum_{r=1}^{n-k} a_{rl}}{n-k}, \quad \bar{a}_{\cdot\cdot} = \frac{\sum_{r,l=1}^{n-k} a_{rl}}{(n-k)^2}, \quad A_{rl} = a_{rl} - \bar{a}_{r\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}. \tag{5}$$

Define $\bar{b}_{r\cdot}$, $\bar{b}_{\cdot l}$, $\bar{b}_{\cdot\cdot}$ and B_{rl} similarly.

Proposition 1.

$$\mathcal{V}_X^n(k) = \frac{1}{(n-k)^2} \sum_{r,l=1}^{n-k} A_{rl} B_{rl}. \tag{6}$$

Proposition 1 is Theorem 1 of SRB with the second sample therein replaced by the lagged observation of the time series. We restate it here because of its fundamental importance in the theory of ADCF. Proposition 1 shows a very interesting algebraic equality which greatly reduces the computation time of $\mathcal{V}_X^n(k)$. Note that the original definition of $\mathcal{V}_X^n(k)$ in (2) involves evaluating an integration on $\mathbb{R}^{2p}$. For a moderately large p such as $p = 5$, direct numerical computation of such integration is formidable. On the other hand, however, Proposition 1 claims that $\mathcal{V}_X^n(k)$ equals $\sum_{r,l=1}^{n-k} A_{rl} B_{rl}/(n-k)^2$, which can be calculated easily with an $O(n^2)$ time complexity.

2.1 Dependence measures

Throughout this paper, we shall assume that the the stationary process $\{X_j\}$ admits the following representation

$$X_j = G(\cdots, \varepsilon_{j-1}, \varepsilon_j), \tag{7}$$

where ε_j, $j \in \mathbb{Z}$, are independent and identically distributed (iid) random variables and $G : \mathbb{R}^\infty \to \mathbb{R}^p$ is a function such that $G(\cdots, \varepsilon_{j-1}, \varepsilon_j)$ converges to an appropriate random vector. The representation (7) can be viewed as a physical system with (ε_j) being the inputs or shocks, (X_j) being the outputs and G being the filter that represents the underlying data generating mechanism. Model (7) covers a very wide range of linear and nonlinear time series models encountered in practice; See Wu (2005) for more discussions and examples.

Define the shift process $\mathcal{F}_j = (\cdots, \varepsilon_{j-1}, \varepsilon_j)$. Let $\{(\varepsilon'_j)\}_{j\in\mathbb{Z}}$ be an iid copy of $\{(\varepsilon_j)\}_{j\in\mathbb{Z}}$. For $j \geq 0$, let the coupled process $\mathcal{F}^*_j = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \cdots, \varepsilon_j)$. Define physical dependence measures for the stationary time series $\{X_j\}$ as follows

Definition 3 (Physical dependence measures)**.** *Assume that* $\|X_j\|_q < \infty$ *with* $q > 0$*, where* $\|\cdot\|_q := [\mathbb{E}(|\cdot|^q)]^{1/q}$*. Define physical dependence measures*

$$\delta(k, q) = \|G(\mathcal{F}_k) - G(\mathcal{F}^*_k)\|_q, \quad k \geq 0. \tag{8}$$

Let $\delta(k, q) = 0$ *if* $k < 0$*. Additionally, write* $\|\cdot\| := \|\cdot\|_2$*.*

Calibrating the idea of coupling, $\delta(k, q)$ measures the dependence of X_k on the input ε_0. The above dependence measures are closely related to the data generating mechanism and therefore are easy to work with theoretically. Wu (2005) contains detailed calculations of $\delta(k, q)$ for a very general class of linear and nonlinear time series models. All the asymptotic results of this paper will be expressed in terms of the physical dependence measures.

Another popular class of time series dependence measures is the mixing coefficients. See for instance Rosenblatt (1956). Generally speaking, the mixing coefficient at lag k measures the strength of dependence between the σ-fields generated by time series observations at least k steps away and a fast decay of the mixing coefficients suggests short range dependence of the series. The mixing coefficients and the physical dependence measures are defined in a purely mathematical fashion and are targeted at theoretical or asymptotic investigations of time series. However, the latter dependence measures generally do not have corresponding sample versions and therefore are not suitable for the purpose of exploring and testing dependence structures of time series.

Proposition 2. *Assume* $\|X_i\| < \infty$ *and* $\delta(j, 2) \to 0$ *as* $j \to \infty$*. Then for any* r *such that* $0 < r < 1$*, we have*

$$\mathcal{V}_X(k) \leq C_0 \sum_{j=0}^{\infty} \{\delta(j, 2)\delta(k + j, 2)\}^{1-r}, \tag{9}$$

where C_0 is a finite constant that only depends on p and r. In particular, if $\delta(j,2) = O(j^{-\beta})$ for some $\beta > 1$, then $\mathcal{V}_X(k) = O(k^{-\beta(1-r)})$ for any $r \in (0, 1-1/\beta)$. And if $\delta(j,2) = O(\chi^j)$ for some $0 < \chi < 1$, then $\mathcal{V}_X(k) = O(\chi_1^k)$ for some $0 < \chi_1 < 1$.

Proposition 2 is proved in Section 7. Proposition 2 shows that the auto distance covariances can be bounded by the physical dependence measures. In particular, if $\delta(j,2)$ decays algebraically (geometrically) to 0, then $\mathcal{V}_X(k)$ also decays algebraically (geometrically) to 0. Generally, however, the physical dependence measures cannot be bounded by the auto distance covariances. As an illustrative example, let us consider the ARMA$(1,1)$ model $(1 - B/2)X_j = (1 - 2B)a_j$, where a_j's are iid standard normal and B denotes the back shift operator. Elementary calculations show that $\mathcal{V}_X(k) = 0$ for all $k > 0$. However, the physical dependence measures are positive at all lags.

3 Examples

3.1 Vector linear models

Consider the vector linear model

$$X_t = A^* + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \tag{10}$$

where ε_j's are iid length-p random vectors with $\mathbb{E}\varepsilon_0 = 0$ and $\mathbb{E}|\varepsilon_0| < \infty$, and A_j's are $p \times p$ coefficient matrices with $\sum_{j=0}^{\infty} |A_j| < \infty$. It is easy to see that $\delta(j,q) = O(|A_j|)$ for all $j \geq 0$ and $q \geq 1$. By Proposition 2, if $|A_j|$ decays algebraically (geometrically) to 0 and $\|\varepsilon_0\| < \infty$, then $\mathcal{V}_X(k)$ and $\mathcal{R}_X(k)$ also decay algebraically (geometrically) to 0. In particular, consider the following vector ARMA(h,q) model:

$$(I - \phi(B))X_t = A^* + (I - \theta(B))\varepsilon_t, \tag{11}$$

where $\phi(B) = \sum_{j=1}^{h} \phi_j B^j$, $\theta(B) = \sum_{j=1}^{q} \theta_j B^j$ and ϕ_j and θ_j are $p \times p$ coefficient matrices. It is easy to see that (11) can be represented in the form of (10) with A_j decaying exponentially if the roots of $|I - \phi(B)|$ are all outside of the unit circle. Therefore for model (11), $\mathcal{V}_X(k)$ and $\mathcal{R}_X(k)$ decay exponentially as well.

We shall first compare the performance of ADCF and ACF for heavy tailed univariate processes. Figure 1 below shows the sample ADCF and sample ACF for model (11) with $p = 1$, $\phi(B) = 0.5B$, $\theta(B) = 2B$ and ε_j's iid $t(2)$. The length of the series is 300. In the upper panel of Figure 1, the dotted (solid) vertical lines are the true (sample) ADCF of the model and the dotted horizontal curve are the critical values of testing $\mathcal{R}_X(k) = 0$ at 95% level. The critical values are obtained by subsampling which will be discussed in detail in Section 5.1. Upper panel of Figure 1 clearly shows the exponential decay of the sample ADCF. Note that, if $\mathbb{E}(\varepsilon_t^2) < \infty$, then X_t is a white noise. Even though the $t(2)$ distribution has infinite variance, the sample ACF behaves as if the series is white noise and fails to capture the dependence structure of the process.

Let $H(p)$ denote the $p \times p$ matrix with $H_{j,j}(p) = 0.6$ and $H_{j,j-1}(p) = 0.2$, where $H_{j,k}(p)$ denotes the (j,k)th entry of $H(p)$. Lower panel of Figure 2 shows the sample ADCF of model (11) with $p = 5$, $\phi(B) = H(5)B$, $\theta(B) = 0$ and ε_j's iid standard 5-dimensional Gaussian. Upper panel of Figure 2 shows the sample ADCF of model (11) with $p = 5$, $\phi(B) = 0$, $\theta(B) = H(5)B$ and ε_j's iid standard 5-dimensional Gaussian. The series lengths are both 300. Figure 2 shows that the ADCF is a very visual-friendly and neat way of summarizing the dependence structure of multivariate time series: The ADCF for the vector AR process decays exponentially and the ADCF for the vector MA process cuts off at lag 1.

3.2 Vector nonlinear moving average processes

We call a multivariate time series (X_t) a nonlinear moving average process of order m (NMA(m)) if it admits the following representation:

$$X_j = W(\varepsilon_{j-m}, \varepsilon_{j-m+1}, \cdots, \varepsilon_j), \tag{12}$$

where W is a (possibly) nonlinear function. Clearly a NMA(m) process is m-dependent. Therefore the ADCF of X_t cuts off at lag m. In fact, if the ADCF of a process (X_t) cuts off at a small lag m, then the NMA(m) model is a good candidate for the process. This is analogous to the linear process case where if the ACF cuts off at lag m, then a MA(m) process should be considered.

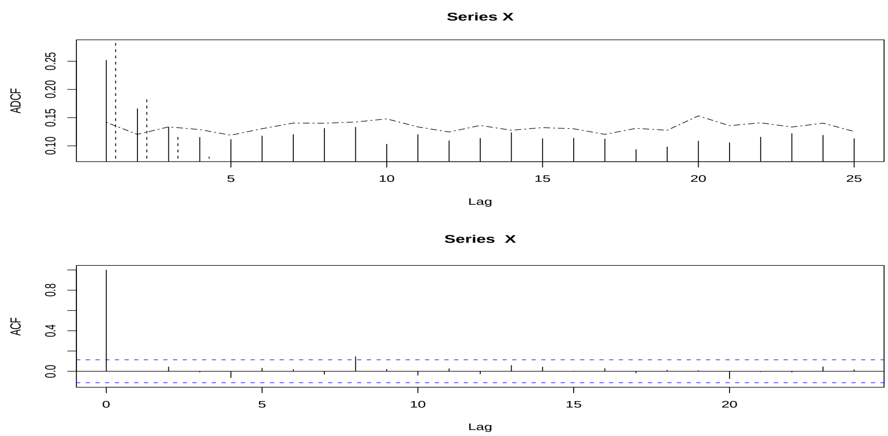

Figure 1: Comparison of the sample ADCF and sample ACF for univariate ARMA$(1,1)$ model : $(1-0.5B)X_j = (1-2B)\varepsilon_j$ with ε_j's iid $t(2)$ and $n = 300$. The dotted (solid) vertical lines are the true (sample) ADCF of the model and the dotted horizontal curve are the critical values at 95% level.

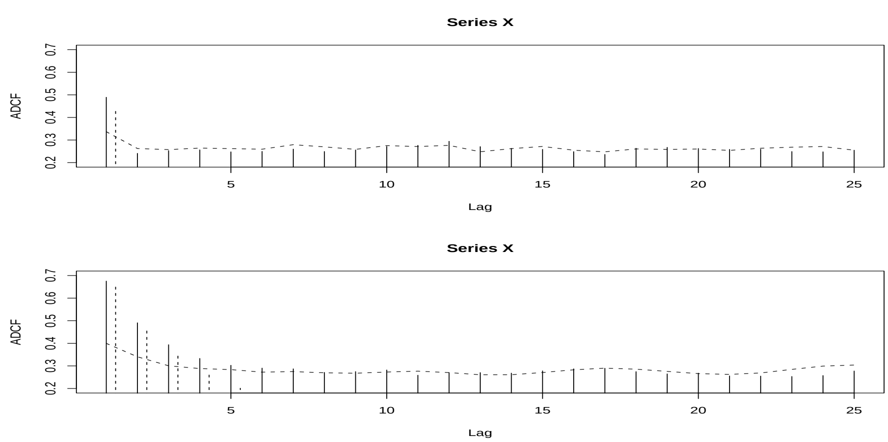

Figure 2: The sample ADCF for vector linear models $(1-H(5)B)X_j = \varepsilon_j$ (lower panel) and $X_j = (1-H(5)B)\varepsilon_j$ (upper panel) with ε_j's iid 5-dimensional standard Gaussian and $n = 300$. The dotted (solid) vertical lines are the true (sample) ADCF of the model and the dotted horizontal curve are the critical values at 95% level.

As an illustrative example, let us consider the following NMA(m) process

$$X_j = \prod_{k=0}^{m} \varepsilon_{j-k}, \tag{13}$$

where ε_k's are iid p vectors with mean 0 and the multiplication in (13) is coordinate-wise. If $\|\varepsilon_j\| < \infty$, then (13) is a multivariate white noise process and thus the ACF cannot reveal the underlying dependence structure. Figure 3 compares the sample ADCF and sample ACF of model (13) with $p = 1$, $m = 2$ and ε_j's iid standard Gaussian. It is clear that the sample ADCF cuts off at lag 2 in this case.

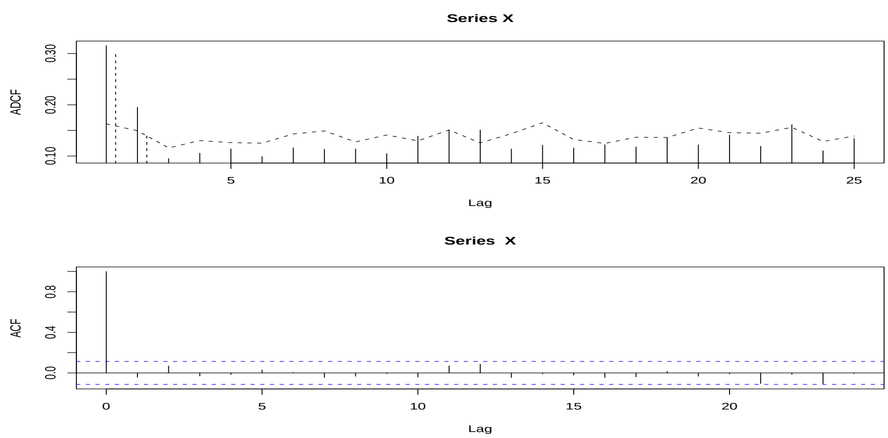

Figure 3: Comparison of the sample ADCF and sample ACF of model (13) with $p = 1$, $m = 2$, ε_j's iid standard Gaussian and $n = 300$. The dotted (solid) vertical lines are the true (sample) ADCF of the model and the dotted horizontal curve are the critical values at 95% level.

3.3 Vector nonlinear auto regressive processes

Many nonlinear time series models used in practice, such as bilinear models (Granger and Andersen, 1978), threshold models (Tong 1990) and (G)ARCH models (Engle (1982) and Bollerslev (1986)), have the following Markovian representation:

$$X_j = M(X_{j-1}, \varepsilon_j). \tag{14}$$

If $X_j = M'(X_{j-1}, \cdots, X_{j-m}, \varepsilon_j)$, then simply let $Y_j = (X_j^T, \cdots, X_{j-m}^T)^T$ and (Y_j) admits representation (14). We shall call (14) a vector nonlinear auto regressive (NAR) model. As shown in Wu (2005), if for some x_0, $\|M(x_0, \varepsilon_j)\| < \infty$, and

$$L < 1, \text{ where } L = \sup_{x \neq y} \frac{\|M(x, \varepsilon_0) - M(y, \varepsilon_0)\|}{|x - y|}. \tag{15}$$

then (14) admits a unique stationary solution, and iterations of (14) lead to $X_i = G(\mathcal{F}_i)$. Furthermore, we have $\delta(j, 2) = O(L^j)$. Therefore Proposition 2 implies that

$$\mathcal{V}_X(k) = O(L_1^k) \quad \text{for some } L_1 \in [0, 1). \tag{16}$$

In other words, the ADCF decays exponentially to 0 for model (14). As an illustrative example, let us consider the following univariate ARCH(2) model:

$$X_j = \sigma_j \varepsilon_j, \text{ where } \sigma_j^2 = 0.5 + 0.8X_{j-1}^2 + 0.1X_{j-2}^2 \tag{17}$$

and ε_j's are iid standard Gaussian. It is well known that many financial returns show no Pearson correlation but the correlation in square returns is strong. The later fact implies that financial returns are nonlinearly correlated. Figure 4 compares sample ACF and sample ADCF of model (17) with $n = 300$. It can be seen that the ACF exhibits no signal for (X_t). However, the ACF shows moderate dependence in (X_t^2). On the other hand, the ADCF shows nice exponential decays for both (X_t) and (X_t^2), which correctly reflects the underlying nonlinear dependence structure.

4 Asymptotic results

Theorem 1. *Suppose that* $\|X_j\|_{1+r_0} < \infty$ *for some* $r_0 > 0$ *and* $\sum_{k=0}^{\infty} [\delta(k, 1 + r_0)] < \infty$. *Then for all* $k \geq 0$

$$\mathcal{V}_X^n(k) \to \mathcal{V}_X(k) \quad \textit{in probability.}$$

Corollary 1. *Under the conditions of Theorem 1, we have that* $\mathcal{R}_X^n(k)$ *is a weakly consistent estimator of* $\mathcal{R}_X(k)$ *for each* $k \geq 1$.

Theorem 1 and Corollary 1 establish the weak consistency of $\mathcal{V}_X^n(k)$ and $\mathcal{R}_X^n(k)$. The consistency only requires that X_j has $(1 + r_0)$th moment for some $r_0 > 0$. The condition

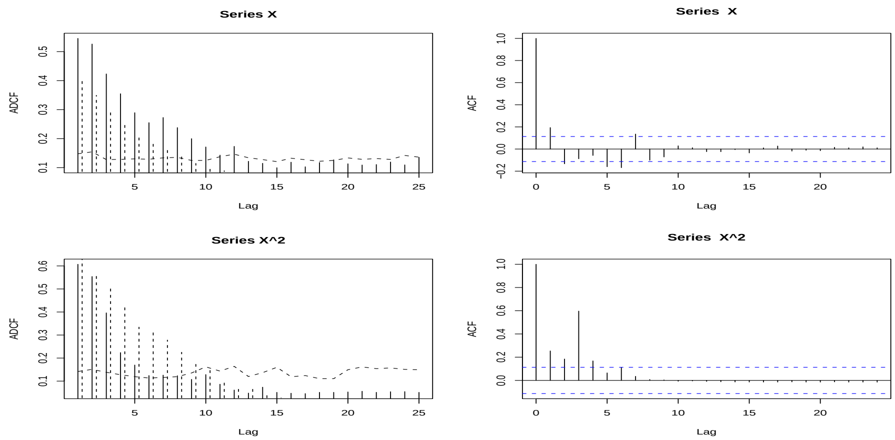

Figure 4: Comparison of the sample ADCF and sample ACF of processes (X_t) and (X_t^2) in model (17) with ε_j's iid standard Gaussian and $n = 300$. The dotted (solid) vertical lines are the true (sample) ADCF of the model and the dotted horizontal curve are the critical values at 95% level.

$\sum_{k=0}^{\infty}[\delta(k, 1 + r_0)] < \infty$ is a very classic short range dependence condition and it means that the cumulative effect of ε_0 in predicting future values of the time series is finite. For the vector linear process (10), the latter condition is equivalent to $\sum_{j=0}^{\infty} |A_j| < \infty$.

Theorem 2. *Let $\xi(t, s)$ denote a complex valued zero-mean Gaussian process with covariance function and relation function*

$$\Gamma^*(g, g_0) = \sum_{j\in\mathbb{Z}} \mathbb{E}[\psi_{j,k}(t, s)\bar{\psi}_{0,k}^T(t_0, s_0)] \text{ and } R^*(g, g_0) = \sum_{j\in\mathbb{Z}} \mathbb{E}[\psi_{j,k}(t, s)\psi_{0,k}^T(t_0, s_0)]$$

respectively, where $\psi_{j,k}(t, s) = [\exp(i\langle t, X_j\rangle) - f_X(t)] \times [\exp(i\langle s, X_{j+k}\rangle) - f_X(s)]$, $g = (t, s)$ and $g_0 = (t_0, s_0)$. Assume that $\Gamma(\cdot, \cdot)$ is positive definite, that $\|X_j\|_{4p+2} < \infty$, that $\sum_{j=0}^{\infty} \delta(j, 4p + 2) < \infty$ and that $\delta(j, 4) = O((j + 1)^{-\beta})$ for some $\beta > 3$. Then under the assumption that X_j and X_{j+k} are independent, we have

$$n\mathcal{V}_X^n(k) \Rightarrow \frac{1}{c_p^2} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \frac{|\xi(t, s)|^2}{|t|^{p+1}|s|^{p+1}}\, dtds, \tag{18}$$

where $\Rightarrow$ denotes weak convergence.

Theorem 2 unveils the limiting distribution of the sample auto distance covariance under the null hypothesis that X_j and X_{j+k} are independent. By Chapter 1 of Kuo (1975) and the proof of Lemma 2 in Section 7, it is easy to show that

$$\frac{1}{c_p^2}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{|\xi(t,s)|^2}{|t|^{p+1}|s|^{p+1}}\,dtds=\sum_{j=1}^{\infty}\lambda_j Z_j^2,$$

where Z_j's are iid standard normal, λ_j's are nonnegative constants with

$$0<\sum_{j=1}^{\infty}\lambda_j=\frac{1}{c_p^2}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{\Gamma(g,g)}{|t|^{p+1}|s|^{p+1}}\,dtds<\infty.$$

It is straightforward to obtain that $n\mathcal{V}_X^n(k)\to\infty$ if X_j and X_{j+k} are dependent. We call $\Gamma(\cdot,\cdot)$ and $R(\cdot,\cdot)$ long-run covariance and long-run relation functions which are due to the dependence of the series. If (X_j) is an iid sequence, then simple calculations show that $\Gamma(g,g_0)=[f_X(t-t_0)-f_X(t)\overline{f_X(t_0)}][f_X(s-s_0)-f_X(s)\overline{f_X(s_0)}]$ and $R(g,g_0)=[f_X(t+t_0)-f_X(t)f_X(t_0)][f_X(s+s_0)-f_X(s)f_X(s_0)]$. In this case the limiting distribution (18) coincides with that in Theorem 5 of SRB. In the limiting distribution (18), the dependence of the series is reflected in the long-run covariance and relation functions. It is worth mentioning that in a short note, Rémillard (2009) suggested using auto distance correlation to measure nonlinear dependence in time series. However, the latter note claimed (without proof) that the asymptotic distribution of the auto distance correlation is the same as that in SRB when the series is a white noise. Based on Theorem 2, we shall point out here that the latter claim of the limiting distribution was in fact incorrect because there are many possible forms of nonlinear dependence in a white noise process which will result in very different asymptotic distributions than that in SRB.

Due to the temporal dependence of the series, the permutation test of SRB is invalid for making inference in the time series case. In Section 5.1 below, a subsampling based test procedure will be introduced.

5 Finite sample performance

5.1 The subsampling

We see from Theorem 2 that the asymptotic null distribution of testing $\mathcal{V}_X(k)=0$ involves the long-run covariance of the series $\{\exp(i\langle t,X_j\rangle)\exp(i\langle t,X_{j+k}\rangle)\}_{j\in\mathbb{Z}}$ and a possibly high dimensional integration. Hence in practice it is difficult to directly use the asymptotic null distribution to perform hypothesis testing, especially when p is large. Here we suggest using the subsampling (Politis et. al, 1999) to avoid directly estimating the asymptotic distributions. Specifically, the following steps can be adopted:

(a). Select a block size l and define subsamples $X_{l,j} = (X_j, X_{j+1}, \cdots, X_{j+l-1})$, $j = 1, 2, \cdots, n - l + 1$.

(b). Calculate $\mathcal{V}^l_{X_{l,j}}(k)$ for $j = 1, 2, \cdots, n - l + 1$.

(c). Let $\mathcal{V}^l(k, 1) \leq \cdots \leq \mathcal{V}^l(k, n - l + 1)$ be the ordered sample auto distance covariances obtained from the subsamples.

(d). For $\alpha \in (0, 1)$, the $100(1 - \alpha)\%$ critical value of testing $\mathcal{V}_X(k) = 0$ can be estimated by $(l - k)\mathcal{V}^l(k, l_\alpha)/(n - k)$, where $l_\alpha = \lfloor (1 - \alpha)(n - l + 1) \rfloor$. Namely reject the null hypothesis if $\mathcal{V}^n_X(k) > (l - k)\mathcal{V}^l(k, l_\alpha)/(n - k)$.

The above subsampling procedure is intuitively plausible since the subsamples preserve the dependence structure of the series. Therefore it is expected that the empirical distribution of the test statistic obtained from the subsamples coincides with the target distribution asymptotically. The following theorem establishes the asymptotic validity of the above subsampling procedure under minimal sufficient conditions for the block size l:

Theorem 3. *Assume that $l \to \infty$ with $l/n \to 0$ and that X_j and X_{j+k} are independent. Then under the assumptions of Theorem 2, we have*

$$\mathbb{P}[(n - k)\mathcal{V}^n_X(k) \leq (l - k)\mathcal{V}^l(k, l_\alpha)] \to 1 - \alpha \text{ as } n \to \infty. \tag{19}$$

We now discuss the choice of the block size l for moderate sample sizes. Here we suggest use the minimum volatility (MV) method advocated in Chapter 9.4 of Politis et. al (1999). The idea behind the MV method is that, if a block size is in a reasonable range, then the estimated critical values for the independence test should be stable when considered as a function of block size. Hence one could first propose a grid of possible block sizes and then choose the block size which minimizes the volatility of the critical values near this size. More specifically, let the grid of possible block sizes be $\{l_1 < \ldots < l_M\}$ and let the estimated critical values be $\{T_{l_j}\}$, $j = 1, 2, \ldots, M$. For each l_j, calculate $se(\bigcup_{r=-3}^{3}\{T_{l_{j+r}}\})$, where se denotes standard error, namely

$$se(\{T_j\}_{j=1}^{k}) = \Big[\frac{1}{k-1}\sum_{j=1}^{k}|T_j - \bar{T}|^2\Big]^{1/2}$$

with $\bar{T} = \sum_{j=1}^{k} T_j/k$. Then one chooses the l_j^* which minimizes the above standard errors. In our simulation studies, the MV method performs reasonably well. We shall refer to Chapter 9.4 of Politis et. al (1999) for a more detailed description and discussion of the latter minimum volatility method.

5.2 A simulation study

In this section we shall conduct a small simulation to study the finite sample accuracy of the auto distance covariance test of independence for multivariate time series. In particular, we are interested in investigating how dimension p and strength of the dependence influence the accuracy of the test. For this purpose, we shall consider the vector ARMA model (11) with $\phi(B) = aI_pB^2$ and $\theta(B) = 0$, $0 \le a < 1$, where I_p is the $p \times p$ identity matrix. The errors ε_j are iid p-dimensional standard Gaussian. Then X_j is independent of X_{j+1} and the dependence of the series gets stronger as a gets closer to 1. We are interested in testing $H_0 : \mathcal{V}_X(1) = 0$. In our studies we choose $n =$ 100,200 and 300. We then perform the subsampling test with the minimum volatility block size selection method to test H_0. Based on 5000 repetitions, the simulated type I error rates with respect to various choice of p and a are reported in Table 1 below.

	$n = 100$			$n = 200$			$n = 300$		
	$a = 0.2$	$a = 0.5$	$a = 0.8$	$a = 0.2$	$a = 0.5$	$a = 0.8$	$a = 0.2$	$a = 0.5$	$a = 0.8$
$p = 1$	11.7%	14.3%	18.1%	11.7%	11.6%	18.2%	9.7%	10.6%	13.3%
$p = 2$	9.7%	14.4%	23.3%	9.1%	12.9%	20.0%	9.7%	10.5%	14.2%
$p = 3$	9.4%	15.0%	24.5%	8.5%	12.1%	21.7%	9.5%	9.6%	14.8%
$p = 4$	10.6%	15.9%	26.7%	9.2%	11.2%	21.9%	9.3%	10.6%	16.5%
$p = 5$	12.5%	18.3%	27.8%	9.3%	11.9%	23.0%	9.0%	11.6%	18.3%
$p = 6$	15.2%	21.2%	30.4%	9.2%	13.2%	24.0%	9.7%	11.6%	18.3%
$p = 7$	17.2%	22.1%	30.9%	9.7%	14.0%	24.5%	9.5%	11.0%	18.4%
$p = 8$	19.9%	25.6%	33.6%	11.9%	15.1%	26.7%	9.0%	10.9%	20.4%
$p = 9$	20.4%	25.9%	37.3%	12.3%	16.1%	26.7%	9.8%	11.1%	21.4%
$p = 10$	22.4%	24.1%	35.6%	12.5%	17.5%	27.9%	10.0%	12.4%	22.4%

Table 1. Simulated type I error rates for model (11) with $\phi(B) = aI_pB^2$ and $\theta(B) = 0$ and nominal level 10%. Series lengths $n = 100$, 200 and 300 with 5000 replicates.

Table 1 shows that the subsampling performs reasonably well in the simulations. We also find that the type I error rate improves as the sample size enlarges. On the other hand, for a fixed sample size, the performance of the subsampling drops as the dependence strengthens. When making inference of the mean and variance functions, it is well known that the accuracy of the subsampling decreases as the dependence of the series gets stronger for a moderate sample size. See for instance Chapter 9.5 of Politis et. al (1999). In our case the same phenomenon is observed. For strongly dependent data, generally we need a relatively large sample size to guarantee the accuracy of the subsampling test. Therefore one should be cautious when exploring temporal dependence structures of strongly dependent time series.

Table 1 shows that the accuracy of the test decreases when the dimension p is large compared to the sample size n. As we mentioned in Section 2, throughout the paper we assume $p \ll n$ and all the asymptotic theory of the paper are established under the latter assumption. In our simulations, we find that, for series of moderate dependence, the performance of the subsampling tests are similar and reasonably accurate for different choices of p as long as $p \ll \sqrt{n}$. Additionally, for many real world applications such as the ones which will be analyzed in Section 6 of this paper, the dimension p is typically much smaller than the series length n and hence the theory and methodology of the paper apply.

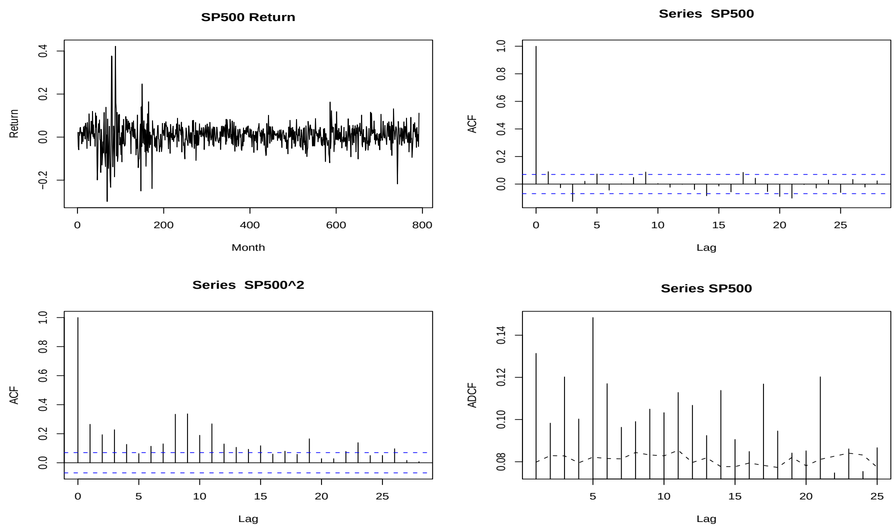

Figure 5: Time series plot of the SP500 monthly excess return data (upper left panel), the ACF of the series and the squared series (upper right and lower left panels), and the ADCF of the series (lower right panel).

6 Real data examples

6.1 The SP500 return data

The (G)ARCH models have been very successful in explaining the volatility clustering phenomenon observed in financial returns. In this section we are interested in testing the adequacy of the (G)ARCH models. As an illustrative example, we shall analyze the SP500 monthly excess return data starting from 1926 for 792 observations. The SP500 index is widely used in the derivative markets. And hence investigating its volatility is of great interest. Figure 5 shows the time series plot of the data as well as the ACF and ADCF plots. The ACF of the original series shows no signal while the ACF plot of the squared returns exhibit moderate level of dependence. The latter phenomenon is classic for time series of financial returns. On the other hand, the ADCF of the original series shows strong signals of dependence compared with the critical values of the independence test. The data set was analyzed in Tsay (2005) and he found an AR(3)-GARCH(1,1) model fits the data well. The upper part of Figure 6 summaries the ACF of the residuals and the squared residuals of the AR(3)-GARCH(1,1) model. From the ACF angle, the residuals show no signs of correlation. However, if we plot the ADCF of the residuals and the squared residuals as shown in the lower panel of Figure 6 (with block size 10), interesting and surprising signals at lags 1 and 3 float to the surface. The p-values of testing $\mathcal{R}_X(k) = 0$ and $\mathcal{R}_{X^2}(k) = 0$ at lags 1 and 3 are all less than 0.01. Ignoring the latter dependencies will lead to incorrect inference of the GARCH parameters and the forecast limits. On the other hand, however, the existence of strong dependencies at lags 1 and 3 of the residuals means that we could possibly utilize the latter dependencies and further improve our forecasts of the volatility. For instance, one could adopt an NMA(3) model for the squared residuals and forecast the volatility accordingly. The detailed modeling and analysis of the residuals is beyond the scope of this paper. In general, we suggest fitting the SP500 monthly excess return data with a GARCH model with *dependent* innovations.

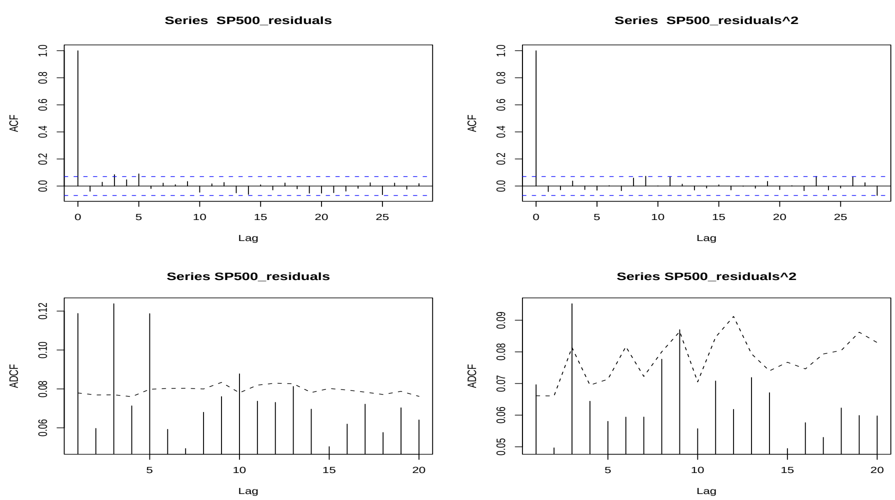

Figure 6: The ACF and ADCF of the residuals and the squared residuals of the SP500 monthly return data.

6.2 The Canadian lynx data

The time series is composed of the annual record of the number of the Canadian lynx "trapped" in the Mackenzie River district of the North-West Canada for the period 1821-1934. Upper panel of Figure 7 plots the data (at $\log_{10}$ scale). The series has become a classic series for threshold auto regressive (TAR) time series models (Tong ,1990) since the latter model successfully mimics the asymmetric cycle of the series with a nice predator (lynx) and prey (hare) interaction interpretation in ecology. Chapter 4.1.4 of Fan and Yao (2003) did a detailed analysis of the series and found that the following TAR model fits the data well:

$$\begin{aligned} X_j &= 0.546 + 1.032X_{j-1} - 0.173X_{j-2} + 0.171X_{j-3} - 0.431X_{j-4} + 0.332X_{j-5} \\ &\quad -0.284X_{j-6} + 0.21X_{j-7} + \varepsilon_j^{(1)}, \text{ if } X_{j-2} \le 3.116 \\ X_j &= 2.632 + 1.492X_{j-1} - 1.324X_{j-2} + \varepsilon_j^{(2)}, \text{ if } X_{j-2} > 3.116. \end{aligned} \tag{20}$$

Fan and Yao (2003) mentioned that the residuals of the above TAR model pass most residual-based tests of autocorrelation comfortably. We shall use the ADCF of the residuals to test if any nonlinear dependence structure is missed from model (20). Lower panel of Figure 7 shows the ADCF plot of the residuals. The block size is chosen as 10 in our analysis. It is clear that the residuals still show strong dependence. In particular, the TAR model (20) did not fully explain the immediate dependence of the adjacent lynx counts and the cyclic dependence. Therefore, if the purpose of the study is to forecast future captures of lynx, then there is plenty of space for model improvement over (20).

7 Proofs

To prove consistency and asymptotic null distribution of the auto distance covariances in Theorems 1 and 2, we first deal with the two singular points 0 and ∞ in the definition (1) of the ADCF. Second half of the proof of Theorem 1 and Lemmas 2 and 3 below achieve the latter task. Then we prove the ergodicity and weak convergence of the empirical characteristic functions using martingale approximation and empirical process techniques. To prove consistency of the subsampling in Theorem 3, The main techniques used are approximating the empirical characteristic functions by m-dependent processes and controlling

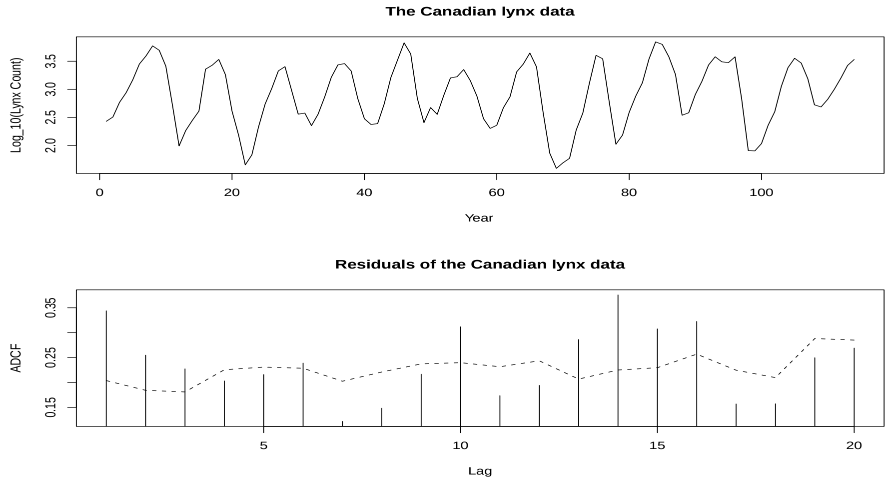

Figure 7: The time series plot of the Canadian lynx data (upper panel) and the ADCF of the residuals of model (20) (Lower panel).

the variation of the empirical process of the corresponding m-dependent processes. See Lemmas 6 and 7.

In this Section the symbol C denotes a finite constant which may vary from place to place. Let

$$\alpha(t,j) := \exp\{i\langle t, X_j\rangle\} \quad \text{and} \quad \beta(t,j) := \exp\{i\langle t, X_{j+k}\rangle\}, j = 1, 2, \cdots, n-k. \tag{21}$$

Write $dw = \frac{1}{c_p^2 |t|^{p+1}|s|^{p+1}} dtds$. Further write

$$\zeta_n(t,s) = f_k^n(t,s) - f^n(t) f^{n,k}(s) \quad \text{and} \quad \zeta(t,s) = f_k(t,s) - f(t)f(s), \tag{22}$$

where $f_k(t,s) = f_{X_i, X_{i+k}}(t,s)$. For $j \in \mathbb{Z}$ define the projection operator

$$\mathcal{P}_j \cdot = \mathbb{E}(\cdot|\mathcal{F}_j) - \mathbb{E}(\cdot|\mathcal{F}_{j-1}). \tag{23}$$

Lemma 1. *For all $j, r \in \mathbb{Z}$, $\eta \in (0,1]$ and $q \geq 1$, we have*

$$\begin{aligned} \|\mathcal{P}_r \alpha(t,j)\|_q &\leq 2^{2-\eta}|t|^\eta [\delta(j-r, \eta q)]^\eta, \\ \|\mathcal{P}_r \alpha(t,j)\beta(s,j)\|_q &\leq 2^{2-\eta}\{|t|^\eta[\delta(j-r,\eta q)]^\eta + |s|^\eta [\delta(j+k-r, \eta q)]^\eta\}. \end{aligned} \tag{24}$$

Proof. We will only prove the first inequality in (24) since the other inequality can be proved using similar arguments. Let $\alpha^*(t,j) = \exp\{i\langle t, X^*_{j,r}\rangle\}$, where $X^*_{j,r} = G(\mathcal{F}^*_{j,r})$ with $\mathcal{F}^*_{j,r} = (\mathcal{F}_{r-1}, \varepsilon'_r, \varepsilon_{r+1}, \cdots, \varepsilon_j)$. By Theorem 1 in Wu (2005), we have

$$\|\mathcal{P}_r\alpha(t,j)\|_q \le \|\alpha(t,j) - \alpha^*(t,j)\|_q \le \|\cos\langle t, X_j\rangle - \cos\langle t, X^*_{j,r}\rangle\|_q + \|\sin\langle t, X_j\rangle - \sin\langle t, X^*_{j,r}\rangle\|_q.$$

Note that

$$\begin{aligned} \|\cos\langle t, X_j\rangle - \cos\langle t, X^*_{j,r}\rangle\|_q &= 2\|\sin\frac{\langle t, X_j + X^*_{j,r}\rangle}{2}\sin\frac{\langle t, X_j - X^*_{j,r}\rangle}{2}\|_q \\ &\le 2\|\sin\frac{\langle t, X_j - X^*_{j,r}\rangle}{2}\|_q \le 2\|\min\{1, \Big|\frac{\langle t, X_j - X^*_{j,r}\rangle}{2}\Big|\}\|_q \\ &\le 2\|\Big|\frac{\langle t, X_j - X^*_{j,r}\rangle}{2}\Big|^\eta\|_q \le 2\|\Big|\frac{|t||X_j - X^*_{j,r}|}{2}\Big|^\eta\|_q \\ &= 2^{1-\eta}|t|^\eta[\delta(j-r, \eta q)]^\eta. \end{aligned}$$

The same inequality holds for $\|\sin\langle t, X_j\rangle - \sin\langle t, X^*_{j,r}\rangle\|_q$. Therefore the lemma follows. $\diamondsuit$

Proof of Theorem 1. For $\gamma > 0$, define

$$D(\gamma) = \{(t,s) \in \mathbb{R}^p \times \mathbb{R}^p : \gamma \le |t| \le 1/\gamma, \gamma \le |s| \le 1/\gamma\}. \tag{25}$$

Let $\mathcal{V}_{n,\gamma}(k) = \int_{D(\gamma)} |\zeta_n(t,s)|^2\, dw$, $\mathcal{V}_\gamma(k) = \int_{D(\gamma)} |\zeta(t,s)|^2\, dw$. For a fixed $(t,s) \in D(\gamma)$,

$$\begin{aligned} ||\zeta_n(t,s)|^2 - |\zeta(t,s)|^2| &= (|\zeta_n(t,s)| + |\zeta(t,s)|)(||\zeta_n(t,s)| - |\zeta(t,s)||) \\ &\le 4|\zeta_n(t,s) - \zeta(t,s)| \\ &\le 4\Big\{|f^n_k(t,s) - f_k(t,s)| + |[f^n(t) - f(t)]f^{n,k}(s)| + |[f^{n,k}(s) - f(s)]f(t)|\Big\} \\ &:= 4\{I + II + III\}. \end{aligned}$$

Note that we used the fact that $|\zeta_n(t,s)| \le 2$ and $|\zeta(t,s)| \le 2$ for all (t,s) and n. Write

$$\Psi_{n,l}(t,s) = \sum_{j=1}^{n-k} \mathcal{P}_{j+k-l}[\alpha(t,j)\beta(s,j)].$$

Then the summands of $\Psi_{n,l}(t,s)$ form a martingale difference series. Without loss of generality, assume $r_0 < 1$. Write $q_0 = 1 + r_0$. By Lemma 1 and Burkholder's inequality, we have

$$(\|\Psi_{n,l}(t,s)\|_{q_0}/B_{q_0})^{q_0} \le \mathbb{E}\Big(\sum_{j=1}^{n-k} |\mathcal{P}_{j+k-l}[\alpha(t,j)\beta(s,j)]|^2\Big)^{q_0/2} \le \sum_{j=1}^{n-k} \mathbb{E}|\mathcal{P}_{j+k-l}[\alpha(t,j)\beta(s,j)]|^{q_0}$$

$$\leq \quad C(\gamma)(n-k)\{\delta(l-k,q_0)+\delta(l,q_0)\}^{q_0},$$

where $B_q = 18q^{3/2}(q-1)^{-1/2}$, $C(\gamma) = (2/\gamma)^{q_0}$. Note that $I = |\sum_{l=0}^{\infty} \Psi_{n,l}(t,s)|/(n-k)$. Therefore

$$(n-k)\|I\|_{q_0} \leq \sum_{l=0}^{\infty} \|\Psi_{n,l}(t,s)\|_{q_0} \leq B_{q_0}[C(\gamma)(n-k)]^{1/q_0} \sum_{l=0}^{\infty}\{\delta(l-k,q_0)+\delta(l,q_0)\} \leq C(n-k)^{1/q_0}.$$

Similarly, it can be shown that $\|II\|_{q_0} + \|III\|_{q_0} \leq C(n-k)^{1/q_0-1}$. Hence

$$\begin{aligned} \|\mathcal{V}_{n,\gamma}(k) - \mathcal{V}_{\gamma}(k)\|_{q_0} &\leq \int_{D(\gamma)} \||\zeta_n(t,s)|^2 - |\zeta(t,s)|^2\|_{q_0}\, dw \\ &\leq 4\int_{D(\gamma)} \|I\|_{q_0} + \|II\|_{q_0} + \|III\|_{q_0}\, dw \leq C(n-k)^{1/q_0-1}. \end{aligned} \quad (26)$$

In particular, $|\mathcal{V}_{n,\gamma}(k) - \mathcal{V}_{\gamma}(k)| \to 0$ in probability for each fixed γ. Clearly $\mathcal{V}_{\gamma}(k)$ converges to $\mathcal{V}_X(k)$ as γ tends to 0. Therefore, to prove Theorem 1, it suffices to show that in probability

$$\limsup_{\gamma\to 0}\limsup_{n\to\infty} |\mathcal{V}_{n,\gamma}(k) - \mathcal{V}_X^n(k)| = 0. \quad (27)$$

Note that, for each $\gamma > 0$,

$$\begin{aligned} |\mathcal{V}_{n,\gamma}(k) - \mathcal{V}_X^n(k)| &\leq \int_{|t|<\gamma} |\zeta_n(t,s)|^2\, dw + \int_{|t|>1/\gamma} |\zeta_n(t,s)|^2\, dw \\ &\quad + \int_{|s|<\gamma} |\zeta_n(t,s)|^2\, dw + \int_{|s|>1/\gamma} |\zeta_n(t,s)|^2\, dw. \end{aligned} \quad (28)$$

Now for $z = (z_1, \cdots, z_p) \in \mathbb{R}^p$, define $H(y) = \int_{|z|<y}\{1-\cos z_1\}/|z|^{1+p}\, dz$.

Then it is clear that $H(y) \leq \lim_{y\to\infty} H(y) \leq c_p$. On the other hand,

$$H(y) = \int_{|z|<y} \frac{2\sin^2[z_1/2]}{|z|^{1+p}}\, dz \leq \int_{|z|<y} \frac{z_1^2}{2|z|^{1+p}}\, dz = C(p)y,$$

where $C(p)$ is a constant that only depends on p. A careful check of the proof of Theorem 2 of Szekely et. al (2007) shows that

$$\int_{|t|<\gamma} |\zeta_n(t,s)|^2\, dw \leq 4\frac{2}{n-k}\sum_{j=1}^{n-k}(|X_{j+k}| + E|X_1|)\frac{2}{n-k}\sum_{j=1}^{n-k}\mathbb{E}_X[|X_j - X|H(|X_j - X|\gamma)],$$

where X is identically distributed and independent of X_j and E_X is taken with respect to X. Now let $q_1 = 1 + r_0/2$. Then

$$\begin{aligned}|X_j - X|H(|X_j - X|\gamma) &\le |X_j - X| \min\{c_p, C(p)|X_j - X|\gamma\} \\ &\le c_p|X_j - X|[C(p)|X_j - X|\gamma/c_p]^{r_0/2} = C^*(p)|X_j - X|^{q_1}\gamma^{r_0/2} \\ &\le C^*(p)2^{q_1-1}\{|X_j|^{q_1} + |X|^{q_1}\}\gamma^{r_0/2},\end{aligned}$$

where $C^*(p) = c_p^{1-r_0/2}C(p)^{r_0/2}$. Therefore,

$$\int_{|t|<\gamma} |\zeta_n(t,s)|^2\, dw \le C\gamma^{r_0/2}\frac{2}{n-k}\sum_{j=1}^{n-k}(|X_{j+k}| + E|X_1|)\frac{2}{n-k}\sum_{j=1}^{n-k}(|X_j|^{q_1} + E|X_1|^{q_1}). \quad (29)$$

Similar to the proof of (26), it can be shown that, in probability

$$\frac{1}{n-k}\sum_{j=1}^{n-k}|X_{j+k}| \to \mathbb{E}|X_1|, \quad \frac{1}{n-k}\sum_{j=1}^{n-k}|X_j|^{q_1} \to \mathbb{E}|X_1|^{q_1}.$$

Hence (29) implies that

$$\limsup_{n\to\infty}\int_{|t|<\gamma}|\zeta_n(t,s)|^2\, dw \le C\gamma^{r_0/2}E|X_1|E|X_1|^{q_1}.$$

Therefore $\limsup_{\gamma\to 0}\limsup_{n\to\infty}\int_{|t|<\gamma}|\zeta_n(t,s)|^2\, dw = 0$ in probability. Again, a careful check of the proof of Theorem 2 of Szekely et. al (2007) shows that

$$\int_{|t|>1/\gamma}|\zeta_n(t,s)|^2\, dw \le 16\gamma\frac{2}{n-k}\sum_{j=1}^{n-k}(|X_{j+k}| + E|X_1|).$$

Hence, in probability $\limsup_{\gamma\to 0}\limsup_{n\to\infty}\int_{|t|>1/\gamma}|\zeta_n(t,s)|^2\, dw = 0$. The other two terms in (28) can be dealt with similarly. Therefore Theorem 1 follows. $\diamondsuit$

Lemma 2. *Assume that $\|X_j\|_4 < \infty$ and $\delta(j,4) = O((j+1)^{-\beta})$ for some $\beta > 3$. Let $\xi_n(t,s) = \sqrt{n-k}\,\zeta_n(t,s)$. If X_j is independent of X_{j+k}, then there exists a finite constant C which does not depend on n, such that*

$$\mathbb{E}\int_{|t|<\gamma}|\xi_n(t,s)|^2\, dw + \mathbb{E}\int_{|s|<\gamma}|\xi_n(t,s)|^2\, dw < C\gamma. \quad (30)$$

Proof. We will only deal with the first summand in (30) since the second summand can be dealt with similarly. Let

$$u_j = \alpha(t,j) - \mathbb{E}\alpha(t,j) \quad \text{and} \quad v_j = \beta(s,j) - \mathbb{E}\beta(s,j), j = 1, 2, \cdots, n-k. \tag{31}$$

Then $\xi_n(t,s) = \sqrt{n-k}\{\sum_j u_j v_j/(n-k) + \sum_j u_j/(n-k)\sum_j v_j/(n-k)\}$. Therefore

$$\begin{aligned}
\mathbb{E}\int_{|t|<\gamma} |\xi_n(t,s)|^2\,dw &\leq \frac{2}{n-k}\mathbb{E}\int_{|t|<\gamma} |\sum_j u_j v_j|^2\,dw + \frac{2}{(n-k)^3}\mathbb{E}\int_{|t|<\gamma} |\sum_j u_j \sum_j v_j|^2\,dw \\
&:= I_0 + II_0.
\end{aligned}$$

Write

$$I_0 = \frac{2}{n-k}\mathbb{E}\int_{|t|<\gamma,|s|\leq 1} |\sum_j u_j v_j|^2\,dw + \frac{2}{n-k}\mathbb{E}\int_{|t|<\gamma,|s|>1} |\sum_j u_j v_j|^2\,dw := I_0' + I_0''.$$

Denote by $\bar{x}$ the complex conjugate of x and $\Re(x)$ the real part of x. Note that

$$\begin{aligned}
I_0' &= \frac{2}{n-k}\int_{|t|<\gamma,|s|\leq 1} \mathbb{E}(\sum_j u_j v_j \sum_r \overline{u_r v_r})\,dw = \frac{2}{n-k}\int_{|t|<\gamma,|s|\leq 1} \mathbb{E}\Re\{\sum_{j,r} u_j v_j u_r v_r\}\,dw \\
&\leq \frac{2}{n-k}\sum_{j,r}\int_{|t|<\gamma,|s|\leq 1} |\mathbb{E}u_j v_j u_r v_r|\,dw.
\end{aligned} \tag{32}$$

Since $\mathbb{E}u_j v_j = \mathbb{E}u_j \mathbb{E}v_j = 0$ for all j and $\mathcal{P}_j$'s are orthogonal, we have

$$\begin{aligned}
|\mathbb{E}u_j v_j u_r v_r| &= |\mathbb{E}\sum_{h\in\mathbb{Z}} \mathcal{P}_h u_j v_j \sum_{l\in\mathbb{Z}} \mathcal{P}_l u_r v_r| = |\mathbb{E}\sum_{h\in\mathbb{Z}} \mathcal{P}_h u_j v_j \mathcal{P}_h u_r v_r| \\
&\leq \sum_{h\in\mathbb{Z}} \mathbb{E}|\mathcal{P}_h u_j v_j \mathcal{P}_h u_r v_r| \leq \sum_{h\in\mathbb{Z}} \|\mathcal{P}_h u_j v_j\| \|\mathcal{P}_h u_r v_r\|.
\end{aligned} \tag{33}$$

Similar to the definition of $X^*_{j,r}$ in Lemma 1, define $u^*_{j,r}$ and $v^*_{j,r}$ by replacing $\mathcal{F}_j$ with $\mathcal{F}^*_{j,r}$. Then Lemma 1 and Theorem 1 in Wu (2005) imply that

$$\begin{aligned}
\|\mathcal{P}_h u_j v_j\| &\leq \|u_j v_j - u^*_{j,h} v^*_{j,h}\| \leq \|(u_j - u^*_{j,h})v_j\| + \|(v_j - v^*_{j,h})u^*_{j,h}\| \\
&\leq \|u_j - u^*_{j,h}\|_4 \|v_j\|_4 + \|v_j - v^*_{j,h}\|_4 \|u^*_{j,h}\|_4 \\
&\leq 2|t|\delta(j-h,4)\|v_j\|_4 + 2|s|\delta(j+k-h,4)\|u^*_{j,h}\|_4.
\end{aligned} \tag{34}$$

Since $v_j = \sum_{h\in\mathbb{Z}} \mathcal{P}_h v_j$, by Burkholder's inequality and Lemma 1, we have

$$(\|v_j\|_4/B_4)^2 \leq \|\sum_{h\in\mathbb{Z}} |\mathcal{P}_h v_j|^2\| \leq \sum_{h\in\mathbb{Z}} \|\mathcal{P}_h v_j\|_4^2 \leq \sum_{h\in\mathbb{Z}} [2|s|\delta(j+k-h,4)]^2 = C|s|^2.$$

Recall that B_q is defined in the proof of Theorem 1. Therefore $\|v_j\|_4 \le C|s|$. Similarly, $\|u^*_{j,h}\|_4 \le C|t|$. Plugging the latter inequalities into (34), we have

$$\|\mathcal{P}_h u_j v_j\| \le C|t||s|(\delta(j-h,4)+\delta(j+k-h,4)). \tag{35}$$

Similarly, $\|\mathcal{P}_h u_r v_r\| \le C|t||s|(\delta(r-h,4)+\delta(r+k-h,4))$. Therefore (33) implies that

$$\begin{aligned}|\mathbb{E}u_j v_j u_r v_r| &\le C|t|^2|s|^2\sum_{h\in\mathbb{Z}}(\delta(j-h,4)+\delta(j+k-h,4))(\delta(r-h,4)+\delta(r+k-h,4))\\ &\le C|t|^2|s|^2(|j-r|+1)^{-\beta}.\end{aligned}$$

Therefore (32) implies that

$$I'_0 \le \frac{C}{n-k}\sum_{j,r}\int_{|t|<\gamma,|s|\le 1}|t|^2|s|^2(|j-r|+1)^{-\beta}\,dw = \gamma\frac{C}{n-k}\sum_{j,r}(|j-r|+1)^{-\beta} \le C\gamma. \tag{36}$$

Now assume $|s|>1$. Note that $\|v_j\|_4 \le 2$. By Lemma 1 and (34), we have

$$\begin{aligned}\|\mathcal{P}_h u_j v_j\| &\le \|u_j - u^*_{j,h}\|_4\|v_j\|_4 + \|v_j - v^*_{j,h}\|_4\|u^*_{j,h}\|_4\\ &\le 4|t|\delta(j-h,4) + 2^{5/3}|s|^{1/3}\delta^{1/3}(j+k-h,4/3)C|t|\\ &\le C|t||s|^{1/3}[\delta(j-h,4)+\delta^{1/3}(j+k-h,4)].\end{aligned}$$

Similarly, $\|\mathcal{P}_h u_r v_r\| \le C|t||s|^{1/3}[\delta(r-h,4)+\delta^{1/3}(r+k-h,4)]$. Hence

$$\begin{aligned}I''_0 &\le \frac{C}{n-k}\sum_{j,r}\int_{|t|<\gamma,|s|>1}|t|^2|s|^{2/3}(|j-r|+1)^{-\beta/3}\,dw\\ &= \gamma\frac{C}{n-k}\sum_{j,r}(|j-r|+1)^{-\beta/3} \le C\gamma. \end{aligned}\tag{37}$$

By (36) and (37), Lemma 2 follows. ◇

Lemma 3. *Assume that $\|X_j\|_4 < \infty$ and $\delta(j,4) = O((j+1)^{-\beta})$ for some $\beta > 3$. If X_j is independent of X_{j+k}, then there exists a finite constant C which does not depend on n, such that*

$$\mathbb{E}\int_{|t|>1/\gamma}|\xi_n(t,s)|^2\,dw + \mathbb{E}\int_{|s|>1/\gamma}|\xi_n(t,s)|^2\,dw < C\gamma^{1/3}. \tag{38}$$

Lemma 3 can be proved using similar arguments as those in the proof of Lemma 2. Details are omitted.

Lemma 4. *Under the assumptions of Theorem 2 and for any fixed $\gamma > 0$, we have that $\xi_n(t,s)$ converges weakly to $\xi(t,s)$ on $D(\gamma)$.*

Proof. We will first prove finite dimensional convergence. Namely for $g_1, g_2, \cdots, g_l \in D(\gamma)$, $l \geq 1$,

$$(\xi_n(g_1), \xi_n(g_2), \cdots, \xi_n(g_l))^T \Rightarrow (\xi(g_1), \xi(g_2), \cdots, \xi(g_l))^T, \tag{39}$$

where $\Rightarrow$ denotes weak convergence. We will only prove the case when $l = 1$, since other cases follow similarly by the Cramer-Wold device. Let $g = (t,s) \in D(\gamma)$. Recall the definition of u_j and v_j in (31). By (35), we have

$$\sum_{h\in\mathbb{Z}} \|\mathcal{P}_h u_j v_j\| \leq C|t||s| \sum_{h\in\mathbb{Z}} (\delta(j-h,4) + \delta(j+k-h,4)) < \infty.$$

Therefore by Theorem 3 in Wu (2005), we have

$$\Big(\Re(\sum_j u_j v_j/\sqrt{n-k}), \Im(\sum_j u_j v_j/\sqrt{n-k})\Big)^T \Rightarrow \Big(\Re(\xi(t,s)), \Im(\xi(t,s))\Big)^T,$$

where $\Im(x)$ denotes the imaginary part of x. Now Similar arguments imply that

$$|\sum_j u_j| = O_{\mathbb{P}}(\sqrt{n-k}) \quad \text{and} \quad |\sum_j v_j| = O_{\mathbb{P}}(\sqrt{n-k}).$$

Therefore $|\sum_j u_j \sum_j v_j/(n-k)^{3/2}| = O_{\mathbb{P}}(1/\sqrt{n-k}) = o_{\mathbb{P}}(1)$. Hence $\xi_n(g) \Rightarrow \xi(g)$.

We will now prove the tightness of $\xi_n(t,s)$ on $D(\gamma)$. For a fixed $g_0 \in D(\gamma)$, similar arguments as those of (35) imply that $\|\xi_n(g_0)\| \leq C$. Hence for each positive η, there exists an a, such that

$$\mathbb{P}(|\xi_n(g_0)| \geq a) \leq \eta, \quad n \geq 1. \tag{40}$$

For $(t,s) \in \mathbb{R}^p \times \mathbb{R}^p$, let $t = (t_1, \cdots, t_p)^T$ and $s = (s_1, \cdots, s_p)^T$. Define

$$D^*(\gamma) = \bigcup_{j=1}^{p} \{(t,s) : \gamma/\sqrt{p} \leq |t_j| \leq 1/\gamma \text{ or } \gamma/\sqrt{p} \leq |s_j| \leq 1/\gamma\}.$$

Obviously $D^*(\gamma)$ covers $D(\gamma)$. For each $\epsilon > 0$, partition each interval $\gamma/\sqrt{p} \leq |t_j| \leq 1/\gamma$ and $\gamma/\sqrt{p} \leq |s_j| \leq 1/\gamma$ into v blocks, namely $\{t_j : t_{j,k} \leq t_j \leq t_{j,k+1}\}$, $\{s_j : s_{j,k} \leq s_j \leq$

$s_{j,k+1}\}$, $k = 1, 2, \cdots, v$, such that $\epsilon \le \min_k(|t_{j,k} - t_{j,k+1}|) \le 2\epsilon$, $\epsilon \le \min_k(|s_{j,k} - s_{j,k+1}|) \le 2\epsilon$ for all j. Note that $v \le C/\epsilon$. Those grid points on each axis further partition $D^*(\gamma)$ into $v^{2p} := v(p)$ cubes. Denote by $c_1, \cdots, c_{v(p)}$ those cubes. Find points $o_1, o_2, \cdots, o_{v(p)}$, such that $o_j \in c_j$, $j = 1, 2, \cdots, v(p)$. Define the quantities

$$w(\epsilon) = \sup_{|g'-g''|\le\epsilon, g', g'' \in D^*(\gamma)} |\xi_n(g') - \xi_n(g'')|, \quad m^*(\epsilon) = \max_{1\le j\le v(p)} \sup_{g', g'' \in c_j} |\xi_n(g') - \xi_n(g'')|.$$

Note that if $|g' - g''| \le \epsilon$, then g' and g'' must lie in the same cube or in adjacent ones. Therefore it is easy to see that

$$w(\epsilon) \le 2m^*(\epsilon) \le 4 \max_{1\le j\le v(p)} \sup_{|g-o_j|\le 2\sqrt{2p}\epsilon} |\xi_n(g) - \xi_n(o_j)| := 4m(\epsilon).$$

Let $I = \{\alpha_1, \alpha_2, \cdots, \alpha_q\} \subset \{1, 2 \cdots, 2p\}$ be a nonempty set and $1 \le \alpha_1 < \cdots < \alpha_q$. For $\mathbf{g} = (g_1, \cdots, g_{2p})^T \in \mathbb{R}^{2p}$, let $\mathbf{g}_I = (g_1 \mathbf{1}_{1\in I}, \cdots, g_{2p}\mathbf{1}_{2p\in I})$ and $X_{j,I} = X_{j,\alpha_1} \times \cdots \times X_{j,\alpha_q}$. For $\theta, y \in \mathbb{R}^{2p}$, define

$$\int_0^{\theta_I} \frac{\partial^q \xi_n(y + \mathbf{g}_I)}{\partial \mathbf{g}_I} \, d\mathbf{g}_I = \int_0^{\theta_{\alpha_1}} \cdots \int_0^{\theta_{\alpha_q}} \frac{\partial^q \xi_n(y + \mathbf{g}_I)}{\partial g_{\alpha_1} \cdots \partial g_{\alpha_q}} \, dg_{\alpha_1} \cdots dg_{\alpha_q}.$$

For the nonempty set I defined above, let $I_1 = I \cap \{1, 2, \cdots, p\}$ and $I_2 = I \cap \{p+1, p+2, \cdots, 2p\} - p$. Let $\chi_j(I) = X_{j,I_1}\alpha(t,j) - \mathbb{E}[X_{j,I_1}\alpha(t,j)]$ and $\lambda_j(I) = X_{j+k,I_2}\beta(s,j) - \mathbb{E}[X_{j+k,I_2}\beta(s,j)]$. If I_1 is empty, then let $X_{j,I_1} = 1$. The same rule applies to X_{j+k,I_2}. For $g = (t^T, s^T)^T \in \mathbb{R}^{2p}$, we have

$$\frac{\partial^q \xi_n(g)}{\partial g_{\alpha_1} \cdots \partial g_{\alpha_q}} = i^{|I|} \Big\{ \sum_{j=1}^{n-k} \frac{\chi_j(I)\lambda_j(I)}{\sqrt{n-k}} - \frac{\sum_{j=1}^{n-k} \chi_j(I) \sum_{j=1}^{n-k} \lambda_j(I)}{\sqrt{(n-k)^3}} \Big\}, \tag{41}$$

where $|I|$ denotes the number of elements in I. Let $r(q) = (2p+1)/(q+1)$. Then similar to the proof of (35), we have

$$\Big\| \frac{\partial^q \xi_n(g)}{\partial g_{\alpha_1} \cdots \partial g_{\alpha_q}} \Big\|_{r(q)} \le C \text{ uniformly over } D^*(\gamma). \tag{42}$$

Since $\| \max_{1\le j\le v(p)} |\frac{\partial^q \xi_n(o_j + \mathbf{g}_I)}{\partial \mathbf{g}_I}| \|_{r(q)}^{r(q)} \le \sum_{j=1}^{v(p)} \| \frac{\partial^q \xi_n(o_j + \mathbf{g}_I)}{\partial \mathbf{g}_I} \|_{r(q)}^{r(q)}$ and $v(p) \le C/\epsilon^{2p}$, we have by (42)

$$\Big\| \max_{1\le j\le v(p)} \Big| \frac{\partial^q \xi_n(o_j + \mathbf{g}_I)}{\partial \mathbf{g}_I} \Big| \Big\|_{r(q)} \le C\epsilon^{-2p/r(q)} \text{ uniformly over } D^*(\gamma). \tag{43}$$

Consequently,

$$\int_{-\epsilon_p}^{\epsilon_p}\cdots\int_{-\epsilon_p}^{\epsilon_p}\Big\|\max_{1\le j\le v(p)}\Big|\frac{\partial^q\xi_n(o_j+\mathbf{g}_I)}{\partial\mathbf{g}_I}\Big|\Big\|_{r(q)}\,d\mathbf{g}_I\le C\epsilon^q\epsilon^{-2p/r(q)}\le C\epsilon^\pi,$$

where $\pi=1/(2p+1)$ and $\epsilon_p=2\sqrt{2p}\epsilon$. Note that

$$\xi_n(g)-\xi_n(o_j)=\sum_{I\subseteq\{1,2,\cdots,2p\}}\int_0^{(g-o_j)_I}\frac{\partial^{|I|}\xi_n(o_j+\mathbf{g}_I)}{\partial\mathbf{g}_I}\,d\mathbf{g}_I.$$

Therefore

$$\begin{aligned}\|m(\epsilon)\| &\le \sum_{I\subseteq\{1,2,\cdots,2p\}}\int_{-\epsilon_p}^{\epsilon_p}\cdots\int_{-\epsilon_p}^{\epsilon_p}\Big\|\max_{1\le j\le v(p)}\Big|\frac{\partial^q\xi_n(o_j+\mathbf{g}_I)}{\partial\mathbf{g}_I}\Big|\Big\|\,d\mathbf{g}_I\\ &\le \sum_{I\subseteq\{1,2,\cdots,2p\}}\int_{-\epsilon_p}^{\epsilon_p}\cdots\int_{-\epsilon_p}^{\epsilon_p}\Big\|\max_{1\le j\le v(p)}\Big|\frac{\partial^q\xi_n(o_j+\mathbf{g}_I)}{\partial\mathbf{g}_I}\Big|\Big\|_{r(q)}\,d\mathbf{g}_I\\ &\le C\sum_{I\subseteq\{1,2,\cdots,2p\}}\epsilon^\pi\le C\epsilon^\pi.\end{aligned}$$

By Markov's inequality, we have

$$\mathbb{P}(m(\epsilon)\ge\epsilon^{\pi/2})\le\frac{\|m(\epsilon)\|^2}{\epsilon^\pi}\le C\epsilon^\pi\to 0. \tag{44}$$

Hence (40), (44) and Theorem 7.3 in Billingsley (1999) imply that $\xi_n(t,s)$ is tight on $D(\gamma)$. ♢

Proof of Theorem 2. By Lemma 4 and the continuous mapping theorem, we have for each fixed $\gamma>0$,

$$n\mathcal{V}_{n,\gamma}(k)\Rightarrow\frac{1}{c_p^2}\int_{D(\gamma)}\frac{|\xi(t,s)|^2}{|t|^{p+1}|s|^{p+1}}\,dtds.$$

Note that

$$\begin{aligned}n|\mathcal{V}_{n,\gamma}(k)-\mathcal{V}_X^n(k)| &\le \int_{|t|<\gamma}|\xi_n(t,s)|^2\,dw+\int_{|t|>1/\gamma}|\xi_n(t,s)|^2\,dw\\ &\quad+\int_{|s|<\gamma}|\xi_n(t,s)|^2\,dw+\int_{|s|>1/\gamma}|\xi_n(t,s)|^2\,dw.\end{aligned}$$

Therefore Lemmas 2 and 3 imply that Theorem 2 holds. ♢

Proof of Proposition 2. Note that $\mathcal{V}_X(k) = \int |\mathrm{Cov}(\alpha(t,0), \beta(s,0))|^2 \, dw$. Furthermore,

$$\begin{aligned} |\mathrm{Cov}(\alpha(t,0), \beta(s,0))| &= |\mathbb{E}\{\sum_{j\in\mathbb{Z}} \mathcal{P}_j\alpha(t,0) \sum_{j'\in\mathbb{Z}} \mathcal{P}_{j'}\beta(s,0)\}| = |\mathbb{E}\{\sum_{j\in\mathbb{Z}} \mathcal{P}_j\alpha(t,0)\mathcal{P}_j\beta(s,0)\}| \\ &\leq \sum_{j\in\mathbb{Z}} \|\mathcal{P}_j\alpha(t,0)\|\|\mathcal{P}_j\beta(s,0)\| = \sum_{j=-\infty}^{0} \|\mathcal{P}_j\alpha(t,0)\|\|\mathcal{P}_j\beta(s,0)\|. \quad (45) \end{aligned}$$

Now by the similar arguments as those in the proof of Lemmas 2 and 3, Proposition 2 follows. Details are omitted. ♢

Proof of Theorem 3. Theorem 3 follows from Proposition 3 below and very similar arguments as those in the proof of Theorem 1 of Beran (1984). See also Theorem 3.2.1 of Politis et. al (1999). Details are omitted. ♢

Proposition 3. *Let U be a random variable following the distribution on the right hand side of* (18)*. Then under the assumptions of Theorem 3, we have for any $x > 0$*

$$L_n(x) \to \mathbb{P}(U \leq x), \;\; \text{as } n \to \infty,$$

where $L_n(x) = \frac{1}{n-l+1}\sum_{j=1}^{n-l+1} I\{(l-k)\mathcal{V}^l_{X_{l,j}}(k) \leq x\}$ and $I\{\cdot\}$ is the indicator function.

Proof. Note that for any $x > 0$, x is a continuous point for the distribution of U. Proposition 3 follows by Lemmas 5, 6, 7 and Theorem 2 when letting $\gamma \to 0$, $m \to \infty$, $m\gamma^{2(p+1)} \to \infty$ and $m/n \to 0$. ♢

Lemma 5. *For $u = 1, 2, \cdots, n-l+1$ and $\gamma > 0$, define*

$$\mathcal{V}_{l,k}(u,\gamma) = \int_{D(\gamma)} \Big|\frac{1}{l-k}\sum_{j=u}^{u+l-k-1} \alpha(t,j)\beta(s,j) - \frac{1}{l-k}\sum_{r=u}^{u+l-k-1}\alpha(t,r)\frac{1}{l-k}\sum_{j=u}^{u+l-k-1}\beta(s,j)\Big|^2 dw.$$

Let $L_n(x,\gamma) = \frac{1}{n-l+1}\sum_{u=1}^{n-l+1} I\{(l-k)\mathcal{V}_{l,k}(u,\gamma) \leq x\}$. Then under the assumptions of Theorem 3, we have

$$L_n(x) \leq L_n(x,\gamma) \quad \text{and} \quad \mathbb{P}[L_n(x) - L_n(x-\gamma^{1/6},\gamma) \leq -\gamma^{1/12}] \leq C_1\gamma^{1/12},$$

where C_1 is a positive constant which does not depend on n and γ.

Proof. Obviously $\mathcal{V}^l_{X_{l,u}}(k) \geq \mathcal{V}_{l,k}(u,\gamma)$, $u = 1, 2, \cdots, n-l+1$. Therefore $L_n(x) \leq L_n(x,\gamma)$. On the other hand, note that, for random variables X and Y and any $x \in \mathbb{R}$ and $\delta > 0$, we have

$$I\{X \leq x\} - I\{Y \leq x - \delta\} \geq -I\{X - Y \geq \delta\}. \tag{46}$$

Therefore

$$L_n(x) - L_n(x - \gamma^{1/6}, \gamma) \geq -\frac{1}{n-l+1} \sum_{u=1}^{n-l+1} I\{(l-k)[\mathcal{V}^l_{X_{l,u}}(k) - \mathcal{V}_{l,k}(u,\gamma)] \geq \gamma^{1/6}\} := -U_n(x,\gamma).$$

Hence by Markov's inequality, it follows that

$$\mathbb{P}[L_n(x) - L_n(x - \gamma^{1/6}, \gamma) \leq -\gamma^{1/12}] \leq \mathbb{P}\{-U_n(x,\gamma) \leq -\gamma^{1/12}\} \leq \frac{\mathbb{E}[U_n(x,\gamma)]}{\gamma^{1/12}}. \tag{47}$$

For each u, by Lemmas 2 and 3 and the Markov's inequality, we have

$$\begin{aligned}
\mathbb{E}[I\{(l-k)[\mathcal{V}^l_{X_{l,u}}(k) - \mathcal{V}_{l,k}(u,\gamma)] \geq \gamma^{1/6}\}] &= \mathbb{P}\{(l-k)[\mathcal{V}^l_{X_{l,u}}(k) - \mathcal{V}_{l,k}(u,\gamma)] \geq \gamma^{1/6}\} \\
&\leq \mathbb{E}\{(l-k)[\mathcal{V}^l_{X_{l,u}}(k) - \mathcal{V}_{l,k}(u,\gamma)]\}/\gamma^{1/6} \\
&\leq C_1\gamma^{1/3}/\gamma^{1/6} = C_1\gamma^{1/6},
\end{aligned} \tag{48}$$

where C_1 is a positive constant which does not depend on n and γ. Therefore $\mathbb{E}[U_n(x,\gamma)] \leq C_1\gamma^{1/6}$. Now by (47), the Lemma follows. ♢

For an integer $m > 0$ and $j = 1, 2, \cdots, n-k$, write

$$\alpha_m(t,j) := \mathbb{E}[\alpha(t,j)|\varepsilon_{j,m}], \quad \beta_m(t,j) := \mathbb{E}[\beta(t,j)|\varepsilon_{j,m}], \quad \eta_m(t,s,j) := \mathbb{E}[\eta(t,s,j)|\varepsilon_{j,m}], \tag{49}$$

where $\eta(t,s,j) = \alpha(t,j)\beta(s,j)$ and $\varepsilon_{j,m} = (\varepsilon_j, \varepsilon_{j-1}, \cdots, \varepsilon_{j-m})$. For $u = 1, 2, \cdots, n-l+1$, let

$$\mathcal{V}_{l,k}(u,\gamma,m) = \int_{D(\gamma)} \Big| \frac{1}{l-k} \sum_{j=u}^{u+l-k-1} \eta_m(t,s,j) - \frac{1}{l-k} \sum_{r=u}^{u+l-k-1} \alpha_m(t,r) \frac{1}{l-k} \sum_{j=u}^{u+l-k-1} \beta_m(s,j) \Big|^2 dw.$$

Write $L_n(x,\gamma,m) = \frac{1}{n-l+1} \sum_{u=1}^{n-l+1} I\{(l-k)\mathcal{V}_{l,k}(u,\gamma,m) \leq x\}$.

Lemma 6. *Under the assumptions of Theorem 3, we have*

$$\begin{aligned}
\mathbb{P}[L_n(x,\gamma) - L_n(x - m^{-1}/\gamma^{2(p+1)}, \gamma, m) \leq -m^{-1/2}/\gamma^{p+1}] &\leq C_2 m^{-1/2}/\gamma^{p+1}, \\
\mathbb{P}[L_n(x + m^{-1}/\gamma^{2(p+1)}, \gamma, m) - L_n(x,\gamma) \leq -m^{-1/2}/\gamma^{p+1}] &\leq C_2 m^{-1/2}/\gamma^{p+1}.
\end{aligned} \tag{50}$$

where C_2 is a positive constant which does not depend on n and γ.

Proof. For $u = 1, 2, \cdots, n-l+1$, write $S^*(u,l) = S_1(u,l) - S_2(u,l)S_3(u,l)$ and $S^*_m(u,l) = S_{1,m}(u,l) - S_{2,m}(u,l)S_{3,m}(u,l)$, where

$$\begin{aligned} S_1(u,l) = \sum_{j=u}^{u+l-k-1} \eta(t,s,j)/(l-k), \quad S_{1,m}(u,l) = \sum_{j=u}^{u+l-k-1} \eta_m(t,s,j)/(l-k), \\ S_2(u,l) = \sum_{j=u}^{u+l-k-1} \alpha(t,j)/(l-k), \quad S_{2,m}(u,l) = \sum_{j=u}^{u+l-k-1} \alpha_m(t,j)/(l-k), \\ S_3(u,l) = \sum_{j=u}^{u+l-k-1} \beta(s,j)/(l-k), \quad S_{3,m}(u,l) = \sum_{j=u}^{u+l-k-1} \beta_m(s,j)/(l-k). \end{aligned}$$

Then for each u, by Cauchy's inequality,

$$\begin{aligned} \|\mathcal{V}_{l,k}(u,\gamma) - \mathcal{V}_{l,k}(u,\gamma,m)\| &= \| \int_{D(\gamma)} |S^*(u,l)|^2 \, dw - \int_{D(\gamma)} |S^*_m(u,l)|^2 \, dw \| \\ &\leq \int_{D(\gamma)} \| |S^*(u,l)|^2 - |S^*_m(u,l)|^2 \| \, dw \\ &\leq \int_{D(\gamma)} \|S^*(u,l) - S^*_m(u,l)\|_4 [\|S^*(u,l)\|_4 + \|S^*_m(u,l)\|_4] \, dw \end{aligned} \tag{51}$$

By Lemma A.1.(i) of Liu and Lin (2009) with $q = 4$ and Lemma 1 with $q = 4$ and $\eta = 1$, we have

$$\sqrt{l-k}\|S_1(u,l) - S_{1,m}(u,l)\|_4 \leq C(p) \sum_{j=m}^{\infty} \{|t|\delta(j,4) + |s|\delta(j+k,4)\} \leq C(p)m^{-2}/\gamma \tag{52}$$

for any $(t,s) \in D(\gamma)$, where $C(p)$ is a constant which only depends on p and can vary from place to place. Analogous results hold for $S_j(u,l) - S_{j,m}(u,l)$, $j = 2, 3$. On the other hand, by the similar Martingale decomposition method used in the proof of Theorem 1, it follows that

$$\max_{j=1,2,3} \sqrt{l-k}\{\|S_j(u,l)\|_4 + \|S_{j,m}(u,l)\|_4\} \leq C(p)/\gamma. \tag{53}$$

Therefore plugging (52) and (53) into (51), we have

$$(l-k)\|\mathcal{V}_{l,k}(u,\gamma) - \mathcal{V}_{l,k}(u,\gamma,m)\| \leq \int_{D(\gamma)} C(p)m^{-2}/\gamma^2 \, dw \leq C(p)m^{-2}/\gamma^{4(p+1)}.$$

Now similar arguments as those in the proof of Lemma 5 leads to (50). Details are omitted. $\diamondsuit$

Lemma 7. *Assume that $m/n \to 0$. Then under the assumptions of Theorem 3, we have*

$$L_n(x,\gamma,m) \to \mathbb{P}\{(l-k)\mathcal{V}_{l,k}(u,\gamma,m) \le x\} \text{ in probability.}$$

Proof. It suffices to show that $\|L_n(x,\gamma,m) - \mathbb{E}L_n(x,\gamma,m)\| \to 0$ as $n \to \infty$. Let $\rho(u,x) := I\{\mathcal{V}_{l,k}(u,\gamma,m) \le x\}$. Since $\{\rho(u,x)\}_{u=1}^{n-l+1}$ is $m+l$-dependent and $0 \le \rho(u,x) \le 1$, we have

$$\begin{aligned}(n-l+1)^2 \mathrm{Var}(L_n(x,\gamma,m)) &= \sum_{|j-r|\le m+l} \mathrm{Cov}(\rho(j,x),\rho(r,x)) + \sum_{|j-r|>m+l} \mathrm{Cov}(\rho(j,x),\rho(r,x)) \\ &= \sum_{|j-r|\le m+l} \mathrm{Cov}(\rho(j,x),\rho(r,x)) \le 2(m+l)^2. \qquad (54)\end{aligned}$$

Since $m/n \to 0$ and $l/n \to 0$, the lemma follows. $\diamondsuit$

REFERENCES

Ashley, R. A., Patterson, D. M. and Hinich, M. J. (1986). A diagnostic test for nonlinear serial dependence in time series fitting errors. *Journal of Time Series Analysis.* **7** 165–178.

Bagnato, L., Punzo, A. and Nicolis, O. (2011). The autodependogram: a graphical device to investigate serial dependences. *Journal of Time Series Analysis.* To appear. DOI: 10.1111/j.1467-9892.2011.00754.x

Beran, R. (1984). Bootstrap methods in statistics. *Jahresberichte des Deutschen Mathematischen Vereins* **86** 14-30.

Billingsley, P. (1999). *Convergence of Probability Measures. second edition.* Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307-327.

Box, G. E. P. and Jenkins, G. M. (1970). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Brillinger, D. R. (1977). The identification of a particular nonlinear time series system. *Biometrika* **64** 509–515.

Brockwell, P. (2007). Beyond linear time series. *Statistica Sinica* **17** 3-7.

Chen, R. and Tsay, R. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association.* **88** 298–308.

Davis, R.A. and Mikosch, T. (1998). The sample ACF of heavy-tailed stationary processes with applications to ARCH. *Annals of Statistics* **26** 2049–2080.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* **50** 987-1007.

Fan, J. and Yao, Q. (2003). *Nonlinear time series, nonparametric and parametric methods.* Springer, New York.

Franses, P. H. and van Dijk, D. (2000). *Nonlinear time series models in empirical fiance.* Cambridge University Press.

Granger, C. W. J. and Andersen, A. (1978). Non-linear time series modelling. In *Applied time series analysis*, ed. D. F. Findley, Academic Press, New York, 25-38.

Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications* **119** 249–280.

Kantz, H. and Schreiber, T. (2000). *Nonlinear time series analysis.* Cambridge University Press.

Kuo, H. H. (1975). Gaussian measures in Banach spaces. *Lecture notes in Math.* **463** Springer, Berlin.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* Springer, New York.

Priestley, M. B. (1988). *Non-linear and non-stationary time series.* Academic Press.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian models in finance.* Wiley, Chichester.

Rémillard, B. (2009). Discussion of: Brownian distance covariance. *Annals of Applied Statistics*, **3** 1295–1298.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences, USA* **42** 43-47.

Szekely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance (with discussions). *Annals of Applied Statistics* **3** 1236-1308.

Szekely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing independence by correlation of distances. *Ann. Statist.* **35** 2769-2794.

Tjøstheim, D. and Auestad, B. H. (1994a). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association.* **89** 1398–1409.

Tjøstheim, D. and Auestad, B. H. (1994b). Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association.* **89** 1410–1419.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press, Oxford, UK.

Tsay, R. S. (2005). *Analysis of financial time series*, second edition. Wiley-Interscience.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences USA*, **102**, 14150–14154.