Journal of Statistical and Econometric Methods, vol.4, no.1, 2015, 63-71
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2015

Some basic properties of cross-correlation functions of n-dimensional vector time series

Anthony E. Usoro[1]

Abstract

In this work, cross-correlation function of multivariate time series was the interest. The design of cross-correlation function at different lags was presented. $\gamma_{X_{it+k}X_{jt+l}}$ is the matrix of the cross-covariance functions, $\gamma_{X_{it}}$ and $\gamma_{X_{jt}}$ are the variances of X_{it} and X_{jt}vectors respectively. Vector cross-correlation function was derived as $\rho_{X_{it+k},X_{jt+l}} = \frac{\gamma_{X_{it+k}X_{jt+l}}}{\sqrt{\gamma_{X_{it}}\gamma_{X_{jt}}}}$. A statistical package was used to verify the vector cross correlation functions, with trivariate analysis as a special case. From the results, some properties of vector cross-correlation functions were established.

Keywords: Vector time series; cross-covariance function and cross-correlation function

[1] Department of Mathematics and Statistics, Akwa Ibom State University, Mkpat Enin, Akwa Ibom State. Email: toskila2@yahoo.com

Article Info: *Received* : January 17, 2015. *Revised* : February 28, 2015.
Published online : March 31, 2015.

1 Introduction

In statistics, the term cross-covariance is sometimes used to refer to the covariance corr(X,Y) between two random vectors X and Y, (where $X = X_1, X_2, \ldots, X_n\ and\ Y = Y_1, Y_2, \ldots, Y_n$). In signal processing, the cross-covariance is often called cross-correlation and is a measure of similarity of two signals, commonly used to find features in an unknown signal by comparing it to a known one. It is a function of the relative time between the signals, and it is sometimes called the sliding dot product. In univariate time series, the autocorrelation of a random process describes the correlation between values of the process at different points in time, as a function of the two times or of the time difference. Let X be some repeatable process, and i be some point in time after the start of that process. (i may be an integer for a discrete-time process or a real number for a continuous-time process.) Then X_i is the value (or realization) produced by a given run of the process at time i. Suppose that the process is further known to have defined values for mean μ_i and variance σ_i^2 for all times i. Then the definition of the autocorrelation between times s and t is

$$\rho(_{s,t}) = \frac{E\{(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s},$$

where "E" is the expected value operator. It is required to note that the above expression is not well-defined for all time series or processes, because the variance may be zero. If the function ρ is well-defined, its value must lie in the range [-1,1], with 1 indicating perfect correlation and -1 indicating perfect anti-correlation. If X is a second-order stationary process then the mean μ and the variance σ^2 are time-independent, and further the autocorrelation depends only on the difference between t and s: the correlation depends only on the time-distance between the pair of values but not on their position in time. This further implies that the autocorrelation can be expressed as a function of the time-lag, and that this would be an even function of the lag $k = s - t$, which implies $s = t + k$. This gives the more familiar form,

$$\rho_k = \frac{E[(X_{t-}\mu)(X_{t+k-}\mu)]}{\sqrt{E[(X_{t-}\mu)^2]E[(X_{t+k-}\mu)^2]}} = \frac{E[(X_{t-}\mu)(X_{t+k-}\mu)]}{\sigma^2}$$

where $X_t$and X_{t+k} are time series process at lag k time difference. Hence, autocovariance coefficient γ_k at lag k, measures the covariance between two values Z_t and Z_{t+k}, a distance k apart. The autocorrelation coefficient $\rho_k$is defined as the autocovariance γ_k at lag k divided by variance $\gamma_{0(k=0)}$. The plot of γ_k against lag k is called the autocovariance function (γ_k), while the plot of ρ_k against lag k is called the autocorrelation function (Box and Jenkins 1976).

In multivariate time series, cross-correlation or covariance involves more than one process. For instance, X_t and Y_t are two processes of which X_t could be cross-correlated with Y_t at lag k. The lag k value return by $ccf(X, Y)$ estimates the correlation between $X(t+k)$ *and* $Y(t)$, Venables and Ripley (2002). Storch and Zwiers (2001) described cross-correlation in signal processing and time series. In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time lag applied to one of them. This is also known as a sliding dot product or sliding inner-product. It is commonly used for searching a -duration signal for a shorter known feature. It also has application in pattern recognition, signal particle analysis, electron tomographic averaging, cryptanalysis and neurophysiology. In autocorrelation, which is the cross-correlation of a signal with itself, there is always a peak at a lag of zero unless the signal is a trivial zero signal. In probability theory and Statistics, correlation is always used to include a standardising factor in such a way that correlations have values between -1 and 1. Let (X_t, Y_t) represent a pair of stochastic process that are jointly wide sense stationary. Then the cross covariance given by Box et al (1984) is

$$\gamma_{xy}(\tau) = E[(X_t - \mu_x)(Y_{t+\tau} - \mu_y)],$$

where μ_x and μ_y are the means of X_t and Y_t respectively. The cross-correlation function ρ_{xy} is the normalized cross-covariance function. Therefore,

$$\rho_{xy}(\tau) = \frac{\gamma_{xy}(\tau)}{\sigma_x \sigma_y}$$

where σ_x and σ_y are the standard deviation of processes X_t and Y_t respectively. If $X_t = Y_t$ for all t, then the cross-correlation function is simply the autocorrelation function for a discrete process of length n defined as $\{X_1, \dots, X_n\}$ which known mean and variance, an estimate of the autocorrelation may be obtained as

$$\hat{R}_{(k)} = \frac{1}{(n-k)\sigma^2} \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu)$$

for any positive integer k<n, Patrick (2005). When the true mean μ and variance σ^2 are known, the estimate is unbiased. If the true mean, this estimate is unbiased. If the true mean and variance of the process are not known, there are several probabilities:

i. if μ and σ^2 are replaced by the standard formulas for sample mean and sample variance, then this is a biased estimate.

ii. if n-k in the above formula is replaced with n, the estimate is biased. However, it usually has a smaller mean square error, Priestly (1982) and Donald and Walden (1993).

iii. if X_t is stationary process, then the following are true

$$\mu_t = \mu_s = \mu \text{, for all t,s} \quad \text{and} \quad C_{xx(t,s)} = C_{xx(s-t)} = C_{xx(T)},$$

where T=s-t, is the lag time or the moment of time by which the signal has been shifted. As a result, the autocovariance becomes

$$C_{xx}(T) = E[(X_{(t)} - \mu)(X_{(t+T)} - \mu)] = E[X_{(t)} X_{(t+T)}] - \mu^2 = R_{xx(T)} - \mu^2,$$

where R_{xx} represents the autocorrelation in the signal processing sense.

$R_{xx}(T) = \frac{C_{xx(T)}}{\sigma^2}$, Hoel (1984).

For X_t and Y_t, the following properties hold:

1. $\rho_{xy(h)} \leq 1$

2. $\rho_{xy(h)} = \rho_{xy(-h)}$

3. $\rho_{xy(0)} \neq 1$

4. $\rho_{xy(h)} = \frac{\gamma_{xy(h)}}{\sqrt{\gamma_{x(0)}\gamma_{y(0)}}}$

Mardia and Goodall (1993) defined separable cross-correlation function as

$$C_{ij}(X_1, X_2) = \rho(X_1, X_2)a_{ij},$$

where $A = [a_{ij}]$ is a $p \times p$ positive definite matrix and $\rho(.,.)$ is a valid correlation function. Goulard & Voltz (1992); Wackernage (2003); Ver Hoef and Barry (1998) implied that the cross- covariance function is

$$C_{ij}(X_1 - X_2) = \sum_{k=1}^{r} \rho_k\,(X_1 - X_2)a_{ik}a_{jk},$$

for an integer 1$\leq r \leq p$, where $A = [a_{ij}]$ is a $p \times r$ full rank matrix and $\rho_{k(.)}$ are valid stationary correlation functions. Apanasovich and Genton (2010) constructed valid parametric cross-covariance functions. Apanasovich and Genton proposed a simple methodology based on latent dimensions and existing covariance models for univariate covariance, to develop flexible, interpretable and computationally feasible classes of cross-covariance functions in closed forms. They discussed estimation of the models and performed a small simulation study to demonstrate the models. The interest in this work is to extend cross-correlation functions beyond a-two variable case, present the multivariate design of vector cross-covariance and correlation functions and therefore establish some basic properties of vector cross-correlation functions from the analysis of vector cross-correlation functions.

2 The design of cross-covariance cross-correlation functions

The matrix of cross-covariance functions is as shown below:

$$\begin{matrix} \gamma_{X_{(1t+k)},X_{(1t+l)}} & \gamma_{X_{(1t+k)},X_{(2t+l)}} & \gamma_{X_{(1t+k)},X_{(3t+l)}} & \cdots & \gamma_{X_{(1t+k)},X_{(nt+l)}} \\ \gamma_{X_{(2t+k)},X_{(1t+l)}} & \gamma_{X_{(2t+k)},X_{(2t+l)}} & \gamma_{X_{(2t+k)},X_{(3t+l)}} & \cdots & \gamma_{X_{(2t+k)},X_{(nt+l)}} \\ \gamma_{X_{(3t+k)},X_{(1t+l)}} & \gamma_{X_{(3t+k)},X_{(2t+l)}} & \gamma_{X_{(3t+k)},X_{(3t+l)}} & \cdots & \gamma_{X_{(3t+k)},X_{(nt+l)}} \\ \vdots & & & & \vdots \\ \gamma_{X_{(mt+k)},X_{(1t+l)}} & \gamma_{X_{(mt+k)},X_{(2t+l)}} & \gamma_{X_{(mt+k)},X_{(3t+l)}} & \cdots & \gamma_{X_{(mt+k)},X_{(nt+l)}} \end{matrix}$$

where $k = 0, \ldots, a, l = 0, \ldots, b.$

The above matrix is a square matrix, and could be reduced to the form,

$$\gamma_{X_{(it+k)},X_{(jt+l)}}$$

where $i = 1, \ldots, m, j = 1, \ldots, n, k = 0, \ldots, a, l = 0, \ldots, b,\ (n = m).$
From the above cross-covariance matrix,

$$\rho_{X_{it+k},X_{jt+l}} = \frac{\gamma_{X_{it+k}X_{jt+l}}}{\sqrt{\gamma_{X_{it}}\gamma_{X_{jt}}}},$$

where, $\gamma_{X_{it+k}X_{jt+l}}$ is the matrix of the cross-covariance functions, $\gamma_{X_{it}}$ and $\gamma_{X_{jt}}$ are the variances of X_{it} and X_{jt}vectors respectively. Given the above matrix, it is required to note that two vector processes X_{it+k} and X_{jt+l} can only be cross-correlated at different lags, if either $k\ lag$ of X_{it} or $l\ lag$ of X_{jt} has a fixed value zero. That is X_{it} can be cross-correlated with $X_{jt+l}(l \pm 1,2, \ldots, b)$, or X_{jt}can be cross-correlated with $X_{it+k}(k \pm 1,2, \ldots, a)$.

3 Analysis of the cross-correlation functions

Given two processes $X_{1t}\ and\ X_{2t}$, $\rho_{(X_{1t},x_{2t+k})}$ is the cross-correlation between $X_{1t}\ and\ X_{2t}\ at\ lag\ k$, while, $\rho_{(x_{2t},x_{1t+k})}$ is the cross-correlation between $X_{2t}\ and\ X_{1t}\ at\ lag\ k$, Box et al (1984). In this work, three vector processes $X_{1t},\ X_{2t}\ and\ X_{3t}$ are used to carry out the cross-correlation analysis. For $k = 0, \pm 1, 2, \ldots, 4$, the following results were obtained with a software:

Lag k	$\rho_{(x_{1t},x_{2t+k})}$	$\rho_{(x_{2t},x_{1t+k})}$	$\rho_{(x_{1t},x_{3t+k})}$	$\rho_{(x_{3t},x_{1t+k})}$	$\rho_{(x_{2t},x_{3t+k})}$	$\rho_{(x_{3t},x_{2t+k})}$
-4	-0.172	0.572	-0.102	0.643	-0.427	-0.350
-3	-0.517	0.405	-0.501	0.410	-0.076	0.042
-2	-0.611	0.098	-0.662	0.067	0.327	0.399
-1	-0.605	-0.290	-0.674	-0.303	0.659	0.697
0	-0.506	-0.506	-0.578	-0.578	0.900	0.900
1	-0.290	-0.605	-0.303	-0.674	0.697	0.659
2	0.098	-0.611	0.067	-0.662	0.399	0.327
3	0.405	-0.517	0.410	-0.501	0.042	-0.076
4	0.572	-0.172	0.643	-0.102	-0.350	-0.427

From the above analysis, the following properties were established:

1. a. $\rho_{X_{it+k},X_{jt+l}} \neq 1, for\ \mathrm{k} = 0, l = \pm 1, \ldots, \pm \mathrm{b},\ i \neq j,$

   b. $\rho_{X_{it+k},X_{jt+l}} \neq 1, for\ l = 0, k = \pm 1, \ldots, \pm \mathrm{a},\ i \neq j.$

2. a. $\rho_{X_{it+k},X_{jt+l}} \neq \rho_{X_{it+k},X_{jt-l}}$, for $\mathrm{k} = 0, l = 1, \ldots, \mathrm{b},\ i \neq j,$

   b. $\rho_{X_{it+k},X_{jt+l}} \neq \rho_{X_{it-k},X_{jt+l}}$, for $l = 0, \mathrm{k} = 1, \ldots, \mathrm{a},\ i \neq j.$

3. a. $\rho_{X_{it+k},X_{jt+l}} = \rho_{X_{jt+k},X_{it-l}}$, for $\mathrm{k} = 0, l = 1, \ldots, \mathrm{b},\ i \neq j,$

   b. $\rho_{X_{it+k},X_{jt+l}} = \rho_{X_{jt-k},X_{it+l}}$, for $l = 0, k = 1, \ldots, \mathrm{a},\ i \neq j.$

4 Conclusion

The motivation behind this research work was to carry out cross-correlation functions of multivariate time series. Ordinarily, cross-correlation compares two series by shifting one of them relative to the other. In the case of X and Y variables, the variable X may be cross-correlated at different lags of Y, and vice versa. In this work, X_{it} and X_{jt} were used as vector time series, using trivariate as a special case of multivariate cross-correlation functions. The design of the cross-covariance functions has been displayed in a matrix form. Estimates obtained revealed some basic properties of vector cross-correlation functions.

References

[1] Tatiyana V. Apanasovich and Marc G. Genton, Cross-Covariance functions for multivariate random fields based on latent dimensions, *Biometrika*, **97**(1), (2010), 15-30.

[2] G.E.P. Box and G.M. Jenkins, *Time Series Analysis; Forecasting and Control*, 1st Edition, Holden-day, San Francisco, 1976.

[3] G.E.P. Box, G.M. Jenkins and G.C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1994.

[4] M. Goulard and M. VolTz, Linear Coregionalization Model: tools for estimation and choice of croo-variogram matrix, *Math. Geol*., **24**, (1992), 269-282.

[5] P.G. Hoel, *Mathematical Statistics*, Wiley, New York, 1984.

[6] K.V. Mardia and C.R. Goodall, Spartial-temporal analysis of multivariate environmental monitoring data. Multivariate environmental Statistics, Northytolland Series in Statistics & Probability 6, Amsterdam, North-Holland, 1993.

[7] Pattrick F. Dunn, *Measurement and Data Analysis for Engineering and Science*, New York, McGraw-Hill, 2005.

[8] Donald B. Percival and Andrew T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, pp. 190-195, 1993.

[9] M.B. Priestly, *Spectral Analysis and Time Series*, Academic press, London, 1982.

[10] H.F. Storch and W. Zwiers, *Statistical Analysis in climate research*, Cambridge University Press, 2001.

[11] W.N. Venables and B.D. Ripley, *Auto and Cross-Covariance and Correlation Function Estimatiom. Modern Applied Statistics*, Fourth Edition, Springer-Verlag, 2002.

[12] J.M. Ver Hoef and R.P. Barry, Constructing and Fitting models for Cokriging and multivariate spatial prediction, *Journal Statistics Plan Inference*, **69**, (1998), 275-294.

[13] H. Wackernagel, *Multivariate Geostatistics: An introduction with Application*, 2nd ed Berlin, Springer, 2003.