5. MODELING THE DEPENDENCY STRUCTURE OF STOCK INDEX RETURNS USING A COPULA FUNCTION APPROACH[1]

Ciprian NECULA*

Abstract

In the present study we assess the dependency structure between stock indexes by econometrically estimating the empirical copula function and the parameters of various parametric copula functions. The main finding is that the t-copula and the Gumbel-Clayton mixture copula are the most appropriate copula functions to capture the dependency structure of two financial return series. With the dependency structure given by the estimated copula functions we quantify the efficient portfolio frontier using as a risk measure CVaR (Conditional VaR) computed by Monte Carlo simulation. We find that in the case of using normal distributions for modeling individual returns the market risk is underestimated no mater what copula function is employed to capture the dependency structure.

Keywords: copula functions, copula mixtures, the efficient portfolio frontier, *Conditional VAR*, Monte Carlo simulation

JEL Classification: C51, C52, G10, G11

1. Introduction

The notion of correlation is central to modern financial theory. The Capital Asset Pricing Model (CAPM) uses the correlation as a measure of dependence between different financial instruments. Furthermore, the correlation is important in the context of pricing derivatives with payoffs depending on multiple assets.

[1] **Acknowledgements:** I would like to thank Professor Moisă Altăr and two anonymous referees for useful comments and suggestions. This work was supported by CNCSIS-UEFISCSU, project number PN II-RU PD_583/2010.

* Academy of Economic Studies, Bucharest; email: ciprian.necula@fin.ase.ro.

It is empirically proved that asset returns from a wide range of markets are leptokurtic, being heavy tailed and overly concentrated around the mean when compared to a Gaussian distribution (Cont, 2001). Many authors have successfully fitted Generalized Hyperbolic Distributions and in particular Normal Inverse Gaussian laws to returns in financial time series. Necula (2009) estimated the parameters of the Generalized Hyperbolic Distribution for the returns of various Eastern European stock indexes, and concluded that the probability density function of the estimated distribution represents an almost exact approximation (at least up to the 4th order term) of the empirical probability distribution function, computed using a non-parametric kernel method. When the return distribution deviates from the elliptic class which includes the Gaussian, a single statistic such as the correlation is not able to correctly assess the dependency structure between the returns. Tail dependence may differ from dependence close to the mean. Also, the dependence between two assets as the market rises may be different than the dependence as the market falls. Furthermore, Blyth (1996) and Shaw (1997) have pointed out that the linear correlation coefficient cannot capture the non-linear dependence relationships that exist between many financial returns series.

There are also more pitfalls of correlation as a dependency measure. Correlation is simply a scalar measure of dependency, and it cannot embody everything one would like to know about the dependence structure of risks. Possible values of correlation depend on the marginal distribution of the risks, all values between -1 and 1 not being necessarily attainable. Perfectly positively dependent risks may not necessarily have a correlation of 1 and perfectly negatively dependent risks may not necessarily have a correlation of -1. A correlation of zero does not imply independence of the risk factors. Correlation is not invariant under nonlinear transformations of the risk factors. Correlation is only defined when the variances of the risks are finite. It is not an appropriate dependence measure for very heavy-tailed distributions like the Generalized Hyperbolic Distribution or α - stable distributions (Belkacem et al., 2000).

Using the Kendall or Spearman definitions of rank correlation, some of the deficiencies of the standard linear correlation can be eliminated. For example, rank correlation coefficients are invariant under increasing nonlinear transformations. Also, for arbitrary marginal distributions, a bi-dimensional distribution can be exactly constructed that has a rank correlation anywhere in the interval [-1, 1]. But, as in the case of the linear correlation, this distribution is not unique. Some of the deficiencies remain. For example, a rank correlation of zero does not imply independence. Moreover, rank correlations cannot be as easily manipulated as the linear correlation.

Copulas represent a way of trying to extract the dependence structure from the joint distribution and to extricate dependence and marginal behavior. The theory of copulas dates back to Sklar (1959), but its application in financial modeling is far more recent and dates back to the late 1990s. A copula is a function that embodies all the information about the dependence structure between the components of a random vector. When it is applied to marginal distributions, which do not necessarily belong to the same distribution family, it results in a proper multivariate distribution. As a consequence, this theory enables us to incorporate a flexible modeling of the dependence structure between different variables, while allowing them to be modeled by different marginal distributions. Using copula functions one can construct a dependency measure characterized by the equivalence between a null value of this measure and the independence of the analyzed data series.

Nelsen (1999) provide an introduction to copula theory, while Cherubini et al. (2004) provide a discussion of copula methods for financial applications. Modeling the dependency structure by copula functions was employed to estimate market risk using VaR methodology (Embrechts et al., 2002; Embrechts et al, 2003), to quantify credit risk (Li, 2000), to quantify operational risk (Ceske and Hernandez, 1999), and to price complex multi-asset derivatives (Bennett and Kennedy, 2004).

In the present study we assess the dependency structure between stock indexes (the pair BUX - Hungary and PX50 - Czech Republic as well as the pair DAX – Germany and SP500 – USA) by econometrically estimating the empirical copula function and the parameters of various parametric copula functions (i.e. three Archimedean copula functions, mixtures between copula functions, the t-Student copula and the Gaussian copula). With the dependency structure given by the estimated copula functions we quantify the efficient portfolio frontier using as a risk measure CVaR (Conditional VaR) computed by Monte Carlo simulation. We employed these particular stock indexes because the present paper is part of series that model and analyze different empirical facts of financial returns. To maintain the consistency of the results across studies we use the same dataset. For example, in a companion paper (Necula, 2010) we develop a new two-dimensional Copula-GARCH model and test it using the dataset.

The rest of the paper is organized in five sections. In the second section we present the main properties of the copula functions. In the third section we present the data and the methodology employed for the analysis. In the fourth section we econometrically estimate the parameters of the copula functions. In the fifth section we compute de efficient frontier for a portfolio of two stock indexes. The final section concludes.

2. The copula functions

A copula (C) is a function with the following properties:

1. $C:[0,1]^2 \to [0,1]$;
2. $C(u,0) = C(0,u) = 0$;
3. $C(u,1) = C(1,u) = u$;
4. $\sum_{i_1=1}^{2}\sum_{i_2=1}^{2}(-1)^{i_1+i_2} C\left(u_{1i_1}, u_{2i_2}\right) \geq 0$.

In fact a copula represents the cumulative distribution function (cdf) of a multidimensional distribution with uniform marginal distributions.

Sklar (1959) proved that a copula function represents the connection between a bi-dimensional distribution and its two marginal distributions, capturing the dependency structure. More precisely, if F is the cdf of the bi-dimensional distribution and F_1 and F_2 are the cdfs of the marginal distributions, there is a unique copula C such that:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

Also, if the cdfs for the bi-dimensional and for the marginal distributions are known, the associated copula function is given by (Sklar, 1959):

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right)$$

An important property of copula functions is the invariance to non linear transforms. Also, the copula functions may be employed to define some dependency measures (Schweitzer and Wolff, 1981):

- the σ dependency coefficient:

$$\sigma = 12 \iint\limits_{[0,1]^2} \left| C(u_1, u_2) - u_1 u_2 \right| du_1 du_2$$

- the Φ dependency coefficient:

$$\Phi = 90 \iint\limits_{[0,1]^2} \left| C(u_1, u_2) - u_1 u_2 \right|^2 du_1 du_2$$

The dependency measures defined using a copula function have the appealing property that a null value of this measures implies the independence of the two random variables.

An important class of copula functions consists of Archimedean copulas (Genest and Rivest, 1993; Nelsen, 1999). An Archimedean copula is generated by a one-dimensional function ψ

$$C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2))$$

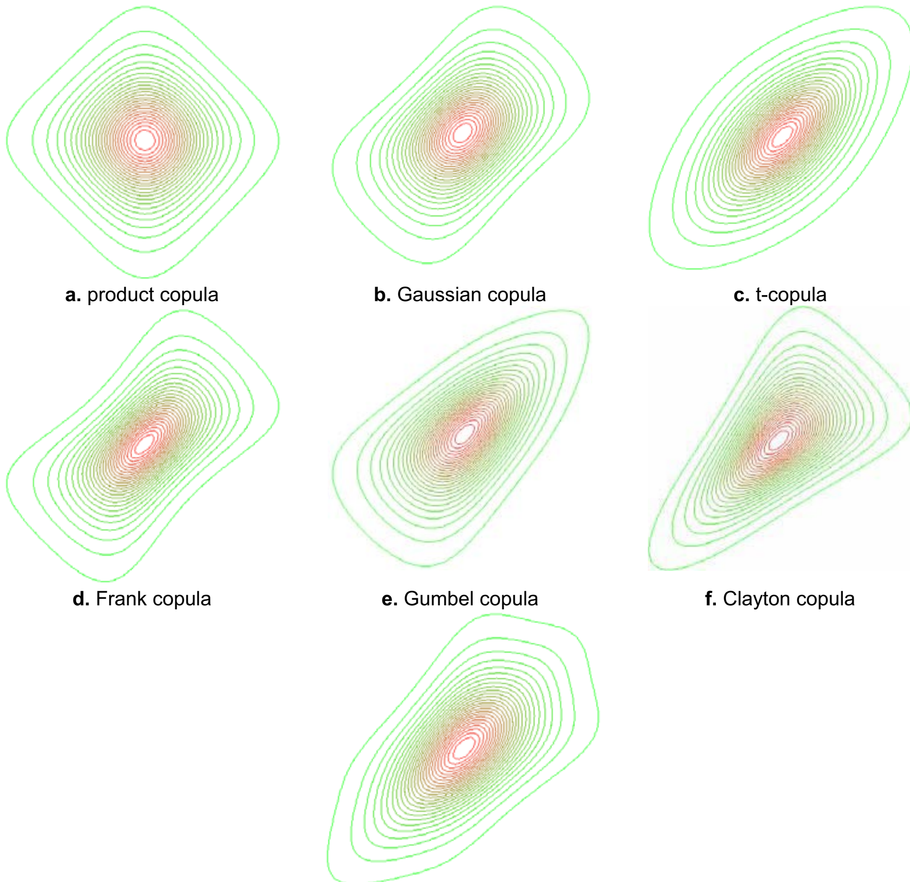
where the generator function ψ must have the following properties:

1. $\psi(1) = 0$;
2. $\psi' > 0$;
3. $\psi'' < 0$.

The most commonly used copulas in finance are the product copula (i.e. the copula that models independence), the Gaussian copula, the t-copula, and three Archimedean copulas: Frank, Gumbel and Clayton. In order to capture the dependency structure of these copulas, Figure 1 depicts the bi-dimensional cumulative distribution function (cdf) generated by each of these copulas providing that the marginal distributions are modeled using the Generalized Hyperbolic Distribution.

Figure 1

Contour plots of bi-dimensional cdf for different copulas



a. product copula **b.** Gaussian copula **c.** t-copula

d. Frank copula **e.** Gumbel copula **f.** Clayton copula

g. Gumbel-Clayton mixture

Source: author's calculations

The Gumbel copula captures dependence in the upper tail of the marginal distributions, while the Clayton copula models the dependence in the lower tail. Financial series usually have dependence both in the lower tail and the upper tail. One way to model such a dependence structure is to employ the t-copula. A more flexible alternative is to use a Gumbel-Clayton mixture. For further information on copula functions one can consult Nelsen (1999) and Cherubini et al. (2004).

3. Data and methodology

The data employed in the study consists of daily returns between January 1998 and December 2008 for stock indexes from four countries: Czech Republic (PX50), Hungary (BUX), Germany (DAX), and USA (SP500). The source of the raw data is Bloomberg Market Data.

The skewness and the kurtosis of the returns together with the Kolmogorov-Smirnov and the Anderson-Darling normality test statistics are presented in Table 1.

Table 1

The distributional characteristics of the returns

Index	Skewness	Kurtosis	KS statistic	AD statistic
PX50	-0.0546	7.0230	0.0502	∞
BUX	-0.4920	11.8510	0.0544	∞
DAX	-0.1399	5.5697	0.0617	4.5234
SP500	-0.2036	6.6866	0.0541	4.4626

Source: Author's calculations.

As expected, the returns are leptokurtic, both the normality tests rejecting the null hypothesis of Gaussian distributed returns.

Using the concept of copula function, the difficult problem of estimating the common two-dimensional probability density function for two financial return series can be decomposed into two more manageable problems: estimating the marginal probability density function and estimating the relevant copula function. The methodology for analyzing the dependency structure between PX50 and BUX, and between DAX and SP500 consists of the following steps:

1. computing the empirical copula (Deheuvels, 1979) using non-parametric econometric techniques (Gijbels and Mielniczuk, 1990; Fermanian and Scaillet, 2003). The kernel empirical copula ($\hat{C}$) is given by:

$$\hat{C}(u_1,u_2)=\frac{1}{T}\sum_{i=1}^{T}G_{u_1,h}\left(\frac{u_1-\hat{F}_1(x_1^i)}{h}\right)G_{u_2,h}\left(\frac{u_2-\hat{F}_2(x_2^i)}{h}\right)$$

   where $G_{u,h}(\cdot)$ is the Gaussian kernel with bandwidth h, and $\hat{F}_1,\hat{F}_2$ are the empirical cdfs of the two marginal distributions, estimated by non-parametric one-dimensional kernel methods. The length of the bandwidth was chosen according to the well-known "rule of thumb" of Silverman (1986).

2. econometrically estimating the parameters of different parametric copulas by using the Canonical Maximum Likelihood (CML) method (Yan, 2006). This method consists of maximizing the following likelihood function:

$$L=\sum_{t=1}^{T}\ln c\left(\hat{F}_1(x_1^t),F_2(x_2^t)\right)$$

where: $c(u_1, u_2) = \frac{\partial^2 C}{\partial u_1 \partial u_2}(u_1, u_2)$ is the copula density, and $\hat{F}_1, \hat{F}_2$ are the empirical cdfs of the two marginal distributions. This method provides consistent estimators for the parameters and it is less computing intensive than the Exact Maximum Likelihood;

3. comparing the Akaike Information Criterion (AIC) statistic for the parametric models in order to select the one that fits the data best;
4. employing the Kormogorov-Smirnov test and the Anderson-Darling test to asses the goodness-of-fit of the estimated copulas (Fermanian, 2005). More specifically, if we define

$$Z_1 = F_1(X_1),\ Z_2 = \frac{\partial C}{\partial u_1}(F_1(X_1), F_2(X_2)) \Big/ F_1(X_1)$$

then the random variables $\phi^{-1}(Z_1), \phi^{-1}(Z_2)$ are independent and normally distributed, where ϕ is the cdf of the standard Gaussian distribution. Hence, the random variable

$$S = \phi^{-1}(Z_1)^2 + \phi^{-1}(Z_2)^2$$

has a $\chi^2(2)$ distribution.

In the second part of the study we use CVaR as a risk measure in order to compute the efficient portfolio frontier. The CVaR is quantified by Monte Carlo simulation. The algorithm for generating two random numbers (x_1, x_2) from a dependency structure given by a copula C consists of the following steps:

- generate two random numbers v_1, v_2 from a uniform $(0,1)$ distribution;
- take $u_1 = v_1$ and $u_2 = g^{-1}(v_2)$ where $g(u) = \frac{\partial C}{\partial u_1}(v_1, u)$;
- take $x_1 = F_1^{-1}(u_1)$ and $x_2 = F_2^{-1}(u_2)$ where F_1, F_2 are the empirical cdfs of the marginal distributions;

The econometric methods employed in this study were implemented in Maple. All the results in the following sections are computed in Maple, according to the presented methodology.

4. Estimation results

The estimated parameters of various parametric copula functions for the two index pairs for which we analyze the dependency structure are presented in Table 2 and in Table 3. The estimations were performed in Maple, using the CML method, as described in the methodology.

Table 2

Estimated parameters for BUX - PX50 returns pair

Copula	Parameters			AIC	GoF statistics KS	GoF statistics AD
Frank		3.1309*** (0.2181)		-203.26	0.0426	3.0129
Clayton		0.7268*** (0.0581)		-212.48	0.0523	4.2524
Gumbel		1.4049*** (0.0364)		-194.29	0.0440	2.9859
G-C mixture	0.6773*** (0.0797)	0.7158*** (0.0834)	1.6396*** (0.1394)	-236.20	0.0391	2.7288
Gaussian		0.4841*** (0.0231)		-227.44	0.0439	3.5328
t	6		0.4691*** (0.0267)	-238.85	0.0325	1.5272

Standard errors parenthesis; *** denotes statistical significance at 1%.

Source: Author's calculations.

Table 3

Estimated parameters for DAX - SP500 returns pair

Copula	Parameters			AIC	GoF statistics KS	GoF statistics AD
Frank		4.6459*** (0.2385)		-378.83	0.0515	5.4849
Clayton		1.0504*** (0.0684)		-336.31	0.0816	15.3852
Gumbel		1.7858*** (0.0485)		-488.79	0.0499	5.7242
G-C mixture	0.2149*** (0.0635)	1.2269*** (0.5234)	1.8933*** (0.0708)	-493.28	0.0499	4.7762
Gaussian		0.6316*** (0.0169)		-436.75	0.0711	9.8665
t	3		0.6212*** (0.0219)	-523.63	0.0424	3.0579

Standard errors parenthesis; *** denotes statistical significance at 1%.

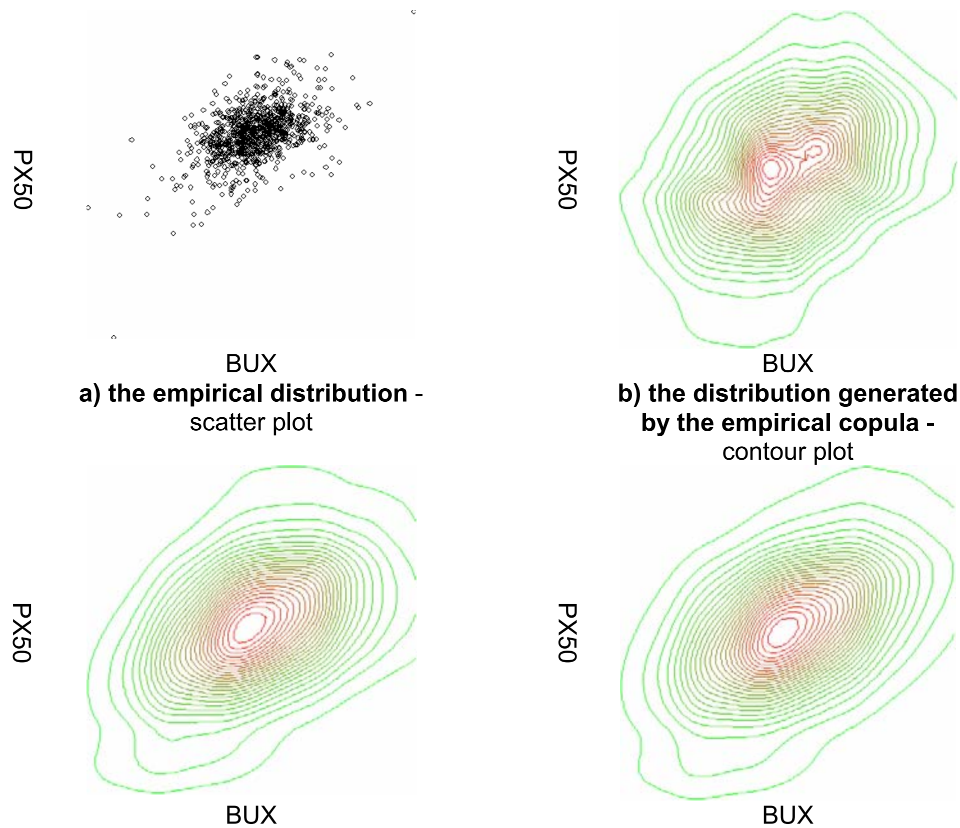
Source: Author's calculations.

Using the Akaike Information Criterion (AIC) and the two goodness-of-fit statistics Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) one can conclude that, in the case of the two returns pairs analyzed in the study, the t-copula and the Gumbel-Clayton mixture copula are the most appropriate to model the dependency structure. In the case of the BUX-PX50 pair more weight is given to the Gumbel copula (i.e. 0.6773), whereas in the case of the DAX-SP500 pair more weight is given to the Clayton copula (i.e. 1 - 0.2149 = 0.7851).

Figure 2 depicts the estimated bi-dimensional distribution for BUX-PX50 pair using the best two parametrical copulas, and the empirical copula, computed using the implemented procedure in Maple as described in the methodology.

Figure 2

Estimated bi-dimensional distributions for BUX-PX50 pair



a) the empirical distribution - scatter plot

b) the distribution generated by the empirical copula - contour plot

b) the distribution generated by the G-C mixture copula - contour plot

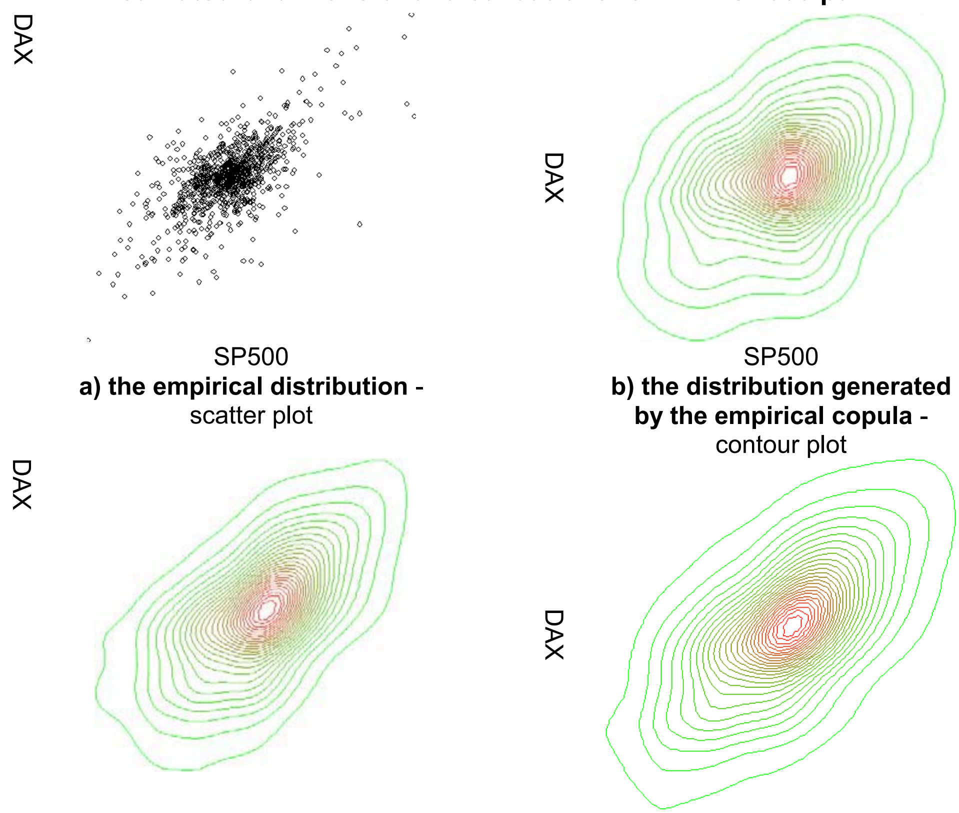
b) the distribution generated by the t- copula - contour plot

Source: Author's calculations.

Figure 3 depicts the estimated bi-dimensional distribution for DAX-SP500 pair modeling the dependency structure by the empirical copula, by the estimated t-copula and by the estimated Gumbel-Clayton mixture.

Figure 3

Estimated bi-dimensional distributions for DAX-SP500 pair



a) the empirical distribution - scatter plot

b) the distribution generated by the empirical copula - contour plot

b) the distribution generated by the G-C mixture copula - contour plot

b) the distribution generated by the t- copula - contour plot

Source: Author's calculations.

In the next section we will construct the efficient portfolio frontier using the dependency structure estimated for the BUX-PX50 pair.

5. The efficient portfolio frontier

Value at risk (VaR) is of central importance in modern financial risk management. Of the various methods that exist to compute the VaR, the most popular are the historical simulation, the variance-covariance method and the Monte Carlo (MC) simulation. While historical simulation is not based on particular assumptions as to the behavior of the risk factors, the two other methods assume some kind of multidimensional normal distribution of the risk factors. Therefore, the dependence structure between different risk factors is described by the covariance or correlation between these factors.

In this study we quantify the market risk modeling the dependency structure of BUX-PX50 pair with copula functions. In order to assess the influence of the marginal probability distribution functions on quantifying market risk, for each copula function we employed two sets of marginal distributions: the empirical marginal distributions (computed using kernel estimation methods), and the normal distributions. First we performed a back-test for VaR 5%. The results are presented in Table 4.

Table 4

VaR 5% back-test

Copula	**Losses greater than VaR 5%**	
	empirical marginals	**Gaussian marginals**
Frank	5.54%	5.91%
Clayton	5.92%	6.68%
Gumbel	5.83%	6.40%
G-C mixture	5.37%	5.86%
Gaussian	4.34%	5.08%
t	3.65%	5.47%

Source: Author's calculations.

As expected, if one uses the empirical marginal distributions then the percent of losses greater than VaR 5% is lower than in the case of using Gaussian marginals.

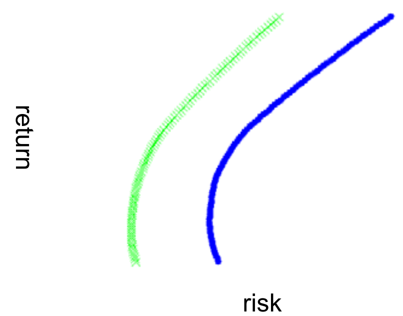
Next we compute the efficient frontier for a two asset portfolio consisting of BUX and PX50 indexes. Since VaR is not a coherent risk measure (Artzner et al., 1998), in estimating the frontier we employed CVaR (Conditional VaR) quantified by Monte Carlo simulation. The computations were performed in Maple.

Figure 4 depicts, in the case of the BUX-PX50 pair, the efficient portfolio frontier if the dependency structure is modeled using four of the copula functions estimated in the previous section: Gaussian copula, Frank copula, t-copula and the Gumbel-Clayton mixture. Also, the computations were performed for the two sets of marginal distributions: the empirical marginal distributions and the Gaussian distributions.

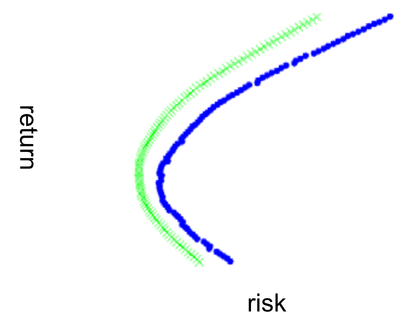
In Figure 4, one may easily notice that no mater the expected return, the associated risk is lower if one employs normal marginal distributions. Therefore, in the case of using Gaussian marginals, the market risk is underestimated no mater what copula function we employed in order to capture the dependency structure between the two assets.

Figure 4

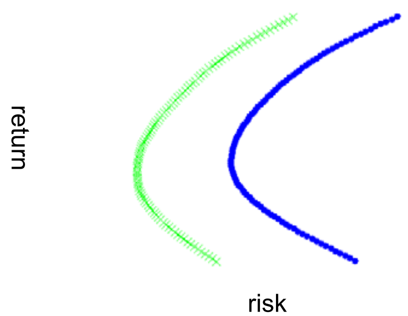
The frontier of efficient BUX-PX50 portfolios



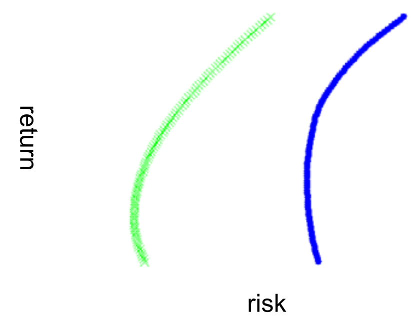
a) the frontier generated by the Gaussian copula



b) the frontier generated by the Frank copula



b) the frontier generated by the G-C mixture copula



b) the frontier generated by the t- copula

Source: Author's calculations.

6. Concluding remarks

In the present study we assessed the dependency structure between two pairs of stock indexes, namely BUX - PX50 and DAX - SP500 by econometrically estimating the empirical copula function and the parameters of various parametric copula functions (i.e. three Archimedean copula functions, mixtures between copula functions, the t-Student copula and the Gaussian copula).

According to the Akaike Information Criterion (AIC) and the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests the t-Student copula and the mixture between the Clayton copula and the Gumbel copula are the appropriate copula functions to capture the dependency structure of two financial return series.

Next, we computed the efficient portfolio frontier using these estimated copula functions. Since VaR is not a coherent risk measure, we quantified the market risk using CVaR (Conditional VaR) estimated by Monte Carlo simulation. In order to assess the influence of the marginal probability distribution functions on quantifying market risk, for each copula function we employed two sets of marginal distributions: the empirical marginal distributions (estimated using kernel estimation methods), and Gaussian distributions.

The main finding is that in the case of using normal distributions for modeling individual financial returns the market risk is underestimated no mater what copula function one employs in order to capture the dependency structure between the two assets.

References

Anderson, T.W. and Darling, D.A. (1954), "A test of goodness of fit," *J. Amer. Statist. Assoc.*, 49: 765-769.

Artzner, P., Delbaen, F., Eber, J-M., and Heath, D. (1998), "Coherent Measures of Risk", *Mathematical Finance*, 9: 203–228.

Belkacem, L., Véhel J., and Walter C. (2000), "CAPM, risk and portfolio selection in α-stable markets", *Fractals*, 8:99–115.

Bennett, M.N. and Kennedy, J.E. (2004), "Quanto pricing with copulas", *Journal of Derivatives,* 12: 26-45.

Blyth, S. (1996), "Out of line", *Risk*, 9(10): 82-84.

Ceske, R., and Hernandez, J. (1999), "Where theory meets practice", *Risk Magazine,* 12: 17-20.

Cherubini U., Luciano E. and Vecchiato W. (2004), *Copula Methods in Finance*, Wiley.

Cont, R. (2001), "Empirical properties of asset returns: Stylized facts and statistical issues", *Quantitative Finance*, 1: 223-236.

Deheuvels, P. (1979), "La fonction de dependance empirique et ses proprietes. Un test non paramtrique d'independance", Acad. Roy. Belg. Bull. Cl. Sci., 65: 274-292.

Embrechts, P., Hoeing, A., and Juri, A. (2003), "Using copulae to bound the Value-at-Risk for functions of dependent risks", *Finance and Stochastics*, 7(2): 145-167.

Embrechts, P., McNeal, A.J., and Straumamm, D. (2002), "Correlation and dependency in risk management: properties and pitfalls". In: M.A.H. Dempster (Ed), *Risk Management: Value at Risk and Beyond*, Cambridge University Press, pp. 176-223.

Fermanian, J. D. (2005), "Goodness-of-fit tests for copulas", *Journal of Multivariate Analysis*, 95: 119-152.

Fermanian, J. D. and Scaillet, O. (2003), "Nonparametric estimation of copulas for time series", *Journal of Risk,* 5: 25–54.

Genest, C., and Rivest, L.P. (1993), "Statistical inference procedures for bivariate Archimedean copulas", *J. Amer. Statist. Assoc.*, 88: 1034-1043.

Gijbels, I., and Mielniczuk, J. (1990), "Estimating the density of a copula function," Comm. Statist. Theory Methods*,* 19: 445-464.

Li, D.X. (2000), “On default correlation: a copula function approach”, *Journal of Fixed Income*, 9(4): 43-54.

Necula, C. (2009), “Modeling Heavy Tailed Stock Index Returns Using the Generalized Hyperbolic Distribution”*, Romanian Journal of Economic Forecasting,* 10(2): 118-131.

Necula, C. (2010), “A Copula-GARCH Model”, *Economic Research (Ekonomska istraživanja),* 23(2): 1-10.

Nelsen, R. B. (1999), *An Introduction to Copulas*, New York: Springer-Verlag.

Schweitzer, B., and Wolff, E. (1981), "On nonparametric measures of dependence for random variables", *Annals of Statistics*, 9: 879-885.

Shaw, J. (1997), “Beyond VAR and Stress Testing". In: S. Grayling (Ed), *VaR: Understanding and Applying Value at Risk*, London: Risk Publications, pp. 211-224.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," Publications de l’Institut de Statistique de l’Université de Paris, 8: 229–231.

Yan, J. (2006), “Multivariate Modeling with Copulas and Engineering Applications". In: H. Pham (Ed.), *Handbook of Engineering Statistics*, Springer, pp. 973-990.