**SOME BASIC PROPERTIES OF CROSS-CORRELATION FUNCTIONS OF N-DIMENSIONAL VECTOR TIME SERIES**

**BY**

**USORO, ANTHONY E.**

**Department of Mathematics and Statistics, Akwa Ibom State University, Mkpat Enin, Akwa Ibom State.**

**Email: toskila2@yahoo.com**

**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. $γ\_{X\_{it+k}X\_{jt+l}}$ is the matrix of the cross-covariance functions, $γ\_{X\_{it}}$ and $γ\_{X\_{jt}}$ are the variances of $X\_{it}$ and $X\_{jt}$vectors respectively. Vector cross-correlation function was derived as $ρ\_{X\_{it+k},X\_{jt+l}}=\frac{γ\_{X\_{it+k}X\_{jt+l}}}{\sqrt{γ\_{X\_{it}}γ\_{X\_{jt}}}}$ . 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.

**KEY WORDS**: Vector time series, cross-covariance function and cross-correlation function.

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 $X=X\_{1},X\_{2},…,X\_{n} and Y=Y\_{1},Y\_{2},…,Y\_{n}).$ 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

 $ρ(\_{s,t})=\frac{E\{(X\_{t}-μ\_{t})(X\_{s}-μ\_{s})]}{σ\_{t}σ\_{s}}$,

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,

 $ρ\_{k}$ = $\frac{E[\left(X\_{t-}μ\right)\left(X\_{t+k-}μ\right)]}{√E[(X\_{t-}μ)^{2}]E[(X\_{t+k-}μ)^{2}]}=\frac{E[\left(X\_{t-}μ\right)\left(X\_{t+k-}μ\right)]}{σ^{2}}$

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 $ρ\_{k}$is defined as the autocovariance $γ\_{k}$ at lag k divided by variance $γ\_{0(k=0)}$. 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 $X\left(t+k\right) and Y(t)$, 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 $γ\_{xy}\left(τ\right)= E[\left(X\_{t}-μ\_{x}\right)\left(Y\_{t+τ}-μ\_{y}\right)]$, 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,

 $ρ\_{xy}\left(τ\right)=\frac{γ\_{xy}(τ)}{σ\_{x}σ\_{y}}$

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 {$X\_{1},…,X\_{n}\}$ which known mean and variance, an estimate of the autocorrelation may be obtained as

$\hat{R}\_{(k)}=\frac{1}{(n-k)σ^{2}}\sum\_{t=1}^{n-k}\left(X\_{t}-μ\right)(X\_{t+k}-μ)$ 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

$μ\_{t}=μ\_{s}=μ $for all t,s and $C\_{xx(t,s)}=C\_{xx(s-t)}=C\_{xx(T)}$, 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 $C\_{xx}\left(T\right)=E[\left(X\_{\left(t\right)}-μ\right)\left(X\_{\left(t+T\right)}-μ\right)]$

 $=E\left[X\_{\left(t\right)}X\_{\left(t+T\right)}\right]-μ^{2}$

 =$ R\_{xx(T)}-μ^{2}$, where $R\_{xx}$ represents the autocorrelation in the signal processing sense.

$R\_{xx}\left(T\right)=\frac{C\_{xx(T)}}{σ^{2}}$,Hoel (1984).

 For $X\_{t}$ and $Y\_{t}$, the following properties hold:

1.$ ρ\_{xy(h)}\leq 1$

2. $ ρ\_{xy(h)}= ρ\_{xy(-h)}$

3. $ ρ\_{xy(0)} \ne 1$

4. $ ρ\_{xy(h)}= \frac{γ\_{xy(h)}}{√γ\_{x(0)}γ\_{y(0)}}$

Mardia and Goodall (1993) defined separable cross-correlation function as $C\_{ij}\left(X\_{1},X\_{2}\right)=ρ\left(X\_{1},X\_{2}\right)a\_{ij},$ where $A=[a\_{ij}]$ is a $p×p$ positive definite matrix and $ρ\left(.,.\right)$ is a valid correlation function. Goulard & Voltz (1992); Wackernage (2003); Ver Hoef and Barry (1998) implied that the cross- covariance function is

$C\_{ij}\left(X\_{1}-X\_{2}\right)=\sum\_{k=1}^{r}ρ\_{k}(X\_{1}-X\_{2})a\_{ik}a\_{jk}$, for an integer 1$\leq r\leq p,$ where $A=[a\_{ij}]$ is a $p×r$ full rank matrix and $ ρ\_{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:

$$γ\_{X\_{(1t+k)}, X\_{\left(1t+l\right) }}γ\_{X\_{(1t+k)}, X\_{\left(2t+l\right) }}γ\_{X\_{(1t+k)}, X\_{\left(3t+l\right) . . . }}γ\_{X\_{(1t+k)}, X\_{(nt+l)}}$$

$$γ\_{X\_{(2t+k)}, X\_{\left(1t+l\right) }}γ\_{X\_{(2t+k)}, X\_{\left(2t+l\right) }}γ\_{X\_{(2t+k)}, X\_{\left(3t+l\right) . . . }}γ\_{X\_{(2t+k)}, X\_{(nt+l)}}$$

$$γ\_{X\_{(3t+k)}, X\_{\left(1t+l\right) }}γ\_{X\_{(3t+k)}, X\_{\left(2t+l\right) }}γ\_{X\_{(3t+k)}, X\_{\left(3t+l\right) . . . }}γ\_{X\_{(3t+k)}, X\_{(nt+l)}}$$

 . .

 . .

 . .

$$γ\_{X\_{(mt+k)}, X\_{\left(1t+l\right) }}γ\_{X\_{(mt+k)}, X\_{\left(2t+l\right) }}γ\_{X\_{(mt+k)}, X\_{\left(3t+l\right) . . . }}γ\_{X\_{(mt+k)}, X\_{(nt+l)}}$$

 $where k=0, …, a, l=0, …, b$

The above matrix is a square matrix, and could be reduced to the form,
$$γ\_{X\_{(it+k)}, X\_{\left(jt+l\right) }}$$

$$where i=1,…,m, j=1,…,n, k=0, …, a, l=0, …, b.(n=m) $$

From the above cross-covariance matrix, $ρ\_{X\_{it+k},X\_{jt+l}}=\frac{γ\_{X\_{it+k}X\_{jt+l}}}{\sqrt{γ\_{X\_{it}}γ\_{X\_{jt}}}}$,

Where, $γ\_{X\_{it+k}X\_{jt+l}}$ is the matrix of the cross-covariance functions, $γ\_{X\_{it}}$ and $γ\_{X\_{jt}}$

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 $X\_{jt+l}(l\pm 1,2,…,b)$, or $X\_{jt}$can be cross-correlated with $X\_{it+k}\left(k\pm 1,2,…,a\right).$

3. ANALYSIS OF THE CROSS-CORRELATION FUNCTIONS

Given two processes $X\_{1t} and X\_{2t}$, $ρ\_{(X\_{1t},x\_{2t+k})}$ is the cross-correlation between $X\_{1t} and X\_{2t} at lag k$, while, $ρ\_{(x\_{2t},x\_{1t+k})}$ 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,…,4, $the following results were obtained with a software:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lag k | $$ρ\_{(x\_{1t},x\_{2t+k})}$$ | $$ρ\_{(x\_{2t},x\_{1t+k})}$$ | $$ρ\_{(x\_{1t},x\_{3t+k})}$$ | $$ρ\_{(x\_{3t},x\_{1t+k})}$$ | $$ρ\_{(x\_{2t},x\_{3t+k})}$$ | $$ρ\_{(x\_{3t},x\_{2t+k})}$$ |
| -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.$ ρ\_{X\_{it+k,}X\_{jt+l}}\ne 1, for k=0,l=\pm 1,…,\pm b, i\ne j,$

 b. $ρ\_{X\_{it+k,}X\_{jt+l}}\ne 1, for l=0,k=\pm 1,…,\pm a, i\ne j.$

1. a.$ ρ\_{X\_{it+k,}X\_{jt+l}}\ne ρ\_{X\_{it+k},X\_{jt-l}}, for k=0,l=1,…,b, i\ne j,$

b.$ ρ\_{X\_{it+k,}X\_{jt+l}}\ne ρ\_{X\_{it-k},X\_{jt+l}}, for l=0,k=1,…,a, i\ne j.$

1. a. $ρ\_{X\_{it+k,}X\_{jt+l}}=ρ\_{X\_{jt+k},X\_{it-l}},$ $for k=0,l=1,…,b, i\ne j,$

b.$ ρ\_{X\_{it+k,}X\_{jt+l}}=ρ\_{X\_{jt-k},X\_{it+l}},$ $for l=0,k=1,…,a, i\ne j$

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

2. Box, G. E. P. And Jenkins, G. M. (1976): Time Series Analysis; Forecasting and Control 1st Edition, Holden-day, san Francisco.

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

4. Goulard, M. & VolTz, M. (1992). Linear Coregionalization Model: tools for estimation and choice of croo-variogram matrix. Math. Geol. 24, 269-82.

5. Hoel, P. G(1984): Mathematical Statistics, Wiley, New York.

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

7. Pattrick F. Dunn(2005): Measurement and Data Analysis for Engineering and Science, New York: McGraw-Hill. ISBN 0-07-282538-3

8. Percival, Donald B.; Andrew T. Walden (1993). Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge University Press. PP. 190-195. ISBN 0-521-43541-2

9. Priestly, M. B (1982): Spectral Analysis and Time Series. Academic press, London.

10. Storch, H. F. and W Zwiers (2001): Statistical Analysis in climate research. Cambridge University Press. ISBN 0-521-01230-9.

11. Venables, W. N. And Ripley, B. D. (2002): Auto and Cross-Covariance and Correlation Function Estimatiom. Modern Applied Statistics, Fourth Edition, Springer-Verlag.

12. Ver Hoef, J. M. & Barry, R. P. (1998). Constructing and Fitting models for Cokriging and multivariate spatial prediction. *Journal Statistics Plan Inference*; 69.275-94.

13. Wackernagel, H. (2003): Multivariate Geostatistics: An introduction with Application. 2nd ed Berlin, Springer.