An experimental comparison of Value at Risk estimates based on elliptical and hierarchical Archimedean copulas

Michael Kubát[1], Jan Górecki[2]

Abstract. In this paper, we estimate Value at Risk for a selected portfolio using elliptical and hierarchical Archimedean copulas, where the latter is based on a recent approach to estimation of hierarchical Archimedean copulas based on the Kendall correlation matrix. The estimates are compared using the Kupiec's test for three periods of time: a period of significant movements of index prices, a period of relative calm in the stock market and a long term period which includes the both situation in the stock market. Our experimental results show that the estimates based on elliptical copulas are more accurate in the periods of relative calm in the stock market, whereas the estimates based on hierarchical Archimedean copulas are more accurate in the period of significant movements in the stock market.

Keywords: copula, elliptical copulas, hierarchical Archimedean copulas, Value at Risk, returns modeling, backtesting

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Modeling portfolio returns and the subsequent estimation of the portfolio risk is one of the basic activities of financial institutions. Value at Risk (VaR) is frequently used to estimate the risk of future portfolio losses. Estimation of VaR is legally mandatory - banks are obliged by Basel II and insurance companies by Solvency II. The VaR is usually estimated based on historical data, and assuming constant development and constant characteristics of the portfolio or assets. Copulas, which appeared in connection with the Sklar's theorem [21], are often used for modeling the portfolio. The theorem states that given a multivariate probability distribution, it can be broken down into two components: i) its univariate marginal distributions and ii) the function describing the dependence among the variables, i.e., its copula.

For modeling i), there exist a number of options, e.g., the normal distribution, the skewed Student's distribution [10], a mixture of normal distributions and the Lévy processes [5]. As financial time series often do not meet the conditions of the normal and the Student's probability distributions of nonzero skewness (in case of the normal distribution also fixed kurtosis), the Lévy's models are frequently used for these purposes, see, e.g., [1]. Hence, in this article, we work with one of the Lévy's models - the normal inverse Gaussian (NIG) distribution - which is, as suggested in [4], appropriate for financial time series modeling.

For modeling ii), *elliptical copulas* (ECs) are the most popular choice, see, e.g., their applications to finance described in [20] or an application dealing with modeling and testing portfolio using ECs described in [14]. However, as ECs are radically symmetric, they are not suitable for specific applications, e.g., they can fail to adequately capture dependence between extreme events, see [2, 18] for financial examples. To overcome some of these restrictions, there appeared several multivariate alternatives, e.g., *Archimedean copulas* (ACs) or their asymmetric generalization, *hierarchical Archimedean copulas* (HACs). For example, a successful application of HACs to collateralized debt obligations is reported in [11], which shows their advantages to ECs.

Our research presented in this paper extends the research presented in [11] and aims to experimentally compare VaR estimates based on ECs and HACs. The VaR estimates are computed for three periods of time: a period of significant movements of index prices, a period of relative calm in the stock market and a long period which includes the both situation in the stock market. The results are compared using the widely known Kupiec's test suggested in [17]. Note that in the estimation processes that involve HACs, we use the approach to estimation of HACs based on the Kendall correlation matrix that was recently proposed in [7, 9], which we implemented in Matlab. We choose this HAC estimation approach as it has shown desirable properties when compared to other HAC estimation approaches, e.g., see the experimental comparisons reported in [6, 9].

[1] Department of Informatics, SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic, O130076@opf.slu.cz
[2] Department of Informatics, SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic, gorecki@opf.slu.cz

The paper is structured as follows. Section 2 recalls the NIG distribution, ECs, HACs and backtesting, Section 3 presents the experimental comparison of VaR estimates based on ECs and HACs, and Section 4 concludes.

2 Methodology

2.1 Normal inverse Gaussian model

Normal inverse Gaussian (NIG) model is one of the Lévy processes with parameters $0 < |\beta| < \alpha$, $-\infty < \mu$ and $\delta > 0$, and its density is given by

$$f_{NIG}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha \exp\big(\zeta + \beta(x-\mu)\big) K_1\left(\alpha\delta q\left(\frac{x-\mu}{\delta}\right)\right)}{\pi q\left(\frac{x-\mu}{\delta}\right)}, x \in \mathbb{R},$$

where $\zeta = \delta\sqrt{\alpha^2 - \beta^2}$, $q(y) = \sqrt{1+y^2}$ and K_1 is the modified Bessel function of the third order and index one. For more details on how to estimate its parameters, see, e.g., [1].

2.2 Copulas

Definition 1 [19] A d-dimensional copula is a d-dimensional multivariate distribution function with standard uniform univariate margins.

Copulas establish a connection between multivariate distribution functions and their univariate margins, which is well-known due to Sklar's Theorem.

Theorem 1 (Sklar's Theorem [21]) Let F be a d-dimensional multivariate distribution function with univariate margins $F_1, \dots, F_d$. Then there exists a copula C: $[0,1]^d \to [0,1]$ such that

$$F(x_1, \dots, x_d) = C\big(F_1(x_1), \dots, F_d(x_d)\big), \quad (1)$$

holds for all $(x_1, \dots, x_d) \in \overline{\mathbb{R}}^d$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Such a function C is uniquely determined, if $F_1, \dots, F_d$ are all continuous. Conversely, if C is a copula and $F_1, \dots, F_d$ are univariate distribution functions, then the function F given by (1) is a multivariate distribution function with margins $F_1, \dots, F_d$.

2.3 Elliptical copulas

ECs are based on existing multivariate elliptical distributions. The Gaussian copula is based on the multivariate normal distribution and the Student t-copula is based on the multivariate Student t-distribution. Formally, a *Gaussian* copula is given by [19]

$$C_{\Sigma}^{Ga}(u_1, \dots, u_d) = \phi_{\Sigma}\left(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)\right),$$

where ϕ is the cumulative density function of the univariate normal distribution and ϕ_{Σ} is the cumulative density function of the multivariate normal distribution with a correlation matrix Σ. A *Student t*-copula is given by [19]

$$C_{v,\Sigma}^{t}\big(u_{1,\dots,}u_d\big) = t_{v,\Sigma}\big(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d)\big),$$

where t_v is the cumulative density function of the univariate Student t-distribution with v degrees of freedom and $t_{v,\Sigma}$ is the cumulative density function of the multivariate Student t-distribution with a correlation matrix Σ and v degrees of freedom.

2.4 Archimedean and hierarchical Archimedean copulas

Definition 2 [11] An *Archimedean generator* (simply, *generator*) is a continuous, non-increasing function $\psi : [0, \infty] \to [0,1]$ that satisfies $\psi(0) = 1, \psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$ and that is strictly decreasing on $[0, \inf\{t \mid \psi(t) = 0\}]$.

Definition 3 [11] Any d-copula C is called *Archimedean copula* (AC), if it admits the form

$$C(\mathbf{u}) := C\,(\mathbf{u};\psi) := \psi\big(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\big), \qquad \mathbf{u} \in [0,1]^d,$$

where ψ is a generator and ψ^{-1}: $[0,1] \rightarrow [0,\infty]$ is defined by $\psi^{-1}(s) = \inf\{t \mid \psi(t) = s\}, s \in [0,1]$.

To derive an explicit form of an AC, we need explicit generators. In this work we use the three popular families of ACs presented in Table 1.

Table 1 The three considered one-parametric Archimedean copula families with the corresponding parameter ranges and forms [19]

Family	θ	$\psi\,(t)$
Clayton (C)	$[-1,\infty)$ / $\{0\}$	$(1+t)^{-1/\theta}$
Frank (F)	$(-\infty,\infty)$ / $\{0\}$	$-\log(1-\left(1-\mathrm{e}^{-\theta}\right)\exp(-t))/\theta$
Gumbel (G)	$[1,\infty)$	$\exp(-t^{1/\theta})$

As it follows from the construction of ACs that all multivariate margins of the same dimensions are equal, which is mostly considered too restrictive in high-dimensional applications, there appeared a generalization of ACs, hierarchical Archimedean copulas, which allow for partial asymmetry in multivariate margins.

Definition 4 [12] A d-dimensional copula C is called a *hierarchical Archimedean copula* (HAC) if it is either an Archimedean copula, or if it is obtained from an Archimedean copula through replacing some of its arguments with other hierarchical Archimedean copulas.

It follows from Definition 4 that every AC is a HAC (but not vice versa). For more details on HACs, see, e.g., [11].

2.5 Backtesting

Backtesting is a procedure in which the ability of a given model that estimates the future loss is evaluated. In the financial industry, the most commonly used model for estimation of the risk of future losses is Value at Risk (VaR). It expresses the maximum potential loss on the dependability of a given VaR confidence level. It is defined by

$$Pr\big(\Delta\Pi_{t+\Delta t} \le -VaR_{c,\Delta t}\big) = 1 - c,$$

where Pr is the probability measure, $\Delta\Pi_{t+\Delta t}$ expresses the change in the value of a time series Π_t at time t over a time period Δt and $VaR_{c,\Delta t}$ is the value of the maximum loss for the time period Δt at a given VaR confidence level $c \in [0,1]$. In this paper, Π_t is portfolio wealth at time t.

Having different $VaR_{c,\Delta t}$ estimators, their comparison is often based on so-called *01-sequence*, see, e.g., [15, 16], given for $t \in \{1,\dots,n\}$ by

$$I_t = \begin{cases} 1, if\ \Delta\Pi_{t+\Delta t} \le -\overline{VaR}_{c,\Delta t} \\ 0, if\ \Delta\Pi_{t+\Delta t} > -\overline{VaR}_{c,\Delta t} \end{cases},$$

where $\overline{VaR}_{c,\Delta t}$ an is an estimated VaR value for time t. A widely known test based on this sequence, the Kupiec's test [17], is a two-sided test that tests the fit of the estimated VaR model with respect to the underestimation and overestimation of the risk. Given a 01-sequence (I_1, …, I_n), the test uses the *Kupiec's likelihood ratio statistic* (LR) given by

$$LR = -2\log\left[\frac{\pi_{ex}^{n_1}(1-\pi_{ex})^{n_0}}{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}\right],$$

where $\pi_{ex} = 1 - c$ is the expected probability of exception occurring, $\pi_{obs} = \frac{n_1}{n_0+n_1}$ is the proportion of exceptions, n_0 is the number of non-exception days (the number of "zeros" in the 01-sequence), n_1 is the number of exception days (the number of "ones" in the 01-sequence) and $n_0 + n_1 = n$. Under the null hypothesis that $\pi_{ex} = \pi_{obs}$, the statistic LR is asymptotically χ^2 (chi-squared) distributed with one degree of freedom. Using this fact and given n, the non-rejection interval for a number of exceptions n_1 can be obtained for a given significance level α. Throughout this paper, we use α = 5%.

3 Experiments

Firstly, we present the data that we analyze in our study. It is a part of the Dow Jones Industrial Average (DJIA) dataset. We arbitrarily selected 6 well-known US companies of the dataset, namely, MSFT (Microsoft Corporation CSCO (Cisco Systems), NKE (Nike, Inc.), MCD (McDonald's), CAT (Caterpillar Inc.) and IBM (International Business Machines Corporation). The comparison is performed for the following three periods:

- Period A - 1998/29/11-2010/31/12 – the long term period (n = 3042 days)
- Period B - 2000/21/11-2004/16/11 – the period of significant movements (n = 1000 days)
- Period C - 1994/24/07-1998/07/07 – the period of relative calm (n = 1000 days)

For each of these periods, given a VaR confidence level *c* and a family of copulas, we used the following procedure based on the Monte Carlo method for t = 501, …, *n+500* to estimate corresponding VaR values:

1. For time *t*, choose the daily returns $\Delta\Pi_{t-500}, \dots, \Delta\Pi_{t-1}$;
2. Estimate the parameters of the six NIG distributions using the method of moments, see [1];
3. Assuming an elliptical copula, estimate its parameters via the Matlab's Statistical toolbox function *copulafit*. Assuming a HAC, estimate its parameters via our implementation of Algorithm 3 proposed in [9];
4. Sample 10 000 observations distributed according to the multivariate distribution with the estimated copula and the estimated univariate marginal distributions using the approach proposed in [11];
5. For each of the observations sampled in the previous step, calculate the portfolio wealth $\Pi_t^i, i = 1, \dots, 10000$ with uneven weights (= 1/6) of the assets;
6. Order $\Pi_t^i, i = 1, \dots, 10000$;
7. Choose Π_t^j as the VaR estimated for *t* such that $j = [10\ 000(1-c)]$ ([*y*] returns the integer part of $y \in \mathbb{R}$);
8. Continue to *t+1*.

After this process, we computed three 01-sequences corresponding to each of the considered time periods, in the way described in Section 2.5.

3.1 Results

All in all, we computed 45 different 01-sequences (5 copulas families (Gaussian, Student, HAC Clayton, HAC rank and HAC Gumbel) * 3 VaR confidence levels c = 85%, 95%, 99% * 3 time periods). Their characteristics corresponding to the Kupiec's test are shown in Figures 1, 2 and 3. The figures show the differences between the expected ($[n\,(1-c)]$) and the estimated (n_1) number of exceptions, and the bounds of the corresponding non-rejection intervals (dashed lines) for the null hypothesis in the Kupiec's test corresponding to the significance level α = 5%, see Section 2.5.

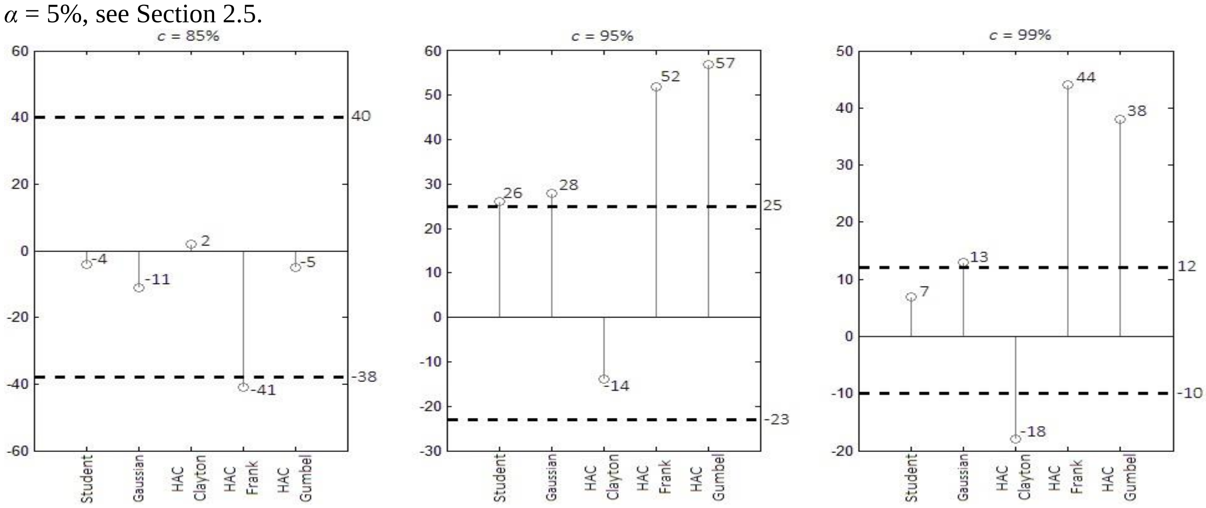

Figure 1 The differences between the expected and the estimated number of exceptions for Period A

In Figure 1, we observe that only the HAC Frank based VaR model is rejected for c = 85%, i.e., the corresponding n_1 - $[n\,(1-c)]$ is outside of the interval (-38, 40). For the remaining VaR models, the differences n_1 - $[n\,(1-c)]$ are close to 0. The HAC Clayton based VaR model, which is the closest to zero for c = 85%, is the only VaR model that is not rejected for c = 95%. The remaining VaR models overestimates the risk for this VaR confidence level. For c = 99%, only the Student based VaR model is not rejected. Also observe, that the HAC Clayton based VaR model is the only model that underestimated the risk for c = 95% and c = 99%. Generally, for this period, the EC based VaR models performed better than HAC based VaR models in the sense that the differences n_1 - $[n\,(1-c)]$ are closer to zero for the EC based VaR models than for the HAC based VaR models.

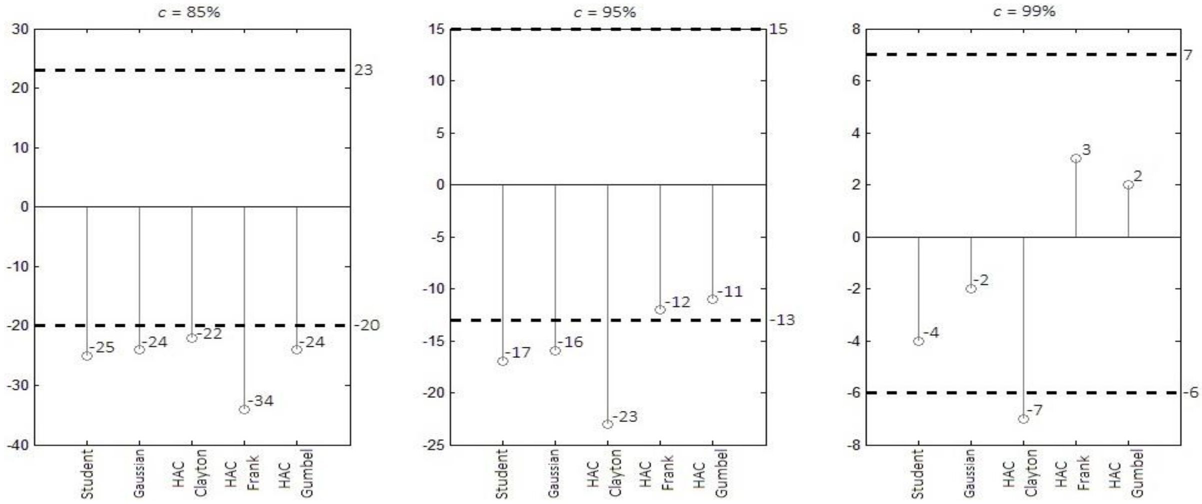

Figure 2 The differences between the expected and the estimated number of exceptions for Period B

In Figure 2, for *c* = 85%, we see that all the VaR models underestimate the risk and rejected. For *c* = 95%, only the HAC Frank and the HAC Gumbel based VaR models are not rejected, but again, all the considered VaR models underestimate the risk. For *c* = 99%, only the HAC Clayton based VaR model is rejected. We also observe that the HAC Clayton based VaR model underestimates the risk for *c* = 95% and *c* = 99%, similarly to Period A. Generally, on the contrary to Period A, the HAC based VaR models performed better than elliptical copula based VaR models, more precisely, the best (the most close to zero) or the second best (for *c* = 99%) VaR model is always a HAC based VaR model.

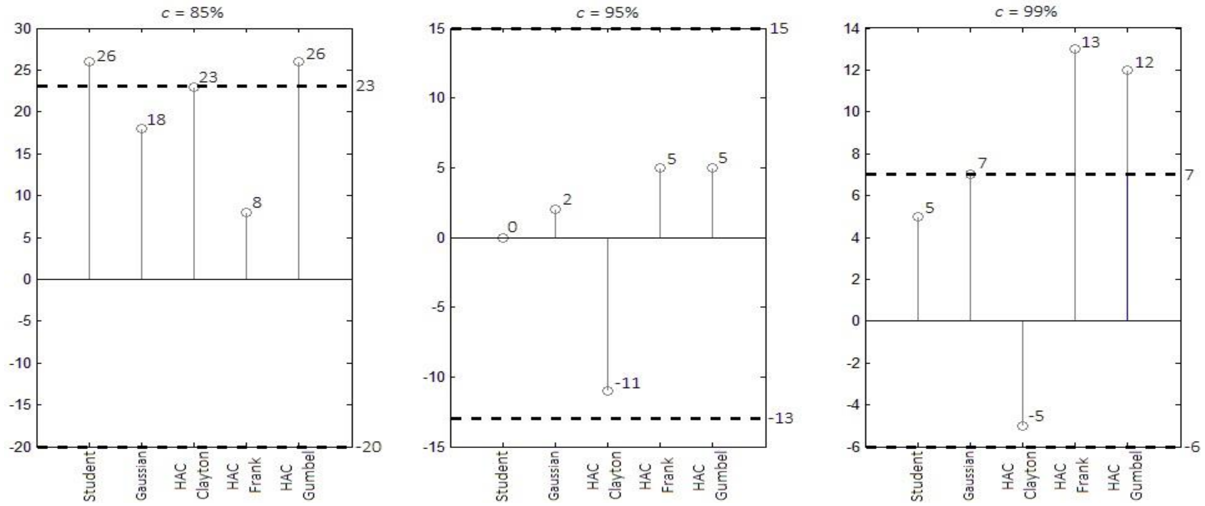

Figure 3 The differences between the expected and the estimated number of exceptions for Period C

In Figure 3, for *c* = 85%, we see that almost all VaR models overestimate the risk. We also observe that only the Gaussian and HAC Frank based VaR models are not rejected. For *c* = 95%, all the models are not rejected and the EC based VaR models score better that the HAC based VaR models. For *c* = 99%, the Gaussian, HAC Frank and HAC Gumbel based VaR models are rejected. The HAC Clayton based VaR model again underestimates the risk for *c* = 95% and *c* = 99%. Generally, for this period, the EC based VaR models performed better than the HAC based VaR models, similarly to Period A.

4 Conclusion

This paper presented an experimental comparison of VaR models based on elliptical and hierarchical Archimedean copulas. The main finding, which confirm the findings reported in [11], is that that the HAC VaR models have been performing better for the period of significant movements (Period B). For example, the HAC Gumbel based VaR model is in average the best performing for this period. This finding, on the one hand, suggests that hierarchical Archimedean copulas are more convenient for modeling VaR in times of significant movements in the stock market. On the other hand, EC based VaR models were more accurate for Period A and Period C, i.e., for the periods in which the days of relative calm dominates over the days of significant moves in the stock market. Another observation is that the HAC Clayton based VaR model more or less underestimates the risk for the two higher VaR confidence levels for all three periods. The question closely related to these conclusion, i.e., how to decide when to use elliptical based and when to use HAC based VaR models, however, remains open, and we will consider it in further research. Also, we want to consider in further research other copula estimation approaches, e.g., the consistent HAC estimator suggested in [8].

Acknowledgement

The research of this work was funded by project SGS/21/2014 – Advanced methods for knowledge discovery from data and their application in expert systems, Czech Republic. This work has been originally published in the proceedings of Mathematical Methods in Economics 2015, Cheb, see http://mme2015.zcu.cz/conference-proceedings. For a BibTex citation, see http://suzelly.opf.slu.cz/~gorecki/en/research.php.

References

[1] Barndorf-Nielsen, O. E.: *Normal inverse Gaussian distributions and the modeling of stock returns*. Aarhus, Dissertation. Aarhus University, 1995.

[2] Bae, K.-H., Karolyi, G. A., and Stulz, R. M.: A new approach to measuring financial contagion. *Review of Financial studies* **16**(3):717–763, 2003.

[3] Cherubini, G., Luciano, E. and Vecchiato, W.: *Copula Methods in Finance*. Chichester : Wiley, 2004.

[4] Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**(2), 223-236, 2001.

[5] Cont, R. and Takov, P.: *Financial modeling with Jump Processes*. Chapman&Hall, London, 2003.

[6] Górecki, J., Holeňa, M.: Structure determination and estimation of hierarchical Archimedean copulas based on Kendall correlation matrix. In: *A. Appice, M. Ceci, C. Loglisci, G. Manco, E. Masciari, Z.W. Ras (eds.) New Frontiers in Mining Complex Patterns*, Lecture Notes in Computer Science, 2014, 132–147. Springer International Publishing.

[7] Górecki, J., Holeňa. M.: An alternative approach to the structure determination of hierarchical Archimedean copulas. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics* (MME 2013), 2013, 201 – 206, Jihlava, Czech Republic

[8] Górecki, J., Hofert, M., Holeňa, M.: On the consistency of an estimator for hierarchical Archimedean copulas. In: *Proceedings of the 32nd International Conference on Mathematical Methods in Economics*, pp. 239 – 234, Olomouc, Czech Republic, 2014.

[9] Górecki, J., Hofert, M. and Holeňa, M..: *An Approach to Structure Determination and Estimation of Hierarchical Archimedean Copulas and Its Application in Bayesian Classification*. Journal of Intelligent Information Systems, pp. 1 - 39, Springer Science+Business Media New York, 2015.

[10] Hansen, B.E.: *Autoregressive conditional density estimation*. International Economic Review, 1994.

[11] Hofert, M.: *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. 1102 Suedwestdeutscher Verlag fuer Hochschulschriften, 2010.

[12] Hofert, M.: A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation* **82**(9), 1239–1255, 2012. DOI http://dx.doi.org/10.1080/00949655.2011.574632

[13] Jarque, C. M. and Bera, A. K. 1987: *A test for normality of observations and regression residuals*. International Statistical Review, 1987, **55**(2), 163-172.

[14] Kresta, A.: Backtesting of market risk estimation assuming various copula functions. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*. Opava: Slezská univerzita v Opavě, 2012, 484-489. ISBN 978-80-7248-779-0.

[15] Kresta, A. and Tichý, T.: Backtesting of portfolio risk in term sof ordinary Lévy copula model.In: *Finance and the performance of firms in science,education and practice.* E. Jiříčková et al (Eds.). Zlín: Univerzita Tomáše Bati ve Zlíně, 2011.

[16] Kresta, A. and Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. Finance a úvěr – *Czech Journal of Economics and Finance* **62**(2), 141–161, 2012.

[17] Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivative* **3**(2), 73-84, 1995.

[18] Longin, F. and Solnik, B.: Extreme correlation of international equity markets. *The Journal of Finance* **56**(2), 649–676, 2001.

[19] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. New York: Springer, 2006.

[20] Rank, J.: Copulas: *From Theory to Application in Finance*. London: Risk Books, 2006.

[21] Sklar, A: *Fonctions de répartition à n dimensions et leurs marges*. Publications de l'Institute de statistique de l'Université Paris, 229-231, 1959.