COMPUTATIONAL ASPECTS ON SERIAL CORRELATIONS IN COHERENT TIME SERIES

Eugen I. SCARLAT[1], Mona MIHĂILESCU[2]

The paper presents the extent to which the temporal gap between processes represented by finite time series can be accurately determined from the analysis of the relative phase function of the complex cross coherence function. The numerical simulations show that the existence of serial correlations lead to significant deviations compared to the theoretical values.

Keywords: fast Fourier transform, cross coherence function, time shift theorem, Hurst exponent, fractional Brownian motion.

1. Introduction

Detecting correlations among processes evidenced by synchronously sampled quantities in the form of time series is of ultimate importance to disclosing deterministic relationships in flocking clusters [1], measuring propagation delays [2], predicting the evolution of economic and financial phenomena [3], or to explain the sensitivity to local, regional or global crises [4,5]. Particularly the short run forecasting could be significantly improved if detecting time shifts below the sampling rates [6].

Such time shifts could be detected by studying the phase synchronization revealed by the relative phase function (RPF) of the complex cross coherence function (CCCF) [7] between pairs of series. The accuracy of the calculi depends on a calibration procedure that involves the computing of the RPF between the series and its time shifted replica. The time shift theorem is known as introducing a phase shift in the Fourier image of a time series [8]. The property can be exploited to calibrating the measurement of time gaps between narrowband processes embedded in discrete time series, provided that the series present significant coherence in the frequency range of interest [9]. However the measuring of time shifts using coherence based techniques is influenced by many factors like the shortness of the series, non-stationarities, or computing shortcomings that hinder the confidence in the final results. If the use of returns is an elegant technique of stationarizing the financial series by preserving the

[1] Applied Sci., Phys. Dept., University *Politehnica* Bucharest, E-mail: egen@physics.pub.ro
[2] Applied Sci., Phys. Dept., University *Politehnica* Bucharest, E-mail: mona_m@physics.pub.ro

economic meaningfulness, in many other cases the study of brute non-stationary series is preferred. For example the gross domestic product time series are correlated and even their variations still exhibit persistence as measured by the Hurst exponent [10].

2. Theory

By denoting $x(t)$ and $y(t)$ two numeric time series, $S_{xx}(f)$, $S_{yy}(f)$ their auto-spectra and $S_{xy}(f)$ the cross spectrum at frequency f, then CCCF is the ratio of the cross-spectrum to the product of the square roots of the auto-spectra [11]:

$$\gamma_{xy}(f) = \frac{S_{xy}(f)}{\sqrt{S_{xx}(f)} \cdot \sqrt{S_{yy}(f)}} = \left|\gamma_{xy}(f)\right| \cdot \exp\left(\mathbf{i}\alpha_{xy}(f)\right), \qquad \mathbf{i}=\sqrt{-1}\,. \tag{1}$$

To be more specific, if the complex Fourier images are

$$X(f)=|X(f)|\cdot \exp\left(\mathbf{i}\varphi_x(f)\right),\; Y(f)=|Y(f)|\cdot \exp\left(\mathbf{i}\varphi_y(f)\right), \tag{2}$$

and if the phase difference is denoted $\Delta\varphi(f)=\varphi_x(f)-\varphi_y(f)$, then the modulus of CCCF is the coherence function (CF)

$$\left|\gamma_{x,y}(f)\right| = \frac{1}{\sqrt{<\left|X(f)\right|^2>_{k_\tau}} \cdot \sqrt{<\left|Y(f)\right|^2>_{k_\tau}}} \times \\ \times\sqrt{\left[<\left|X(f)\right|\cdot\left|Y(f)\right|\cdot\cos(\Delta\varphi)>_{k_\tau}\right]^2 + \left[<\left|X(f)\right|\cdot\left|Y(f)\right|\cdot\sin(\Delta\varphi)>_{k_\tau}\right]^2} \tag{3}$$

while the argument is the corresponding RPF

$$\alpha_{x,y}(f) = \mathrm{Arg}\{<\left|X(f)\right|\cdot\left|Y(f)\right|\cdot\exp\left(\mathbf{i}\Delta\varphi(f)\right)>_{k_\tau}\}. \tag{4}$$

The average "$< >_{k_\tau}$" is computed according to the principles given in [12] using the moving average technique of a window of width τ over the full length T of the numeric series [13], and k_τ is the running index of the window.

Hereafter the theoretically results are discussed in the case $y(t)=x(t+\Delta t)$ where Δt is a constant time shift. According to the time shift theorem one has

$$x(\Delta t)\leftrightarrow X(f)\cdot \exp\left[-\mathbf{i}\left(2\pi\frac{f}{\tau}\cdot\Delta t\right)\right]. \tag{5}$$

By denoting $\alpha_{x(0),x(\Delta t)}(f)$ the RPF of the pair consisting in the genuine and the shifted series given by Eq.(4) becomes

$$\alpha_{x(0),x(\Delta t)}(f) = \mathrm{Arg}\left\{ <\left|X(f)\right|^2 \cdot \exp\left[-\mathbf{i}\left(2\pi \frac{f}{\tau}\cdot \Delta t\right)\right]>_{k_\tau} \right\}. \quad (6)$$

Since the exponential is the same whatever the window, the argument is:

$$\alpha_{x(0),x(\Delta t)}(f) = -2\pi \frac{\Delta t}{\tau} \cdot f \quad \text{for } f=0, \ldots, 2/\tau. \quad (7)$$

The theoretical conclusion is the RPF scales linearly with frequency with the slope:

$$b = -2\pi \frac{\Delta t}{\tau}, \quad f=0, \ldots, 2/\tau. \quad (8)$$

Therefore the RPF of any pair of series could be regressed over the investigated frequency range and assuming the estimated slope $\hat{b}$ would be statistically significant, the value of the time shift $\Delta\hat{t}$ can be computed as:

$$\Delta\hat{t} = -\frac{\tau}{2\pi} \cdot \hat{b}. \quad (9)$$

In the particular case of $\Delta t=1$ Eq.(9) takes the form:

$$\Delta\hat{t} = -\frac{\tau}{2\pi} \cdot \hat{b}\Big|_{\Delta t=1}. \quad (9')$$

Since it should be unity $\Delta\hat{t}\Big|_{\Delta t=1} = 1$, Eq.(9') can be used to calibrate Eq.(9) [14].

The computational results are in disagreement with theory when the serial correlations are present in the series. The following section is presenting such results using synthesized series.

3. Synthesized series

The series where synthesized using MatLab facilities as described in [15]. In Table 1 are presented the main characteristics of the synthesized series. The computed results are estimations over ten samples of 2^{12}=4096 points each extracted at random from synthesized series of 2^{14}=16384 points with a designed Hurst exponent ranging from 0 to 1. Since the resulting series are not rigorously monofractal, in the rightmost column is indicated the deviation from monofractality as the difference between the generalized Hurst exponents $h(0)$ and $h(2)=H$ [16].

Table 1

Hurst exponent in synthesized series

H (designed)	H (estimated)	Multifractality $\|h(0)-H\|$
0.0	0.011±0.004	0.334
0.1	0.045±0.001	0.337
0.2	0.165±0.018	0.260
0.3	0.283±0.019	0.195
0.4	0.392±0.039	0.142
0.5	0.507±0.020	0.077
0.6	0.615±0.024	0.043
0.7	0.723±0.024	0.002
0.8	0.829±0.030	0.048
0.9	0.929±0.033	0.075
1.0	1.020±0.041	irrelevant

4. Computational results

In Table 2 are presented the results of the computed time shift $\left.\Delta\hat{t}\right|_{\Delta t=1}$ according to Eq.(9'). Where not shown in the cells, the standard errors are less than $5\cdot10^{-4}$. The statistical significance of the estimated slope given by Eq.(8) is analyzed using the variance table for the fitted model (ANOVA); the subsequent p-values and R^2 coefficients are indicated in the table. For the scope of the present study the statistical significance is fulfilled for the usually accepted thresholds p<1%, R^2>0.750.

Table 2

Computed time shift as function of Hurst exponent and window size for Δt=1

H / Window size τ	0.0	0.1	0.2	0.3	0.4	0.5	H≥0.6
8	0.921	0.934±0.008	0.939±0.025	0.919±0.055	0.811±0.084	0.380±0.035	Not significant
16	0.963	0.971±0.005	0.975±0.015	0.961±0.035	0.880±0.064	0.376±0.047	Not significant
32	0.983	0.989±0.004	0.994±0.009	0.987±0.019	0.929±0.039	0.438±0.021	Not significant
64	0.994	0.998±0.002	1.003±0.004	1.000±0.010	0.957±0.024	0.617±0.041	Not significant
128	0.998	1.001±0.001	1.005±0.002	1.006±0.005	0.981±0.014	0.625±0.031	Not significant
256	1.000	1.002	1.005±0.001	1.007±0.003	0.991±0.008	0.727±0.028	Not significant
	p<0.1%, R^2>0.990		p<1%, R^2>0.960			p<1%, R^2>0.830	p, R^2 irrelevant

One should remark three cases:

i) For $0 \le H < 0.5$ (white columns in Table 2), RPF scales linearly across the whole frequency range and the estimated time shift approach the theoretical value $\Delta\hat{t}\big|_{\Delta t=1} \cong 1$;

ii) For H=0.5 (light gray column) the estimated time shift is still statistically significant but it differs from the theoretical value;

iii) For $H \ge 0.5$ (dark gray columns), the estimations are not significant and the RPF does not scale with frequency across the entire band.

When significant – in cases i) and ii) –, the values of the estimated time shift are more accurate for larger sizes of the windows. In the case H=0.5 Eq.(9') can be used with corrective factors [14].

5. Discussion

To explain this behavior, the implementation of fast Fourier transforms (FFT) in Mathematica package should be considered. If the k_τ is the running index of the window with step P, then the FFT of the series is given by:

$$X(f) = \; < \sum_{t=0}^{\tau-1} x(k_\tau P + t) \cdot \exp\left(-\mathbf{i}2\pi \frac{f}{\tau} t \right) >_{k_\tau}, f=0,\ldots 2/\tau . \tag{10}$$

The RPF between the genuine series and its one-step shifted replica becomes:

$$\begin{aligned}
\alpha_{x(0),x(1)}(f) = \operatorname{Arg}\Bigg\{ &< \underbrace{\left(\sum_{t=\tau-1}^{\tau-1} x(k_\tau P + t - \tau + 1) \cdot x(k_\tau P + t) \right)}_{\substack{\text{one term of two-factor product} \\ \text{shifted with } \tau-1 \text{ steps each other}}} >_{k_\tau} \exp\left(-\mathbf{i}2\pi \frac{f}{\tau} (\tau - 1) \right) + \\
&+ \ldots + < \underbrace{\left(\sum_{t=1}^{\tau-1} x^2(k_\tau P + t) \right)}_{\substack{\tau-1 \text{ terms of two-factor product} \\ \text{zero shifted each other}}} >_{k_\tau} \exp\left(-\mathbf{i}2\pi \frac{f}{\tau} \cdot 1 \right) + \\
&+ < \underbrace{\left(\sum_{t=0}^{\tau-1} x(k_\tau P + t) \cdot x(k_\tau P + t + 1) \right)}_{\substack{\tau \text{ terms of two-factor product} \\ \text{shifted with one step each other}}} >_{k_\tau} \exp\left(-\mathbf{i}2\pi \frac{f}{\tau} \cdot 0 \right) + \\
&+ < \underbrace{\left(\sum_{t=1}^{\tau-1} x(k_\tau P + t - 1) \cdot x(k_\tau P + t + 1) \right)}_{\substack{\tau-1 \text{ terms of two-factor product} \\ \text{shifted with two steps each other}}} >_{k_\tau} \exp\left(\mathbf{i}2\pi \frac{f}{\tau} \cdot 1 \right) + \ldots + \\
&+ < \underbrace{\left(\sum_{t=\tau-1}^{\tau-1} x(kP + t - \tau + 1) \cdot x(kP + t + 1) \right)}_{\substack{\text{one term of two-factor product} \\ \text{shifted with } \tau \text{ steps each other}}} >_{k_\tau} \exp\left(\mathbf{i}2\pi \frac{f}{\tau} (\tau - 1) \right) \Bigg\} .
\end{aligned} \tag{11}$$

By comparing Eq.(11) to Eq.(6) one should remark they would be the same if all averages cancel out excepting the one containing the quadratic sum i.e. the one associated with the argument of interest $\exp\left(-\mathbf{i}2\pi\frac{f}{\tau}\cdot 1\right)$. If the fully cancellation does not occur then Eqs.(7-8) do not hold and the result deviates from theory. The coefficients of the exponentials are the autocorrelation function (ACF) $r(\theta)$ in the interval $-(\tau-1)$ to $+\tau$. One should note the averages operate only on ACF, not on the exponentials.

Because of the existence of conjugated exponentials in Eq.(11) the cancellations of the imaginary parts would be complete in the case of infinite series; for finite series the cancellations of the imaginary parts of $\exp\left(\pm\,\mathbf{i}2\pi\frac{f}{\tau}\cdot t\right)$ are not complete. However the main reason of non-complying of the results with theory emerges from the real part of the terms in Eq.(11) so that the further focus is on their contributions to the total real part while the residuals of the imaginary part is of little importance.

ACF is strongly depending on the serial correlations in the series. Assuming the truncated series in the window as particular realizations of a fractional Brownian motion [17], the two point ACF between θ-spaced past and future increments normalized to the variance $\sigma^2_{\tau-\theta}$ of length $\tau-\theta$ is the same at all time and depends only on the H parameter [18]:

$$r(\theta)=\begin{cases}\frac{1}{2}\left((\theta+1)^{2H}-2\theta^{2H}+|\,\theta-1|^{2H}\right)\cdot\sigma^2_{\tau-\theta}\ \text{ for small}\,\theta,\\ H(2H-1)\,\theta^{2H-2}\cdot\sigma^2_{\tau-\theta}\ \text{ for large}\,\theta.\end{cases}\tag{12}$$

According to Eq.(12) ACF is negative for $0\le H<0.5$ and positive for $H>0.5$; in the case $H=0.5$ the autocorrelation is zero. Since the series are not strictly monofractal (see Table 1) and therefore there is not a unique H value along the series, the cases presented in Sec.4 are explained only in a semi-quantitative approach as follows.

i) The case $0\le H<0.5$, the series is anti-persistent. The numbers of positive and negative terms in any serial summation are comparable and consequently the total real component is small (positive or negative) such as the quadratic sum highly dominates in Eq.(11) and RPF behaves linearly with frequency.
ii) The case $H=0.5$, the series is of a pure random walk type such that the total real component in Eq.(11) is moderate as compared to the quadratic sum. Adding a pure real number to the complex number

$\exp\left(-\mathbf{i}2\pi\frac{f}{\tau}\cdot 1\right)$ reduces the absolute value of the argument. The higher the frequency, the greater the influence (see Fig.1).

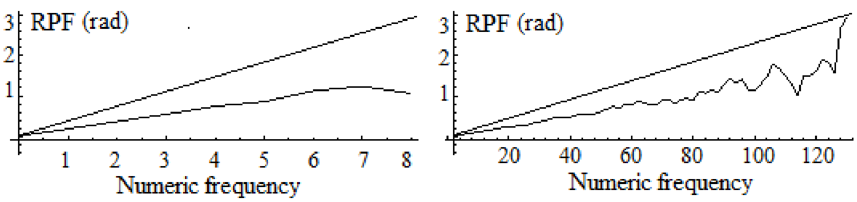

Fig.1. Relative phase function vs. frequency for H=0.6 and window sizes τ=16 (left) and τ=256 (right); the straight lines represent the theoretical scaling for H=0.

iii) The case H>0.5, the series is serially correlated. The terms in any summation are more likely to be positive and consequently the total real component is positive and of the same order of magnitude with the quadratics. The effect at higher frequency is enough to repeal the statistical significance because the RPF scaling is no more linear (see Fig.2).

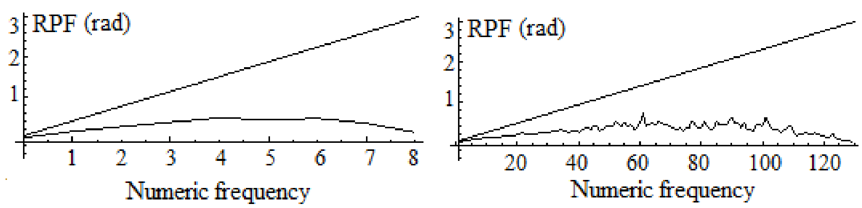

Fig.2. Relative phase function vs. frequency for H=0.6 and window sizes τ=16 (left) and τ=256 (right); the straight lines represent the theoretical scaling for H=0.

6. Conclusions

The temporal gaps between processes represented by finite time series can be determined by analyzing the relative phase function of the complex cross coherence function. The determination depends on a calibration procedure that involves the computing of the relative phase of the complex cross coherence function between the series and its time shifted replica.

The computational results are fully compliant with theory in the case of anti-persistent series characterized by the Hurst exponent H<1/2, can be cautiously accepted with corrective factors in the case of pure random walk series with H=1/2, and deviate from the theory in the case of persistent series with H>1/2.

When significant, the values of the estimated time shift are more accurate as the size of the window increases.

Acknowledgement

The research is supported by HOLCOMM PN-II-PT-PCCA no. 203/2012.

REFERENCES

[1] *C.A. Yates, R. Erban, C. Escudero, I.D. Couzin, J. Buhl, I.G. Kevrekidis, P.K. Maini, D.J. Sumpter*, "Inherent noise can facilitate coherence in collective swarm motion", Proc. National Acad. Sci. USA. 106 (14), 5464–9, 2009.

[2*] M. Mihailescu*, "Natural quasy-periodic binary structure with focusing property in near field diffraction pattern", Optics Express 18 (12), 12526-12536, 2010.

[3] *T. Aste, W. Shaw, T. Di Matteo*, "Correlation Structure and Dynamics in Volatile Markets", New Journal of Physics 12 085009, 21p. 2010.

[4] *V. Dedu, I. Mihai, F. Neagu*, "Trends of the Contagion Risk in Sovereign Spreads for Emerging European Countries", Rom. J. Ec. Forecasting, 13 (2), 265-279, 2010.

[5] *L. Preda, M. Mihailescu*, „A Synthesis of JPY-USD Exchange Rate Based on the <<Generalized Devil Staircase Singular Function>>", in Predictability in non-linear dynamical systems. The economic crises, Politehnica Press, Buch. 70-82, 2011.

[6] *E.I. Scarlat*, "Sub-Daily Time Shift Detection Between Daily Sampled Time Series", UPB Sci. Bull. A, **76** (2), 223-232, 2014.

[7] *L. Mandel, E. Wolf*, "Spectral coherence and the concept of cross-spectral purity", JOSA **66** (6), 529-535, 1976.

[8] *B. Saff, A. Snider*, Fundamentals of Complex Analysis for Mathematics, Science and Engineering, Prentice Hall, 1993.

[9] *E.I. Scarlat, C.M. Cristescu, C. Stan, C.P. Cristescu*, „The Magnitude Coherence Function of ROL/USD-ROL/EUR Exchange Rates", UPB Sci. Bull. A, **76** (1), 247-254, 2014.

[10] *H.E. Hurst*, "Long Term Storage Capacity of Reservoirs", Transactions of the American Society of Civil Engineers, 116, 770-799, 1951.

[11] *J.W. Goodman*, Statistical Optics, John Wiley & Sons, 1985.

[12] *P.D. Welch*, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Transactions on Audio Electroacoustics, AU-15, 70-73, 1967.

[13] *S-Y. Wang, M-X. Tang*, "Exact Confidence Interval For Magnitude-Squared Coherence Estimates", IEEE Signal Processing Letters, **11**, (3) 326-329, 2004.

[14] *E.I. Scarlat*, "Coherence-Based Method to Detect Time Shifts Smaller than the Sampling Rate of Time Series", Rom. J. Phys., Nos 3-4, inprint, 2015.

[15] *D. Xie, L. Wan, Y. Zhu*, "The Effectiveness Evaluation of Two Kinds of Fractal Sequences on Detrended Fluctuation Sequences on Detrended Fluctuation Analysis", Mathematical Computation, **3**, (2), 58-62, 2014.

[16*] E.I. Scarlat, L. Preda, M. Mihăilescu*, „A Generalized Devil Staircase-Based Generator for the JPY-USD Exchange Rate?", J. Int. Fed. Autom. Contr., **2** (1), 321-326, 2009.

[17] *B.B. Mandelbrot, J.W. Van Ness*, „Fractional Brownian Motions, Fractional Noises and Applications", SIAM Review, **10**, 4, 422-437, 1968.

[18] *J. Gao, Y. Cao, W-W. Tung, H. Jing*, Multiscale analysis of complex time series, Wiley-Interscience, Hoboken, NJ, 2007.