STRICT STATIONARY TIME SERIES AND AUTOCOPULA

Daniel CIUIU

Technical University of Civil Engineering, Bucharest;
Romanian Institute for Economic Forecasting (dciuiu@yahoo.com)

DOI: 10.19062/1842-9238.2018.16.2.6

Abstract: *In this paper we consider not only the classical weakly stationarity of time series (same expectation, same variance and same correlations). We also aim to consider the strict stationarity of a time series. Therefore, each observation* X_i *from the time series* $X_1,..., X_n$ *has the same cumulative distribution function.*

We consider that the common cdf F *is the common marginal distribution of* $X_1,..., X_n$*, and the dependence is expressed by a copula* C *of order* n*. An Archimedean copula is used, and the parameters (of the marginal cdf and the copula parameter θ) are estimated using the maximum likelihood method.*

Keywords: *Time series, stationarity, smoothing.*

1. INTRODUCTION

The definition of the term "copula" can be easily found in literature [2,15,10]. For time series, we use the theorem of Sklar, which establishes that every multivariate cumulative distribution function H (in our case the multivariate cdf of $\left(X_1,...,X_n\right)$) can be written

$$H\left(x_1,...,x_n\right)=C\left(F_1\left(x_1\right),...,F_n\left(x_n\right)\right), \tag{1}$$

where F_i is the marginal distribution of X_i.

For the copula C we use in this paper Archimedean copula, for which it is proved [2,8,11] that there exists $\varphi:[0,1]\to R$ decreasing and convex with $\phi(1)=0$ having the pseudo-inverse g ($g\left(y\right)=x$ if x exists in such a way that $\phi\left(x\right)=y$, otherwise $g\left(y\right)=0$) so that

$$C\left(u_1,...,u_n\right)=g\left(\sum_{i=1}^{n}\phi\left(u_i\right)\right). \tag{2}$$

For the Clayton copula, we have

$$\begin{cases}\phi\left(u\right)=\frac{u^{-\theta}-1}{\theta}\\ g\left(x\right)=\left(\theta\cdot x+1\right)^{-\frac{1}{\theta}}\end{cases},\ with\ \theta>0. \tag{3}$$

For the Frank family of copulas we have [2,15]

$$\begin{cases}\phi\left(u\right)=\ln\left(\frac{1-e^{-\theta}}{1-e^{-\theta u}}\right)\\ g\left(x\right)=-\frac{1}{\theta}\ln\left(\gamma e^{-x}+1\right),\ where\ \gamma=e^{-\theta}-1\end{cases},\ with\ \theta\in R. \tag{4}$$

For the Gumbel-Hougaard copula we have [10,11]

$$\begin{cases} \phi(u) = (-\ln u)^{\theta} \\ g(x) = e^{-x^{\frac{1}{\theta}}} \end{cases}, \ with\ \theta \geq 1. \tag{5}$$

For the Gumbel-Barnet copula we have [15,11]

$$\begin{cases} \phi(u) = \frac{\ln(1-\theta\ln u)}{\theta} \\ g(x) = e^{\frac{1-e^{\theta x}}{\theta}} \end{cases}, \ with\ 0 < \theta \leq 1. \tag{6}$$

For the Ali-Mikhail-Haq copula we have [15,2]

$$\begin{cases} \phi(u) = \frac{1}{1-\theta} \cdot \ln\left(\theta + \frac{1-\theta}{u}\right) \\ g(x) = \frac{1-\theta}{e^{(1-\theta)x} - \theta} \end{cases}, \ with\ -1 \leq \theta \leq 1. \tag{7}$$

The Frank, Gumbel-Hougaard and Ali-Mikhail-Haq copulas contain the product copula (independence case) for $\theta = 0$, $\theta = 1$, $\theta = 0$, respectively. For the other copula families (Clayton and Gumbel-Barnett), the product copula is in both cases the limit case $\theta \to 0$.

Some copulas have been simulated in [15], and methods for simulating random variables and Monte Carlo methods can be found in [14].

When we consider the Gaussian time series, we test the weak stationarity using the Dickey-Fuller unit root test [7], and we stationarize the time series. Next, we find the ARMA model using the Box-Jenkins methodology [4,9,12]. If after we have found the ARMA model we have non-significant autocorrelations and partial autocorrelations, it means that in the Gaussian approach the model is correct. But, if we apply the BDS (Brock, Deckert and Skheinkman) test [3] and we obtain that the errors are not mutually independent and with the same distribution, and in the Gaussian approach the model is correct, it means that we must use a non-Gaussian approach. This test is as follows. First, we compute the probability

$$P_{k;\varepsilon} = P\left(\left\|\left(X_i, ..., X_{i+k-1}\right) - \left(X_j, ..., X_{j+k-1}\right)\right\| < \varepsilon\right) \tag{8}$$

for a given ε empirically, where the above norm is the infinite norm (the maximum absolute value). If the values are independent with the same distribution, we have

$$P_{k;\varepsilon} = P_{1;\varepsilon}^{k}. \tag{8'}$$

Therefore we compute

$$Z_{k;\varepsilon} = \frac{P_{k;\varepsilon} - P_{1;\varepsilon}^{k}}{\sigma_{k;\varepsilon}}, \tag{8''}$$

where $\sigma_{k;\varepsilon}^2$ is the variance of the above numerator. The computation of $\sigma_{k;\varepsilon}^2$ is presented in [3], where it is mentioned that $Z_{k;\varepsilon}$ is asymptotically normal. Therefore, we have to compare these values to the cuantiles of the standard normal distribution, for $k = \overline{2, k_{\max}}$.

2. THE MODEL

In order to apply the copula for strict stationary time series, we first have to take into account that strict stationarity yields to the same cdf for X_i. Therefore, in (1) we have $F_i = F$ (the same marginal cdf). The copula C models the dependence between the time series values at different moments. We call this copula autocopula, by analogy to the classical use of autocorrelations for classical ARMA models.

For estimating the parameters of the model, we take into account [15,2] that the multivariate pdf $h\left(x_1,...,x_n\right)=\frac{\partial^n H}{\partial x_1...\partial x_n}$ can be written

$$h\left(x_1,...,x_n\right)=\left|g^{(n)}\left(\sum_{i=1}^{n}\phi\left(F_i\left(x_i\right)\right)\right)\right|\cdot\prod_{i=1}^{n}\left|\phi'\left(F_i\left(x_i\right)\right)\right|\cdot\prod_{i=1}^{n}f_i\left(x_i\right), \tag{9}$$

where f_i are the marginal pdfs.

Denote now by α the vector of parameters for the common marginal distribution having the cdf *F*, and by $\tilde{V}=\prod_{i=1}^{n}f_i\left(x_i;\alpha\right)$ the likelihood in the independence case (when X_1,..., X_n are independent identical distributed, with the pdf *f* and cdf *F*). In the time series case, the above common pdf *h* is in fact the likelihood *V*, which must have a maximal value of 5 (we apply the maximum likelihood method). By computation, we obtain

$$\ln V=\ln\left(\left|g^{(n)}\left(\sum_{i=1}^{n}\phi\left(F\left(X_i;\alpha\right)\right)\right)\right|\right)+\sum_{i=1}^{n}\ln\left(-\phi'\left(F\left(X_i;\alpha\right)\right)\right)+\ln\tilde{V}. \tag{10}$$

For solving the system of equations given by derivative on α components (α can be multiple, as in the normal case, when we have two parameters – the expectation and the variance), $\frac{\partial \ln V}{\partial\alpha_k}=0$ and the derivative on θ, $\frac{\partial \ln V}{\partial\theta}=0$, we try to express first the log-likelihood, because it is possible to separate α and θ. For instance, in the case of Clayton family we obtain

$$\ln V=\sum_{i=1}^{n-1}\ln\left(i\cdot\theta+1\right)-\left(\tfrac{1}{\theta}+n\right)\ln\left(\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)-n+1\right)-\left(\theta+1\right)\left(\sum_{i=1}^{n}\ln F\left(X_i;\alpha\right)\right)+\ln\tilde{V}. \tag{11}$$

We notice that the last term, $\ln\tilde{V}$ does not depend on θ, the first sum does not depend on α, and the sum of logarithms multiplied by $\theta+1$ does not depend on θ. These make the computations easier, and we obtain for $\theta>0$

$$\begin{cases}\frac{\partial\ln V}{\partial\alpha_k}=\left(n\theta+1\right)\frac{\sum_{i=1}^{n}F^{-\theta-1}\left(X_i;\alpha\right)\cdot\frac{\partial F}{\partial\alpha_k}\left(X_i;\alpha\right)}{\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)-n+1}-\left(\theta+1\right)\sum_{i=1}^{n}\frac{\frac{\partial F}{\partial\alpha_k}\left(X_i;\alpha\right)}{F\left(X_i;\alpha\right)}+\frac{\partial\ln\tilde{V}}{\partial\alpha_k}\\ \frac{\partial\ln V}{\partial\theta}=\sum_{i=1}^{n-1}\frac{i}{i\theta+1}+n\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)\ln F\left(X_i;\alpha\right)-\sum_{i=1}^{n}\ln F\left(X_i;\alpha\right)+\frac{\theta\left(\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)\ln F\left(X_i;\alpha\right)\right)+\left(\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)-n+1\right)\ln\left(\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)-n+1\right)}{\theta^2\left(\sum_{i=1}^{n}F^{-\theta}\left(X_i;\alpha\right)-n+1\right)}\end{cases} \tag{12}$$

For the limit case, $\theta\to 0$ we obtain, using l'Hôpital

$$\begin{cases}\frac{\partial\ln V}{\partial\alpha_k}=\frac{\partial\ln\tilde{V}}{\partial\alpha_k}\\ \frac{\partial\ln V}{\partial\theta}=C_n^2+\left(n-1\right)\left(\sum_{i=1}^{n}\ln F\left(X_i;\alpha\right)\right)+\frac{\left(\sum_{i=1}^{n}\ln F\left(X_i;\alpha\right)\right)^2}{2}-\frac{\sum_{i=1}^{n}\ln^2 F\left(X_i;\alpha\right)}{2}.\end{cases} \tag{12'}$$

We can also prove that the Hessian is in the general case $\theta\neq 0$

$$\begin{cases} \frac{\partial^2 \ln V}{\partial \alpha_j \partial \alpha_k} = \frac{n\theta+1}{\left(\sum\limits_{i=1}^{n} F_i^{-\theta} - n+1\right)^2} \cdot \left(-(\theta+1) \left(\sum\limits_{i=1}^{n} F_i^{-\theta-2} \frac{\partial F_i}{\partial \alpha_j} \frac{\partial F_i}{\partial \alpha_k} \right) \left(\sum\limits_{i=1}^{n} F_i^{-\theta} - n + 1 \right) + \left(\sum\limits_{i=1}^{n} F_i^{-\theta-1} \frac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_k} \right) \left(\sum\limits_{i=1}^{n} F_i^{-\theta} - n \right. \right. \\ \left. \left. \left(\sum\limits_{i=1}^{k} F_i^{-\theta} \frac{\partial F_i}{\partial \alpha_j} \right) \left(\sum\limits_{i=1}^{k} F_i^{-\theta} \frac{\partial F_i}{\partial \alpha_k} \right) \right) - (\theta+1) \left(\sum\limits_{i=1}^{n} \frac{\frac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_k} - \frac{\partial F_i}{\partial \alpha_j} \frac{\partial F_i}{\partial \alpha_k}}{F_i^2} \right) + \frac{\partial^2 \ln \tilde{V}}{\partial \alpha_j \partial \alpha_k} \\ \frac{\partial^2 \ln V}{\partial \theta \partial \alpha_k} = \frac{n\left(\sum\limits_{i=1}^{n} F_i^{-\theta} - n+1\right) - (n\theta+1)\left(\sum\limits_{i=1}^{n} F_i^{-\theta} \ln F_i\right)}{\left(\sum\limits_{i=1}^{n} F_i^{-\theta} - n+1\right)^2} \cdot \left(\sum\limits_{i=1}^{n} F_i^{-\theta-1} \frac{\partial F_i}{\partial \alpha_k} \right) - \frac{n\theta+1}{\sum\limits_{i=1}^{n} F_i^{-\theta} - n+1} \cdot \left(\sum\limits_{i=1}^{n} F_i^{-\theta-1} \frac{\partial F_i}{\partial \alpha_k} \right) - \sum\limits_{i=1}^{n} \frac{\frac{\partial F_i}{\partial \alpha_k}}{F_i} \\ \frac{\partial^2 \ln V}{\partial \theta^2} = -S_1 - nS_2 - \frac{\theta^2 S_2 (S_0-n+1) + 2\theta S_3 (S_0-n+1) - \theta^2 S_3^2 + 2(S_0-n+1)\ln(S_0-n+1)}{\theta^3 (S_0-n+1)^2} \end{cases} \tag{13}$$

, where

$$\begin{cases} S_0 = \sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha,\theta) \\ S_1 = \sum\limits_{i=1}^{n-1} \frac{i^2}{(i\theta+1)^2} \\ S_2 = \sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha,\theta) \ln^2 F(X_i;\alpha,\theta) \\ S_3 = \sum\limits_{i=1}^{n} F^{-\theta}(X_i;\alpha,\theta) \ln F(X_i;\alpha,\theta) \end{cases} . \tag{13'}$$

In the independence case, $\theta = 0$, in the first equation, all the terms except $\frac{\partial^2 \ln \tilde{V}}{\partial \alpha_j \partial \alpha_k}$ vanish, and we can say the same about the right side of the second equation: $\lim\limits_{\theta \to 0} \frac{\partial^2 \ln V}{\partial \alpha_k \partial \theta} = 0$. In the case of the Clayton copula, the Hessian is negatively defined. For this reason, the non-linear system

$$\begin{cases} \frac{\partial \ln V}{\partial \alpha_k} = 0 \\ \frac{\partial \ln V}{\partial \theta} = 0 \end{cases} \tag{14}$$

is solved by means of the Newton-Raphson method, solving the involved linear system having the matrix of this system the Hessian (the Jacobean in the general case of non-linear systems) and the right sides given by the actual values of the left sides in (14) with the inverse sign by the Cholesky method. Before applying the Cholesky method, we multiply first the linear system by (-1) in order to obtain a positively defined matrix of the linear system.

3. APPLICATIONS

Consider the ROBOR rate between January 1, 2017 and April 3, 2018 (313 data points, observed daily, five days/ week).

First, we apply the classical Box-Jenkins approach. If we apply the Dickey-Fuller test, model III, we obtain the coefficients for β, Φ and γ 0.00545, -0.00635 and $5.92 \cdot 10^{-5}$, with the Student statistics 1.4531, -1.64969, respectively 2.5674. The time series is not stationary. After removing the moving average [9] of order $q = 2$, we obtain the new coefficients for β, Φ and γ -0.08792, -0.88554 and 0.000494, with the Student statistics -0.08976, -15.6655, respectively 0.091049. In fact, due to small values, the residues after removing the moving average is multiplied by the constant number 1000.

The SARMA stationary model is

$$X_t = 0.55063X_{t-1} - 0.33823X_{t-2} - 0.2505X_{t-5} + a_t + 0.5a_{t-1}$$

The maximum correlation in absolute value for the white noise a_t among the first 36 is $\rho_{33} = 0.181$, and the maximum partial correlation is $\hat{\rho}_{28} = 0.159$. For the first ten, the maximum autocorrelation and partial autocorrelation in absolute value are for lag 7: -0.081 and -0.073. This means that we have obtained the correct Gaussian model.

But, if we apply to the obtained white noise the BDS (Brock, Deckert and Sheinkman) test, we obtain the Z statistics for maximum dimension 6 and $\varepsilon = 0.7$ between 5.8352 (dimension=2) and 9.23336 (dimension=6). Therefore, we reject the null hypothesis of independence and same distribution, with the 1% threshold.

In the non-Gaussian case we first apply the Mann-Kendal test for initial data, and we obtain the statistics $Z = 17.19643$, which means that ROBOR data follows an increasing trend. After that, we apply the same linear transformations as in the Gaussian case, the Z statistics of Mann-Kendall test becomes $Z = -0.31851$, which means a non-significant decreasing trend (in fact we accept the null hypothesis of lack of trend).

Consider now the Clayton copula and the exponential marginal distribution. We obtain the following results, after 10 iterations using Newton-Raphson method.

Table 1. Results if we use the Clayton copula and the exponential marginal distribution

Value	Initial	Final
λ	0.0101	0.0206
θ	0	10.33141
ln V	-1751.25231	-89.08146
$\begin{pmatrix} \frac{\partial \ln V}{\partial \lambda} \\ \frac{\partial \ln V}{\partial \theta} \end{pmatrix}$	$\begin{pmatrix} 0 \\ -124723.4167 \end{pmatrix}$	$\begin{pmatrix} 0.00927 \\ -0.03483 \end{pmatrix}$
Hessian	$\begin{pmatrix} -3067317.013 & 0 \\ 0 & -10172519.95 \end{pmatrix}$	$\begin{pmatrix} -173360.23 & -20655.083 \\ -20655.083 & -4959.455 \end{pmatrix}$

CONCLUSIONS

In [5] we have presented a heavy tail smooth for non-stationary long memory time series. In this paper, we use auto-copula for stationary long memory time series: an Archimedean copula is the same dependence between X_n and X_{n-1}, and between X_n and X_1.

Before applying a model for a stationary time series, we have to test first the stationarity. In the Gaussian case, we apply the Dickey-Fuller unit root test, as we have mentioned before. But in the non-Gaussian case of our paper, we have to use other tests. For instance, we apply the Mann-Kendall test for lack of trend, used in [13] for discharges of Danube River. After we have found an increasing trend for the initial data, we have made the same transformations as in the Gaussian case to obtain stationary time series. Nevertheless, the Mann-Kendall test confirms the stationary of the transformed time series. An open problem is to check other transformations for non-normal distributions, for instance the ratio in exponential case. Such transformation avoids negative values, and we do not need to subtract the minimum value for obtaining positive values of last time series.

Another open problem is to solve the non-linear system (14) for other types of copula for which we can not separate the parameter θ from the marginal parameters, as in the Clayton case.

For instance, we can use the recurrence formulae obtained in [6] for the Frank copula, Gumbel-Hougaard copula, Gumbel-Barnet copula and Ali-Mikhail-Haq copula. For this, we need an analogue recurrence formula, but for mixed derivatives including u, and θ.

In the exponential case discussed in our paper, the non-linear system (14) has two variables: λ and θ. Therefore, another open problem is to consider other marginal distributions as well. The difficulty is not the number of variables, but the computation of the involved marginal cdf in (10) - (13), (12') and (13').

REFERENCES

[1] *** Piaţa Monetară Interbancară, ROBID - ROBOR: serii zilnice, *The Interactive Database of the National Bank of Romania*, www.bnr.ro, accessed on 3 Apr. 2018;

[2] G. Dall' Aglio, *Fréchet classes: the beginning*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 1-12, Kluwer Academic Publishers, 1991;

[3] W.A. Brock, W.P. Deckert and J.A. Sheinkman, A Test for Independence Based on the Correlation Dimension, *Department of Economics, University of Wisconsin at Madison, University of Houston and University of Chicago, grants no. SES-8420872, 144-AH01 and SES-8420930*, 1987;

[4] P.J. Brockwell and R.A. Davis, *Springer Texts in Statistics. Introduction to Time Series and Forecasting*, Springer-Verlag, 2002;

[5] D. Ciuiu, *Heavy Tail Smooth and Application to Long Memory Time Series*, in L. Chivu, C. Ciutacu, V. Ioan-Franc and J.-V. Andrei (Eds) *Economic Dynamics and Sustainable Development - Resources, Factors, Structures and Policies*, pp. 157-164, Proceedings of The 3rd International Conference "Economic Scientific Research - Theoretical, Empirical and Practical Approaches ESPERA 2015", December 3-4 2015, Bucharest, Peter Lang, 2016;

[6] D. Ciuiu, Simulation of Queueing Systems with Many Stations and of Queueing Networks Using Copulas, *Mathematical Modeling in Civil Engineering*, no. 3, pp. 72-87, 2010;

[7] D.A. Dickey and W.A. Fuller, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica* vol. 49, no. 4, pp. 1057-1072, 1981;

[8] C. Genest, Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of American Statistical Association*, vol. 19, no. 1, pp. 1034-1043, 1993;

[9] D. Jula and N.M. Jula, *Prognoza economică*, Ed. Mustang, Bucharest, 2015;

[10] S. Kotz and J.P. Seeger, *A new approach to dependence in multivariate distributions*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 113-127, Kluwer Academic Publishers, 1991;

[11] R. Nelsen, *Copulas and association*, in G. Dall' Aglio, S. Kotz and G. Salinetti (Eds) *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, pp. 51-74, Kluwer Academic Publishers, 1991;

[12] Th. Popescu, *Serii de timp. Aplicaţii în analiza sistemelor*, Technical Publishing House, Bucharest, 2000;

[13] R. Trandafir, D. Ciuiu, R. Drobot, *Testing Some Hypotheses for the Discharges of the Danube River*, in G. Păltineanu, P. Matei and G. Groza (Eds) *Proceedings of the 11th Workshop of the Department of Mathematics and Computer Science, Technical University of Civil Engineering Bucharest, May 27 2011*, *TUCB*, pp. 95-98, Matrix Rom, Bucharest, 2011;

[14] I. Văduva, *Modele de simulare*, Bucharest University Printing House, 2004;

[15] I. Văduva, Simulation of some multivariate distributions, *Analele Universităţii Bucureşti*, no. 1, pp. 127-140, 2003.