On Auto-Distance Correlation Matrix

Maria Pitsillou *[1] and Konstantinos Fokianos[1]

[1] *Department of Mathematics & Statistics, University of Cyprus*

We introduce the notions of auto-distance covariance and correlation matrices for multivariate time series and give their consistent estimators. In addition, a testing methodology for testing the i.i.d. hypothesis for multivariate time series data is developed. The resulting test statistic is compared with the related multivariate Ljung-Box statistic in a real data example.

Keywords: Characteristic function ; Correlation ; Stationarity ; U-statistic ; Wild Bootstrap

1 Introduction

There has been a considerable recent interest in measuring dependence by employing the concept of distance covariance function, a new measure of dependence for random variables, introduced by Székely et al. (2007). This tool has been recently defined to the context of multivariate time series by Zhou (2012), but without exploring the interrelationships between the various time series components. In this paper, we extend the notion of distance covariance to multivariate time series by defining its matrix version. Based on this new concept, we develop a multivariate testing methodology for testing independence.

2 Auto-distance covariance matrix

We denote by $\{\mathbf{X}_t : t = 0, \pm 1, \pm 2, \dots\}$ a d-dimensional time series process, with components $X_{t;r}$, $r = 1, \dots, d$. Suppose we have available a sample of size n, that is $\{\mathbf{X}_t, t = 1, \dots, n\}$. We define the pairwise auto-distance covariance function as a function of the joint and marginal characteristic functions of the pair $(X_{t;r}, X_{t+j;m})$, for $r, m = 1, \dots, d$. Denote by $\phi_j^{(r,m)}(u, v)$ the joint characteristic function of $X_{t;r}$ and $X_{t+j;m}$; that is

$$\phi_j^{(r,m)}(u, v) = E\left[\exp\left(i(uX_{t;r} + vX_{t+j;m})\right)\right], \quad j \in \mathbb{Z},$$

*Corresponding author: pitsillou.maria@ucy.ac.cy

and the marginal characteristic functions of $X_{t;r}$ and $X_{t+j;m}$ as $\phi^{(r)}(u) := \phi_j^{(r,m)}(u,0)$ and $\phi^{(m)}(v) := \phi_j^{(r,m)}(0,v)$ respectively, where $(u,v) \in \mathbb{R}^2$, and $i^2 = -1$. The *pairwise auto-distance covariance function (ADCV)* between $X_{t;r}$ and $X_{t+j;m}$, $V_{rm}(j)$, is defined as the positive square root of

$$V_{rm}^2(j) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{\left|\phi_j^{(r,m)}(u,v) - \phi^{(r)}(u)\phi^{(m)}(v)\right|^2}{|u|^2 \, |v|^2} du dv, \quad j \in \mathbb{Z}.$$

The auto-distance covariance matrix, $V(j)$, is then defined by

$$V(j) = \left[V_{rm}(j)\right]_{r,m=1}^d, \quad j \in \mathbb{Z}.$$

The *pairwise auto-distance correlation function (ADCF)* between $X_{t;r}$ and $X_{t+j;m}$, $R_{rm}(j)$, is a coefficient that lies in the interval $[0,1]$ and also measures dependence and is defined as the positive square root of

$$R_{rm}^2(j) = \frac{V_{rm}^2(j)}{\sqrt{V_{rr}^2(0)}\sqrt{V_{mm}^2(0)}},$$

for $V_{rr}(0)V_{mm}(0) \neq 0$ and zero otherwise. The auto-distance correlation matrix of $\mathbf{X}_t$, is then defined as

$$R(j) = \left[R_{rm}(j)\right]_{r,m=1}^d, \quad j \in \mathbb{Z}.$$

When $j \neq 0$, $V_{rm}(j)$ measures the dependence of $X_{t;r}$ on $X_{t+j;m}$. In general, $V_{rm}(j) \neq V_{mr}(j)$ for $r \neq m$, since they measure different dependence structure between the series $\{X_{t;r}\}$ and $\{X_{t;m}\}$ for all $r, m = 1, 2, \ldots, d$. Thus, $V(j)$ and $R(j)$ are non-symmetric matrices, but $V(-j) = V'(j)$ and $R(-j) = R'(j)$. More properties can be found in Fokianos and Pitsillou (2017b). The empirical pairwise ADCV, $\widehat{V}_{rm}(j)$, for $j \geq 0$, is the non-negative square root of

$$\widehat{V}_{rm}^2(j) = \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} A_{ts}^r B_{ts}^m,$$

where $A^r = A_{ts}$ and $B^m = B_{ts}$ are Euclidean distance matrices given by

$$A_{ts}^r = a_{ts}^r - \bar{a}_{t.}^r - \bar{a}_{.s}^r + \bar{a}_{..}^r,$$

with $a_{ts}^r = |X_{t;r} - X_{s;r}|$, $\bar{a}_{t.}^r = \left(\sum_{s=1}^{n-j} a_{ts}^r\right)/(n-j)$, $\bar{a}_{.s}^r = \left(\sum_{t=1}^{n-j} a_{ts}^r\right)/(n-j)$, $\bar{a}_{..}^r = \left(\sum_{t,s=1}^{n-j} a_{ts}^r\right)/(n-j)^2$. B_{ts}^m is defined analogously in terms of $b_{ts}^m = |X_{t+j;m} - X_{s+j;m}|$.

Fokianos and Pitsillou (2017b) showed that for a d-dimensional strictly stationary and ergodic process $\{\mathbf{X}_t\}$ with $E\left|X_{t;r}\right|^2 < \infty$, for $r = 1, \ldots, d$, then for all $j \in \mathbb{Z}$,

$$\widehat{V}(j) \to V(j),$$

almost surely, as $n \to \infty$. In addition, under pairwise independence it holds that

$$n\widehat{V}_{rm}^2(j) \to Z := \sum_k \lambda_k Z_k^2,$$

in distribution, as $n \to \infty$, where $\{Z_k\}$ is an i.i.d sequence of $N(0,1)$ random variables, and (λ_k) is a sequence of nonzero eigenvalues.

3 The testing problem

In this section, we develop a test statistic for testing the null hypothesis that $\{\mathbf{X}_t\}$ is an i.i.d. sequence. Following Hong's (1999) generalized spectral domain methodology, we first consider the generalized spectral density matrix

$$F(\omega, u, v) = \left[f^{(r,m)}(\omega, u, v)\right]_{r,m=1}^{d},$$

where

$$f^{(r,m)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(r,m)}(u, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi],$$

with p denoting the bandwidth parameter. Under the null hypothesis of independence, $F(\cdot, \cdot, \cdot)$ reduces to

$$F_0(\omega, u, v) = \frac{1}{2\pi}\left[\sigma_0^{(r,m)}(u, v)\right]_{r,m=1}^{d}.$$

Thus, comparing the Parzen's (1957) kernel-type estimators $\hat{F}(\omega, u, v)$ and $\hat{F}_0(\omega, u, v)$ via a Frobenious norm we result to a test statistic based on the ADCV matrix, given by

$$\widetilde{T}_n = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \mathrm{tr}\{\widehat{V}^*(j)\widehat{V}(j)\}, \tag{1}$$

where $k(\cdot)$ is a univariate kernel function satisfying some standard properties. Moreover, $\widehat{V}^*(\cdot)$ denotes the complex conjugate matrix of $\widehat{V}(\cdot)$ and $\mathrm{tr}(A)$

denotes the trace of the matrix A. Fokianos and Pitsillou (2017b) also formed a similar test statistic in terms of the ADCF matrix, given by

$$\overline{T}_n = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \mathrm{tr}\{\widehat{V}^*(j) \widehat{D}^{-1} \widehat{V}(j) \widehat{D}^{-1}\}. \quad (2)$$

Under the null hypothesis of independence and some further assumptions about the kernel function $k(\cdot)$, the standardized version of the test statistics $\widetilde{T}_n$ and $\overline{T}_n$ given in (1) and (2) were proved to follow $N(0,1)$ asymptotically and they are consistent. Fokianos and Pitsillou (2017a) developed a similar testing methodology based on ADCV/ADCF for testing serial dependence in a univariate strictly stationary time series setting.

4 Real data example

In this section we apply the proposed testing methodology to the monthly log returns of the stocks of IBM and the S&P 500 composite index starting from 29 May 1936 to 28 November 1975 for 474 observations. A larger data set and the aforementioned testing methodology are included in the R package **dCovTS** (Pitsillou and Fokianos, 2016). Assuming that the bivariate series follows a VAR model and employing the AIC to choose its best order, we obtain that a VAR(2) model fits well the data. Figure 1 shows the ADCF plot of the residuals after fitting a VAR(2) model to the original series. Based on this plot, the residuals of VAR(2) model do not have any strong dependence. The shown critical values (dotted horizontal line) are the 95% simultaneous critical values computed based on an algorithm suggested by Fokianos and Pitsillou (2017b) using the independent wild bootstrap approach (Dehling and Mikosch, 1994; Shao, 2010; Leucht and Neumann, 2013). To formally confirm the adequacy of this model fit, we perform tests of independence among the residuals for the following bandwidth values, $p = 6$, 11 and 20. The proposed statistic $\overline{T}_n$ and the related multivariate Ljung-Box statistic (Hosking, 1980) both give large p-values (0.254, 0.190, 0.098 and 0.958, 0.809, 0.811 respectively) suggesting the absence of any serial dependence among the residuals. The calculation of the statistic $\overline{T}_n$ is based on the Bartlett kernel. The computation of the p-values is based on 499 independent wild bootstrap realizations.

Acknowledgements: Financial support from a University of Cyprus research grant is greatly acknowledged.

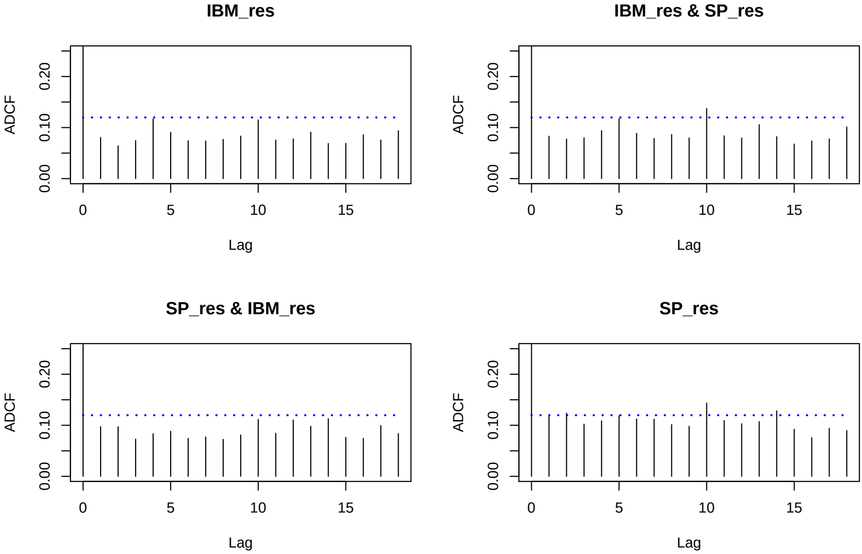

Figure 1: The sample ADCF of the residuals after fitting VAR(2) model to the bivariate series IBM and S&P500.

References

Dehling, H. and T. Mikosch (1994). Random quadratic forms and the boostrap for U-statistics. *Journal of Multivariate Analysis* **51**, 392–413.

Fokianos, K. and M. Pitsillou (2017a). Consistent testing for pairwise dependence in time series. *Technometrics* **59**, 262–270.

Fokianos, K. and M. Pitsillou (2017b). Testing pairwise independence for multivariate time series by the auto-distance correlation matrix. Under revision.

Hong, Y. (1999). Hypothesis testing in time series via the emprical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* **94**, 1201–1220.

Hosking, J. R. M. (1980). Multivariate Portmanteau statistic. *Journal of the American Statistical Association* **75**, 349–386.

Leucht, A. and M. H. Neumann (2013). Dependent wild bootstrap for degenerate U- and V- statistics. *Journal of Multivariate Analysis* **117**, 257–280.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* **28**, 329–348.

Pitsillou, M. and K. Fokianos (2016). dCovTS: Distance covariance and correlation for time series analysis. *The R Journal* **8**, 324–340.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association* **105**, 218–235.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis* **33**, 438–457.