Central European Review of Economic Issues
EKONOMICKÁ REVUE

Backtesting AVaR and VaR with a simulated copula

Matteo MALAVASI[a*], Roberto PREVITALI[a], Sergio ORTOBELLI LOZZA[a], Tomas TICHY[b]

[a] *Department of Management, Economics and Quantitative Methods, University of Bergamo, via dei Caniana 2, 24127 Bergamo, Italy.*
[b] *Department of Finance, Faculty of Economics, VŠB – Technical University of Ostrava, Sokolská tř. 33, 702 00 Ostrava, Czech Republic.*

Abstract

The aim of this study is to verify whether the average value at risk (AVaR) can be a good alternative to the value at risk (VaR) for estimating portfolio losses, especially regarding tail events. To achieve this aim, we use a copula framework to estimate the dependence between the stock returns of a portfolio composed of 94 components of the S&P100 index to compute the AVaR and VaR and compare the results with respect to the Gaussian exponentially weighted moving average (EWMA). To compute the simulated returns, we employ the algorithm used by Biglova et al. (2014) in portfolio selection problems and then backtest the model with Kupiec's and Christoffersen's tests. The results are coherent with the literature; in particular, the VaR computed both via the copula and via the EWMA seems to fail to provide an accurate risk measurement while the AVaR with the copula and EWMA appears to be more reliable.

Keywords

Average value at risk, backtesting, copula, EWMA model, value at risk.

JEL Classification: G17, G15, G12

[*] matteomalavasi2@gmail.com (corresponding author)
The research was supported by an SGS project of VSB-TU Ostrava under No. SP2016/11 and the Czech Science Foundation through project No. 13-13142S. The support is greatly acknowledged.

© 2016 Published by VŠB-TU Ostrava. All rights reserved.
ISSN 1212-3951 (Print), 1805-9481 (Online)
ER-CEREI, Volume 19: 15–24 (2016).
doi: 10.7327/cerei.2016.03.02

Backtesting AVaR and VaR with a simulated copula

Matteo MALAVASI, Roberto PREVITALI, Sergio ORTOBELLI LOZZA, Tomas TICHY

1. Introduction

In this paper we backtest the portfolio average value at risk (AVaR)[1] and value at risk (VaR) to verify whether the AVaR can be a good alternative to the value at risk for estimating portfolio losses, especially regarding tail events. We consider 10 years of daily observation of stocks belonging to the Standard & Poor's 100 from 1 December 2005 to 18 March 2016. The sample used is reduced to those that are present throughout the entire 10-year period considered; thus, the number of remaining stocks is 94 (with respect to the original 100).

Many different approaches have arisen recently in the literature regarding AVaR and VaR backtesting. Berkovitz (2001) proposes an approach based on a censored likelihood ratio test to backtest the tail of the empirical distribution with the predicted one. Kerkhof and Melenberg (2004) extend the censored likelihood ratio procedure to a more general class of risk measure. Both the approaches rely on the asymptotic distribution of the test statistic, which might not be accurate for small samples.

Since AVaR backtesting usually relies on small samples, Wong (2008) suggests a saddle point technique to approximate the required *p*-value under the null hypothesis. A model-independent and non-parametric backtest is proposed by Acerbi and Szekely (2014). They also solve the issue of elicitability by arguing that elicitability is relevant only to model selection and not to model testing.[2] In view of the fact that the AVaR and VaR are spectral measures, Costanzino and Curran (2015) develop a test statistic valid for all spectral measures that makes backtesting for AVaR and VaR comparable.

A different backtesting procedure can be found in the study by Marinelli et al. (2007). They backtest the VaR and the AVaR in a univariate framework considering Pareto stable distribution versus Gaussian and extreme value theory. They follow Embrechts et al. (2005), who define two measures to test the AVaR. The first is the negative of the VaR, and the second is the conditional average of the difference between the AVaR and the realized ex post return, conditioned on the set in which the ex post return is lower than the negative of the estimated VaR.

We propose to overcome the small-sample issues by relying on a simulation approach. In particular, we use a methodology from portfolio optimization that allows us to generate *realistic* simulated scenarios and return series (see Rachev et al., 2007; Biglova et al., 2010, 2014). In portfolio optimization we have to combine some well-known stylized facts: typical features of return series (such as asymmetric fat tails and volatility clusters) with asymmetry in the dependence structure (lower-tail dependence stronger than upper-tail dependence) and high dimensionality (see for example Cont, 2001; Papp et al., 2005; Hong et al., 2007; Kondor et al., 2007). According to Cont and Tankov (2003) and Schoutens (2003), excess kurtosis and skewness in return distributions are dealt with using other distributions than those of the Gaussian law. Cherubini et al. (2004) assert that asymmetry in the dependence structure can be handled with an appropriate copula. Moreover, high dimensionality might affect the unbiasedness of the estimators that are generally used in portfolio selection problems (see Rachev et al., 2005; Sun et al., 2008). In particular, the estimation of the impact of a tail event requires a high number of observations (see Papp et al., 2005; Kondor et al., 2007).

For these reasons we model the risk of a portfolio with proper simulated scenarios. The algorithm used for generating returns is the one provided by Rachev et al. (2007) and Biglova et al. (2010, 2014), which simulates 5000 scenarios (other studies that implement a similar algorithm are for example Patton (2006), Ortobelli et al. (2010) and González-Pedraz et al. (2015)). To justify the use of this algorithm, we propose an ex post empirical analysis based on the maximization of the Sharpe ratio both on simulated and on historical data. The empirical analysis confirms that the optimal portfolio based on simulated data performs better than the former one. Since the algorithm seems

[1] Often called the conditional value at risk (CVaR).

[2] For elicitability and its relation to the VaR, AVaR and backtesting, see among others Gnetting (2011) and Ziegel (2014).

to capture return features reasonably, we claim that it can also be used for estimating portfolio losses. The generated returns are then employed in estimating the daily value at risk (VaR) and average value at risk (AVaR) at the 95%, 96%, 97%, 98% and 99% levels of confidence for every portfolio considered. We propose an empirical graphical comparison between the different VaRs (and AVaRs) estimated and the historical logarithmic return series of a randomly chosen portfolio and compute the daily means of the VaR and AVaR. An analogous procedure is applied using a Gaussian EWMA model (see RiskMetrics, Longerstaey and Zangari, 1996) as a benchmark.[3]

Then we backtest the VaR and AVaR interval forecasts using Kupiec's and Christoffersen's tests to investigate which risk measure better covers the risk of tail events and whether the copula model can be a good alternative to the EWMA. By comparing the percentages of acceptance of the two tests for the AVaR and VaR computed using the copula and the EWMA framework, we observe that the VaR does not seem to be adequate for measuring risk in both the copula and the EMWA model, while the AVaR tends to overestimate risk in the former and underestimate it in the latter.

The entire process is computed using the MatLab® software, and some of the algorithms used, especially the ones necessary for generating skewed *t* copulas, are provided by the mathworks.com site.

In Section 2 we describe the algorithm that we use, basic definitions of VaR and AVaR, the EMWA models, Sharpe ratio maximization and the backtesting methods. In Section 3 we present some graphical results and report the outcome of Kupiec's and Christoffersen's tests.

2. The framework

Considering the profit/loss $W_{t+\tau} - W_t$, a random variable defined on a probability space $(\Omega, \Im, \mathbb{P})$, obtained from t to $t + \tau$ and a level of confidence θ, the VaR and AVaR are defined (see for example Pflug, 2000; Rockafellar and Uryasev, 2000; Lamantia et al., 2006) as:

$$VaR_{\theta,[t,t+\tau]}(W_{t+\tau} - W_t) = \\ = -inf\{q|\mathbb{P}(W_{t+\tau} - W_t \le q) \\ > 1 - \theta\}$$

and

$$AVaR_{\theta,[t,t+\tau]} = \frac{1}{1-\theta}\int_0^{1-\theta} VaR_{q,[t,t+\tau]}(W_{t+\tau} - W_t)dq.$$

If $W_{t+\tau} - W_t$ have continuous distribution, then (see Pflug, 2000; Rockafellar and Uryasev, 2000):

$$AVaR_{\theta,[t,t+\tau]}(W_{t+\tau} - W_t) \\ = -\mathbb{E}\left[W_{t+\tau} - W_t | W_{t+\tau} - W_t \le VaR_{\theta,t+\tau}\right].$$

Let w be the vector of n portfolio asset weights $w = [w_1, \ldots, w_n]'$ and $r_{i,t+1} = \log(P_{i,t+1}/P_{i,t})$ the log return of each i-th asset between times t and $t + 1$; then the portfolio return at time $t + 1$ is computed following Lamantia et al. (2006) as

$$z_{(p),t+1} = \sum_{i=1}^{n} w_i r_{i,t+1}.$$

To fit the VaR and AVaR of a large number of portfolios, we vary the weights using the following scheme. The weight of the i-th stock is $0.0i$, while all the other stocks are weighted $(1 - i\ 0.01)/93$, for $i = 1, \ldots, 100$.

The procedure is repeated for the 2nd to the 94th stock to obtain all the $94 \cdot 100 = 9400$ portfolios in a 94×9400 matrix.

We suggest comparing the portfolio VaR and AVaR either by considering simulated future scenarios in a copula framework or by assuming the returns following the classic EWMA model. In the latter case, following Longerstaey and Zangari (1996), we assume that the vector of returns $r_{t+1} = [r_{1,t+1}, \ldots, r_{n,t+1}]'$ follows a conditional joint Gaussian distribution with a zero mean and dispersion matrix $Q_{t+1|t} = \left[\sigma^2_{ij,t+1|t}\right]$ of which the elements $\sigma^2_{ij,t+1|t}$ are computed through the exponentially weighted moving average (EWMA) model (Lamantia et al., 2006):

$$\sigma^2_{ii,t+1|t} = E_t\left(r^2_{i,t+1}\right) = \lambda\sigma^2_{ii,t|t-1} + (1-\lambda)r^2_{i,t}$$
$$\sigma^2_{ij,t+1|t} = E_t\left(r_{i,t+1}r_{j,t+1}\right) \\ = \lambda\sigma^2_{ij,t|t-1} + (1-\lambda)r_{i,t}r_{j,t},$$

where λ is the optimal decay factor.

As an alternative to the EWMA, we test whether the multivariate scenario generation approach proposed by Rachev et al. (2007) or Biglova et al. (2010, 2014) is able to identify tail events.

2.1 An algorithm for simulating copulas

In this section we recall briefly the algorithm that we use for generating realistic scenarios (for further details see Rachev et al., 2007; Biglova et al., 2010, 2014).

The general idea is to assume that each log return $r_{j,t}$ follows an ARMA(1,1)–GARCH(1,1):

$$\begin{aligned} \boldsymbol{r_{j,t}} &= \boldsymbol{a_{j,0}} + \boldsymbol{a_{j,1}}\boldsymbol{r_{j,t-1}} + \boldsymbol{b_{j,1}}\boldsymbol{\varepsilon_{j,t-1}} + \boldsymbol{\varepsilon_{j,t}} \\ \varepsilon_{j,t} &= \sigma_{j,t}u_{j,t} \\ \sigma^2_{j,t} &= c_{j,0} + c_{j,1}\sigma^2_{j,t-1} + f_{j,1}\varepsilon^2_{j,t-1} \end{aligned} \quad . \qquad (1)$$

[3] EWMA models are probably one of the most-used tools in risk evaluation by financial institutions.

Then we generate the future scenarios according to the following algorithm:

Step 1: Margins of the innovation of each log-return series

Perform the maximum likelihood estimation of the model in (1). Then approximate the standardized innovations $\hat{u}_{j,t} = \frac{\hat{\varepsilon}_{j,t}}{\sigma_{j,t}}$ with an α_j-stable distribution $S_{\alpha_j}(\mu_j, \sigma_j, \beta_j)$. Since we have no closed form formula for stable distribution function, we simulate *M* scenarios from a stable distribution for each of the future scenarios and compute the sample distribution for each $\hat{u}_{j,T+1}$, given by:

$$F_{\hat{u}_{j,T+1}}(x) = \frac{1}{M}\sum_{i=1}^{M} I_{\{\hat{u}_{j,T+1}^{(i)} \le x\}}.$$

Step 2: Dependence structure

Perform maximum likelihood estimation to fit the random vector $\hat{u} = [\hat{u}_1, \dots, \hat{u}_{94}]$ with an asymmetric *t*-distribution with ν degrees of freedom defined as:

$$\hat{\boldsymbol{u}} = \boldsymbol{\mu} + \boldsymbol{\gamma} \boldsymbol{g}(\boldsymbol{W}) + \sqrt{\boldsymbol{W}}\boldsymbol{Z}, \qquad (2)$$

with g:[0, ∞)→:[0, ∞)[4], γ vector of skewness parameters, $Z \sim N(0,\Sigma)$ with $\Sigma = [\sigma_{ij}]$ and $W \sim \mathrm{IG}(\nu/2, \nu/2)$. The density function of $\hat{u}$ is

$$f(x) = c\frac{K_{\frac{d+n}{2}}\left(\sqrt{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma}\right)\exp((x-\mu)'\Sigma^{-1}\gamma)}{\left(\sqrt{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma}\right)^{-\frac{\nu+n}{2}}\left(1+\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\nu}\right)^{\frac{\nu+n}{n}}}$$

with the normalizing parameter:

$$c = \frac{2^{\frac{2-(\nu+n)}{2}}}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}},$$

where K_λ is a Bessel function of the third kind, which is a function that solves the differential equation (see Abramowitz and Stegun, 1965; Demarta and McNeil, 2005):

$$z^2\frac{d^2y}{dz^2} + z\frac{dy}{dz} + (z^2 - \nu^2)y = 0.$$

We impose $\nu = 5$, since the mean of the degrees of freedom of the marginals of the fitted asymmetric *t*-distribution can be approximated by 5.

Then with $(\nu, \hat{\mu}_i, \hat{\sigma}_{ii}, \hat{\gamma}_i)$ we can compute the variance covariance matrix of $\hat{u}$:

$$\hat{\Sigma} = \left(cov\,(X) - \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\hat{\gamma}\hat{\gamma}'\right)\frac{\nu-2}{\nu}.$$

Thus, we generate *N* scenarios for the vector $\hat{u}$ from (2) and call them $[\hat{V}_1^i, \dots, \hat{V}_{94}^i]$ for $i = 1, \dots, N$ and let $F_{V_j}(x)$ be the marginal of the estimated asymmetric *t*-distribution. Moreover, considering $U_j^i = F_{V_j}(\hat{V}_j^i)$, we can generate *N* scenarios (U_1^i, U_{94}^i), $(i = 1, \dots, N)$ from a uniform random vector (U_1, ..., U_{94}) (with support on the 94-dimensional unit cube), the joint distribution of which is given by the copula $C(t_1, \dots, t_{94}) = F_V\left(F_{V_1}^{-1}(t_1), \dots, F_{V_{94}}^{-1}(t_{94})\right)$ (see Sklar, 1959). For further details see Biglova et al. (2014). Consequently, we can take into account the dependence structure of the standardized innovation vector at time $T + 1$ with its N scenarios $u_{T+1}^i = \left[u_{T+1}^{1,i}, \dots, u_{T+1}^{94,i}\right], i = 1, \dots, N$, obtained considering $u_{T+1}^{j,i} = \left(F_{\hat{u}_{j,T+1}}\right)^{-1}(U_j^i)$ by the probability integral transform (see Casella and Berger, 2002).

Step 3: Log-return generation

By (1) generate the vector of the model's residuals as: $\varepsilon_{T+1}^i = \left[\varepsilon_{T+1}^{1,i}, \dots, \varepsilon_{T+1}^{94,i}\right] = [\sigma_{1,T+1}u_{T+1}^{1,i}, \dots, \sigma_{94,T+1}u_{T+1}^{94,i}]$. Then finally by (1) generate the vector of future returns: $r_{T+1}^i = \left[r_{T+1}^{1,i}, \dots, r_{T+1}^{94,i}\right]$.

In this work we simulate 5000 scenarios to evaluate the VaR and AVaR of 9400 portfolios. The product of the simulated returns (in a 5000 × 94 matrix) and the portfolios determines 5000 scenarios of the 9400 portfolios. Then the outputs are sorted into ascending order for each portfolio from the worst scenario to the best one. The VaR at α = 1%, 2%, 3%, 4% and 5% is computed by selecting the fiftieth, one hundredth, one hundred and fiftieth, two hundredth, and two hundred and fiftieth worst values for each portfolio. The AVaR at α = 1%, 2%, 3%, 4% and 5% is computed by making the mean of the worst 50, 100, 150, 200 and 250 values for each portfolio.

2.2 On the motivation for using simulated returns

In this section we propose an empirical analysis to show the advantages of this algorithm. In particular, we consider a portfolio selection strategy based on the

[4] In this work $g(W) = W$.

maximization of the Sharpe ratio, $SR(x'r) = \frac{E(x'r-r_b)}{STD(x'r-r_b)}$; that is (Sharpe, 1994):

$$\max_x SR(x'r)$$

$$s.t. \ \ \forall i \ x_i \geq 0; \sum_{i=1}^{n} x_i = 1,$$

where x is the vector of the portfolio weights, r is the vector of the returns of the portfolio's assets and r_b is a null risk-free return or a benchmark's return. Here we perform the Sharpe ratio optimization both on historical and on simulated data to verify the adaptability of the algorithm in Biglova et al. (2014) to our data. As we can see from Figure 1, the strategy with historical data performed badly by losing 18% of its initial value.

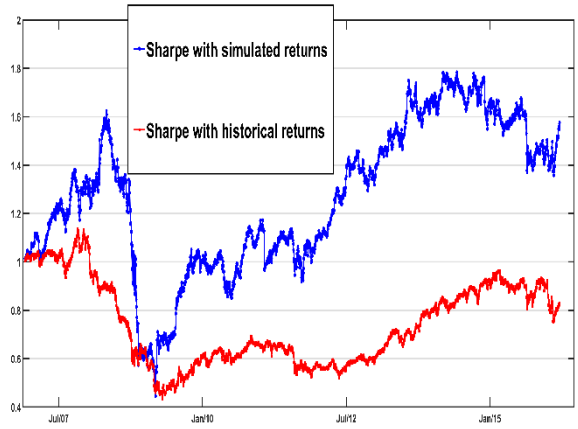

Figure 1 Comparison between Sharpe optimization strategies based on historical and simulated data

Moreover, it loses the wealth for most of the intervals considered: once the crisis started, the strategy based on historical data always registered negative performance with respect to the initial investment, despite the improvement in the last three years. What arises from these graphs is that there is a considerable difficulty in providing a strategy that is able to increase wealth during crisis booms, since systemic risk cannot be avoided. Furthermore, the ex post wealth obtained from the strategy of simulated returns appears to perform much better than the strategy based on historical data. This is the reason why we propose such algorithm for generating future scenarios. A similar algorithm is used by Patton (2006) and González-Pedraz et al. (2015). González-Pedraz et al. (2015) impose a conditional asymmetric student-*t* copula with skewed and fat-tailed marginals to investigate portfolio selection with commodities for investors with time-varying three-moment preferences.

They achieve better investment ratios and a better relative performance measure with different specifications than the classical Gaussian models. Patton (2006) applies an analogous algorithm in the context of exchange rates, finding that the correlation between exchange rates tends to vary across periods of appreciation and depreciation. Since the algorithm seems to capture return features reasonably, we claim that it can also be used for estimating losses.

2.3 Kupiec's and Christoffersen's tests

Kupiec's proportion of failure (POF) test is an unconditional coverage test that verifies that the number of exceedances x is equal to the number of out-of-confidence levels forecasted by the VaR model; in other words, it checks whether the actual probability of exceeding the limit imposed by the VaR model is equal to the probability forecasted by the VaR model's confidence level. Its null and alternative hypotheses are expressed as $H_0 : p = \hat{p} = \frac{x}{T}, H_1 : p \neq \frac{x}{T}$, with T being the number of observations, while its likelihood ratio is given by Kupiec (1995):

$$LR_{POF} = -2 \ln \left(\frac{(1-p)^{T-x} p^x}{\left(1-\frac{x}{T}\right)^{T-x} \left(\frac{x}{T}\right)^x} \right),$$

which follows a χ^2 distribution with one degree of freedom. Its critical value at the 5% significance level is 3.84, so, if LR_{POF} > 3.84, the null hypothesis is rejected. Depending on the probability levels chosen, there is a specific critical value for each of them (Kupiec, 1995).

Instead of computing point forecasts, Christoffersen (1998) suggests the study of interval forecasts that consider a range of outcomes. First, a testing criterion and an efficient sequence of interval forecasts are defined (see Lamantia et al., 2006):

$$\left\{VaR_{\theta,t+1|t}\left(z_{(p),t+1}\right), \infty\right\}_{t=1}^{T},$$

where the variable I_{t+1} indicates losses greater than the estimated ones (see Lamantia et al., 2006):

$$I_{t+1}\left(z_{(p),t+1}\right) = \\ = \begin{cases} 1, & if \quad z_{(p),t+1} \geq VaR_{\theta,t+1|t}\left(z_{(p),t+1}\right) \\ 0, & if \quad z_{(p),t+1} < VaR_{\theta,t+1|t}\left(z_{(p),t+1}\right) \end{cases}.$$

Then the forecasted interval for the VaR is efficient if given an information set Ψ_t, $\mathbb{E}[I_t|\Psi_{t-1}] = \theta$.[5]

Lamantia et al. (2006) show that testing the sequence of forecasted intervals for $\left\{VaR_{\theta,t+1|t}\left(z_{(p),t+1}\right), \infty\right\}$ has a correct conditional coverage that is equivalent to testing

[5] For example, consider the following information set: $\Psi_t = \{I_t, I_{t-1}, \dots, I_1\}$.

$\{I_{t+1}\} \sim^{IID} Bernoulli(\theta)$. The correct coverage can be tested conditionally or unconditionally.

The unconditional version is a likelihood ratio test in which we test H_0: $\mathbb{E}[I_{t+1}] = \theta$ against the alternative H_1: $\mathbb{E}[I_{t+1}] \neq \theta$ under the hypothesis that $I_{t+1} \sim^{IID} Bernoulli(\theta)$. The likelihood function under the null is $L(\theta|I_n, \dots, I_1) = (1-\theta)^{n_0}\theta^{n_1}$, while under the alternative it is $L(\hat{\theta}|I_n, \dots, I_1) = (1-\hat{\theta})^{n_0}\hat{\theta}^{n_1}$, where $\hat{\theta} = arg\max_q L(q|I_n, \dots, I_1)$, n_0 is the number of times that we have $z_{(p),t+1} \geq VaR_{\theta,t+1|t}(z_{(p),t+1})$ and n_1 is the number of times that we have $z_{(p),t+1} < VaR_{\theta,t+1|t}(z_{(p),t+1})$. The test statistic LR_{uc} is asymptotically distributed as a χ_1^2. The unconditional version, in other words, counts how many times the portfolio falls beyond the lower threshold given by the VaR over the entire time period without grants for variance dynamics (see Lamantia et al., 2006).

The conditional version tests whether the sequence $\{I_{t+1}\}$ presents correct unconditional coverage and serial independence. Christoffersen (1998) shows that the test statistic is given by $LR_{cc} = LR_{uc} + LR_{ind}$, where LR_{ind} is the likelihood ratio for testing serial independence against a first-order Markov alternative given by the transition matrix:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix},$$

where $\pi_{ij} = \mathbb{P}[I_{t+1} = j | I_t = i]$. Christoffersen (1998) also shows that LR_{cc} is approximately distributed as a χ_2^2.

Christoffersen's test for the AVaR is a little more delicate, since it might not be clear whether the conditions in the null hypothesis of the unconditional test also hold for the AVaR.

To have a comparable test for AVaR and VaR backtesting, we propose the following extension of Kupiec's and Christoffersen's tests. Consider a continuous random variable X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the sigma algebra $\mathcal{G}$ formed by the sets $\mathcal{G} = \{A, A^C, \Omega, \emptyset\}$, where $A = \left[x \leq {F_X}^{-1}(\theta)\right]$, where ${F_X}^{-1}$ is the quantile function of the random variable X. Then, given the random variable $Y = \mathbb{E}[X|\mathcal{G}]$, we can write:

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|A]\chi_A + \mathbb{E}[X|A^C]\chi_{A^C},$$

where χ_A is the characteristic function of set A. In this particular case:

$$Y = \begin{cases} \mathbb{E}[X|A] & with\ probability\ \theta \\ \mathbb{E}[X|A^C] & with\ probability\ 1-\theta \end{cases}.$$

Then let $\phi(\theta) = \mathbb{E}[X|X \leq {F_X}^{-1}(\theta)]$. Under the assumption of continuity of X, the function $\phi(\theta)$ is continuous and monotone in θ. Since, for a continuous random variable, $AVaR_\theta(X) = -\mathbb{E}\left[X|x \leq {F_X}^{-1}(\theta)\right]$, this appears to be a natural method to extend Kupiec's and Christoffersen's tests for the AVaR backtest. Therefore, when we evaluate the forecasted interval for $\left\{\mathbb{E}[z_{(p),t+1|t}|z_{(p),t+1|t} \leq {F_z}^{-1}(\theta)], \infty\right\}_{t=1}^{T}$, it is the same as testing:

$$\mathbb{E}[I_{t+1}(Y)] = 1-\theta,$$

where

$$I_{t+1}(Y) = \begin{cases} 1, & if\ \ \mathbb{E}\left[z_{(p),t+1|t}|z_{(p),t+1|t} \leq {F_z}^{-1}(\theta)\right] \geq \mathbb{E}\left[z_{(p),t}|z_{(p),t} \leq {F_z}^{-1}(\theta)\right] \\ 0, & if\ \ \mathbb{E}\left[z_{(p),t+1|t}|z_{(p),t+1|t} \leq {F_z}^{-1}(\theta)\right] < \mathbb{E}\left[z_{(p),t}|z_{(p),t} \leq {F_z}^{-1}(\theta)\right] \end{cases}.$$

Therefore, testing $\mathbb{E}\left[z_{(p),t}|z_{(p),t} \leq {F_z}^{-1}(\theta)\right]$ is equivalent to testing $-AVaR_\theta\left(z_{(p),t}\right)$. Furthermore, we can proceed to backtest the AVaR by simply counting the number of times in which the mean of the ex post AVaR falls outside the forecasted interval given by the predicted AVaR. Even though we could test the AVaR with other methodologies, we backtest the models using the aforementioned tests because the results are immediately comparable with Christoffersen's and Kupiec's VaR backtesting procedures. We can say that a risk measure measures the risk better if the percentage of acceptance of Christoffersen's and Kupiec's tests is higher.

3. Results

We compare the performances of the VaR and AVaR in the copula-based framework described in the previous section with the EMWA model. Figure 2 shows a graphical comparison between randomly[6] chosen ex post portfolios and the VaR and AVaR. As we can see, there is a significant number of times for which all the portfolio returns fall under the limits provided by the VaRs. Differently, the AVaR violations are considerably fewer than the VaR ones. Just by looking at these graphical representations, we presume that the computed AVaR can be a good indicator of risk, while it is apparent that the VaR is not able to secure the portfolio accurately from potential losses since it underestimates the risk. This underestimation is proved by the fact that the number of exceedances over VaR is much larger than expected. With only 5000 scenarios computed, the use of the VaR may be misleading, while the AVaR shows a better

[6] The randomly chosen portfolio is composed of Mondelez International (MDLZ) for 4% and the other stocks weighted equally at (1-4%) / 93 = 0.01032258.

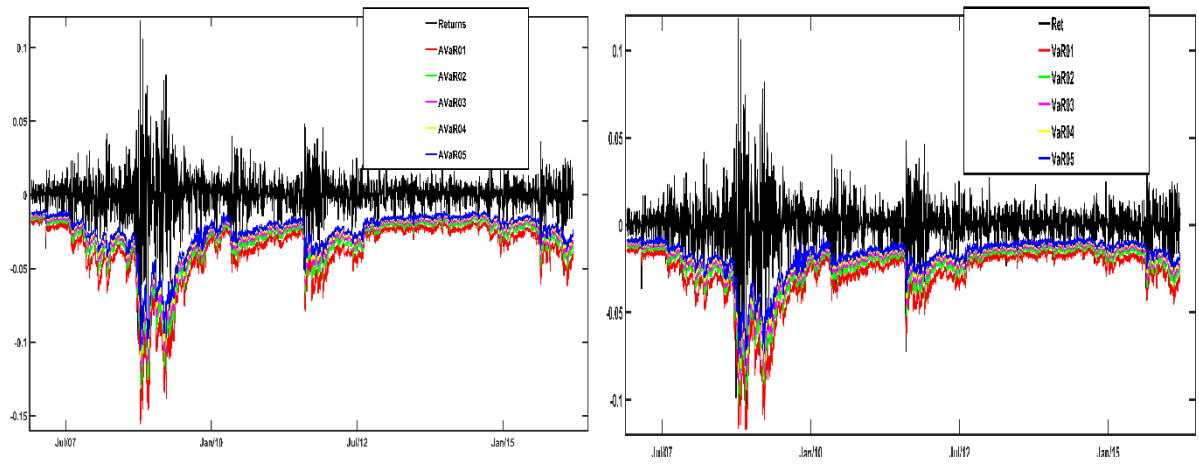

Figure 2 Graphical comparison between the AVaR and VaR and the ex post portfolio

approximation because it is a mean value of exceedances.

3.1 Kupiec's and Christoffersen's tests

To check whether the VaRs and AVaRs computed with the copula model are accurate, we perform Kupiec's and Christoffersen's tests for all the 9400 portfolios; in the case that the *i*-th portfolio's VaR or AVaR is accepted by the test, then the output of the test gives 1 and otherwise it gives 0.

$$test_i = \begin{cases} 1 & VaR_{i,\theta} \text{ (or } AVaR_{i,\theta}\text{) is accepted} \\ 0 & VaR_{i,\theta} \text{ (or } AVaR_{i,\theta}\text{) is rejected} \end{cases}.$$

Thanks to this, we can easily compute the percentage of times that the risk measures are accepted by determining the mean of the results for both Kupiec's and Christoffersen's tests. In the table below, we present the percentage of tests that accept the copula model's VaR and AVaR.

Since the aim of the model is to estimate the risk of extreme values along the tails, we consider the mean values for the first and the second percentile (α = 1%, 2%) as well as computing the mean for the fifth percentile to describe the behaviour of the model for lower levels of confidence. What we can see from Table 1 is that most of the VaRs are not acceptably accurate for measuring risk. The percentage of successfully tested AVaRs at 1% and 2% is higher, but only the AVaR at 1% seems to provide an acceptable measure of risk given its 83.89% success rate in Kupiec's test and 92.72% in Christoffersen's test. For the AVaR at 5%, both tests are always rejected.

Table 1 Percentages of VaRs and AVaRs accepted by Kupiec's and Christoffersen's tests

Copula model	VaR			AVaR		
α	1%	2%	5%	1%	2%	5%
Kupiec	0.023	0.001	0.001	0.839	0.407	0
Christoffersen	0.058	0.005	0.007	0.927	0.578	0

We conduct the same calculation and graphical inspection of the EMWA model and obtain the same result: the EWMA AVaR seems to be better than the EWMA VaR, which is not able to capture potential losses precisely since it underestimates the risk. In this case the exceedances over the EWMA VaR are also much greater than expected. Table 2 reports the results of Kupiec's and Christoffersen's tests for the EMWA VaR and AvaR. Even in this case, most of the VaRs provided by the model are not acceptably accurate for measuring risk. The percentage of successfully tested AVaRs at 1% and 2% is considerably higher. In particular, the AVaR at 2% provides an acceptable measure of risk given its 98.83% success in Kupiec's test and 99.31% in Christoffersen's test. Looking at Table 1 and Table 2, we can see that the VaR at 1% and 2% perform badly in both backtesting methods, while the AVaR at 2% performs better. As suggested by an anonymous reviewer, a possible explanation for this fact could be that, since the AVaR is computed as the average of the worst portfolio values, the VaR at 1% overestimates risk and the VaR at 2% tends to underestimate risk, then the AVaR results are correct.

Table 2 Percentages of EWMA VaRs and AVaRs accepted by Kupiec's and Christoffersen's tests

EMWA model	VaR			AVaR		
α	1%	2%	5%	1%	2%	5%
Kupiec	0.004	0.034	0.219	0.675	0.988	0.043
Christoffersen	0.014	0.065	0.288	0.753	0.993	0.169

Similar results for the AVaR can be obtained with different backtesting methods (see among others Bergovitz, 2001; Kerkhof and Melenberg, 2004; Marinelli et al., 2007; Wong, 2008; Acerbi and Szekely, 2014; Costanzino and Curran, 2015). Then, to compare the performance of the copula model and the EMWA model, we examine the mean VaR and mean AVaR (Figure 3).

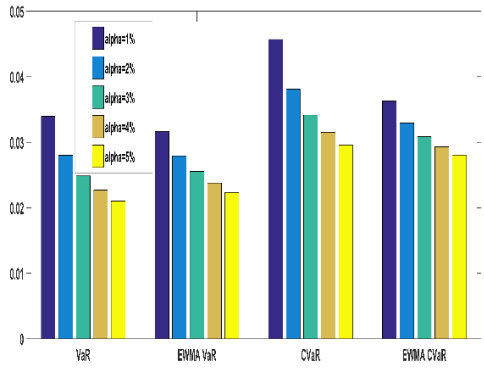

Figure 3 Mean VaRs and AVaRs for the copula and EWMA models

Comparing the mean VaRs and AVaRs of the two models, it appears that there is no great difference between the VaRs and the EWMA VaRs. Only the EWMA first-percentile VaR is higher than the copula one, while the AVaRs are higher than the EWMA AVaRs for all the percentiles. Then we perform Kupiec's and Christoffersen's tests. The results are presented in Table 3.

Table 3 Comparison of the copula and EWMA test results

	Kupiec		Christoffersen	
	Copula	EMWA	Copula	EWMA
VaR $\theta = 1\%$	0.023	0.004	0.058	0.014
VaR $\theta = 2\%$	0.001	0.034	0.005	0.065
VaR $\theta = 5\%$	0.001	0.219	0.007	0.288
AVaR $\theta = 1\%$	0.839	0.675	0.927	0.753
AVaR $\theta = 2\%$	0.407	0.988	0.579	0.993
AVaR $\theta = 5\%$	0	0.043	0	0.169

The copula model VaR at the first percentile provides a better estimation of risk than the corresponding EWMA one. On the other hand, at higher percentiles, the EWMA model performs considerably better than the copula one. Analogous results can be derived from the AVaR comparisons: at the first percentile, the copula AVaR is a more accurate risk measure than the EWMA AVaR, but at the second and fifth percentiles, its accuracy is remarkably lower than that of the EWMA. Given the results of the tests, it is possible to derive the following suppositions: since all the copula AVaRs are greater than the EWMA AVaRs and their accuracy is better for the first percentile but worse for the second percentile, we can presume that the EWMA AVaR at α = 1% underestimates risk while the AVaR at α = 2% and more overestimates risk.

4. Conclusion

An accurate measure of risk in risk management should be able to capture extreme values of risk along the distribution tail belonging to the first percentile. We focus on market risk estimation by applying a skewed *t*-copula model to portfolios composed of 94 components of the S&P100 to compute the VaR and AVaR, and then we compare their results with the outcomes of a classic Gaussian EWMA model. To compute the return simulations, we use a copula algorithm provided by Rachev et al. (2007), Sun et al. (2008) and Biglova et al. (2014). As shown in Section 2.2, the use of returns generated by this algorithm in Sharpe ratio optimization significantly increases the portfolio's wealth with respect to the historical one.

Comparing the average VaRs and AVaRs of the two models, it appears that there is no large difference between the performance of copula VaRs and EWMA VaRs, except for the first percentile, for which we can see that the copula VaR clearly underestimates the risk less than the EWMA VaR. We also observe that the copula AVaRs are higher than the EWMA AVaRs. By conducting Kupiec's and Christoffersen's tests, we discover that, for both the copula and the EWMA model, the VaRs are not acceptably accurate for measuring risk while the copula AVaR at α = 1 and the EWMA AVaR at α = 2% are acceptable measures of risk. Moreover, it is possible to presume that the EWMA AVaR at α = 1% underestimates risk while the copula AVaR at α = 2% and more overestimates risk. In other words, the VaR systematically fails to measure the risk of tail events. The AVaR with $\alpha = 1\%$ predicts risk better in the copula framework, while higher values of α work better with EWMA.

The results of this study indicate that the VaR is not a good risk measure while the AVaR provides a better estimate of risk, especially for the first two percentiles.

References

ABRAMOWITZ, M., STEGUN, I.A. (1965). Chapter 9. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover Publications.

ACERBI C., SZEKELY B. (2014). Backtesting expected shortfall. *Risk Magazine* 47: 371–384.

BERKOWITZ, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19(4): 465–474. http://dx.doi.org/10.1198/07350010152596718

BIGLOVA, A., ORTOBELLI, S., FABOZZI, F.J. (2014). Portfolio selection in the presence of systemic risk. *Journal of Asset Management* 15(5): 285–299. http://dx.doi.org/10.1057/jam.2014.30

BIGLOVA, A., ORTOBELLI, S., RACHEV, S., FABOZZI, F. (2010). Modeling, estimation, and optimization of equity portfolios with heavy-tailed distributions. In: Satchell, S.: *Optimizing Optimization:*

The next Generation of Optimization Applications and Theory. Amsterdam: Academic Press, 117–141.

CASELLA, G., BERGER, R.L. (2002). *Statistical Inference* (Vol. 2). Pacific Grove, CA: Duxbury Press.

CHERUBINI, U., LUCIANO, E., VECCHIATO, W. (2004). *Copula Methods in Finance*. New York: Wiley. http://dx.doi.org/10.1002/9781118673331

CHRISTOFFERSEN, P.F. (1998). Evaluating interval forecasts. *International Economic Review* 39(4): 841–862. http://dx.doi.org/10.2307/2527341

CONT, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–236. http://dx.doi.org/10.1080/713665670

CONT R., TANKOV, P. (2003). *Financial Modelling with Jump Processes*. Boca Ranton: CRC press. http://dx.doi.org/10.1201/9780203485217

COSTANZINO, N., CURRAN, M. (2015). Backtesting general spectral risk measures with application to expected shortfall. *Journal of Risk Model Validation* 9(1): 21–33. http://dx.doi.org/10.21314/JRMV.2015.131

DEMARTA, S., MCNEIL, A.J. (2005). The t copula and related copulas. *International Statistical Review/ Revue Internationale de Statistique* 73: 111–129.

EMBRECHTS, P., KAUFMANN, R., PATIE, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications* 32(1–2): 61–90. http://dx.doi.org/10.1007/s10589-005-2054-7

GNEITING, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* 106: 746–762. http://dx.doi.org/10.1198/jasa.2011.r10138

GONZÁLEZ-PEDRAZ, C., MORENO, M., PEÑA, J.I. (2015). Portfolio selection with commodities under conditional copulas and skew preferences. *Quantitative Finance* 15(1): 151–170. http://dx.doi.org/10.1080/14697688.2014.935463

HONG, Y., TU, J., ZHOU, G. (2007). Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20(5): 1547–1581. http://dx.doi.org/10.1093/rfs/hhl037

KERKHOF J., MELENBERG B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance* 28(8): 1845–1865. http://dx.doi.org/10.1016/j.jbankfin.2003.06.007

KONDOR, I., PAFKA, S., NAGY, G. (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance* 31(5): 1545–1573. http://dx.doi.org/10.1016/j.jbankfin.2006.12.003

KUPIEC, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3(2): 74–88. http://dx.doi.org/10.3905/jod.1995.407942

LAMANTIA, F., ORTOBELLI, S., RACHEV, S. (2006). An empirical comparison among VaR models and time rules with elliptical and stable distributed returns. *Investment Management and Financial Innovation* 3: 8–29.

LONGERSTAEY, J., ZANGARI, P. (1996). RiskMetrics. *Technical Document. J.P. Morgan*. Fourth edition. New York: J.P. Morgan Guaranty Trust Company.

MARINELLI, C., D'ADDONA, S., RACHEV, S.T. (2007). A comparison of some univariate models for value-at-risk and expected shortfall. *International Journal of Theoretical and Applied Finance* 10(6): 1043–1075. http://dx.doi.org/10.1142/S0219024907004548

ORTOBELLI, S., BIGLOVA, A., RACHEV, S.T., STOYANOV, S. (2010). Portfolio selection based on a simulated copula. *Journal of Applied Functional Analysis* 5(1): 177–194.

PAPP, G., PAFKA, S., NOWAK, M.A., KONDOR I. (2005). Random matrix filtering in portfolio optimization. *Acta Physica Polonica B* 36(9): 2757–2765.

PATTON, A.J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review* 47(2): 527–556. http://dx.doi.org/10.1111/j.1468-2354.2006.00387.x

PFLUG, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev, S. (ed.): *Probabilistic Constrained Optimization*. US: Springer, 272–281. http://dx.doi.org/10.1007/978-1-4757-3150-7_15

RACHEV, S.T., MENN, C., FABOZZI, F.J. (2005). *Fat-tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Hoboken: Wiley.

RACHEV, S.T., MITTNIK, S., FABOZZI, F.J., FOCARDI, S.M., JASIC, T. (2007). *Financial Econometrics: From Basics to Advanced Modeling Techniques*. New York: Wiley.

ROCKAFELLAR, R.T., URYASEV, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk* 2(3): 21–42. http://dx.doi.org/10.21314/JOR.2000.038

SCHOUTENS, W. (2003). *Lévy Process in Finance. Pricing Financial Derivatives.* New York: Wiley. http://dx.doi.org/10.1002/0470870230

SHARPE, W.F. (1994). The Sharpe ratio. *The Journal of Portfolio Management* 21(1): 49–58. http://dx.doi.org/10.3905/jpm.1994.409501

SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications del'Institut de Statistique de L'Université de Paris* 8: 229–231.

SUN, W., RACHEV, S., STOYANOV, S.V., FABOZZI, F.J. (2008). Multivariate skewed Student's t copula in the analysis of nonlinear and asymmetric dependence in the German equity market. *Studies in Nonlinear Dynamics & Econometrics* 12(2/3): 1–35. http://dx.doi.org/10.2202/1558-3708.1572

WONG, W.K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance* 32(7): 1404–1415. http://dx.doi.org/10.1016/j.jbankfin.2007.11.012

ZIEGEL, J.F. (2014). Coherence and elicitability. *Mathematical Finance*. DOI: 10.1111/mafi.12080 (to appear)