SIS 2017
Statistics and Data Science: new challenges, new generations

28–30 June 2017
Florence (Italy)

Proceedings of the Conference of the Italian Statistical Society

edited by
Alessandra Petrucci
Rosanna Verde

FIRENZE UNIVERSITY PRESS
2017

SIS 2017. Statistics and Data Science: new challenges, new generations : 28-30 June 2017 Florence (Italy) : proceedings of the Conference of the Italian Statistical Society / edited by Alessandra Petrucci, Rosanna Verde. – Firenze : Firenze University Press, 2017.
(Proceedings e report ; 114)

http://digital.casalini.it/9788864535210

ISBN 978-88-6453-521-0 (online)

Peer Review Process
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

SOCIETÀ ITALIANA DI STATISTICA
Sede: Salita de' Crescenzi 26 - 00186 Roma
Tel +39-06-6869845 - Fax +39-06-68806742
email: sis@caspur.it web:http://www.sis-statistica.it

La Società Italiana di Statistica (SIS), fondata nel 1939, è una società scientifica eretta ad Ente morale ed inclusa tra gli Enti di particolare rilevanza scientifica. La SIS promuove lo sviluppo delle scienze statistiche e la loro applicazione in campo economico, sociale, sanitario, demografico, produttivo ed in molti altri settori di ricerca.

SIS2017 Committees

Scientific Program Committee:
Rosanna Verde (chair), Università della Campania "Luigi Vanvitelli"
Maria Felice Arezzo, Sapienza Università di Roma
Antonino Mazzeo, Università di Napoli Federico II
Emanuele Baldacci, Eurostat
Pierpaolo Brutti, Sapienza Università di Roma
Marcello Chiodi, Università di Palermo
Corrado Crocetta, Università di Foggia
Giovanni De Luca, Università di Napoli Parthenope
Viviana Egidi, Sapienza Università di Roma
Giulio Ghellini, Università degli Studi di Siena
Ippoliti Luigi, Università di Chieti-Pescara "G. D'Annunzio"
Matteo Mazziotta, ISTAT
Lucia Paci, Università Cattolica del Sacro Cuore
Alessandra Petrucci, Università degli Studi di Firenze
Filomena Racioppi, Sapienza Università di Roma
Laura M. Sangalli, Politecnico di Milano
Bruno Scarpa, Università degli Studi di Padova
Cinzia Viroli, Università di Bologna

Local Organizing Committee:
Alessandra Petrucci (chair), Università degli Studi di Firenze
Gianni Betti, Università degli Studi di Siena
Fabrizio Cipollini, Università degli Studi di Firenze
Emanuela Dreassi, Università degli Studi di Firenze
Caterina Giusti, Università di Pisa
Leonardo Grilli, Università degli Studi di Firenze
Alessandra Mattei, Università degli Studi di Firenze
Elena Pirani, Università degli Studi di Firenze
Emilia Rocco, Università degli Studi di Firenze
Maria Cecilia Verri, Università degli Studi di Firenze

Supported by:
Università degli Studi di Firenze
Università di Pisa
Università degli Studi di Siena
ISTAT
Regione Toscana
Comune di Firenze
BITBANG srl

Index

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online),

Preface

The 2017 SIS Conference aims to highlight the crucial role of the Statistics in Data Science. In this new domain of "meaning" extracted from the data, the increasing amount of produced and available data in databases, nowadays, has brought new challenges. That involves different fields of statistics, machine learning, information and computer science, optimization, pattern recognition. These afford together a considerable contribute in the analysis of "Big data", open data, relational and complex data, structured and no-structured. The interest is to collect the contributes which provide from the different domains of Statistics, in the high dimensional data quality validation, sampling extraction, dimensional reduction, pattern selection, data modelling, testing hypotheses and confirming conclusions drawn from the data. In the mention that statistics is the "grammar of data science", statistics has become a basic skill in data science: it gives right meaning to the data. Still, it isn't replaced by newer techniques from machine learning and other disciplines but it complements them. The Conference is also addressed to the new challenges of the new generations: the native digital generations, who are called to develop professional skills as "data analyst", one of the more request professionality of the 21st Century, crossing the rigid disciplinary domains of competence. In this perspective, all the traditional statistical topics are admitted with an extension to the related machine learning and computer science ones. The present volume includes the short papers of the contributions that will be presented in the 4 invited speaker sessions; in the 19 specialized sessions; in the 11 solicited sessions; in the 6 foreign societies sessions and in the 17 contributed sessions as well as, in the panel session.

Rosanna Verde
President of the Scientific Programme Committee

Alessandra Petrucci
President of the Local Organizing Committee

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

A G.E.D. method for market risk evaluation using a modified Gaussian Copula

Un metodo G.E.D. per la valutazione del rischio di mercato usando una Copula Gaussiana modificata

Massimiliano Giacalone and Demetrio Panarello

Abstract In this paper, we show some results regarding the evaluation of Value-at-Risk (VaR) of some portfolios using a Gaussian Copula, modified by introducing the Generalized Correlation Coefficient, and assuming a Generalized Error Distribution (G.E.D.) for the single returns in the portfolios. In the literature, various authors considered the Copula function approach to evaluate market risk. In our proposal we consider a Lp_{min} algorithm to estimate p, the shape parameter of the distribution. Finally, we compare the classical RiskMetrics method with our G.E.D. method based on a modified Gaussian Copula.

Abstract *In questo lavoro vengono mostrati alcuni risultati riguardanti la valutazione del Valore a Rischio (VaR) di alcuni portafogli utilizzando una Copula Gaussiana, modificata introducendo il Coefficiente di Correlazione Generalizzato, ed assumendo che i singoli rendimenti dei portafogli siano distribuiti secondo una Generalized Error Distribution (G.E.D.). Nella letteratura, vari autori hanno affrontato il tema della valutazione del rischio di mercato considerando l'approccio della funzione Copula. Nella nostra proposta consideriamo un algoritmo* Lp_{min} *per stimare* p*, il parametro di forma della distribuzione. Infine, confrontiamo il classico metodo RiskMetrics con il nostro metodo G.E.D. basato su una Copula Gaussiana modificata.*

Key words: Value-at-Risk, Gaussian Copula, RiskMetrics Method, Generalized Error Distribution, Generalized Correlation Coefficient.

Massimiliano Giacalone
Department of Economics and Statistics, University of Naples 'Federico II', e-mail: massimiliano.giacalone@unina.it

Demetrio Panarello
Department of Economic and Legal Studies, Parthenope University of Naples, e-mail: demetrio.panarello@uniparthenope.it

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online),

1 Introduction

One of the most important issues in finance is to correctly measure the riskiness of a portfolio, which is fundamental to preserve its value over time. Since asset returns are usually fat-tailed, the use of Gaussian processes leads to an underestimation of the risk (Rachev et al., 2005).
Value-at-Risk (VaR) is used to quantify the risk of loss of an asset or a portfolio. The most straightforward method to calculate the (1-c)% Value-at-Risk is the Risk-Metrics one (Longerstaey et al., 1996), where it is hypothesized that the returns R_i, with i=1,2,...,N, of the N assets of a portfolio are jointly distributed according to a Gaussian multivariate (Kasch & Caporin, 2013). However, this hypothesis is simplifying, since it only considers the first two moments, neglecting the fact that the variations in asset returns usually have a leptokurtic and asymmetric behavior (Caporin, 2003).
For a better calculation of the risk, one of the proposals (e.g. Malevergne & Sornette, 2003) is to model the returns' interdependence of the assets in a portfolio by means of Copula functions. Here, the problem is to identify the marginal distributions that best model the returns of the single assets, and to define the Copula which is more suitable to represent the returns' interdependence structure.

2 The Generalized Error Distribution and the Gaussian Copula

The Generalized Error Distribution (G.E.D.) family was introduced by Subbotin (1923) and has been employed by various authors with different names and parameterizations. In our paper, we will use the Vianelli (1963) parameterization, which is:

$$f(x;\mu,\sigma_p,p) = \frac{1}{2\sigma_p p^{p-1}\Gamma(1/p)} exp(-\frac{1}{p}|\frac{x-\mu}{\sigma_p}|^p) \quad for \quad -\infty < x < \infty \qquad (1)$$

where μ is the location parameter, $\sigma_p = [E|x-\mu|^p]^{1/p}$ is the scale parameter and $p \geq 1$ is the shape parameter.
A Copula is a simple function that associates univariate marginal distributions to their joint ones (Jaworski, 2010). There are various Copula functions in the literature (McNeil et al., 2015) and others can be introduced, in order to capture the different dependence structures among stochastic variables.
In the bivariate case, the Gaussian Copula is:

$$C(u,v|\rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp\{\frac{-(r^2-2\rho rs+s^2)}{2(1-\rho^2)}\} dr\ ds$$

where Φ^{-1} is the inverse of Gaussian distribution function and ρ is the Pearson's correlation coefficient.
The Gaussian Copula considered here is indicated with $C(\rho_p)$, since the ρ parame-

ter is replaced by the Generalized Correlation Coefficient ρ_p, introduced by Taguchi (1974) as the correlation parameter of a bivariate Generalized Error Distribution, and defined as (Agrò & Martorana, 2002):

$$\rho_p = \frac{codisp^{(p)}(X,Y)}{\sigma_p(X)\sigma_p(Y)}, with -1 \leq \rho_p \leq 1$$

where

$$|codisp^{(p)}(X,Y)|^p = |E[(Y-\mu_Y)|X-\mu_X|^{p-1} sign(X-\mu_X)]| \cdot |E[(X-\mu_X)|Y-\mu_Y|^{p-1} sign(Y-\mu_Y)]|,$$

$$\sigma_p(X) = [E|X-\mu_X|^p]^{1/p},$$

$$\sigma_p(Y) = [E|Y-\mu_Y|^p]^{1/p},$$

μ_X and μ_Y power means of order p.

3 The algorithm

The joint density required for calculating portfolio's Value-at-Risk is obtained as (Agrò, 2008):

$$1-c = \int\int_{s+t\leq -VaR} f(s,t)ds\ dt$$

The parameters μ, p and σ are estimated using the Lp_{min} method (Giacalone, 1996; Giacalone & Richiusa, 2006).
The ρ_p parameter is estimated using the Exponentially Weighted Moving Average recursive formula:

$$\rho_p = \frac{codisp^{(p)}_{t+1}(x,y)}{D^{(p)}_{t+1}(x) D^{(p)}_{t+1}(y)} \tag{2}$$

where

$$codisp^{(p)}_{t+1}(x,y) = \left(|(\mu^{(p)}_{y/x})_{t+1}|x|(\mu^{(p)}_{x/y})_{t+1}|\right)^{1/p} \text{ x } sign[(\mu^{(p)}_{y/x})_{t+1} + (\mu^{(p)}_{x/y})_{t+1}]$$

$$(\mu^{(p)}_{y/x})_{t+1} = (1-\lambda)\sum_{i=1}^{n} \lambda^{n-i} y_i |x_i|^{p-1} sign(x_i)$$

$$(\mu^{(p)}_{x/y})_{t+1} = (1-\lambda)\sum_{i=1}^{n} \lambda^{n-i} x_i |y_i|^{p-1} sign(y_i)$$

$$[D^{(p)}_{t+1}(x)]^p = (1-\lambda)\sum_{i=1}^{n} \lambda^{i-1} |x|^p_{t-1}$$

$$[D^{(p)}_{t+1}(y)]^p = (1-\lambda)\sum_{i=1}^{n} \lambda^{i-1} |y|^p_{t-1}$$

The fundamental steps in the algorithm are:

- estimation of the parameters μ_i, p_i, σ_i for the two series of returns;
- estimation of the ρ_p parameter, with $p = \sum_{i=1}^{2} p_i/2$;
- generation of (x, y) pairs, which is the realization of the double stochastic variable (X, Y) having G.E.D. marginals and relation of dependence expressed by a $C(\rho_p)$;
- calculation of the distribution function of the returns in the portfolio;
- identification of Value-at-Risk of the return distribution.

4 Application and results

In order to evaluate and compare the performances of the two considered VaR methods, two portfolios were constructed, as is described below:

1. a Bond-ETF Portfolio, made up of a BTP-1FB37 4% Italian Bond, a BTP-1MZ21 3.75% Italian Bond and a LYXOR Exchange-Traded Fund. The data used are the daily prices for the years 2012-2016, for a total of 1267 data (data source: Teleborsa.it);
2. an Exchange indices Portfolio, made up of three indices on stock exchanges: the Euro-US Dollar (EUR-USD), the Pound Sterling-US Dollar (GBP-USD) and the Swiss Franc-Yen (CHF-JPY). The data used in this case refer to the daily quotations 2012-2016 for a total of 1305 data (data source: Investing.com).

Each time series of daily prices p_t was transformed into a series of logarithmic returns according to the relationship:

$$R_t = log(p_{t+h}) - log(p_t)$$

where h, the temporal interest interval, is set as one day.
The estimates of the p shape parameter were $\hat{p} < 2$; hence the distributions of the returns are leptokurtic and more fat-tailed than the Gaussian one. Figure 1 shows the returns of the two portfolios, with the adaptation of a Gaussian distribution and a G.E.D. one.
The reliability of the methods for calculating Value-at-Risk is evaluated by means of a backtest, the (1-c)% VaR prediction data being compared with the values of profits and losses effectively recorded in the market.

The RiskMetrics method (VaR-R.M.) and the G.E.D. method, here called VaR-G.E.D., were applied to the two portfolios.
The backtest applied to the Bond-ETF and Exchange indices portfolios to predict 2.5% VaR and 0.5% VaR highlighted the better predictive capacity of the method which considers the leptokurtosis of marginal distributions. The G.E.D. VaR gives

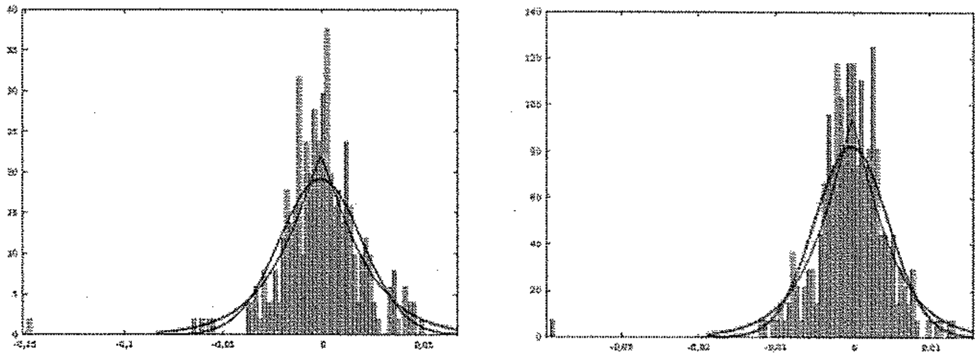

Fig. 1 Bond-ETF (left) and Exchange indices (right) Portfolios, $\hat{p} = 1.4$

predictions which are closer to the real losses, also in relation to the extreme losses that are present in the time series of returns. Moreover, the extreme event, i.e. the exceptional loss or profit, is not somatized in a short time but influences subsequent predictions, which hence are of a cautionary type (overestimation of the risk).
The number of VaR violations can be seen as a binomial stochastic variable in which the probability of success p is the percentage of VaR violations predicted (for example 5%) and the number of trials m is the number of days used for the backtest.

	Bond-ETF Port.		Exchange indices Port.	
	2.5%	0.5%	2.5%	0.5%
VaR-R.M.	6	4	5	3
Var-G.E.D.	5	2	3	2
Confidence intervals	1-9	0-3	1-9	0-3

Table 1 VaR violations and 95% confidence intervals

Tab. 1 gives the number of VaR violations recorded for the two methods and the relative 95% confidence intervals in a number of observations m = 200. It shows that, for both portfolios, the VaR violations with the VaR-G.E.D. method are lower than the ones with the VaR-R.M. method and are within the confidence range.

5 Conclusions

In our application, we made a comparison between the Gaussian and the G.E.D. Copula. The reason is that the variation of the p shape parameter allows the G.E.D. to represent all the symmetric distributions that are described in the literature. That is, after estimating p, we are able to use the Copula which best fits our data: all the

Copula functions can be obtained as particular cases of the G.E.D. Copula.
Among the different methods proposed in the literature for calculating Value-at-Risk, we took into account the well-known RiskMetrics method. We proposed a G.E.D. method and evaluated its performance compared to the RiskMetrics one. The two methods were evaluated by backtest, in order to examine the ability of predicting the potential loss of a portfolio.
The results obtained confirm the higher performance of the G.E.D. method, while the assumption of normality of the returns' distribution determines confidence intervals with the lowest predictive power. The assumption of normality, subject to verification, was rejected as the returns of all stocks examined have kurtosis characteristics which are neglected by the RiskMetrics method. It does seem that the VaR-G.E.D. method can constitute a valid generalization of the VaR-R.M., which it is close to in the case of Gaussian marginal distributions, while it moves away from it if the distributions are more fat-tailed.
All the necessary calculations have been implemented and processed on the statistical environment R.

References

1. Agro', G.: On VaR using modified Gaussian Copula. Annali della Facolta di Economia dell'Universita di Palermo (2008)
2. Agro', G., Martorana, G.: VaR e Copula Normale di ordine p. Atti del XXVI Convegno AMASES, Verona, 19-22 (2002)
3. Caporin, M.: The trade off between complexity and efficiency of VaR measures: a comparison of RiskMetric and GARCH-type models (Vol. 3). GRETA, working papers (2003)
4. Giacalone, M.: Parameter evaluation of an exponential power function by simulation study. In: Shorts Communications and Posters. Compstat 96, Barcelona (1996)
5. Giacalone, M., Richiusa, R.: Lp-norm estimation: some simulation studies in presence of multicollinearity. Student, 5, 235-246 (2006)
6. Jaworski, P., Durante, F., Hardle, W. K., Rychlik, T.: Copula theory and its applications. Springer (2010)
7. Kasch, M., Caporin, M.: Volatility Threshold Dynamic Conditional Correlations: An International Analysis. J.Financial Econom., 11, 706-742 (2013)
8. Longerstaey, J., Zangari, P.: RiskMetrics Technical Document. J.P. Morgan, Fourth Edition, New York (1996)
9. Malevergne, Y., Sornette, D.: Testing the Gaussian Copula hypothesis for financial assets dependences. Quantitative Finance, 3, 231-250 (2003)
10. McNeil, A. J., Frey, R., Embrechts, P.: Quantitative risk management: Concepts, techniques and tools. Princeton University Press (2015)
11. Rachev, S. T., Menn, C., Fabozzi, F. J.: Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing. John Wiley & Sons (2005)
12. Subbotin, M.: On the Law of Frequency of Error. Mathematicheskii Sbornik (1923)
13. Taguchi, T.: On Fechner's thesis and statistics with Norm-p. Annals of the Institute of Statistical Mathematics, 26(1), 175-193 (1974)
14. Vianelli, S.: La misura della variabilita condizionata in uno schema generale delle curve normali di frequenza. Statistica, 33, 447-474 (1963)