Stockholms
universitet

Value at Risk Estimation. A GARCH-EVT-Copula Approach

Ngoga Kirabo Bob

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2013:6
Matematisk statistik
Oktober 2013

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2013:6**
http://www.math.su.se

Value at Risk Estimation
A GARCH-EVT-Copula Approach

Ngoga Kirabo Bob*

October 2013

Abstract

Value at Risk (VaR) is one of the most widely used risk measure in risk management. It is defined as the worst loss to be expected of a portfolio over a given time horizon at a given confidence level. In this thesis we estimate portfolio VaR using an approach combining Copula functions, Extreme Value Theory (EVT) and GARCH models. We apply this approach to a portfolio consisting of stock indices from Germany, Spain, Italy and France. To estimate the VaR of this portfolio, we first use an asymmetric GARCH model and an EVT method to model the marginal distributions of each log returns series and then use Copula functions (Gaussian, Student's t, Clayton, Gumbel and Frank) to link the marginal distributions together into a multivariate distribution. We then use Monte Carlo Simulation (MCS) approach to find estimates of the portfolio VaR. To check the goodness of fit of the approach we use Backtesting methods. From the results, we conclude that, in general the GARCH-EVT-Copula approach performs well and specifically the GARCH-EVT-Student's t Copula outperforms all other GARCH-EVT-Copulas and traditional methods such as Historical Simulation (HS) and Variance Covariance (VC).

Keywords: Value at Risk (VaR), Copula, GARCH, Extreme Value Theory (EVT), Backtesting.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: ngbo5379@su.se. Supervisor: Joanna Tyrcha.

Acknowledgements

I am grateful to the supervisor of the thesis, Prof. Ph.D. Joanna Tyrcha, for giving valuable comments and guidance; and for assistance and insightful reviews.

This work would not have been possible without the help of many parents, friends and colleagues. I refrain from listing all their names here for fear that this section might otherwise exceed the others in length.

Stockholm, October 12^{th}, 2013.

Ngoga K. Bob.

Contents

List of Tables

List of Figures

1 Introduction

The Basel III framework developed by the Basel Committee on Banking Supervision requires that financial institutions such as banks and investment firms, set aside a minimum amount of capital to cover potential losses from their exposure to credit risk, operational risk and market risk [2]. For measuring market risk they recommend to use VaR, which is the worst loss to be expected of a portfolio over a given time horizon at a given confidence level. Mainly three approaches are used to estimate portfolio VaR; Historical Simulation (HS), Variance Covariance (VC) and Monte Carlo Simulation approaches. A key element in the VaR estimation is the distribution assumed for the financial log returns of the assets constituting the portfolio. In practice it is assumed that assets' log returns are normally distributed. However, this is not in accordance with the stylized fact that distributions of financial log returns series are fat-tailed and leptokurtic. Thus the normal distribution based VaR models (like the VC) tend to underestimate risk [4].

Extreme Value theory (EVT) which is a branch of statistics that studies rare or extreme events is well suited to describe the above mentioned fat-tailed property. It is important to mention that some EVT methods assumes that the data to be studied are independently and identically distributed (i.i.d.), which is not always the case for most financial log returns series. In this thesis, in order to estimate portfolio VaR with assets' log-returns which are not i.i.d. we adopt an approach proposed by McNeil and Frey [11]. They use GARCH models to estimate the current volatility of the log-returns series and EVT for estimating the tail of the innovations' distribution of the GARCH model before estimating VaR. They find that this approach gives better estimates than methods which ignore the fat tails of the innovations or the stochastic nature of the volatility. Nyström and Skoglund [12] combine ARMA-(asymmetric) GARCH and EVT method to estimate quantiles of univariate portfolio risk factors. They find that for high quantiles (between 97% and 98%) the use of EVT does indeed give a substantial contribution and the Generalized Pareto distribution(GPD) is more able than the normal distribution to accurately model the empirically observed fat tails.

Another key element in portfolio VaR estimation is the dependence structure between financial assets in the portfolio. The most common measure of dependence in the Pearson Correlation. However, Poon et al. [16] argue that Pearson Correlation is not a good measure of dependency in cases where the extreme realizations are important. This has resulted in the introduction of Copula based models in the last few years. Copulas are defined as functions that links univariate marginals to form multivariate distributions. Therefore, the key characteristic of Copula based models is the separation of the joint distribution into two components; the marginal distributions and the dependence structure [4]. Applications of Copula based models to portfolio VaR estimation have combined GARCH (or ARMA-GARCH-EVT) filtered marginals with various Copula functions. Hsu et al. [14] estimates VaR using a Copula-EVT approach on Asian markets. They find that Clayton Copula-EVT have the best performance. Ghorbel and Trabelsi [17] proposes a method for estimating VaR using an ARMA-GARCH-EVT Copula approach. For a given multivariate financial data, they find that their approach provide a better presentation of the dependence structure of the multivariate data and produce accurate estimates of VaR.

In this thesis we select stock indices from Germany, Spain, Italy and France to form a portfolio to study. To estimate the portfolio VaR we combine GARCH models, EVT and Copula functions in an approach we refer to as GARCH-EVT-Copula approach. In this approach we use an asymmetric GARCH model and an EVT method to model the marginal distributions of each log return series and then use Copula functions (Gaussian, Student's t, Clayton, Gumbel and Frank) to link the marginal distributions together to form a multivariate distribution. We then use Monte Carlo Simulation (MCS) approach to find estimates of the portfolio VaR which are backtested to check the performance of the approach compared to other popular approaches of VaR estimation. The structure of this thesis is as follows; the next Section describes the theoretical framework used, that is, a review of VaR as a risk measure, GARCH models, EVT and Copula theory. Section 3 describes how marginal distributions and dependence structure are modeled and the Backtesting procedure. Empirical results are presented in Section 4 and, finally, Section 5 concludes.

2 Theoretical Framework

2.1 Value at Risk (VaR)

2.1.1 Backgrounds and Definitions

Value at Risk is one of the most (if not the most) widely used risk measure in the field of risk management. It has been adapted in the regulatory Basel framework for banks as the major determinant of the risk capital required for covering potential losses arising from market risks [2]. Its greatest advantages are that , it summarizes risk in a single, easy to understand number and that it does not depend on a specific kind of distribution and therefore, in theory, can be applied to any kind of financial asset.

Definition:(Value at Risk (VaR)) The VaR at level $p \in (0, 1)$ of a portfolio with value X at time $t + 1$ is given by

$$VaR_p(X) = \min\{m : P(mR_t + X \leq 0) \leq p\}, \tag{1}$$

where R_t is the percentage return of a risk free asset at time t.

This means that the VaR of a position with value X at time $t + 1$ is the smallest amount of money that if added to the position now (at time t)and invested in the risk free asset ensures that the probability of a strictly negative value at time $t + 1$ is not greater than p.

One can show that the minimum m in equation (1) is attained. In fact,
$\{m : P(mR_t + X \leq 0) \leq p\} = \{m : P(-\frac{X}{R_t} > m) \leq p\}$ =$\{m : 1 - P(-\frac{X}{R_t} \leq m) \leq p\}$
= $\{m : P(-\frac{X}{R_t} \leq m) \geq 1 - p\}$.
Since a distribution function F is right continuous ($F(x) \downarrow F(x_0)$ as $x \downarrow x_0$) and increasing, then $\{m : F(m) \geq 1 - p\} = [m_0, \infty)$ for some m_0, and therefore there exists a smallest element.

Hult et al. [1] suggest that, for VaR_p to be a sensible choice of risk measure for typical asset portfolios with mainly long positions, one can have the following view: at the current time t, one take a risk free loan of size V_t (which is the current portfolio value), use the capital to purchase the asset portfolio, and end up with the net value $X = V_{t+1} - V_t R_t$ at time $t + 1$. Therefore, the portfolio is classified as acceptable if the difference between the actual future portfolio value and the value that would be obtained by instead investing the current portfolio value in a risk free asset is VaR_p acceptable.

Define, $L_t = -\frac{X}{R_t}$. Considering X as the net gain from the investment, where the current portfolio value V_t is viewed as a Liability gives us

$$L_t = -\frac{X}{R_t} = V_t - \frac{V_{t+1}}{R_t} \tag{2}$$

which has a natural interpretation of the natural discounted loss. Because V_{t+1} is unknown to us L_t is random from the perspective of time t. The distribution of L_t will be referred to as the *loss distribution*. Note that the definition of loss presented here implicitly assumes that the portfolio composition is constant over the considered time interval. In terms of L_t, VaR_p is given by

$$VaR_p(X) = \min\{m : P(L_t \leq m) \geq 1 - p\}. \tag{3}$$

One can then interpret $VaR_p(X)$ as the smallest value m such that the probability of the discounted loss being at most m is at least $1-p$; or, it is the smallest amount of money that, if put aside and invested in a risk free asset at time t, will be sufficient to cover a potential loss at time $t+1$ with a probability at least $1-p$.

We can see that VaR_p has two basic parameters: the significance level p (or the confidence level $1-p$) and the length in physical time of the time period over which the discounted loss is modeled which we will refer as the risk horizon h. Often used values of p are $5\%, 1\%$ and 0.5%, which shows that VaR_p describes (to some extent) the right tail of the loss distribution. The risk horizon h is the period over which we measure the potential discounted loss. Different risks are naturally assessed over different time periods according to their liquidity. For example in market risk measurement h is equal to 1 day or 10 days (under Basel III) and 1 year for credit and insurance risk measurement.

Definition (Quantile): The u-quantile of a random variable L with distribution function F_L is defined as $F_L^{-1}(u) = \min\{m : F_L(m) \geq u\}$ where F_L^{-1} is just the ordinary inverse if F_L is strictly increasing.

If F_L is both continuous and strictly increasing, then $F_L^{-1}(u)$ is the unique value m such that $F_L(m) = u$ [1]. For a general F_L, the quantile value $F_L^{-1}(u)$ is obtained by plotting the graph of F_L and setting $F_L^{-1}(u)$ to be the smallest value m for which $F_L(m) \geq u$. From the definition above, $VaR_p(X)$ is defined as the $1-p$ quantile of L, that is,

$$VaR_p(X) = F_L^{-1}(1-p). \tag{4}$$

2.1.2 Empirical Quantiles and VaR

Consider observations $x_1, ..., x_n$ from independently and identically distributed variables $X_1, ...X_n$ with a common unknown distribution function F defined on the real line $\Re$. The empirical quantile function F_n^{-1} is the quantile function of the empirical distribution function F_n and therefore given by $F_n^{-1}(p) = \min\{x : F_n(x) \geq p\}$. Similarly, the empirical quantile function $F_{n,X}^{-1}$ is the quantile function of $F_{n,X}$. It is shown in [1] that the empirical quantile $F_{n,X}^{-1}(p)$ is the k^{th} largest of the sample points $x_1, ...x_n$ where $k = [n(1-p)] + 1$; this means that the empirical quantile function is given by

$$F_{n,X}^{-1}(p) = X_{[n(1-p)]+1,n} \quad p \in (0,1) \tag{5}$$

So, given a sample $L_1, ...L_n$ of independent copies of L, the empirical estimate of

$VaR_p(X)$ is given by

$$\widehat{VaR_p(X)} = L_{[np]+1,n} \tag{6}$$

where $L_{1,n} \geq ... \geq L_{n,n}$ is the ordered sample and $[y]$ means the integer part of y. Note that $\widehat{VaR_p(X)}$ is just the empirical $1-p$ quantile of L. So, to compute the empirical VaR estimate from a sample of historical prices, one need to first transform the prices into a sample $L_1, ...L_n$ and then compute the VaR estimate as an empirical quantile.

2.1.3 Approaches for VaR Estimation

Historical Simulation (HS) Approach

The idea behind the historical simulation approach is to use the historical distribution of assets' prices (or log returns) in our portfolio to estimate the portfolio's VaR, assuming that we held this portfolio over the period of time covered by our historical data set. To apply this approach, we first identify the different instruments in our portfolio and collect a sample of their historic log returns over some observation period. We then use the weights in our current portfolio to simulate the discounted loss distribution (L_t). Thus, the HS approach does not rely on any parametric model assumptions. It does however rely on the stationarity of the historical data set to ensure convergence of the empirical discounted loss distribution to the true discounted loss distribution [4].
We assume that this historical distribution of log returns is a good approximation of the distribution of log returns we face over the next holding period. This means that we assume that the history will repeat itself in the future. Finally, the relevant quantile from the distribution of historical log returns will lead us to the expected portfolio VaR. Suppose we have $P_{i,j}$ (Price of asset i at time j) observations, j running from period 1 to period N. Let $r_{i,j}$ be the log return of asset i over period j, W_i be the relative weight of asset i and that we have n assets in our portfolio. The historically simulated loss series $\{L_{t,j}\}_{j=1}^{N}$ is generated using equation (2). Where, $V_t = \sum_{i=1}^{n} W_i P_{i,t}$ and $V_{t+1,j} = \sum_{i=1}^{n} W_i P_{i,t} \mathrm{e}^{R_{i,j}}$. Where the $R_{i,j}$ are the historically simulated log returns. Most often in practice, they are simulated by drawing with replacement from the original sample of $r_{i,j}$. The portfolio VaR is inferred from the discounted loss distribution.
The HS approach has a number of attractions; it is conceptually simple and easy to implement. Moreover, it does not depend on assumption about the distributions of log returns because it allows the data to speak for itself and determine the distribution. To ensure sufficient estimation precision, HS requires a large amounts of data [4]. However, it is not always practically feasible to obtain such large appropriate sample data; and even if it is the history may not repeat itself or contains sufficient extreme

observations for the VaR estimation.

Variance Covariance Methods

Here we illustrate the Variance-Covariance approach, which is the most well known approach for VaR estimation and which has been first fully explained in detail by J.P. Morgan's (1996) RiskMetrics Technical Document. Under this approach, VaR for individual positions and portfolios can be easily derived by estimating the variance and the covariance (or alternatively, the standard deviations and correlations) of some predefined risk factors' log returns and the sensitivity of the portfolio to those risk factors. The most basic assumption in the model is that risk factor log returns are independently and identically distributed (i.i.d.) with a multivariate normal distribution. Suppose we have a portfolio consisting of n assets with log returns of asset i as r_i and relative weights W_i; the portfolio return R_{PF} and variance σ^2_{PF} at time t are defined as:

$$R_{PF} = \sum_{i=1}^{n} W_i r_i \tag{7}$$

$$\sigma^2_{PF} = \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \sigma_{ij} \tag{8}$$

where σ_{ij} is the covariance between asset i and j at time t. Using vector notation we have that

$$\sigma^2_{PF} = W^T \Sigma W \tag{9}$$

where W is the portfolio weights and Σ the covariance matrix of log returns. Having $X = V_{t+1} - V_t R_t$ and assuming that t is one day we can suppose that the influence of interest rates for such a short period can be neglected, then we have $X = V_{t+1} - V_t$. Assuming that X is normally distributed $X \sim N(\mu, \sigma^2)$ and using the ellipticity property of the normal distribution [1] we have that $X =^d \mu + \sigma Z$ with $Z \sim N(0,1)$, and

$$VaR_p(X) = -\mu + \sigma\Phi^{-1}(1-p) \tag{10}$$

where $\mu = E[R_{PF}]$, $\sigma = \sigma_{PF}$ and Φ is the distribution function of Z. The advantage of Variance Covariance model is its simplicity. VaR computation is relatively easy if normality is assumed, as standard mathematical properties of the normal distribution can be used to calculate VaR at different levels. In addition, normality assumption allows easy translatability between different levels and holding periods [6]. VaR can be adjusted for different time horizon by $VaR_{t_2} = \sqrt{\frac{t_2}{t_1}} VaR_{t_1}$ and for different confidence levels by $VaR_p = \frac{\Phi^{-1}(p)}{\Phi^{-1}(p1)} VaR_{p_1}$. Note that the Variance Covariance approach is only applicable to a portfolio whose return(or loss) is a linear function of its assets returns. Thus, it gives poor estimate for portfolios with non linear instruments such as options. However the most predominant disadvantage of the approach is the normality assumption. Most financial assets are known to have fat tailed log return distributions, meaning that in reality extreme outcome are more probable than the normal distribution would suggest. As a result, VaR estimate might be understated.

Monte Carlo Simulation (MCS) Methods

The basic idea behind MCS approach is to simulate repeatedly random process governing the prices of all financial instruments in our portfolio. Each simulation gives us a possible value for our portfolio at the end of our target horizon; and if we take enough of these simulations, the simulated distribution of portfolio values will converge to the portfolio's unknown "true" distribution, and we can use the simulated distribution to infer the VaR of the "true" one. Dowd [6] summarizes the approach in the following steps:

1. Select a model for the stochastic variables of interest.
2. Construct simulated paths for the stochastic variables.
3. Repeat the simulations in (2) enough times to be confident that the simulated distribution is sufficiently close to the "true" distribution of actual values.
4. Infer the VaR from the distribution in (3).

We illustrate the above steps by first assuming that we have one financial asset in our portfolio and that the price of the asset is described by the following process:

$$dP_t = \sigma dZ_t^1 \tag{11}$$

This equation gives the changes in the current price dP_t in terms of the volatility parameter σ (which is assumed to be known or can be estimated) and a Wiener process Z_t^1. Equations of this type are usually made more tractable by taking a discretised approximation over some short time interval $\triangle t$. Let us therefore take such an approximation while also choosing our time units so that $\triangle t = 1$. Hence:

$$P_t = P_{t-1} + \sigma Z_t \sqrt{\triangle t} = P_{t-1} + \sigma Z_t \tag{12}$$

which gives the current price P_t in terms of the price for the previous period P_{t-1}, as well as σ and Z_t (with Z_t a standard normal variable). We now wish to simulate the asset price over an interval $[t, T]$. We do so by first leading equation (12) by one period to give us an expression for P_{t+1} in terms of P_t and Z_{t+1}; we then substitute equation (12) into this equation to get

$$P_{t+1} = P_t + \sigma Z_{t+1} = P_{t-1} + \sigma Z_t + \sigma Z_{t+1} = P_{t-1} + \sigma(Z_t + Z_{t+1}) \tag{13}$$

We repeat the above to get an expression for P_{t+2}, then for P_{t+3}, and so on until we have an expression of P_T

$$P_T = P_{t-1} + \sum_{i=t}^{T} Z_i \tag{14}$$

The price at T depends on the initial price P_{t-1} and the sum of the realised values of Z_i for $i = t, t+1, ...T$. We now use a random number generator to produce a series of the realised values of $Z_t,\ Z_{t+1}, ...\ Z_T$ and substitute in equation (14) to produce a final simulated price, P_T. We then multiply this price by the number of shares we have to determine the portfolio value. If we repeat the process enough times, the distribution of simulated portfolio values produced in this way will converge to the "true" probability density function underlying the portfolio values process [6]. We can then use the simulated portfolio values to construct the corresponding loss distribution and hence infer the VaR.

Suppose now that we have in our portfolio more than one financial asset. This require us to simulate more than one Z-path, except for two special cases in which the procedure is much the same as with a single asset. The first of these is where the prices are independent, in which case we have equations similar to equation (14). With two assets, we have:

$$\begin{bmatrix} P_{1,t} \\ P_{2,t} \end{bmatrix} = \begin{bmatrix} P_{1,t-1} \\ P_{2,t-1} \end{bmatrix} + \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \tag{15}$$

The portfolio value at time t is found by multiplying equation (15) by $\begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix}$ the vector of relative weights of the two assets in the portfolio, to get

$$V_t = \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} P_{1,t-1} \\ P_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \tag{16}$$

which depends on time t's drawings of $Z_{1,t}$ and $Z_{2,t}$. The portfolio value at time T is given by:

$$V_T = \sum_{i=t}^{T} V_i \tag{17}$$

which depends on the drawings of both Z variables from t to T. All we now need to do is run enough iterations to simulate the true distribution of portfolio values, switch over to the loss distribution and then infer the VaR.

The other case is where the assets prices are perfectly correlated with each other, either positively or negatively. In this case,

$$V_t = \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} P_{1,t-1} \\ P_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ \pm\sigma_2 & 0 \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \tag{18}$$

Use equation (17) to solve for the final portfolio value V_T, and proceed as before to repeat the exercise as many times as required to produce a reliable VaR estimate.
The difficulties arise when the two prices are imperfectly correlated with each other. In this case, we have

$$V_t = \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} P_{1,t-1} \\ P_{2,t-1} \end{bmatrix} + \begin{bmatrix} W_{1,t} & W_{2,t} \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \tag{19}$$

where the $a_{i,j}$ are functions of the underlying variances and correlations. We now have to solve for these, and the usual procedure is by means of a Choleski decomposition. To see what is involved, let's first rewrite equation (19) in first difference form by subtracting the lagged P terms from each side.

$$\begin{bmatrix} \triangle P_{1,t} \\ \triangle P_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} \tag{20}$$

Now denote the vector of $\triangle P_{i,t}$ terms by $\triangle P_t$, the matrix of $a_{i,j}$ terms by A and the vector of $Z_{i,t}$ terms by Z_t. We then multiply each side of equation (20) by its transpose denoted by $Z_t^T A^T$ and $\triangle P_t^T$ to get:

$$\triangle P_t \triangle P_t^T = A Z_t Z_t^T A^T \tag{21}$$

Taking the expectation of equation (21) and since the $Z_{i,t}$ variables are independent standard normal, the expectation of the matrix $Z_t Z_t^T$ is the identity matrix with ones along the diagonal and zero elsewhere we get:

$$\Sigma = AA^T \tag{22}$$

Equation (22) tells us that the matrix A is the 'square root matrix' of the expected variance covariance matrix Σ. The actual values of the terms in A then turn out to be:

$$\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{(1-\rho^2)} \end{bmatrix} \tag{23}$$

Once we have the matrix A, we can find prices at time t; (16) and (17) give us the portfolio values at time t and T respectively. We then do repeated iterations as before, estimate the loss distribution and infer the VaR. A potential weakness of the MCS models is Model Risk, which arises due to wrong assumptions about the pricing models and underlying stochastic processes. This will lead to wrong VaR estimates, if not properly specified [7]. In practice, for equity positions the Geometric Brownian Motion assumption is used. This is a generalization of equation (11) and this assumes that the stock price is governed by the process $\frac{dS_t}{S_t} = \mu dt + \sigma dWt$ where Wt is a Wiener process. Moreover, Dowd [6] points out that complicated procedures associated with MCS methods require special expertise. Senior management may therefore have hard time following how VaR estimates are calculated when using MCS methods.

2.2 GARCH Models

In this section, we briefly describe time series models which are able to explain a number of important features common to most financial data including:

- **Leptokurtosis**: that is, the tendency for financial asset log returns to have distributions that exhibit fat tails and excess peakedness at the mean.
- **Volatility Clustering**: the tendency for volatility in financial markets to appear in bunches. Thus large log returns (of either sign) are expected to follow large log returns, and small log returns (of either sign) are expected to follow small log returns.
- **Leverage effects**: the tendency for volatility to rise more following a large price fall than following a price rise of the same magnitude [19].

To capture the volatility clustering, Engle (1982) suggested the autoregressive conditional heteroscedasticity (**ARCH**) model in which the conditional variance is modeled as a linear function of past squared innovations. The general $ARCH(q)$ model has the form:

$$\sigma_t^2 = w + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \tag{24}$$

where $w > 0$ and $\alpha_j \geq 0$, for $j = 1, ..., q$ in order to keep the conditional variance positive. Unfortunately, we often need q to be large in order to fit the data. As a way to model persistent movements in volatility without estimating a large number of parameters, Bollerslev and Taylor (1986) proposed a more parsimonious model; the generalized autoregressive conditional heteroscedasticity (**GARCH**) model. The general $GARCH(p, q)$ model is given by:

$$\sigma_t^2 = w + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \tag{25}$$

where $w > 0$, $\alpha_j \geq 0$, for $j = 1, ..., q$ and $\beta_i \geq 0$ for $i = 1, ..., p$. The model is a generalized version of the ARCH model in the sense that the squared conditional volatility σ_t^2 is a linear function of past squared conditional volatilities as well as squared innovations of the process. The ARCH and GARCH models are symmetric in the sense that negative and positive shocks have the same effect on volatility, this means that, the signs of the innovations or shocks have no effect on the conditional volatility, only the squared innovations enter the conditional variance equation. This is, however, inconsistent with the stylized fact that negative shocks tend to have a large impact on volatility than the positive shocks of the same magnitude (leverage effect). In the following we consider two extensions of the GARCH model that take this asymmetry into account.

The exponential GARCH (**EGARCH**) model proposed by Nelson (1991) explicitly allows for asymmetries in the relationship between return and volatility. The general $EGARCH(p, q)$ model can be expressed as follows:

$$\ln(\sigma_t^2) = w + \sum_{i=1}^{p} \beta_i \ln(\sigma_{t-i}^2) + \sum_{j=1}^{q} \gamma_j \frac{\varepsilon_{t-j}}{\sigma_{t-j}} + \sum_{j=1}^{q} \alpha_j [\frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} - E(\frac{|\varepsilon_{t-j}|}{\sigma_{t-j}})] \tag{26}$$

By parameterizing the natural logarithm of the conditional variance as opposed to the conditional variance, no inequality constraints are needed to ensure positive conditional variances. The expected value of the standardized innovations depends on the assumed innovation distribution which can be normally or Student's t distributed. The EGARCH model differs from the GARCH model by the fact that, it allows positive and negative shocks to have a different impact on volatility and also allows large shocks to have greater impacts on volatility. Finally, in contrast to the GARCH model which allows for volatility clustering through a combination of the β_i and α_j terms, volatility clustering is entirely captured by β_i in the EGARCH model.

The Glosten-Jagannathan-Runkle GARCH (**GJR-GARCH**) model proposed by Glosten (1993) [3] also allows the conditional variance to respond differently to past negative and positive innovations, but the manner in which it does so is different from the EGARCH model. In the EGARCH model the coefficients γ_j are applied to the actual innovations ε_{t-j}, while in the GJR-GARCH model this coefficient enters the model through a Boolean indicator. The general $GJR - GARCH(p, q)$ model is as follows:

$$\sigma_t^2 = w + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j}^2 I_{\varepsilon_{t-j}<0} \tag{27}$$

where $w > 0$, $\beta_i \geq 0$ for $i = 1, ..., p$, $\alpha_j \geq 0$ and $\alpha_j + \gamma_j \geq 0$ for $j = 1, ..., q$. $I_{\{.\}}$ denotes the indicator function which returns one if the threshold level is satisfied and zero otherwise.

2.3 Extreme Value Theory (EVT)

In this section we briefly discuss Extreme Value Theory (EVT) with applications in financial risk management focusing on the methods we will be using in this thesis. Within the context of EVT, there are mainly two approaches for modeling extremal events. The Block Maxima Models, which are based on directly modeling the distribution of extreme realizations, and the Peaks Over Threshold (POT) models, which models the exceedences of a particular threshold. McNeil et al [4], acknowledge that the POT method uses data more efficiently and is therefore considered as the most useful for practical applications. Thus we will focus only on the POT method. EVT rests on the assumption of independently and identically distributed (i.i.d.) data. In this thesis, however we are using financial time series data which have log returns which are known to have some mild correlations and exhibit dependence in the second moment. Consequently, when EVT is applied to financial time series data we need to take temporal dependencies into account. If not we will produce estimators with non optimal performances.

2.3.1 Generalized Pareto Distribution (GPD)

The GPD describes the limiting distribution for modeling excesses over a certain threshold. If X is a random variable (say daily portfolio losses) which is generalized Pareto distributed, then its distribution function has the form

$$G_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 - \frac{\gamma x}{\beta}\right)^{\frac{-1}{\gamma}} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{if } \gamma = 0 \end{cases} \tag{28}$$

where $\beta > 0$ and $x \geq 0$ when $\gamma \geq 0$ and $0 \leq x \leq \frac{-\beta}{\gamma}$ when $\gamma < 0$. The parameters γ and β are referred to as the shape and scale parameters respectively. The GPD is generalized in the sense that it contains a number of specific distributions under its parametrization. When $\gamma > 0$, the distribution function $G_{\gamma,\beta}$ is that of a heavy tailed ordinary Pareto distribution; when $\gamma = 0$ we have a light tailed exponential distribution and when $\gamma < 0$ we have a short tailed Pareto type II distribution. Moreover for fixed x the parametric form is continuous in γ, so, $\lim_{\gamma \to 0} G_{\gamma,\beta}(x) = G_{0,\beta}(x)$ [4]. The GPD family can be extended by adding a location parameter $\mu \in \Re$, that is

$$G_{\gamma,\mu,\beta}(x) = G_{\gamma,\mu,\beta}(x - \mu) \tag{29}$$

The support has to be adjusted accordingly. When $\mu = 0$ and $\beta = 1$, the representation is known as the standard GPD. The GPD density function has the form

$$g_{\gamma,\beta}(x) = \begin{cases} \frac{1}{\beta}\Big(1 + \frac{\gamma x}{\beta}\Big)^{\frac{-\gamma-1}{\gamma}} & \text{if } \gamma \neq 0 \\ \frac{1}{\beta}\exp\Big(-\frac{x}{\beta}\Big) & \text{if } \gamma = 0 \end{cases} \tag{30}$$

The tail of the density fattens and the peaks are sharpening with increasing γ, while with increasing β the central part of the density gets more flat [4].

2.3.2 The Peak Over Threshold (POT) Method

Let's consider a series $\{X_i\}_{i=1}^{n}$ of independently and identically distributed random variables representing financial losses with a common but unknown distribution function F and upper endpoint $x_F = \sup\{x \in \Re : \ F(x) < 1\} \leq \infty$. Let u be a certain threshold and denote N_u by $N_u =$ Card$\{i : X_i > u, \ i = 1, ..., n\}$ the number of exceedences of u by $\{X_1, ..., X_n\}$.

Definitions:

- A distribution function F has a heavy tail if the tail probability $F(x)$ decays slowly as x decreases, that is

$$\lim_{x \to -\infty} \frac{F(x)}{\exp(-\lambda(-x))} = \infty \text{ for every } \lambda > 0 \tag{31}$$

  and a heavy right tail if

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{\exp(-\lambda x)} = \infty \text{ for every } \lambda > 0 \tag{32}$$

- A distribution function F has a regularly varying right tail $\overline{F} = 1 - F$ if there exists a number ρ such that

$$\lim_{t \to \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{\rho} \text{ for every } x > 0 \tag{33}$$

  and a regularly varying left tail if

$$\lim_{t \to -\infty} \frac{F(tx)}{F(t)} = x^{\rho} \text{ for every } x > 0 \tag{34}$$

Since $\overline{F}(x)$ is decreasing in x, it holds that $\rho \leq 0$, and we may set $\rho = -\alpha$ for $\alpha \geq 0$. Then (33) can be written as

$$\lim_{t\to\infty} P(X > tx|X > t) = x^{-\alpha} \text{ for every } x > 1. \tag{35}$$

Suppose now that the common and unknown distribution function F of the $\{X_i\}_{i=1}^n$ has a regularly varying right tail. The following result from Theorem 7 (Pickands-Balkema-de Haan) in [4], states that for large u approaching x_F, the excesses $y = X_k - u$ with distribution function F_u are well approximated by the GPD. That is

$$\lim_{u\to x_F} \sup_{0\leq y\leq x_F-u} |F_u(y) - G_{\gamma,\beta(u)}(y)| = 0 \tag{36}$$

for some positive function $\beta(u)$. This result is used by the POT method to construct estimates of tail probabilities and quantiles. For $\gamma > 0$ and $\beta > 0$ the GPD is defined by

$$G_{\gamma,\beta}(x) = 1 - (1 + \frac{\gamma x}{\beta})^{\frac{-1}{\gamma}} \text{ for } x \geq 0 \tag{37}$$

Suppose that X is a random variable with distribution function F that has a regularly varying right tail so that $\lim_{u\to\infty} \frac{\overline{F}(\lambda u)}{\overline{F}(u)} = \lambda^{-\alpha}$ for all $\lambda > 0$ and some $\alpha > 0$. Then

$$\begin{aligned}
\lim_{u\to\infty} P\Big(\frac{X-u}{\frac{u}{\alpha}} > x|X > u\Big) &= \lim_{u\to\infty} \frac{P(X > u(1+\frac{x}{\alpha}))}{P(X > u)} \\
&= (1 - \frac{x}{\alpha})^{-}\alpha \\
&= \overline{G}_{\frac{1}{\alpha},1}(x)
\end{aligned} \tag{38}$$

The excess distribution function of X over the threshold u is given by

$$F_u(x) = P(X - u \leq x|X > u) \; for \; x \geq 0 \tag{39}$$

This can be written in terms of F as:

$$\begin{aligned}
\overline{F}_u(x) &= P(X > u + x|X > u) = \frac{P(X > u + x, X > u)}{P(X > u)} \\
&= \frac{\overline{F}(u + x)}{\overline{F}(u)} = \frac{\overline{F}(u(1 + \frac{x}{u}))}{\overline{F}(u)}
\end{aligned} \tag{40}$$

Since $\overline{F}$ is regularly varying with index $-\alpha < 0$, it holds that $\frac{\overline{F}(\lambda u)}{\overline{F}(u)} \to \lambda^{-\alpha}$ uniformly in $\lambda \geq 1$ as $u \to \infty$ [1],this means that

$$\lim_{u\to\infty} \sup_{\lambda\geq 1} |\overline{F}(u) - \lambda^{-\alpha}| = 0$$

Hence from equation (40) and using the result in (36) we have that :

$$\lim_{u\to\infty}\sup_{x>0}|\overline{F}_u(x)-\overline{G}_{\gamma,\beta(u)}(x)|=0 \tag{41}$$

where $\gamma=\frac{1}{\alpha}$ and $\beta(u)\sim\frac{u}{\alpha}$ as $x\to\infty$. As we are interested in finding $\overline{F}(u+x)$ and $F^{-1}(p)$ for large u and p; recall from (40) that:

$$\overline{F}(u+x)=\overline{F}(u)\overline{F}_u(x) \tag{42}$$

If u is not too far into the tail, then the empirical approximation $\overline{F}(u)\approx\overline{F}_n(u)=\frac{N_u}{n}$ is accurate. Moreover, from (41) we have that

$$\overline{F}_u(x)\approx\overline{G}_{\gamma,\beta(u)}(x)\approx\overline{G}_{\widehat{\gamma},\widehat{\beta}}(x)=(1+\widehat{\gamma}\frac{x}{\widehat{\beta}})^{\frac{-1}{\widehat{\gamma}}} \tag{43}$$

where $\widehat{\gamma}$ and $\widehat{\beta}$ are the estimated parameters of γ and β respectively. Relation (42) suggests estimating the tail of F by estimating $\overline{F}_u(x)$ and $\overline{F}(u)$ separately. We consider the estimator for $\overline{F}(u+x)$ given by:

$$\widehat{\overline{F}}(u+x)=\frac{N_u}{n}(1+\widehat{\gamma}\frac{x}{\widehat{\beta}})^{\frac{-1}{\widehat{\gamma}}} \tag{44}$$

Using the quantile definition and the equation (43) leads to,

$$\begin{aligned}
\widehat{\overline{F}^{-1}}(p) &= \min\{x:\widehat{\overline{F}}(x)\le 1-p\}\\
&= \min\{u+x:\widehat{\overline{F}}(u+x)\le 1-p\}\\
&= u+\min\{x:\frac{N_u}{n}(1+\widehat{\gamma}\frac{x}{\widehat{\beta}})^{\frac{-1}{\widehat{\gamma}}}\le 1-p\}\\
&= u+\frac{\widehat{\beta}}{\widehat{\gamma}}\Big(\big(\frac{n}{N_u}(1-p)\big)^{-\widehat{\gamma}}-1\Big) \qquad (45)
\end{aligned}$$

Hult at al. [1] summarizes the POT method for estimating tail probabilities and quantiles by the following steps:

1. Choose a high threshold u and count the number N_u of exceedences $X_k>u$
2. Given the sample $Y_1,...,Y_{N_u}$ of excesses, estimate the GPD parameters γ and β.
3. Combine step (1) and (2) to get tail probabilities and quantile estimates using relations (44) and (45).

2.4 Copula Theory

The problem of modelling asset log returns is one of the most important issue in Finance. In general it is assumed that log returns are normally distributed; however empirical research has shown that asset log returns are leptokurtic and fat tailed. Another issue in Finance which has been getting more attention after the financial crisis is the capital allocation within Banks. Banks are urged by regulatory institutions to build sound internal models to measure risks (mostly credit and market risks) for all their activities. These internal models used to measure risks face an important problem which is the modeling of the joint distributions of different risks. These two issues can be treated as a problem of copula. A copula is a function that links univariate marginals to their multivariate distribution. Copula theory was first developed in Sklar (1959) [10]. It is a powerful tool as it does not require any assumptions on the selection distribution function and it allows the risk manager to decompose any n-dimensional joint distribution function into n marginals and a copula. In this section we briefly describe copula theory by highlighting important properties and examples.

2.4.1 Definitions and Basic Properties

Before defining a copula we recall an important proposition from [1] (Proposition 6.1 page 166). We denote the uniform distribution on an interval $(0,1)$ by $U(0,1)$, i.e., the probability of a random variable U satisfying $P(U \leq u) = u$ for $u \in (0,1)$.
Proposition: Let F be a distribution function on $\Re$. Then

(i) $u \leq F(x)$ if and only if $F^{-1}(u) \leq x$.

(ii) If F is continuous, then $F(F^{-1}(u)) = u$.

(iii) (Quantile Transform) If $U \sim U(0,1)$ then $P(F^{-1}(U) \leq x) = F(x)$.

(iv) (Probability Transform) If X has a distribution function F, then $F(X) \sim U(0,1)$ if and only if F is continuous.

Definition (Copula): A d-dimensional copula is the distribution function C of a random vector U whose components U_k are uniformly distributed. That is:

$$C(u_1, ..., u_d) = P(U_1 \leq u_1, ..., U_d \leq u_d), \ \ (u_1, ..., u_d) \in (0,1)^d \tag{46}$$

Let $(X_1, ..., X_d)$ be a random vector with distribution function $F(x_1, ..., x_d) = P(X_1 \leq x_1, ..., X_d \leq x_d)$ and suppose that $F_k(x) = P(X_k \leq x)$ is a continuous function for every k. The probability transform from the statement (iv) of the proposition implies that the components of the vector $U = (U_1, ..., U_d) = (F_1(X_1), ..., F_d(X_d))$ are uniformly distributed. In particular, the distribution function C of U is a copula and we call it the copula function of X. Using the statement (i) from the proposition we have that

$$\begin{aligned} C(F_1(x_1), ..., F_d(x_d)) &= P(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ &= P(F_1^{-1}(U_1) \leq x_1, ..., F_d^{-1}(U_d) \leq x_d) \\ &= F(x_1, ..., x_d) \end{aligned} \tag{47}$$

Equation (47) is the result from the famous Sklar's theorem [8]. It is the representation of the joint distribution function F in terms of the copula C and the marginal distributions $F_1, ..., F_d$, which explains the name Copula; a function that couples the joint distribution function to its univariate marginal distribution functions.

Definition (Copula Density function): The density $c(u_1, ..., u_d)$ associated to the copula $C(u_1, ..., u_d)$ is defined as:

$$c(u_1, ..., u_d) = \frac{\partial C(u_1, ..., u_d)}{\partial u_1 ... u_d} \tag{48}$$

For continuous random variables, the copula density is related to the density function F, denoted as f, by the following so called canonical copula representation [8]:

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \prod_{j=1}^{d} f_j(x_j) \tag{49}$$

where f_j are the densities of the marginals $f_j = \frac{dF_j(x_j)}{dx_j}$.

2.4.2 Examples of Copulas

There are mainly two families of copulas used in financial applications: Elliptical copulas and Archimedean copulas.

Elliptical Copulas

Elliptical copulas are derived from multivariate elliptical distributions. The most important copulas in this family are the Gaussian (or normal) copula and the Student's t copula.

The Gaussian Copula

The Gaussian copula C_ρ^{Ga} of a d-dimensional standard normal distribution, with linear correlation matrix ρ, is the distribution function of the random vector $(\Phi(X_1), ..., \Phi(X_d))$, where Φ is the univariate standard normal distribution function and $X \sim N_d(0, \rho)$. Hence,

$$C_\rho^{Ga} = P(\Phi(X_1) \leq u_1, ..., \Phi(X_d) \leq u_d) = \Phi_\rho^d(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)) \tag{50}$$

where Φ_ρ^d is the distribution function of X.

The Student's t Copula

The Student's t copula $C_{\nu,\rho}^t$ of a d-dimensional standard student's t distribution with $\nu \geq 0$ degrees of freedom and linear correlation matrix ρ, is the distribution of the random vector $(t_\nu(X_1), ..., t_\nu(X_d))$, where X has a $t^d(0, \rho, \nu)$ distribution and t_ν is the univariate standard student's t distribution function. Hence,

$$C_{\nu,\rho}^t = P(t_\nu(X_1) \leq u_1, ..., t_\nu(X_d) \leq u_d) = t_{\nu,\rho}^d(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)) \tag{51}$$

with $t_{\nu,\rho}^d$ the distribution function of X.

Archimedean Copulas

Luciano et al. [8] define a d-variate Archimedean copula as the function

$$C(u_1, ..., u_d) = \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_d)) \tag{52}$$

where $\varphi(u)$ is called the generator of the copula and is such that $\varphi(u) \in C^2$ function with $\varphi(1) = 0$, $\varphi\prime(u) < 0$ (φ is strictly decreasing) and $\varphi\prime\prime(u) > 0$ (φ is convex) for all $0 \leq u \leq 1$. The inverse of φ, φ^{-1} must be completely monotonic on $[0, \infty)$[8].

Clayton Copula

The generator is given by $\varphi(u) = u^{-\alpha} - 1$, hence $\varphi^{-1}(t) = (t+1)^{\frac{-1}{\alpha}}$, it is completely monotonic if $\alpha > 0$. The Clayton d-copula is therefore:

$$C(u_1, ..., u_d) = \Big[\sum_{i=1}^{d} u_i^{-\alpha} - d + 1\Big]^{\frac{-1}{\alpha}} \text{ with } \alpha > 0 \tag{53}$$

Gumbel Copula

The generator is given by $\varphi(u) = (-\ln(u)^{\alpha})^{\alpha}$, hence $\varphi^{-1}(t) = \exp(-t^{\frac{1}{\alpha}})$, it is completely monotonic if $\alpha > 1$. The Gumbel d-copula is therefore:

$$C(u_1, ..., u_d) = \exp\{-\Big[\sum_{i=1}^{d}(-\ln(u_i)^{\alpha})\Big]^{\frac{1}{\alpha}}\} \text{ with } \alpha > 1 \tag{54}$$

Frank Copula

The generator is given by

$$\varphi(u) = \ln\Big(\frac{\exp(-\alpha u) - 1}{\exp(-\alpha) - 1}\Big)$$

hence

$$\varphi^{-1}(t) = -\frac{1}{\alpha}\ln\Big(1 + \exp(t)(\exp(-\alpha) - 1)\Big)$$

it is completely monotonic if $\alpha > 0$. The Frank d-copula is therefore:

$$C(u_1, ..., u_d) = -\frac{1}{\alpha}\ln\{1 + \frac{\prod_{i=1}^{d}(\exp(-\alpha u_i) - 1)}{(\exp(-\alpha) - 1)^{d-1}}\} \text{ with } \alpha > 0 \text{ when } n \geq 3 \tag{55}$$

2.4.3 Dependence Measures

In this section we explore ways in which copulas can be used in the study of dependence or association between random variables. The usual ways used to measure dependence are the ones fulfilling the scale invariance property. This means that, they remain unchanged under strictly increasing transformation of the random variables. The most widely known scale invariant measures of association are the Kendall's tau and Spearman's rho, both of which measure a form of dependence known as concordance [10]. Informally, a pair of random variables are concordant if large values of one tend to be associated with large values of the other and small values of one with small values of the other.

Definition (Concordance): Let (x_i, y_i) and (x_j, y_j) be two observations from a vector (X, Y) of continuous random variables. We say that (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. Similarly, we say that (x_i, y_i) and (x_j, y_j) are discordant if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$. Note the alternative formulation [10]: (x_i, y_i) and (x_j, y_j) are concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$.

Kendall's Tau

Let (X_1, Y_1) and (X_2, Y_2) be i.i.d. random vectors, each with joint distribution function H. Then the kendall's tau is defined as the probability of concordance minus the probability of discordance:

$$\tau = \tau_{X,Y} = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0) \tag{56}$$

Define a concordance function Q, which is the difference of the probabilities of concordance and discordance between the two vectors (X_1, Y_1) and (X_2, Y_2) of continuous random variables with (possibly) different joint distributions H_1 and H_2, but with common margins F and G. It is shown in [10] (Theorem 5.1.1 page 159) that Q depends on the (X_1, Y_1) and (X_2, Y_2) through their copulas

$$Q = Q(C_1, C_2) = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1 \tag{57}$$

where C_1 and C_2 are the copulas of (X_1, Y_1) and (X_2, Y_2) respectively so that $H_1(x, y) = C_1(F(x), G(y))$ and $H_2(x, y) = C_2(F(x), G(y))$.

Spearman's Rho

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be three independent random vectors with a common joint distribution function H (whose margins are again F and G) and copula C. Spearman's rho is defined to be the probability of concordance minus the probability of discordance for the two vectors (X_1, Y_1), (X_2, Y_3), i.e., a pair of vectors with the same margins but one vector has distribution function H, while the components of the other are independent.

$$\rho_{X,Y} = 3(P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0)) \tag{58}$$

(the pair (X_3, Y_3) could be used equally as well). It is shown in [10] (Theorem 5.1.6 page 167), that with continuous random variables X and Y whose copula is C that

$$\rho_{X,Y} = \rho_C = 12 \iint_{I_2} C(u, v) du dv - 3 \tag{59}$$

2.4.4 Estimation of Copula Parameters

In this section we briefly describes approaches for estimating copula parameters.

Exact Maximum Likelihood Estimation Method(MLE)

Recall the canonical copula representation

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \prod_{j=1}^{d} f_j(x_j)$$

Let $\{x_{1t}, ..., x_{nt}\}_{t=1}^{T}$ be the sample data matrix. The log-likelihood function is given by:

$$l(\theta) = \sum_{t=1}^{T} \ln c(F_1(x_{1t}), ..., F_n(x_{nt})) + \sum_{t=1}^{T} \sum_{j=1}^{n} \ln f_j(x_{jt}) \tag{60}$$

where θ is the set of all parameters of both the marginals and the copula. Hence, given a set of marginal probability density functions and a copula the previous log-likelihood may be written, and by maximization we obtain the maximum likelihood estimator:

$$\widehat{\theta}_{MLE} = \max_{\theta} l(\theta) \tag{61}$$

A drawback of the exact MLE is that it could be computationally intensive in the case of high dimension because it estimates jointly the parameters of the margins and the parameters of the copula.

Inference for Margins (IFM)

To overcome the exact MLE drawback, IFM estimates the parameters in the log-likelihood function in two steps:

1. Estimate the margins' parameters θ_1 by performing the estimation of the univariate marginal distributions

$$\widehat{\theta}_1 = \text{argmax}_{\theta_1} \sum_{t=1}^{T} \sum_{j=1}^{n} \ln f_j(x_{jt}; \theta_1) \tag{62}$$

2. Given $\widehat{\theta}_1$, perform the estimation of the copula parameter θ_2

$$\widehat{\theta}_2 = \text{argmax}_{\theta_2} \sum_{t=1}^{T} \ln c(F_1(x_{1t}), ..., F_n(x_{nt}); \theta_2, \widehat{\theta}_1) \tag{63}$$

The IFM estimator is defined as $\widehat{\theta}_{IFM}(\widehat{\theta}_1, \widehat{\theta}_2)$.

Canonical Maximum Likelihood (CML)

The idea behind the CML is that, the copula parameters can be estimated without specifying the marginal distributions. The CML consists in transforming the sample data $\{x_{1t}, ..., x_{nt}\}_{t=1}^{T}$ into uniform variates $\{u_{1t}, ..., u_{nt}\}_{t=1}^{T}$ and then estimating the copula parameters as follows:

1. First estimate the margins using the empirical distributions (without assumptions on the parametric form for each of them); i.e., $\widehat{F}_j(x_{jt})$ with $j = 1, ..., n$.

2. Estimate via MLE the copula parameters

$$\widehat{\theta}_2 = \text{argmax}_{\theta_2} \sum_{t=1}^{T} \ln c(\widehat{F}_1(x_{1t}), ..., \widehat{F}_n(x_{nt}); \theta_2) \tag{64}$$

2.4.5 Simulation from Copulas

One of the main applications of copula related to this thesis is the VaR estimation using Monte Carlo Simulation approach. In this section we describe a general method to simulate draws from a chosen copula using a conditional approach (Conditional Sampling). We first describe the simulation principle in a bivariate case then we extend it in the multivariate case. Assume a bivariate copula in which all of its parameters are known. Our task is to generate pairs (u, v) of observations of $(0, 1)$ uniformly distributed random variables U and V whose joint distribution is C. To do so, we use the conditional distribution

$$c_u(v) = P(V \leq v | U = u) \tag{65}$$

for the random variable V at a given value u of U. From probability theory, we know that,

$$c_u(v) = P(V \leq v | U = u) = \lim_{\triangle u \to 0} \frac{C(u + \triangle u, v) - C(u, v)}{\triangle u} = \frac{\partial C}{\partial u} = C_u(v) \tag{66}$$

where $C_u(v)$ is the partial derivative of the copula. It is shown that $c_u(v)$ is a non-decreasing function and exists for almost all $v \in (0, 1)$ [8]. Thus we can generate the random pair (u, v) in the following steps:

1. Generate two independent random variables u and t from $U(0, 1)$;

2. Set $v = C_u^{-1}(t)$, where C_u^{-1} is the inverse function of c_u;

3. The pair (u, v) is just the random numbers from the copula.

The idea is the same when extending the simulation to a multivariate case. The goal in multivariate case is to simulate $U_1, ..., U_d$ from the copula $C(u_1, ..., u_d)$. We do it in the following steps:

1. Generate $u_1 \sim U(0, 1)$

2. Set
$$G_2(U_2|U_1 = u_1) = P(U_2 \leq u_2|U_1 = u_1) = \frac{\partial C(u_1, u_2, 1..., 1)}{\partial u_1} \tag{67}$$
we put $u_2 = G_2^{-1}(u_2|u_1)$, where $u_2 \sim U(0, 1)$

3. In general,
$$\begin{aligned} G_k(U_k|u_1, ..., u_{k-1}) &= P(U_k \leq u_k|U_1 = u_1, ..., U_{k-1} = u_{k-1}) \\ &= \frac{\partial C(u_1, ..., u_k, 1, ..., 1)}{\partial u_1 ... \partial u_{k-1}} / \frac{\partial C(u_1, ..., u_{k-1}, 1, ..., 1)}{\partial u_1 ... \partial u_{k-1}} \end{aligned} \tag{68}$$
we put $u_k = G_k^{-1}(U_k|u_1, ..., u_{k-1})$ where $U_k \sim U(0, 1)$.

The conditional approach is very elegant but is may not be possible to calculate the inverse function analytically. In this case one has to do it numerically, and this procedure might be computationally intensive. In the case of Archimedean Copulas this method maybe rewritten using theorem 6.1 page 189 in [8], which gives that:
$$C_k(u_k|u_1, ..., u_{k-1}) = \frac{\varphi^{-1(k-1)}(\varphi(u_1) + ... + \varphi(u_k))}{\varphi^{-1(k-1)}(\varphi(u_1) + ... + \varphi(u_{k-1}))} \tag{69}$$
with $C(u_1, ..., u_d) = \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_d))$ is an Archimedean copula with generator $\varphi(u)$.

3 Methodology

As the title of the thesis states, in this thesis we are interested in estimating portfolio VaR using GARCH methods, EVT theory and Copula theory. We call the resulting procedure the GARCH-EVT-Copula approach. From Copula theory described in section (2) (more precisely the canonical copula representation), we can say that, in general, a statistical problem for copulas could be decomposed into two steps: the identification of marginal distributions and the definition of an appropriate copula function. We refer to the first step as modeling the marginal distributions and the second as modeling the dependence structure. In this section we describe these two steps.

For this Thesis we choose to use an hypothetical Portfolio consisting of indices from Germany, Spain, Italy and France. The data consists of 3961 daily closing prices of DAX, IBEX 35, FTSE MIB and CAC 40 downloaded from Yahoo Finance. The Figure 1 below illustrates the relative price movements of each Index. (Initial level of each Index have been normalized to unity to facilitate the comparison of their relative performances).

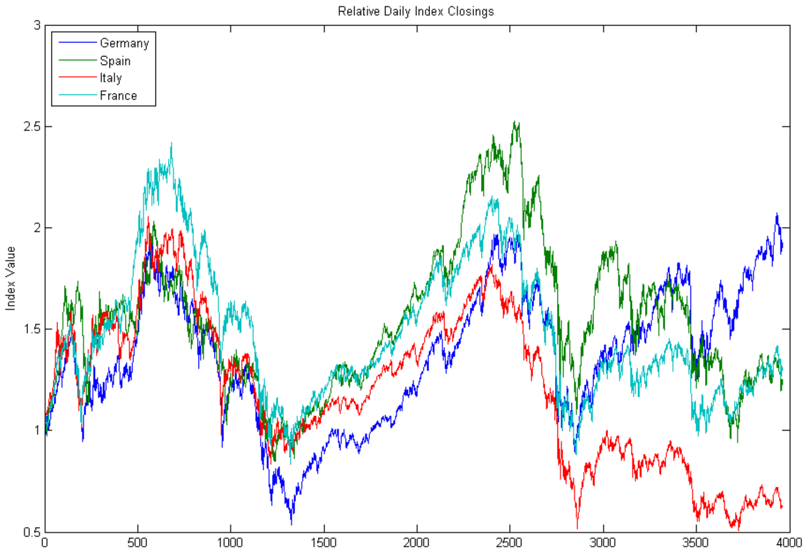

Figure 1: Relative Price movements of each Index.

Since stock prices are mostly non-stationary, it is common in time series to model related changes of prices, that is the log return series. The log returns of the indices are defined as:

$$r_{i,j} = \ln\left(\frac{P_{i,j}}{P_{i,j-1}}\right), i = 1, \ldots 4. \tag{70}$$

Where $P_{i,j}$ is the i^{th} index price at time j; $i = 1, \ldots, 4$ corresponding to stock index from Germany, Spain, Italy and France respectively. Figure 2 shows the log returns for the four indices.

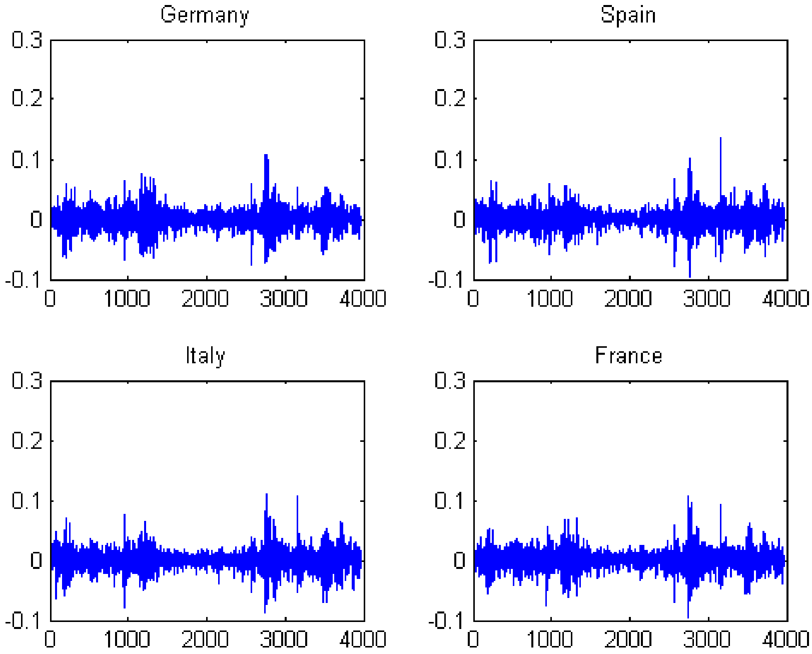

Figure 2: Plot of the Indices Log Returns.

Table 1: Summary Statistics of daily Log returns.

Statistics	Germany	Spain	Italy	France
Mean	0.0001623	0.0000522	-0.0001182	0.0000653
Std	0.016	0.0158	0.0158	0.0154
Skewness	-0.0289	0.0401	-0.0663	-0.0034
Kurtosis	6.6839	7.4636	7.0859	7.2419
Jarque-Bera	1	1	1	1
Ljung-Box(Returns)	1	1	1	1
Ljung-Box(Squared Returns)	1	1	1	1
Ljung-Box(Residuals)	0	0	0	0
Ljung-Box(Squared Residuals)	1	0	1	0
Engle's ARCH(Returns)	1	1	1	1
Engle's ARCH(Squared Returns)	1	1	1	1
Engle's ARCH(Residuals)	1	0	1	0
Engle's ARCH(Squared Residuals)	0	0	0	0

Before proceeding to the margins and dependence structure modeling, we first verify some typical assumed properties of log returns, namely normality and i.i.d. properties.

To check if the indices are normally distributed, one can use QQ-Plots which are plots of the empirical quantiles against the quantiles of a reference distribution (the Normal distribution in this case). If the QQ-plot looks linear then we say that the suggested distribution gives a good fit; otherwise one should reconsider the choice of the parametric family. From Figure 3 one can see that the tails are not well modeled by the normal distribution.

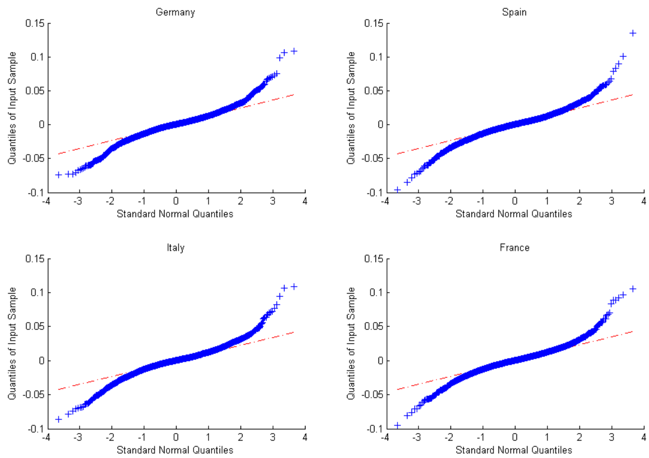

Figure 3: QQ Plots of the Returns.

To confirm the observations from the QQ-plots, we perform a Jarque Bera test statistic [15], of the null hypothesis that the sample come from a normal distribution. Table 1 shows the results with values 1 meaning that the test rejects the null hypothesis at a 5% significance level. Moreover, from the Skewness and Kurtosis in Table 1 we see that the series are nearly symmetric (or slightly left skewed), and fat tailed (Kurtosis > 3).

To check if the log returns are independently and identically distributed, first we look at the plots in Figure 2. From an eyeball inspection it is unclear whether the observed log returns can be seen as outcomes of independently and identically distributed random variables. What we can see is that all returns have clustered volatilities (some periods have higher volatility than others). For instance we observe a high volatility in the period between 2008 and 2010 (Points between 2500 and 3000 in Figure 3) representing the time of the financial crisis.

Plots of the Sample ACF of the log returns in Figure 4 show virtually no significant correlation. Performing a Portmanteau test of Ljung and Box [15] to check whether the autocorrelations with different lags are zero, the test reject the null hypothesis , that log returns exhibits no autocorrelation at 40 lags with a 5% significance level. This suggests

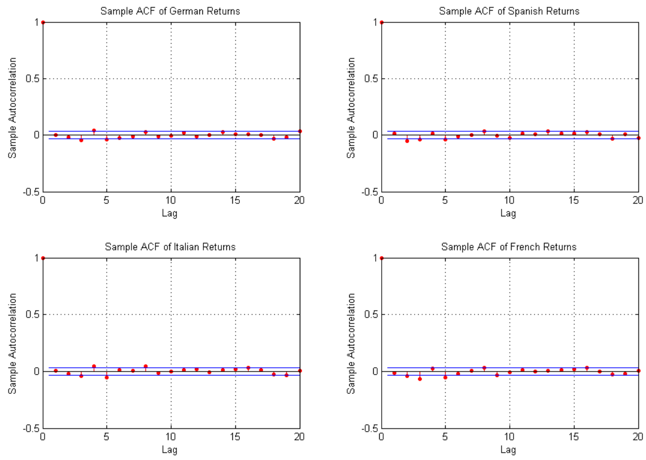

Figure 4: Sample ACF of the Log Returns

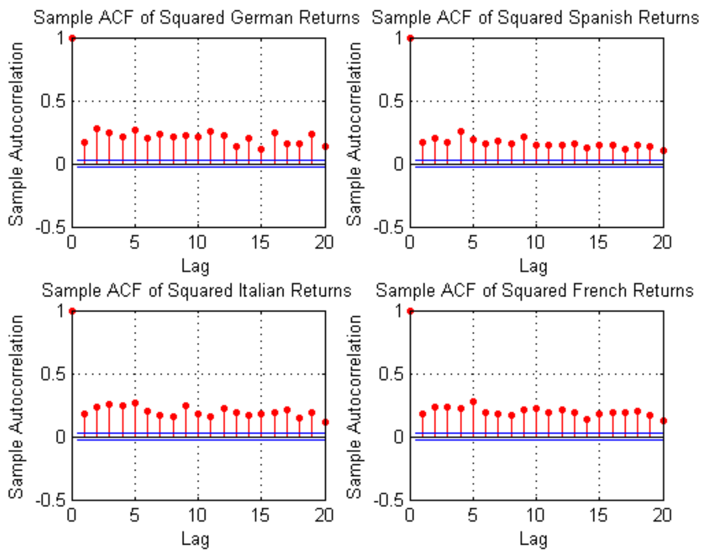

Figure 5: Sample ACF of the Squared Log Returns.

that a conditional mean model is needed for these log return series. To check the log return series for conditional heteroscedasticity, we plot the sample ACF of the squared log returns series. From Figure 5 we can see significant autocorrelation. This suggests that a GARCH model with lagged variances and lagged squared innovations might be appropriate for modeling these log returns series. Engle's ARCH test rejects the null hypothesis (Values 1 in Table 1) of no ARCH effects in favor of the alternative ARCH model with two lagged squared innovations.

3.1 Modeling Marginal Distributions

We have seen that the log returns series are not independently and identically distributed. We solve this by fitting a first order autoregressive model($AR(1)$) to the conditional mean of the log returns of each index and an asymmetric GARCH ($GJR(1,1)$) model to the conditional variance. The first order autoregressive model compensates for autocorrelation and the GJR model for heteroscedasticity. Additionally, the standardized residuals of each index are modeled as standardized Student's t distribution to compensate for the fat tails of the log return series.The $AR(1)$ and $GJR(1,1)$ choice is motivated by McNeil and Frey (2000) [11] and Huang et al. [18].

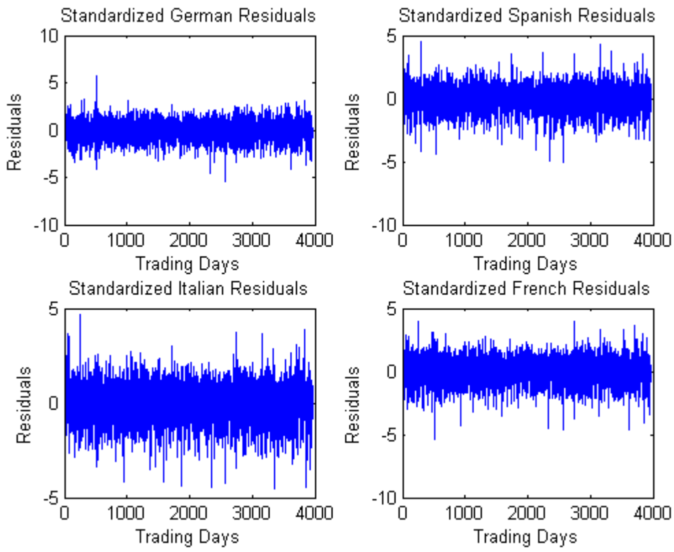

Figure 6: Standardized Residuals.

Figure 6 shows the standardized residuals ($AR(1) - GJR(1,1)$). We examine the sample ACFs of the standardized residuals and the squared standardized residuals in Figure 7 and 8 respectively; comparing these figures to Figure 4 and 5 we see that the standardized residuals are now approximately independent and identically distributed.

The Ljung-Box test statistic and the Engle's ARCH test (Table 1) confirms the analysis above (with an exception of German and Italian standardized squared residuals which fails the tests at 5%). We conclude that the AR(1)-GJR(1,1) model sufficiently explains the autocorrelation and heteroscedasticity effects in each log return series and leads to standardized residuals which represent the underlying zero mean and unit variance independently and identically distributed series . From the standardized residuals series obtained we can now model the marginal distributions. We choose to use a semi para-

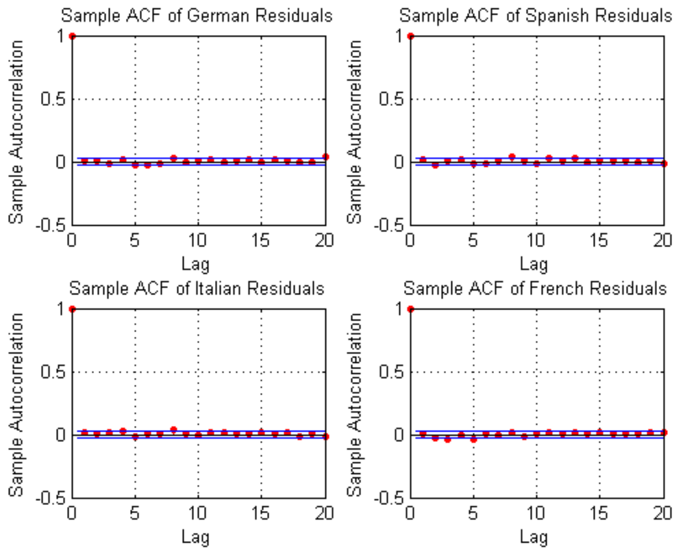

Figure 7: Sample ACF of the Standardized Residuals.

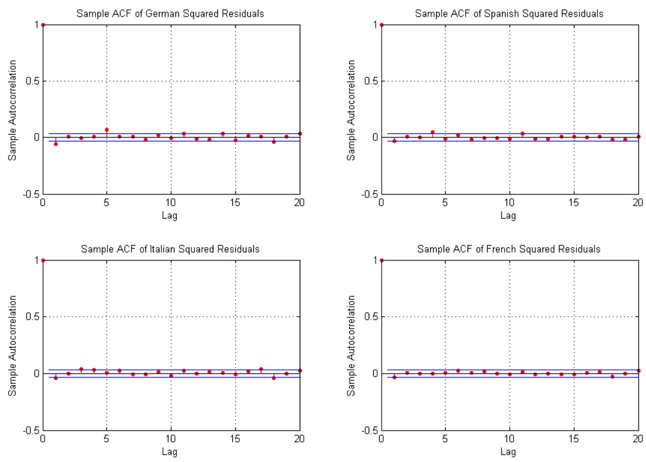

Figure 8: Sample ACF of the Squared Standardized Residuals.

metric distribution which consists of using a Kernel Density estimation method (with a normal density as kernel function) for the interior of the distribution, and a GPD in the tails as proposed by Carol (2008) [5].

Briefly, the univariate Kernel Density Estimation method can be described as follows: Assume we have an i.i.d. random sample $X_1, ..., X_n$ from a continuous density function

f. Denote by $\widehat{f}(x;h)$ the kernel density estimator of f which is defined as:

$$\widehat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big) \tag{71}$$

The function K is called "kernel" and it satisfies $\int K(x)dx = 1$, and h is called "bandwidth" and is a positive number. The following rescaled version of the kernel is useful

$$K_h(u) = \frac{1}{h}K(\frac{u}{h}) \tag{72}$$

For example if the kernel is $N(0,1)$, then the scaled kernel is $N(0,h^2)$ as

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-u^2}{2}) \text{ and} \tag{73}$$

$$K_h(u) = \frac{1}{\sqrt{2\pi h}}\exp(\frac{-u^2}{2h^2}) \tag{74}$$

This gives the following formula for the kernel density estimator

$$\widehat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \tag{75}$$

McNeil and Frey (2000) and Nyström and Sloglund conducted Monte Carlo simulation experiments in order to evaluate the properties of the MLE in estimating the GPD parameters for various distributions and sample sizes. The results show that the MLE estimates are almost invariant to the threshold value u when this is set between 5% and 13% of the sample data. Carol (2008) argues that, provided that the historical sample is sufficiently large (at most 2000 observations) there will be enough log returns in the 10% tail to obtain a reasonably accurate estimate of the GPD scale and tail parameters. Thus, we estimate the marginal distributions using the kernel density estimator in the interior of the distribution and the POT method in the tails using 10% of the data points for each tail. Figure 9 shows the obtained semi parametric distributions for the four log return series.

3.2 Modeling Dependence Structure

To model the dependence between the returns of the indices we use Copulas. GARCH-EVT-Copula parameters are estimated using Canonical Maximum Likelihood estimation method described in Section 2. Table 2 shows the different Copula parameters obtained. To check the fit of these Copulas in value at risk estimation, backtesting methods are implemented in later sections.

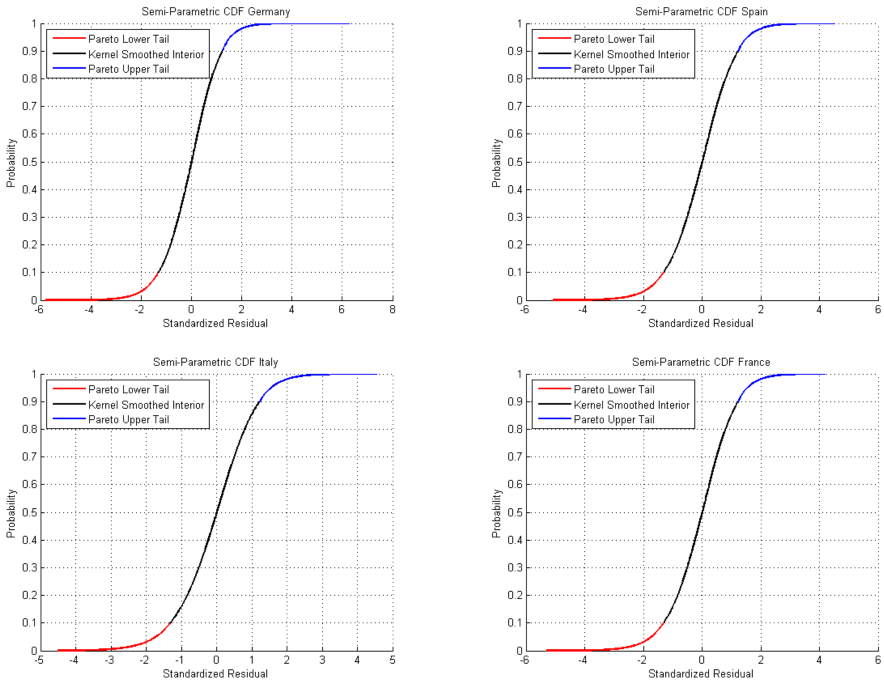

Figure 9: Semi Parametric CDFs.

Table 2: Copula Parameters.

Names	Gaussian	Student's t	Clayton	Gumbel	Frank
Parameters	ρ_G	ρ_t,28.4299	1.11805	1	5.9609E-5

Where ρ_G and ρ_t are given by: $\begin{pmatrix} 1 & 0.0048 & 0.0222 & 0.0298 \\ 0.0048 & 1 & 0.0469 & 0.1405 \\ 0.0222 & 0.0469 & 1 & 0.0724 \\ 0.0298 & 0.1405 & 0.0724 & 1 \end{pmatrix}$

$$\begin{pmatrix} 1 & 0.0032 & 0.0218 & 0.0294 \\ 0.0032 & 1 & 0.0445 & 0.1427 \\ 0.0218 & 0.0445 & 1 & 0.0689 \\ 0.0294 & 0.1427 & 0.0689 & 1 \end{pmatrix}$$

3.3 Backtesting Methodology

VaR models are useful only if they predict future risks accurately. In order to evaluate the quality of the estimates, the models should always be backtested with appropriate methods. Backtesting is a statistical procedure where actual losses are systematically compared to corresponding VaR estimates. We say that we have an exception (or a violation or a breach) if the VaR has been underestimated, that is the portfolio has experienced a loss greater than the estimated VaR. In the backtesting process we could statistically examine whether the frequency of exceptions over some specified time interval is in line with the selected confidence level. These types of tests are known as tests of *unconditional coverage*. They are straightforward tests to implement since they do not take into account for when the exception occur [7]. In theory, however, a good VaR model not only produces the correct amount of exceptions but also exceptions that are evenly spread over time, i.e., they are independent of each other. Clustering of exceptions indicates that the model does not accurately capture the changes in market volatility and correlations. Tests of *conditional coverage* therefore examine also time variation in the data [7]. In the following we will briefly describe the Kupiec's proportion of failures test (which is an unconditional coverage test), and the Christoffersen's tests (Christoffersen's independence test and Christoffersen's interval forecast test) which are conditional coverage tests.

3.3.1 Kupiec's Proportion of Failures (POF) Test

The most common test of VaR model is to count the number of exceptions and if the number of exceptions is less that the selected confidence level would indicate then we say that the VaR model overestimate risk. Otherwise we say we have an underestimation of risk. In practice it is rarely the case that we observe the same number of exceptions as suggested by the confidence level. It therefore comes down to statistical analysis to study whether the number of exceptions is reasonable or not. Let x be the total number of exceptions and T the total number of observations. We define the *failure rate* as $\frac{x}{T}$. In an ideal situation, this rate would reflect the selected confidence level. For instance, if a confidence level of 99% is used we have a null hypothesis that the frequency of tail losses is equal to $p = 1 - c = 1 - 0.99 = 1\%$. Assuming that the model is accurate, the observed failure rate should act as an unbiased measure of p, and thus converge to 1% as the sample size is increased [7]. Each trading outcome either produces an exception or not. This sequence of successes and failures is commonly known as Bernoulli trial. The number of exceptions x follows a binomial distribution

$$f(x) = \left[\begin{array}{c} T \\ x \end{array} \right] p^x (1-p)^{T-x}$$

The POF test measures whether the number of exceptions is consistent with the confidence level.

$$H_0 : \; p = \widehat{p} = \frac{x}{T}$$

The idea is to find out whether the observed failure rate $\widehat{p}$ is significantly different from p, the failure rate suggested by the confidence level. According to Kupiec, the POF test is best conducted as a likelihood ratio (LR) test. The test statistic takes the form

$$LR_{POF} = -2\ln\Big(\frac{(1-p)^{T-x}p^x}{[1-(\frac{x}{T})]^{T-x}(\frac{x}{T})^x}\Big) \tag{76}$$

Under the null hypothesis that the model is correct, LR_{POF} is asymptotically χ^2 distributed with 1 degree of freedom. If LR_{POF} is greater than the critical value of the $\chi^2(1)$ then the null hypothesis is rejected and the model is said to be inaccurate. An important drawback of the POF test is that it considers only the frequency of exceptions and not the time when they occur. As a result, it may fail to reject a model that produces clustered exceptions.

3.3.2 Christoffersen's Tests

Christoffersen uses the same likelihood testing framework as Kupiec, but extends the test to include also a separate statistic for independence of exceptions. In addition to the correct rate of coverage, his test examines whether the probability of an exception on any day depends on the outcome of the previous day. The testing procedure described below is explained in [7] and [13]. The test is carried out by first defining an indicator variable that gets value 1 if VaR is exceeded and 0 if not.

$$I_t = \begin{cases} 1 & \text{if violation occurs} \\ 0 & \text{if no violation occurs} \end{cases}$$

Then define n_{ij} the number of days when condition j occurred assuming that condition i occurred on the previous day. The outcome can be displayed in a contingency table:

-	$I_{t-1}=0$	$I_{t-1}=1$	-
$I_t=0$	n_{00}	n_{10}	$n_{00}+n_{10}$
$I_t=1$	n_{01}	n_{11}	$n_{01}+n_{11}$
-	$n_{00}+n_{01}$	$n_{10}+n_{11}$	$n_{00}+n_{01}+n_{10}+n_{11}$

In addition, let π_i be the probability of observing an exception on state i on the previous day.

$$\pi_0 = \frac{n_{01}}{n_{00}+n_{01}}$$
$$\pi_1 = \frac{n_{11}}{n_{10}+n_{11}}$$
$$\pi = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}} \tag{77}$$

If the model is accurate, then an exception today should not depend on whether or not an exception occurred yesterday. In other words, under the null hypothesis the probabilities π_0 and π_1 should be equal. The relevant test statistic for independence of exception is a likelihood ratio:

$$LR_{IND} = -2\ln\Big(\frac{(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\Big) \tag{78}$$

By combining this independence statistic test with Kupiec's POF test we obtain a joint test that examines both properties of a good VaR model, the correct failure rate and the independence of exceptions or conditional coverage:

$$LR_{CC} = LR_{POF} + LR_{IND} \tag{79}$$

$LR_{CC} \sim \chi^2$ with 2 degrees of freedom since there are 2 separate LR statistics in the test. If the LR_{CC} statistic is less than the critical value of the $\chi^2(2)$ then the model passes the test. Note that we can calculate the p-value associated with our test statistic by, $P_{\text{value}} = 1 - F_{\chi_1^2(LR)}$ where $F_{\chi_1^2(\ldots)}$ denotes the cumulative density function of a $\chi^2(1)$. If the P-value is below the desired significance level, then we reject the null hypothesis.

4 Empirical Results

4.1 Value at Risk Estimation

Including modeling marginal distributions and the dependence structure separately, the GARCH-EVT-Copula approach for a portfolio value at risk estimation can be described as follows:

Step1 Having index prices, transform them to log returns (r_i) with $i = 1, ..., 4$ representing Germany, Spain, Italy and France respectively). Then filter the returns by fitting an AR(1)-GJR(1,1) model to the returns to get the residuals and standardized them by the corresponding standard deviations to get the standardized residuals.

Step2 Model the marginal distributions F_i. From the standardized residuals, estimate the empirical CDF with a Gaussian kernel method for the interior of the distribution and POT methods for the tails.

Step3 Transform the standardized residuals of each asset i from step (1) to the Copula scale by taking F_i(residuals).

Step4 Fit a Copula to the transformed margins from Step 3 and estimate the Copula parameters .

Step5 Use the estimated Copula parameters to generate N (10000 in our case) random numbers from the estimated joint probability distribution. As we have ($i = 1, ..., 4$) indices we form a matrix $U = [U_1, ..., U_4]$ where each U_i in an $N \times 1$ vector of marginal simulated numbers.

Step6 Transform U_i to the original scales of the log returns using the inverse quantile function of the marginals, i.e $F_i^{-1}(U_i)$.

Step7 Reintroduce the autocorrelations and heteroscedasticity observed in the original returns using again the AR(1)-GJR(1,1) model to get R_i which corresponds to the simulated returns for each corresponding marginal distribution.

Step8 Since $r_{i,j} = \ln(\frac{P_{i,j}}{P_{i,j-1}})$, we have that $P_{i,j} = P_{i,j-1}\mathrm{e}^{r_{i,j}}$ where $P_{i,j}$ is the i^{th} stock index price at time j. We choose in our portfolio that indices have equal weights ($W_i = 1/4$). The portfolio value at time t is defined as: $V_t = \sum_{i=1}^{4} W_i P_{i,t}$. From this, the portfolio value at time $t+1$ is given by: $V_{t+1,j} = \sum_{i=1}^{4} W_i P_{i,t}\mathrm{e}^{R_{i,j}}$. We define the portfolio discounted loss at time t as $L_{t,j} = V_t - V_{t+1,j}$. The distribution of the series $\{L_{t,j}\}_{j=1}^{N}$ is the distribution of the discounted loss function of the portfolio between time $[t, t+1]$.

Step9 The one day value at risk at time t with confidence level α (99%, 95% and 90%), $VaR_t(\alpha)$ is just the $1-\alpha$ quantile of the distribution of the discounted loss series $\{L_{t,j}\}_{j=1}^{N}$.

We have in Table 3 estimated VaR on July 3^{rd} 2013 from the different GARCH-EVT-Copulas, the Historical Simulation (HS) and the Variance Covariance (VC) methods.

Table 3: One day VaR Estimates.

-	Gaussian	Student's t	Clayton	Gumbel	Frank	HS	VC
VaR(99%)	181.7708	184.3487	313.9214	183.4489	183.3647	228.9704	167.9644
VaR(95%)	125.4743	124.4993	193.7815	120.7332	120.6514	128.9491	118.8915
VaR(90%)	95.0177	95.5305	141.6696	93.5284	93.4345	89.1899	92.7049

4.2 Value at Risk Backtesting

In this Section, backtesting procedures are conducted by comparing daily losses with daily VaR estimates. We use the Kupiec's Proportion of Failures Test, the Christoffersen's Independence Test and the Christoffersen's Interval forecast Test.

Let $L_{[t,t+1]}$ denote the portfolio loss over a one day interval. The corresponding VaR estimate VaR_t is calculated at the beginning of the period, i.e using closing prices of day t, and is based on historical data. We fix an estimation period consisting of 2000 observations which defines the sample used to estimate the VaR model parameters. Then we use a rolling window approach as follows. The estimation sample is rolled over almost the entire data period, keeping the estimation period constant, starting at the beginning of the data set. We fix the length of the risk horizon to one day, and the test sample starts at the end of the estimation period. The result of this procedure is two time series covering the sample from 2001^{th} to 3960^{th} observation, i.e covering all the consecutive rolling test periods. One series is the one day VaR estimates and the other is the one day realized portfolio losses (we do this for each VaR estimation method). The backtest is based on these two series. Figure 10, 11 and 12 show these two series for the different methods used and the different confidence levels.

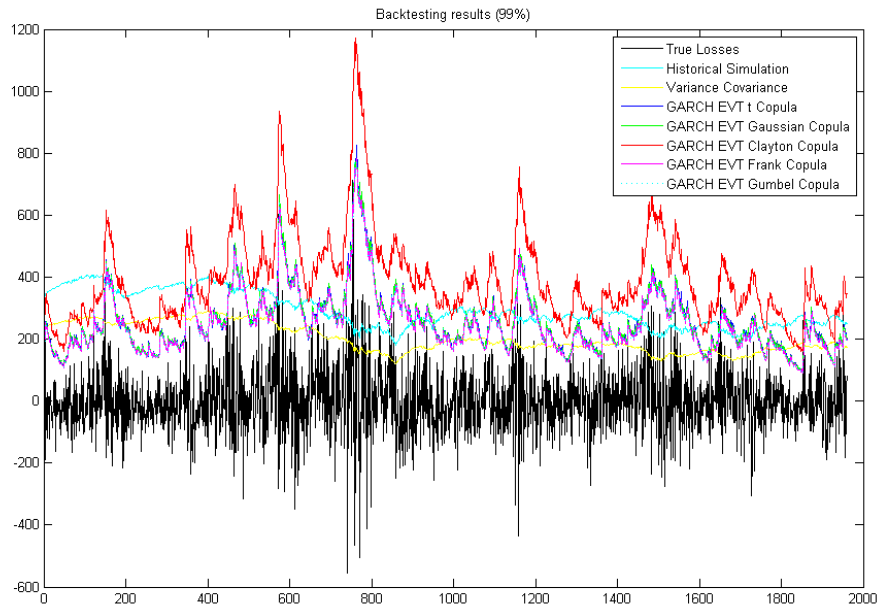

Figure 10: Backtesting Results (99%).

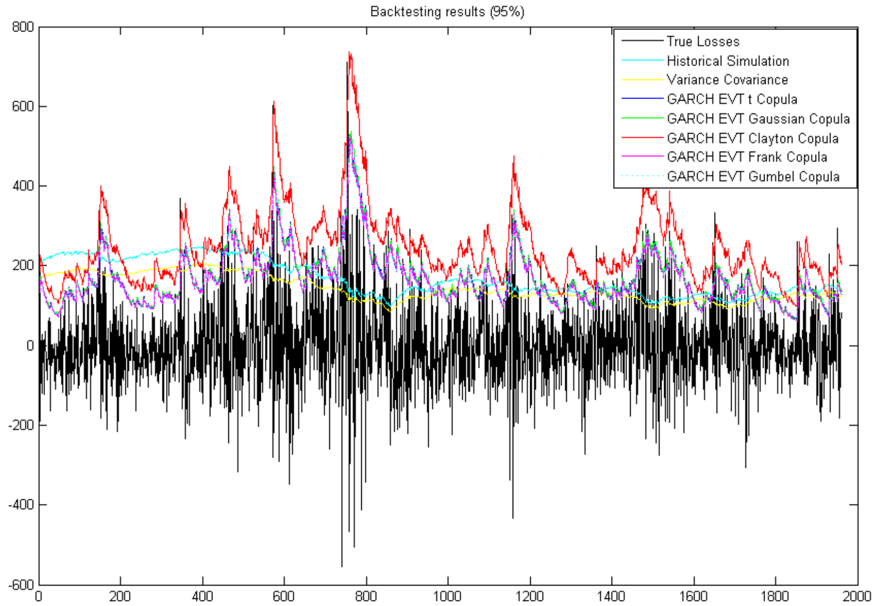

Figure 11: Backtesting Results (95%).

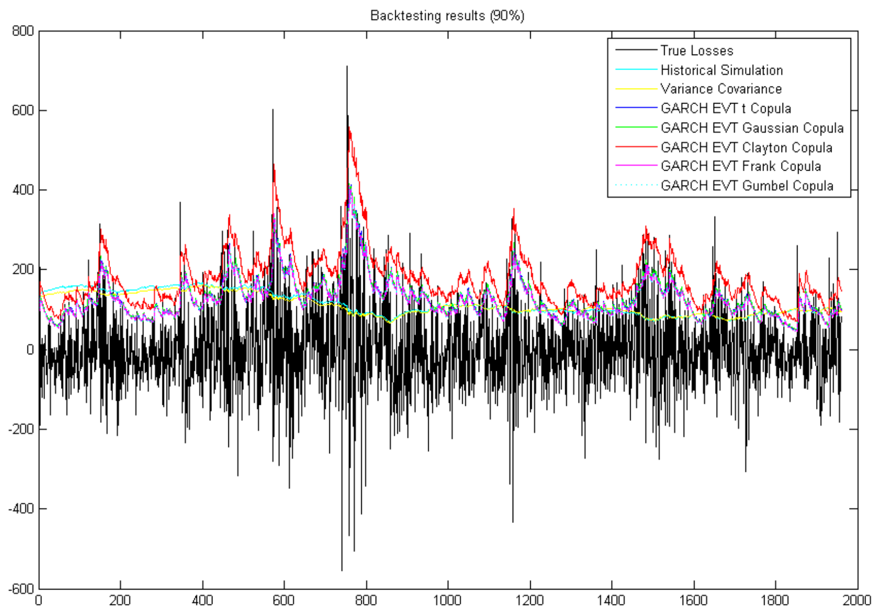

Figure 12: Backtesting Results (90%).

Table 4: Backtesting Tests Results.

HS	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	34	148	262	-
LR_{UC} p value	0.0030	0.00000129	0	(R R R)
LR_{IND} p value	0.6204	0.000616	0	(A R R)
LR_{CC} p value	0.011	0.000000023	0	(A R R)
VC	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	84	200	267	-
LR_{UC} p value	0	0	0	(R R R)
LR_{IND} p value	0.2245	0.0001677	0	(A R R)
LR_{CC} p value	0	0	0	(R R R)
Gaussian Copula	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	18	98	206	-
LR_{UC} p value	0.7143	0.9959	0.4503	(A A A)
LR_{IND} p value	0.5634	0.3329	0.6868	(A A A)
LR_{CC} p value	0.7914	0.6258	0.6933	(A A A)
Student's t Copula	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	18	98	202	-
LR_{UC} p value	0.7143	0.9959	0.6474	(A A A)
LR_{IND} p value	0.5634	0.3329	0.8387	(A A A)
LR_{CC} p value	0.7914	0.6258	0.8822	(A A A)
Clayton Copula	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	5	19	86	-
LR_{UC} p value	0.00007687	0	0	(R R R)
LR_{IND} p value	0.8729	0.5418	0.2971	(A A A)
LR_{CC} p value	0.0003978	0	0	(R R R)
Gumbel Copula	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	20	115	217	-
LR_{UC} p value	0.9261	0.0850	0.1176	(A R A)
LR_{IND} p value	0.5207	0.7548	0.9932	(A A A)
LR_{CC} p value	0.8101	0.2162	0.2939	(A A A)
Frank Copula	VaR_{99}	VaR_{95}	VaR_{90}	Tests results
Number of Exceptions	20	115	218	-
LR_{UC} p value	0.9261	0.0850	0.1015	(A R A)
LR_{IND} p value	0.5207	0.7548	0.7721	(A A A)
LR_{CC} p value	0.8101	0.2162	0.2508	(A A A)

Empirical backtesting results based on the different GARCH-EVT-Copulas, the Historical Simulation (HS) and the Variance Covariance (VC) methods for our portfolio are summarized in Table 4.

LR_{UC}, LR_{IND} and LR_{CC} represents the Kupiec's Proportion of Failure test, the Christoffersen's Independence test and the Christoffersen's Interval Forecast test respectively. The notation (A R A) in Table 4 means that the first test accept the null hypothesis that the model used for VaR estimation performs well on average, the second test rejects the null hypothesis and the last accept the null hypothesis. Note that, from Table 4 most of the number of exceptions are close to the expected exceptions (20 for the 99% confidence level, 98 for the 95% confidence level and 196 for the 90% confidence level). If the obtained number of exceptions are much larger than the expected ones, the models are said to have a poor performance in predicting the VaR. Otherwise (when they are much less) we say that the models fails to capture the information of historical observations (GARCH-EVT-Clayton Copula in our case). The results in the last Column of Table 4 are obtained by considering a significance level of 10%, and check if a test statistics' p value is less than the significance level to reject the null hypothesis, that the model is not accurate on average. From the results we conclude that in general the GARCH-EVT-Copula approaches outperforms the commonly used Variance Covariance method and the Historical Simulation method with the exception of the GARCH-EVT-Clayton Copula case which passes only the Christoffersen's Independence test.

5 Summary and Conclusions

Accurate VaR estimates for portfolios are essential for financial institutions' risk management teams and for regulators like the Basel Committee on Banking Supervision. In the past years, traditional VaR estimation approaches like Historical Simulation and Variance Covariance have been supplemented by more flexible approaches based on Copula functions. In this thesis we used a GARCH-EVT-Copula approach to estimate the VaR of a portfolio consisting of stock indices from Germany, Spain, Italy and France.

We started the estimation by first, performing a preliminary analysis on the four stock index log return series to verify some typical assumed properties of log returns, namely normality and i.i.d. properties. We found that the series were heavy-tailed and leptokurtic and moreover they were not i.i.d. To solve this we fitted an AR(1)-GJR(1,1) model to the series and divided the obtained residuals by their corresponding volatilities to get standardized residuals which were approximately i.i.d. The second step was of estimating the marginal distribution of each series. We achieved this by combining the Kernel Density Estimation method and the Peak Over Threshold (POT) method. That is, having the standardized i.i.d. residuals we estimated the interior of the empirical cumulative density functions (CDF) using the Kernel Density method and the POT method in the tails. We thus obtained semi-parametric empirical CDF for each series. The third step was the modeling of dependence structure using Copulas (Gaussian, Student's t, Clayton, Gumbel and Frank) by the Canonical Maximum Likelihood method. It consists of, first transform the standardized residuals to uniform variates by the semi-parametric empirical CDF and then fit the different Copulas to the transformed data to obtain Copula parameters. The last step of the approach was the portfolio VaR estimation using Monte Carlo Simulation method. Following this method, we generated 10000 observations using the different estimated Copula parameters and the VaR estimates were inferred. To check the performance of the approach we used Backtesting procedures.

Using Kupiec's proportion of failures test, Christoffersen's independence test and Christoffersen's interval forecast test, we found that GARCH-EVT-Gaussian Copula and GARCH-EVT-Student's t Copula passed all the test with the former being better as at the 90% level it had fewer number of exceptions. GARCH-EVT-Gumbel Copula and GARCH-EVT-Frank Copula performed quite well passing all the tests except at the 95% level where they didn't pass the Kupiec's test. GARCH-EVT-Clayton Copula passed only the Christoffersen's independence test. In general, the GARCH-EVT-Copula approach outperformed the Historical Simulation and Variance Covariance approaches. Eventhough the GARCH-EVT-Copula approach performs well in general, further works could be done to improve the approach. One could check the performance of the approach in high dimension, and also, one could use time-varying copulas instead of static copula for modelling the dependence structure.

References

[1] Risk and Portfolio Analysis. Principles and Methods. Henrik Hult, Filip Lindskog, Ola Hammarlid and Carl Johan Rehn (2012). Springer.

[2] Basel Committee on Banking Supervision (2011).

[3] On the Relation Between the Expected value and the Volatility of the nominal excess return on stocks. Lawrence R. Glosten, Ravi Jagannathan and David E. Runkle. The Journal of Finance, VOL XLVIII No. 5 (1993)

[4] Quantitative Risk management: Concepts, Techniques and Tools. Alexander J. McNeil, Rudiger Frey and Paul Embrechts. Princeton Series in Finance (2005)

[5] Market Risk Analysis (Volume IV). Value at Risk Models. Carol Alexander. John Wiley and Sons(2008).

[6] Beyond value at Risk. The new Science of risk management. Kevin Dowd. John Wiley and Sons (1998).

[7] Value at Risk. The new benchmark for managing financial risk. (Third edition). Phillipe Jorion. The McGrawHill Companies (2007).

[8] Copula Methods in Finance. Umberto Cherubini, Elisa Luciano and Walter Vecchiato. John Wiley and Sons (2004).

[9] Estimation of Multivariate models for Time Series of possibly different lengths. Patton, A. J. Journal of applied Econometrics, Vol 21, pp 147-173, (2006).

[10] An Introduction to Copulas (Second Edition). Roger B. Nelsen. Springer Series in Statistics (2006).

[11] Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Alexander J. McNeil and Rudiger Frey. Journal of Empirical finance, Vol 7 pp 271-300 (2001).

[12] Univariate EVT, GARCH and measure of Risk. Kaj Nyström and Jimmy Skoglund. Swedbank, Group Financial Risk Control (2002).

[13] Evaluating Interval Forecast. Peter F. Christoffersen. International Economic review, Vol 39, No. 4 pp 841-862.

[14] Effectiveness of Copula-EVT in estimating VaR: Empirical evidence from Asian emerging markets. Chun-Pin Hsu, Chin-Wen Huang, Wan-Jiun and Paul Chiou. Review of Quantitative Finance. Springer (2011).

[15] Introduction to Time Series and Forecasting (Second Edition). Peter J. Brockwell and Richard A. Davis. Springer Text in Statistics (2002).

[16] Modeling Extreme Value Dependence in international stock markets. Ser-Huang Poon, Michael Rockinger and Jonathan Tawn. Statistica Sinica 13 929-953(2003).

[17] Measure of Financial risk using Conditional extreme value Copulas with EVT margins. Ahmed Ghorbel and Abdelwahed Trabelsi. The Journal of Risk Vol.11, No. 4, 51-85 (2009).

[18] Applying GARCH-EVT-Copula Models for Portfolio Value at risk on G7 Currency Markets. Shian-Chang Huang, Yi-Hsin Chien and Ruei-Ci Wang. International research Journal of Finance and Economics, Issue 74(2011).

[19] Introductory Econometrics for Finance (Second Edition). Chris Brooks. Cambridge University Press (2008).