CHAPTER 2

Theory and Literature Review

The idea and theory applied in this thesis consist of the return on asset, unit root test, risk management, portfolio optimization, ARMA with GARCH model, Extreme Value Theory, and Copula model. These theories have different property, advantage, and disadvantage which can be explained as follows:

2.1 Theory

2.1.1 The return on asset

In this section, return is used to study because it is able to assess the investment efficiency or to compare a number of different investments efficiency. Campbell et al. (1997) gave a reason in using returns for investors, which a return has a completeness and the investment opportunity has scale-free summary.

From the definition of Campbell et al., let P_t be an asset price at time t. Assume at the moment that the asset pays no dividends. R_t is one-period simple net return that hold the asset for one period from $t-1$ to t, and would result in a simple return as follows:

$$1+R_t = \frac{P_t}{P_{t-1}} \quad \text{or} \qquad (1)$$
$$P_t = P_{t-1}(1+R_t)$$

The simple net return of one-period can be defined as:

The arithmetic return is $$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad (2)$$

The logarithmic return is $$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \qquad (3)$$

This thesis uses logarithmic return because this can be useful for the statistical hypothesis, and this value is close to the raw return.

2.1.2 Unit root and stationarity test

The unit root test is Augmented Dickey-Fuller and Phillips-Perron and the stationarity test is Kwiatkowski, Phillips, Schmidt and Shin. For the properties of these tests, the unit root tests are used to investigate data trend, the data should be regressed on time deterministic functions to display the data as first differenced, and the stationarity tests are aimed to complement unit root tests which test both the the unit root and stationarity hypothesis. We can separate the series that appears to be stationary, the series that appears have a unit root, and the series being not enough informative to be sure that they are stationary. The theory of unit root and stationarity test can be explained as follows:

1) Dickey-Fuller test (DF test)

The Dickey-Fuller test was summarized by Dickey and Fuller (1979) which this test is the testing by using an autoregressive model. DF test can be defined as follows:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{4}$$

where y_t and y_{t-1} are variable time series at time t and $t-1$, ρ is coefficient of autocorrelation, and ε_t is random error. If assumed that $\rho = 1$, the model would be non-stationary which we can define the hypotheses in this case as:

$$H_0 : \rho = 1$$
$$H_a : |\rho| < 1; -1 < \rho < 1$$

If H_0 is accepted, this means that y_t has unit root and is non-stationary. Equation (4) can be adjusted subtracting y_{t-1} on both sides:

$$\begin{aligned} y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \varepsilon_t \\ &= (\rho - 1) y_{t-1} + \varepsilon \end{aligned} \tag{5}$$

and we can rewrite equation (5) as:

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \tag{6}$$

where $\delta = \rho - 1$. If $\delta = 0$, then $\rho = 1$ and H_0 is accepted. In contrast, if $\delta = 1$ then $\rho = 0$ and H_a is accepted that they are stationary. Then the hypotheses can be written as:

$$H_0 : \delta = 0$$
$$H_a : \delta < 0$$

If $\delta = 0$ or $\rho = 1$, from equation (4); $\Delta y_t = \delta y_{t-1} + \varepsilon_t$ we can get:

$$\Delta y_t = \varepsilon_t \tag{7}$$

For the above steps, there are three main aspects of the test as:

1. Test for a random walk:

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \tag{8}$$

2. Test for a random walk with drift:

$$\Delta y_t = \alpha_0 + \delta y_{t-1} + \varepsilon_t \tag{9}$$

3. Test for a random walk with drift and deterministic time trend:

$$\Delta y_t = \alpha_0 + \alpha_1 t + \delta y_{t-1} + \varepsilon_t \tag{10}$$

where $\alpha_0, \alpha_1, \delta$ are the parameters and t is the trend.

The unit root tests discussed above studied in time series characteristics using AR(1) with white noise. However, there is a complicated structure in many time series of finance. Said and Dickey (1984) augmented the autoregressive unit root test to facility general ARMA(*p*, *q*) models which their test was improved to be augmented Dickey-Fuller (ADF). Therefore, the ADF test is used in this study which will be conducted by adding the lagged the dependent variable values Δy_t in three equations.

There are three Augmented Dickey-Fuller tests consisting of:

1. Test for a random walk:

$$\Delta y_t = \delta y_{t-1} + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t \tag{11}$$

2. Test for a random walk with drift:

$$\Delta y_t = \alpha_0 + \delta y_{t-1} + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t \qquad (12)$$

3. Test for a random walk with drift and deterministic time trend:

$$\Delta y_t = \alpha_0 + \alpha_1 t + \delta y_{t-1} + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t \qquad (13)$$

In equations (11), (12), and (13), terms are added that is $\sum_{i=1}^{m} \beta_i \Delta y_{t-i}$. This term is a lagged difference terms which is a test that was developed to solve a problem of serial correlation. Moreover, ADF test has asymptotic distribution the same as DF-test, and can use the same critical values.

2) Phillips Perron test (PP test)

The PP test has difference from the ADF tests because it can deal with a serial correlation and heteroskedasticity. Particularly, in the test regression, the PP tests ignore any serial correlation, but the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors. The regression for the PP tests is $\Delta y_t = \alpha y_{t-1} + x_t' \delta + \varepsilon_t$, and modifies the α coefficient *t*-ratio to the serial correlation that does not affect the asymptotic distribution of the statistic test. The PP test can be shown as the formula based on the statistics as follows:

$$\tilde{t}_\alpha = t_\alpha \left(\frac{\gamma_0}{f_0}\right)^{\frac{1}{2}} - \frac{T(f_0 - \gamma_0)(se(\hat{\alpha}))}{2 f_0^{1/2} s} \qquad (14)$$

where t_α the ratio of α and $\hat{\alpha}$ is the estimate, $se(\hat{\alpha})$ is the standard error coefficient, and s is the standard error. Moreover, γ_0 is a consistent estimate of the error variance in $\Delta y_t = \alpha y_{t-1} + x_t' \delta + \varepsilon_t$ (calculated as $(T-k)s^2 / T$, where *k* is the regressors number) and the f_0 is the residual estimation at frequency zero. There are two choices which we have to making when conducting the PP test. In the test regression, we must first choose that to have a constant, a constant and a time trend, or neither. Second, we have to pick a approach for estimating f_0.

3) Kwiatkowski–Phillips–Schmidt–Shin test (KPSS test)

The unit root tests are used to test for the null hypothesis that a time series y_t is $I(1)$. Besides, the stationarity tests are used to test for the null that y_t is $I(0)$.

The KPSS test was purposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992) and started with the formula as follows:

$$\begin{aligned} y_t &= \beta' D_t + \mu_t + u_t \\ \mu_t &= \mu_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \end{aligned} \tag{15}$$

where D_t has a defined components consisting of a constant and a constant with a time trend which u_t represents $I(0)$ and may have heteroskedastic. μ_t is random walk with variance σ_ε^2. We can show the null hypothesis that y_t is $I(0)$ in the formula as $H_0 : \sigma_\varepsilon^2 = 0$, which means that μ_t is a constant. This null hypothesis means a unit root of moving average in the ARMA form of Δy_t. The KPSS test is the Lagrange multiplier statistic for testing $\sigma_\varepsilon^2 = 0$ with the alternative that $\sigma_\varepsilon^2 > 0$ and is given by

$$KPSS = \left(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2 \right) / \hat{\lambda}^2 \tag{16}$$

where $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$, $\hat{u}_t$ is residual in y_t regression on D_t, and $\hat{\lambda}^2$ is an estimating consistent for u_t the long-run variance by using $\hat{u}_t$. Based on the null that y_t is $I(0)$, KPSS converges to a standard Brownian motion function that relies on the deterministic terms D_t but not their coefficient values β. Especially, if $D_t = 1$ then

$$KPSS \xrightarrow{d} \int_0^1 V_1(r)dr \tag{17}$$

where $V_1(r) = W(r) - rW(1)$ and $W(r)$ is a Brownian motion for $r \in [0,1]$. If $D_t = (1,t)'$ then

$$KPSS \xrightarrow{d} \int_0^1 V_2(r)dr \tag{18}$$

where $V_2(r) = W(r) - r(2-3r)W(r) + 6r(r^2-1)\int_0^1 W(s)ds$. The critical values in equation (17) and (18) must be obtained by simulation approachs. So that, from the optimal asymptotic distribution (17) or (18), we rejects the null of stationarity at the $100 \cdot \alpha\%$ level if the KPSS test statistic (16) is greater than the $100 \cdot (1 - \alpha)\%$ quantile.

2.1.3 Risk Management

The risk management is the evaluation and prioritization for risks applying the coordination and resource allocation in the economy to investigate, control, and minimize the probability of poor events or to maximize the positive opportunities. The risk management purpose to assure that an unstable does not deflect the effort from the business goal (Hubbard, 2009).

There are various sources of risk consisting of threats from project failures, uncertainty in financial markets, accidents, natural causes and disasters, or events of uncertainty etc. There are two events types, namely positive events classified as opportunities while negative events classified as risks.

For risk management, there are many methods to asses or investigate the risk for hedges that may occur in the future. This study uses value at risk (VaR) and expected shortfall (ES) estimates to manage risk which can be explained as follows:

1) Value at Risk (VaR)

VaR is widely used in risk management to measure risk and has been applied to market risks. The advantages of this model is able to summarize single character of risk which does not depend on an identified kind of distribution or an easy to understand number, and any kind of financial asset applied this model (Bob, 2013).

For the VaR definition, McNeil et al. (2005) gave a confidence level $\alpha \in (0,1)$. We defined the portfolio VaR at confidence level α by the smallest number l in the probability that the loss L exceeds l is no larger than $(1-\alpha)$. Formally,

$$VaR_{\alpha} = inf\left\{l \in \mathbb{R} : P(L > l) \leq 1-\alpha\right\} = inf\left\{l \in F_L(l) \geq \alpha\right\} \quad (19)$$

The equation (19) is the definition of VaR and presents the profit and loss probability distribution, where the cumulative distribution function (cdf) is $F_L(l)$ and the minus sign presents VaR only as a positive value. The approaches for VaR estimation can be represented as follows:

1.1) The Historical Simulation (HS)

The HS approach is the use of the historical distribution of assets' prices or log returns in portfolio to estimate the VaR of portfolio, assuming that we hold portfolio over the period of time covered by our historical data set. To apply this approach, first the different instruments are identified in portfolio and a sample of their historic log returns is collected over some observation period. Then, use the weights in our current portfolio to simulate the discounted loss distribution (L_t). Thus, the HS approach does not depend on any parametric model assumptions. However, this approach depends on the stationarity of the historical data set to ensure convergence of the empirical discounted loss distribution to the true discounted loss distribution.

Assume this historical distribution is a good approximation of the log returns distribution that 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 historical log returns' distribution will lead us to the expected portfolio VaR.

Moreover, it does not rely on assumption about the distributions of log returns because it allows the data to explain and determine the distribution. To ensure sufficient estimation precision, HS requires large amounts of data. However, this method is not always practically feasible to obtain such large appropriate sample data, and even if it is, the history may not repeat itself or contain sufficient extreme observations for the VaR estimation.

1.2) The Variance - Covariance Method (VC)

The VC method is the most well-known approach for VaR estimation. Based on this method, VaR for portfolios can be derived by estimating the variance and the covariance of some predefined risk factors' log returns and the sensitivity of the portfolio to those risk factors. The log returns of risk factor are independently and identically distributed (*i.i.d.*) is the most basic assumption in the model with a multivariate normal distribution.

Note that the VC method is only appropriate to a portfolio whose gain or loss being assets returns in a linear function. Therefore, it gives a poor estimate for portfolios with non-linear instruments such as options. However, the outstanding disadvantage of the method is the common assumption. Most financial assets are recognized to have log return distributions being fat tailed, implies that in truth, extreme results are more feasible than the normal distribution would suggest. Consequently, the VaR estimate might be understated.

1.3) The Monte Carlo Simulation (MCS)

Based on the MCS approach aimed to repeatedly simulate a random process that governs all financial instruments' prices in the portfolio. Each simulation gives a feasible value for portfolio at the end of target horizon. The simulated portfolio values distribution will concentrate to the portfolio's unknown "true" distribution, if we take enough of these simulations, and we can use the distribution simulated to compute the "true" VaR.

2) Expected Shortfall (ES)

ES is a measure that produces better motivation for traders than VaR. Sometimes, this method is also called as conditional VaR (CVaR) which Artzner et al. (1997) have suggested the ES use to deal with the problems in VaR. ES is a tail VaR that combines the VaR aspects with more information in the returns distribution (Kevin Sheppard, 2013). The definition and condition of ES can be explained as follows:

2.1) Expected Shortfall Definition

ES is defined as all losses average or the expected value of portfolio loss which are greater or equal than a VaR that has computed. For the unconditional ES can be defined as

$$\begin{aligned} ES &= E\left[\frac{W_1 - W_0}{W_0} \middle| \frac{W_1 - W_0}{W_0} < -VaR\right] \\ &= E\left[r_t \middle| r_t < -VaR\right] \end{aligned} \tag{20}$$

where $W_1, t = 0,1$ is the assets value in the portfolio, and 1 and 0 is a time length such as one day, one year, etc.

2.2) Conditional Expected Shortfall

We can estimate ES with the return that we obtain from prediction, which the conditional ES can be defined as:

$$ES_{t+1} = E_t\left[r_{t+1} \middle| r_{t+1} < -VaR_{t+1}\right]. \quad (21)$$

where r_{t+1} is a portfolio return at time $t+1$. Since t refer to a time measure, and $t+1$ refers to a time unit in the future such as one day, one month, or five years, etc.

2.1.4 Portfolio Optimization

The optimal portfolio is the selecting processes of various assets proportions that are held in a portfolio in order to make the best portfolio. The criterion will be considered directly or indirectly in the portfolio's rate return in the expected value and maybe other measures in financial risk.

For this section we first explain the terminology that will be used by assuming that we have *n* assets with random rates of returns $\xi_1, \xi_2, \ldots, \xi_n$. The expected rates of return are $E[\xi_1], E[\xi_2], \ldots, E[\xi_n]$. Suppose that there is a portfolio *n* assets, and w_i is the asset weight *i* in the portfolio, by $\sum_{i=1}^{n} w_i = 1$. Therefore, the portfolio returns are $\sum_{i=1}^{n} w_i \xi_i$ and the expected portfolio return is $\sum_{i=1}^{n} w_i E[\xi_i]$.

Risk is often specified as the variance of portfolio, σ^2. The random variable's variance is its second central moment, and its mathematical definition is

$$\begin{aligned} \sigma^2 &= E[(\xi - \bar{\xi})^2] = E(\xi^2) - 2E(\xi)\bar{\xi} + \bar{\xi}^2 \\ &= E(\xi^2) - (\bar{\xi})^2 \end{aligned}. \quad (22)$$

Given the individual assets' variance, the portfolio's variance can be computed by using the covariance between asset *i* and asset *j* - σ^2.

Let σ_i^2 be the variance of asset *i* and σ_p^2 be the variance of portfolio. The variance of portfolio can be computed as follows:

$$\begin{aligned} \sigma_p^2 &= E[(\xi_p - \bar{\xi}_p)^2] = E[(\sum_{i=1}^{n} w_i \xi_i - \sum_{i=1}^{n} w_i \bar{\xi}_i)^2] \\ &= E[(\sum_{i=1}^{n} w_i (\xi_i - \bar{\xi}_i))^2] = E[(\sum_{i=1}^{n} w_i (\xi_i - \bar{\xi}_i))(\sum_{i=0}^{n} w_i (\xi_i - \bar{\xi}_i))] \\ &= E[(\sum_{i=1}^{n} w_i w_j (\xi_i - \bar{\xi}_i)(\xi_j - \bar{\xi}_j))^2] \\ &= \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \end{aligned}. \quad (23)$$

On the other way, let W be the vector that represents the each asset weights, Σ be the variance-covariance matrix, and P be the position of current portfolio. The variance of portfolio can be shown as

$$\sigma_p^2 = (\mathrm{W}^{\mathrm{T}}\,\Sigma\,\mathrm{W}) \tag{24}$$

1) Problem of Maximum Expected Returns

The problem of maximizing the expected return of the portfolio for an investor can be shown as:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{n} w_i E[\xi_i] \\ &\text{subject to} && \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{25}$$

2) Minimum Variance Portfolio

Although Portfolio's maximizing returns might be attractive, the model in equation (25) does not take into account the portfolio risk. Therefore, investors being risk-averse might wish to minimize the portfolio risk. For the problem given the minimum variance in portfolio, we can define the formula as:

$$\begin{aligned} &\text{minimize} && \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \\ &\text{subject to} && \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{26}$$

3) Mean-Variance Efficient Portfolio

The minimum variance and maximum returns portfolio give the returns and risk in two extremes. Most investors wish to have portfolio being a balance between the minimum variance and maximum returns. The mean-variance in optimal portfolio can be regulated as follows:

$$\begin{aligned} &\text{minimize} && \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \\ &\text{subject to} && \sum_{i=1}^{n} w_i = 1 \\ & && \sum_{i=1}^{n} w_i E[\xi_i] \geq R \end{aligned} \tag{27}$$

In the equation (27) provide the optimal portfolio namely minimum risk for the specified minimum required wealth, R, which mean that other portfolios cannot give higher return with a lower risk. On the other way, we could set the maximize returns problem given a risk level σ, which can be illustrated in the equation below.

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{n} w_i E[\xi_i] \\ &\text{subject to} && \sum_{i=1}^{n} w_i = 1 \\ & && \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \leq \sigma \end{aligned} \tag{28}$$

4) Efficient Frontier

Finding and plotting all optimum portfolios on a returns and risk diagram, we can get the plotting of efficient frontier. This figure means that portfolios proving the maximum returns for a given risk level. In order to get the efficient frontier, firstly, we solved the problem of minimum risk, ignoring any constraints of returns, consequently, we will obtain a minimum expected return W_{min}. Then, we solved the problem of maximum returns, ignoring any constraints of the risk level, so obtaining a maximum expected return W_{max}. Hence, we can solve the problem of minimum risk for a plot set $W \in [W_{min}, W_{max}]$. Therefore, the resulting shows a plot sets form the efficient frontier. In Figure 2.1 show the sample of the efficient frontier.

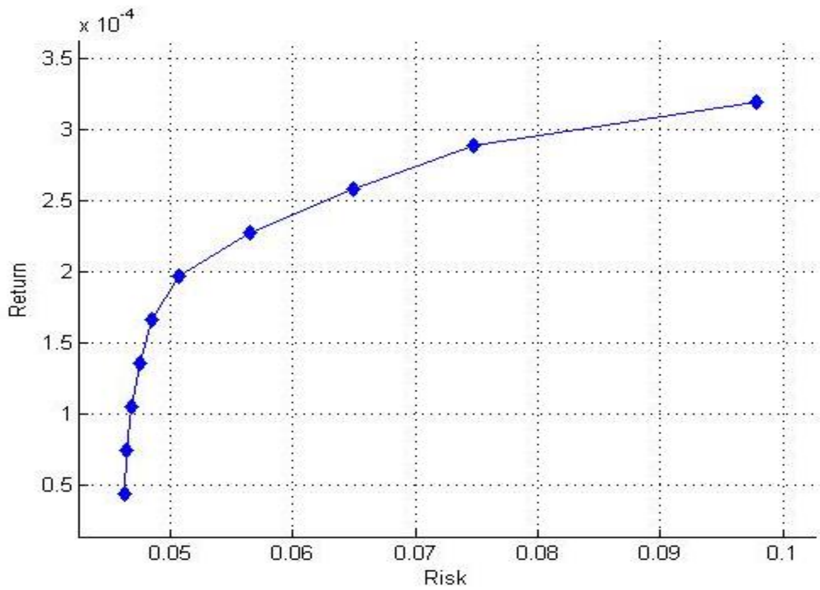


Figure 2.1 Efficient frontier

2.1.5 ARMA and GARCH model

The asymmetric ARMA and GARCH model was used to filter the time series in order to obtain "independent and identically distributed" (*i.i.d.*) in residual or random variables being necessary for implementation of Extreme Value Theory in next step.

1) Autoregressive Moving Average model (ARMA)

The ARMA model was found by Box and Jenkins (1970) to fit data removing the linear dependence and to obtain the random variable or residuals uncorrelated. This model is used widely in time series of finance modeling and is a linear model consisting of two parts, AR(*p*) and MA(*q*) process, which the linear processes can be conducted as follows:

1.1) The AR (*p*) process

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t \qquad (29)$$

where $\phi_1,...,\phi_p$ are the model's parameter, μ is constant, and ε_t is white noise. Moreover, the AR (*p*) process can also be written as: $\phi(L)X_t = \varepsilon_t$ (without constant), where $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ which can write the AR (*p*) as the MA (∞) process: $r_t = \gamma(L)\varepsilon_t$, where $\gamma(L) = (1 - \phi_1 L - ... - \phi_p L^p)^{-1}$.

1.2) The MA (*q*) process

$$r_t = d + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \qquad (30)$$

where $\psi_1,...,\psi_q$ are the parameter of model, d is a constant, and ε_t is white noise. If we set the condition $|\psi| < 1$, then we can show r_t as an AR (∞) process. This condition is known as invertibility condition. Because the MA models are feebly stationary by construction, they are finite the white noise in linear combination.

1.3) The ARMA (*p, q*) model is

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \qquad (31)$$

We can call the ARMA (*p, q*) model that a causal process if $r_t = \sum_{i=0}^{\infty} \varphi_i \varepsilon_{t-i}$ where φ_i should satisfy $\sum_{i=0}^{\infty} |\varphi_i| < \infty$. If we write the characteristic equation of ARMA (*p, q*), it can be illustrated that:

$$\begin{cases} \tilde{\phi}(a) = 1 - \phi_1 a - ... - \phi_p a^p \\ \tilde{\psi}(a) = 1 - \psi_1 a - ... - \psi_q a^q \end{cases} \tag{32}$$

and given that $\tilde{\phi}(a)$ and $\tilde{\psi}(a)$ have no general roots. The ARMA processes are causal if and only if $\tilde{\phi}(a)$ does not has roots in a unit circle $|a| \leq 1$. So in the stationary ARMA process, we should have $(\phi_1 + ... + \phi_p) < 1$.

2) GARCH model

Because of the volatilities clustering and the effect of leverage, so the ARMA equation cannot account for the effects of heteroskedastic in the time series process generally observed in the fat tails form. Engle (1982) proposed the Autoregressive Conditional Heteroskedastic model, namely ARCH model. This model can observe that there are the same volatility effects in negative and positive shocks. However, in empirical testing, these effects are asymmetric, and the ARCH model reacts slowly to large isolated shocks. Bollerslev (1986) presented the Generalized Autoregressive Conditional Heteros-kedastic (GARCH) model that is the ARCH model extension. The GARCH (*k*,*l*) is described by the equations set as follows:

$$\varepsilon_t = \sigma_t z_t \tag{33}$$

$$z_t \sim N_\vartheta(0,1)$$

$$\sigma_t^2 = \varpi + \sum_{i=1}^{k} \alpha_i \varepsilon_{i-t}^2 + \sum_{i=1}^{l} \beta_i \sigma_{t-i}^2 \tag{34}$$

where z_t is an *i.i.d.* process with mean 0 and variance 1. Although ε_t is uncorrelated by definition that its variance condition is σ_t^2. We consider in the following differ only in functional form for the conditional variance in all the GARCH models. $N_\vartheta(0,1)$ is the function of probability density of the innovations. ϑ is the additional distributional parameter to describe the skew and the shape of the distribution. GARCH model can be reduced to the ARCH model, if the coefficients β is zero.

Therefore, from the explanation above ARMA (*p*,*q*) - GARCH (*l*,*k*) model can be set as:

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \quad (35)$$

$$\varepsilon_t = \sigma_t z_t \quad (36)$$

$$\sigma_t^2 = \varpi + \sum_{i=1}^{k} \alpha_i \varepsilon_{i-t}^2 + \sum_{i=1}^{l} \beta_i \sigma_{t-i}^2 \quad (37)$$

3) Distribution assumptions

In this section, we focus on the normal and the student-*t* distribution to consider the return distributions in the skewness, excess kurtosis, and heavy-tails. For the detail, it can be illustrated as follows:

3.1) Normal Distribution

The normal distribution is widely used in predicting GARCH models. The standard normal distribution in log-likelihood function is given by

$$L_T = \ln \prod_t \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\varepsilon_t^2}{2\sigma_t^2}} = -\frac{1}{2}\sum_{t=1}^{T}[\ln(2\pi) + \ln(\sigma_t^2) + z_t^2] \quad (38)$$

if the error term follows a normal, where $z_t = \varepsilon_t / \sigma_t$ is *i.i.d* and the $E[z_t] = 0$, $\mathrm{var}[z_t] = 1$, and T is the observation number with regards to the return series.

3.2) Student-*t* Distribution

Bollerslev (1986) suggested student-*t* distribution being used to fit the GARCH model for the standardized error in order to better capture the return series having the observed fat tails. The symmetric is around mean 0 and function of log-likelihood can be formulated as follows:

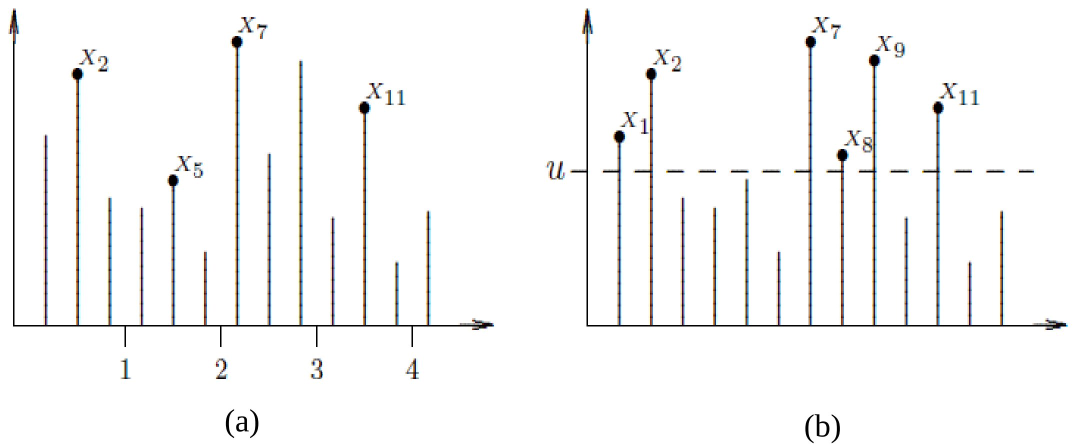
$$L_T = T\left\{\ln\left[\Gamma\left(\frac{v+1}{2}\right)\right]-\ln\left[\Gamma\left(\frac{v}{2}\right)\right]-\frac{1}{2}\ln\left[\pi\left(v-2\right)\right]\right\} -\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left(\sigma_t^2\right)+\left(1+v\right)\ln\left(1+\frac{z_t^2}{v-2}\right)\right] \quad (39)$$

where $v>2$ is parameter of shape, $2<v\leq\infty$ and $\Gamma(.)$ is function of gamma. If v is lower than 2, there will be fatter tails.

2.1.6 Extreme Value Theory (EVT)

LeBaron and Samanta, 2005 supported EVT which aimed to forecast the occurrence of exceptionally rare events. This theory is widely used in weather disasters management (such as modeling the maximum damages from floods, storms, etc.), in insurance, and is particularly well adapted to model financial crashes.

EVT can be divided into two approaches when it model extreme events. First is the Block Maxima model, and second is the Peak Over Threshold (POT). In Block maxima model, the point of maximum data in periods is the extremes value or we can call these values that the generalized extreme value distribution (GEV) (Fisher and Tippett, 1928). On the other hand, the POT method focuses only on the sample of observations that exceeds a certain threshold.



Source: Gilli and KÄellezi (2006)

Figure 2.2 Block-maxima and excesses over a threshold *u*

The Block-maxima diagnoses the maximum variable taken in successive time periods, for instance monthly or annual. These chosen data include the extreme events, can called that a block maxima. For the Figure 2.2 (a), the observations X_2, X_5,

X_7 and X_{11} represents that the three observations in each block have the block maxima for four time periods.

The Peak Over Threshold method focuses on a realization exceeding a given threshold (calculated). As Figure 2.2 (b), the observations following will be considered as extreme events, that are X_1, X_2, X_7, X_8, X_9 and X_{11}, all exceeding the threshold u.

In this thesis, we focused on the the Peak Over Threshold (POT) method and Generalized Pareto Distributions (GPD) which can be explain as follows:

Nystrom and Skoglund (2002b) and McNeil et al. (2005) state that POT is choice method when applying EVT to find tails of financial returns distributions. The starting point is a sample of *i.i.d.* observations $(X_t)_{t\in\mathbb{N}}$ which focus on the distribution followed by observations in Figure 2.2 on a certain threshold.

Following Jondeau et al., 2006, let u be the threshold in the support of X_t, which there is the formula as follows:

$$\begin{aligned} F_u(y) &= P_r[X_t - u \le y \mid X_t \ge u] \\ &= \frac{F_X(y+u) - F_X(u)}{1 - F_X(u)}, 0 \le y \le x_F - u \end{aligned} \quad (40)$$

This formula is the excess distribution function or *edf* of the random variables X_t over the threshold u, and the formula $e(u) = E[X_t - u \mid X_t \ge u]$ is the mean-excess function or *mef*.

Moreover, we can approximate the asymptotic distribution of the scaled excess over threshold by the generalized Pareto distribution (Balkema and De Haan, 1974 and Pickands III, 1975). Theorem can be shown as follows:

We can approximate the excess distribution $F_u(y)$, if F_X is in the attraction domain of the extreme value distribution H_ξ, which there must be a u large enough, as follows:

$$F_u(y) \approx G_{\xi,\beta}(y), u \to \infty$$

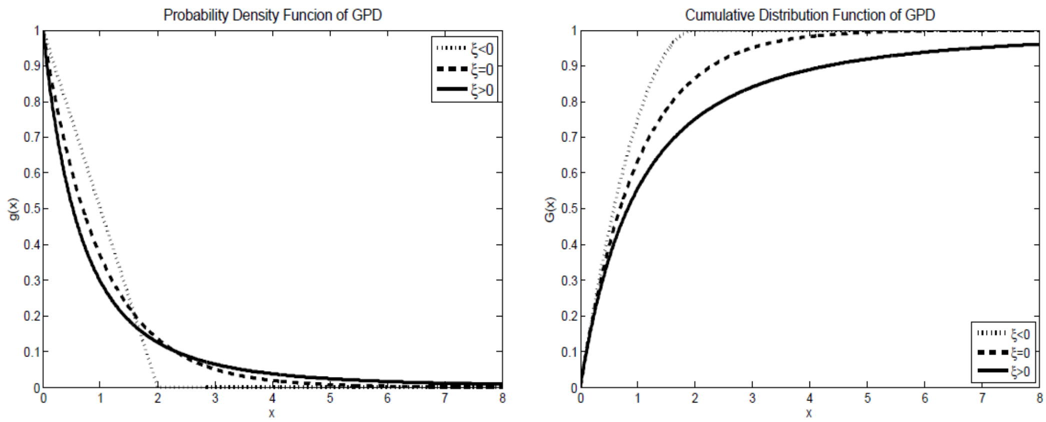
where

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\beta} y\right)^{-1/\xi}, & if\ \xi \ne 0 \\ 1 - exp\left(-\dfrac{y}{\beta}\right), & if\ \xi = 0 \end{cases} \quad (41)$$

with $y \in \begin{cases} [0,\infty], & if\ \xi \geq 0 \\ [0,-1/\xi], & if\ \xi \leq 0 \end{cases}$ for $0 \leq y \leq x_F - u$, is generalized Pareto distribution. The preceding formula can be rewritten as:

$$G_{\xi,u,\beta}(x) = \begin{cases} 1-\left(1+\dfrac{\xi}{\beta}(x-u)\right)^{-1/\xi}, & if\ \xi \neq 0 \\ 1-exp\left(-\dfrac{(x-u)}{\beta}\right), & if\ \xi = 0 \end{cases} \quad (42)$$

if we define *x* = *y* + *u*, which, ξ is the tail index and identically to GEV distributions.



(a) (b)

Source: Avdulaj, 2010

Figure 2.3 Probability density functions and cumulative distribution function. $\xi = 0$ exponential, $\xi = -0.5$ Pareto type II and $\xi = 0.5$ Pareto distribution. In all case $\beta = 1$ and $\nu = 0$.

In the above description, the the POT and block maxima method provide a parametric density of distribution in tails, which is important for risk managers. In order to get a tails distribution estimation for *y*, refer back to the excess distribution function (*edf*) definition:

$$\begin{aligned} F_u(y) &= P_r[Y \leq y | Y \geq u] \\ &= \frac{F(y)-F(u)}{1-F(u)} \\ 1-F(y) &= (1-F(u))(1-F_u(y-u)) \end{aligned} \quad (43)$$

the tail of y was estimated by applying EVT. Nystrom and Skoglund (2002b) assume the observations that are sorted in ascending (or minimum to maximum) defining $Y_{n,n} \le \ldots \le Y_{1,n}$, and $u = Y_{k+1,n}$ as the threshold. We then analyze the following observations: $Y_{k,n} \le \ldots \le Y_{n,n}$, and an estimator for $1 - F(u)$ is k/n. The estimator for the $F(y)$ tail is obtained by using the GPD as follows:

$$\hat{F}(y) = 1 - k/n\left(1 + \hat{\xi}\left(\frac{y - Y_{k+1,n}}{\hat{\beta}}\right)^{-1/\xi}\right) \tag{44}$$

with $\hat{\xi}$ and $\hat{\beta}$ estimates of the GPD parameters. Now we have to calculate the high quantiles $q = F(z_q)$ of the y distribution using the above equation, we obtain

$$\hat{y}_q = \hat{y}_{q,k} = Y_{k+1,n} + \frac{\hat{\beta}}{\xi}\left(\left(\frac{1-q}{k/n}\right)^{-\xi} - 1\right) \tag{45}$$

This clearly shows that the quantiles values estimates are dependent on the sample size n and the threshold u and also clear that the GPD parameters need to be estimated before obtaining the quantiles estimates. Additionally, the maximum likelihood is recommended by most studies and scholars and is the choice method here.

Selecting the optimal threshold *u*

Although the threshold u already demonstrated to the importance, there are some tricky to find its optimal value: u should be high enough, so that the observations distribution is larger than the threshold but the higher the u converges to the GPD.

There are several methods to estimate u such as a the mean-excess plot or simple graphical method (Jondeau et al., 2006). Assuming the underlying GPD distribution, the mean excess formula can be shown as:

$$\begin{aligned} e(u) &= E[X_t - u \mid X_t \ge u] \\ &= \frac{\beta + \xi u}{1 - \xi} \end{aligned} \tag{46}$$

where $\xi \le 1$, which is clear that a linear function in u is the *mef*. T observations (x_1, ...,x_T) are assumed to be above the threshold u. The *mef* estimate can be computed as the sum of the excess observations that divided by the number of observations above the threshold N_u.

$$\hat{e}(u) = \frac{1}{N_u}\sum_{t=1}^{T}(x_t - u)1_{\{x_t \geq u\}}(x_t), u \geq 0 \tag{47}$$

The mean-excess plot can be shown as:

$$\left\{\left[x_{t,T}, \hat{e}(x_{t,T})\right] \quad where\ t = 1,..,T\right\} \tag{48}$$

and where $\{x_{t,T}\}^T$ is the ordered observations sample. The u was selected so that the plot of $\hat{e}(x_{t,T})$ is approximately linear for $x_{t,T} \geq u$. The ME plot has a negative slope for thin tailed distributions and the extremes are not far from center, while fat-tailed distributions exhibit an increasing ME plot.

For another way to choose an arbitrary threshold, u is picked at k observations exceeding it in an n sample of observations. So, the integer value $(k+1)/n$ is often selected as u. An MLE is fitted to $\{x_{1,n} - u, ..., x_{k,n} - u\}$ to estimate ξ and β. There are still ways to choose a threshold such as Nystrom and Skoglund (2002) argued that the percentage of observations that exceed u should be between 5-12% of the data. The exceedances were choosed to be the 10^{th} percentile of the sample for the upper and lower tail because the 10^{th} percentile is the appropriateness to choose in the generalized Pareto model, and there isthe same biases occurred as in the stable law analysis (Dumouchel, 1983).

2.1.7 Copula Model

Copula model is a statistical measure that illustrates a multivariate uniform distribution or a function combining univariate distributions to obtain a joint distribution, which examines the relation or dependence between many random variables. This model is flexible because it distinguishes the choice of dependence among variables from the marginal distributions choice of each variable. Copulas are useful in portfolio or risk management by helping the marginal modeling and dependence structure modeling of a multivariate probability model. The basic and main ideas in copula model can be explained as follows:

1) Pearson's correlation

After Markowitz (1952) wrote the article on mean-variance in portfolio, Pearson's correlation coefficient has become the most popular to model dependencies. The definition of Pearson's coefficient can be demonstrated as follows:

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}, \quad (49)$$

with $\rho \in [-1,1]$. The correlation coefficient of Pearson essentially captures the linear relationship between two variables, and its usage is represented when modeling the dependencies between jointly normally distributed returns. However, this coefficient has been used to model dependence structures between variables being not assumed to be normally distributed. For example credit risk, leading to a widespread misjudgment of risks that eventually contributed to the financial meltdown in 2008.

Embrechts et al. (1999) stated that the random variables that are analyzed should have elliptical distributions. These distributions have densities which are constant on ellipsoids. The multivariate normal and student-*t* distributions are examples of elliptical distributions and this idea can be used with mean-variance portfolio, VaR, and justified correlation.

2) Copulas and Sklar's theorem

Copulas can be used to model the dependence structure between random variables following identical or different distributions. The copulas usefulness can be understood using this example:

Define X and Y, two random variables with marginal distributions $F(x) = P[X \le x]$ and $G(x) = P[Y \le y]$. The joint distribution function is noted: $H(x, y) = P[X \le x, Y \le y]$. When the joint distribution exists like the multivariate normal distribution, H has a clear expression. In many cases, the margins are easy to obtain, while the joint distribution is very elusive. Copulas link the margins with the multivariate distribution function, and allow the dependence of risk factors to be modified independently of the marginal distributions which are a very useful tool for stress testing.

Definition: Schmidt (2006)

A d-dimensional copula is a function C: $[0, 1] \times [0, 1] \to [0, 1]$ being a cumulative distribution function (*cdf*) with uniform marginal distributions. Thus, copulas are noted: $C(u) = C(u_1,..., u_d)$ and have three properties as follows:

1. *cdf* are always increasing; any increase in each u_i implies an increase of $C(u)$

2. The i^{th} uniform marginal distribution is extracted by positing $u_j = 1$ for $i \neq j$ and:

$$C(1,...,1,u_i,...,1) = u_i \tag{50}$$

3. For $a_i \leq b_i$, the probability $P(U_1 \in [a_1,b_1],...,U_d \in [a_d,b_d])$ must be positive, which can be written as follows:

$$\sum_{i_1=1}^{2} ... \sum_{i_d=1}^{2} (-1)^{i_1+...+i_d} C(u_{1,i_1},...,u_{d,i_d}) \geq 0 \tag{51}$$

Schmidt (2006) showed the results on transformation of quantile explain how a copula can link marginals and dependence structure. The generalized distribution function *F* inverse is defined by:

$$F^{\leftarrow}(y) = inf\{x : F(x) \geq y\} \tag{52}$$

Let *U* be a random variable uniformly distributed on [0,1], and *F* be *cdf*. We have:

$$P(F^{\leftarrow}(U) \leq x) = F(x) \tag{53}$$

On the other hand, if the real-valued random variable *Y* follows the continuous distribution function *F*, then:

$$F(Y) \sim U[0,1] \tag{54}$$

Thus, we can simulate random variables with any *cdfs* using uniformly distributed variables. Consequently, we obtain a multivariate distribution function which each distribution can be linked to a copula, by choosing a copula and some marginal.

Theorem: Sklar (1959)

For a joint *cdf F* with marginal distributions F_1,..., F_d, There exists a copula *C*, such that:

$$F(x_1,...x_d) = C(F_1(x_1),...,F_d(x_d)) \tag{55}$$

for all x_i in $[-\infty,\infty]$, $i = 1, ...,d$. C is unique, if the marginal distributions are continuous for all i. Conversely, let us consider a copula C, and marginals $F_1,..., F_d$, then F as defined in the equation (55) is a joint distribution based on the abovementioned marginal.

A copula density can be computed as:

$$C(u) = \frac{\partial^d C(u_1,...,u_d)}{\partial_{u_1}...\partial_{u_d}} \tag{56}$$

Densities are often used in practice which they are easy to interpret.

3) Non-parametric copulas

This approach can be used in order to model a non-linear dependence. This method has the advantage of not requiring any additional assumptions on the non-linear dependence. However, it also has serious drawbacks, namely the non-linear dependence patterns can be complicated to interpret, and the results are often unstable even in the bivariate case. Deheuvels (1979) introduced the non-parametric copula, which is defined:

$$C\left(\frac{t_1}{T},\frac{t_2}{T}\right) = \frac{1}{T}\sum_{t=1}^{T} 1\{x_t \le x_{t_1,T}, y_{t_2,T}\}\, with 1 \le t_1, t_2 \le T, \tag{57}$$

where $x_{1,T} \le ... \le x_{T,T}$ are sorted observations. The empirical copula frequency is defined as:

$$c_T\left(\frac{t_1}{T},\frac{t_2}{T}\right) = \begin{cases} \frac{1}{T} & if\ (x_{t_1,T}, y_{t_2,T})\, belong\ to\ the, sample \\ 0 & otherwise \end{cases} \tag{58}$$

We can link c_T and C_T through the equation:

$$C\left(\frac{t_1}{T},\frac{t_2}{T}\right) = \sum_{p=1}^{t_1}\sum_{q=1}^{t_2} c_T\left(\frac{p}{T},\frac{q}{T}\right) \tag{59}$$

4) Copula families in Elliptical copula

In this thesis we examine in Value at Risk. So, in the copula model, we focused on determining Normal or student-*t* distributed copulas (Elliptical copula) because they are better suited for risk management purposes.

One can extract the Gaussian and student-*t* copula from the multivariate Normal and *t*-distribution, two elliptical ones.

Normal copula

In the bivariate case, we can define by the *cdf*:

$$C_\rho = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left(-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)} \right) ds_1 ds_2 \qquad (60)$$

with Σ the 2 × 2 correlation matrix: $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, Φ_Σ the bivariate standard normal *cdf*, and Φ^{-1} the inverse of the univariate Gaussian distribution and ρ denotes the Pearson correlation and belongs to the interval [-1,1]. Independence is equivalent to a correlation of zero, and in that case the normal copula is the independence copula. For $\rho = 1$, the normal copula is the comonotonicity copula, and for $\rho = -1$, it becomes the counter monotonicity copula. ρ is the only parameter needed to characterize the copula dependence structure which makes the normal copula a comprehensive one.

The normal copula density can be written as:

$$c_\rho = \frac{1}{|\Sigma|^{1/2}} exp\left(-\frac{1}{2}\psi'(\Sigma^{-1} - I_2)\psi \right) \qquad (61)$$

with $\psi = (\Phi^{-1}(u_1), \Phi^{-1}(u_2))'$. Spearman's rho and Kendall's tau are given by:

$$\rho_s(C_\rho) = \frac{6}{\pi} arcsin(\rho) \\ \rho_T(C_\rho) = \frac{2}{\pi} arcsin(\rho / 2) \qquad (62)$$

In the multivariate case, the Gaussian copula is given by:

$$C_\Sigma = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)) \qquad (63)$$

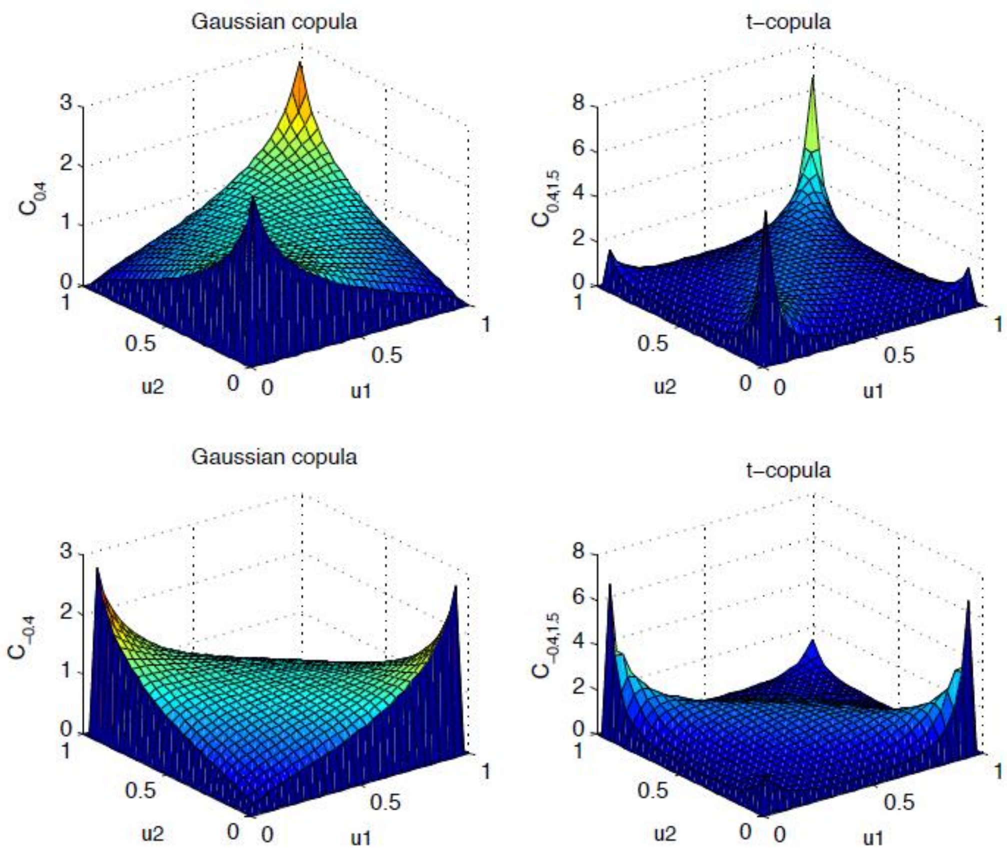
Student-*t* copula

The student-*t* can be defined as

$$C_{\rho,n}(u_1, u_2) = t_{\rho,n}(t_n^{-1}(u_1), t_n^{-1}(u_2)) \\ = \int_{-\infty}^{t_n^{-1}(u_1)} \int_{-\infty}^{t_n^{-1}(u_1)} \frac{\Gamma\left((n+2)/2\right)}{\Gamma(n/2)\pi n\sqrt{1-\rho^2}} \left(1 + \frac{\psi'\Sigma^{-1}\psi}{n} \right)^{\frac{-n+2}{2}} d\psi, \qquad (64)$$

with $\psi = (t_n^{-1}(u_1), t_n(u_2))', t_{\rho,n}$ is the student-*t* *cdf* with *n* degrees of freedom (*df*), and correlation *ρ*, and $t_{\rho,n}^{-1}$ is the *cdf* inverse (Jondeau, Poon, and Rockinger, 2006).

Figure 2.4 plots sets of normal and student-*t* copulas (*df*=1.5). The first row shows the copulas for positive correlation *ρ*, while the second row reveals copulas for negatively correlated variables. Moreover, there are some differences between between the two types of copulas.



Source: Mudry, 2013

Figure 2.4 Gaussian and *t*-copula with *ρ* = +/−0.4 and DoF = 1.5

The student-*t* copula exhibits stronger tail dependence than the normal copula and another interesting point is the student-*t* copula represents spikes in all four corners and can be explained by its mixing nature.

Further, the student-*t* copula clearly displays tail dependence, even for a correlation of zero. Assuming $\rho \geq 1$, then:

$$\lambda_t = \lambda_u = 2t_{\upsilon+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right) \tag{65}$$

Therefore, student-*t* copula is more able than normal copula to model joint extreme positive and negative returns. Moreover, the dependence structure linked to these copulas enables to reproduce a feature observed during the financial crisis 2008.

5) Monte Carlo simulations using copulas

Copulas allow to model dependency structure between random variables more precisely than the other tools, so that copulas are great for using Monte Carlo simulations in financial portfolios. The idea of the simulation can be explain as:

We firstly need to simulate the uniform margins vector $U = (U_1, ..., U_d)$ and choosing the desired copula. Then we obtain the arbitrary marginal by using the quantile transformation. We repeated the algorithm *n* times to obtain a *n* independent sample *i.i.d.* multivariate pseudo random variables with the dependence structure which is defined by the copula. The procedure of Monte Carlo simulations using copulas can be concluded as follows:

<u>Normal copula</u>

1. A matrix of arbitrary covariance $\tilde{\Sigma}$, deduct a correlation matrix Σ by resizing each cell components by a variance of 1.

2. Conduct a Cholesky-decomposition $\Sigma = A'A$.

3. Build standard normal variables $\tilde{X}_1, ..., \tilde{X}_d$ being *i.i.d.*

4. Obtain $(X_1, ..., X_d)' = X = A\tilde{X}$, with $\tilde{X} = (\tilde{X}_1, ..., \tilde{X}_d)$.

5. Calculate $U = (\Phi(X_1), ..., \Phi(X_d))$ with Φ the cumulative standard normal distribution.

<u>Student-*t* copula</u>

1. Perform the above steps 1 to 4 from the procedure of normal copula.

2. Generate $\xi = \sum_{i=1}^{\nu} Y_i^2$, with y_i *i.i.d.* $\sim N(0,1) \xi \sim \chi_\nu^2$.

3. Obtain $U_i = t_\nu(X_i / \sqrt{\xi/\nu}), i = 1, ..., d$ with t_ν the univariate *cdf* student-*t* distribution with a *df* of *ν*.

2.2 Literature Review

Leonard (2007) studied about measuring market risk in a case study about copula and extreme value approach. This study presents a method for measuring the portfolio risk that is the assets' component with heteroscedastic return series. To obtain a better estimate for VaR and ES, this model tried to capture the data generating for each return series and also the dependence structure that exists at the portfolio level. The return series was modeled by using GARCH model, and the dependence structure was defined using a student-*t* copula. These techniques was used to simulate a portfolio return distribution allowing risk measures which this method was applied to five portfolio Romanian stocks, and then the accuracy of risk measures was tested using a backtesting approach.

Chan (2009) studied the forecasting value at risk applying maximum entropy density which the usefulness of this method would be assessed empirically using S&P 500 data. Moreover, VaR forecasts will also be constructed by ARMA-GARCH and ARMA-GJR models along with different assumptions of distribution. The empirical result showed that the estimated parameter under different assumptions of distribution was similar except ARMA-GARCH model under the student-*t* distribution. Therefore, the estimates could not be modeled in special case and the assessment on the VaR estimate was not conducted with the ARMA-GARCH model in the assumption on student-*t* distribution. The violation percentage is the ratio between the violations relative numbers that the total forecasts number should be approximate to the significance level. The performed MED is the best, followed by GARCH with GED distribution and GJR with normal distribution.

Hammoudeh et al. (2011) studied risk management in the precious metals. They examined the correlation dynamics and volatility in precious metals consisting of gold, silver, platinum, and palladium. They studied the risk management relating to market risk and hedging. VaR was used to analyze the poor situation in market associated with investments in financial market, and plan the optimum risk management. VaR was computed using three approaches consisting of, the different GARCH models, the calibrated RiskMetrics and the semi-parametric filtered historical simulation approach. The result showed that the economic importance was highlighted for assessing the daily capital charges which analyzed with the VaR estimate.

Tesarov (2012) studied value at risk GARCH versus stochastic volatility models. This work compares GARCH volatility and stochastic volatility (SV) models with the assumption of student-*t* distribution in errors and its empirical forecasting performance of Value at Risk on five stock price indices: S&P, NASDAQ Composite, CAC, DAX and FTSE. It introduced in detail the problem of SV models Maximum Likelihood examinations and suggested a newly developed approach called efficient importance sampling (EIS). EIS is a procedure that provides an accurate Monte Carlo evaluation of likelihood function which depends on high-dimensional numerical integrals. Comparison analysis was separated to in-sample and out-of-sample forecasting performance and evaluated using standard statistical probability backtesting methods as conditional and unconditional coverage. Based on empirical analysis, the thesis showed that SV models can perform at least as good as GARCH models if not superior in forecasting volatility and parametric VaR.

Bob (2013) studied VaR estimation with GARCH-EVT-Copula approach. This study applied this approach with the stock index data from Germany, Spain, Italy and France. The marginal distributions was modeled by using an asymmetric GARCH and EVT method for each log returns series, and then used Copula functions to capture the dependent structure. The Monte Carlo Simulation approach was used to estimates the portfolio VaR. Backtesting methods were used to check the goodness of the approach. The results was concluded that, the GARCH-EVT-Copula approach was better performances than other approach and could estimate accurate VaR.

Tang et al. (2014) studied the risk estimate of natural gas portfolios by using GARCH-EVT-Copula approach. Firstly, the univariate ARMA-GARCH was used to model and extract residuals from each natural log return series. Second, they fitted the residuals tails by using EVT to model marginal distributions. Third, they capture the copula parameter with multivariate normal and student-*t* copula to explain the dependence structure of natural gas portfolio. Finally, they simulated *N* portfolios to estimated VaR and ES. The empirical results showed that, the VaR and ES value of the five natural gases in equally weighted portfolio obtained from the student-*t* copula are higher than value obtained from the normal copula. Additionally, in the minimizing portfolio risk, the natural gas portfolio optimization weights found that the multivariate normal copula and student-*t* copula are similar and different confidence levels.

Table 2.1 Literature Review Methodology

Author	Topic	Variables used in the study	Methodology
Stanga Alexandru Leonard (2007)	Measuring market risk: a copula and extreme value approach	A portfolio of five Romanian equities traded on the Bucharest Stock Exchange.	1) GARCH models 2) Extreme Value Theory (EVT) models 3) Copula models 4) Measures of risk
Felix Chan (2009)	Forecasting Value-at-Risk using Maximum Entropy Density	S&P 500	1) Parametric Volatility Models and Value-at-Risk 2) Maximum Entropy Density
Burcak Bulut (2010)	Forecasting the Prices of Non-Ferrous Metals with GARCH Models and Volatility Spillover from World Oil Market to Non-Ferrous Metal Markets	The mean three-month futures prices of three commonly traded non-ferrous metals (copper, aluminum, lead, nickel, tin, and zinc)	GARCH family is GARCH, TGARCH and EGARCH models
Henrik Skaarup Andersen and David Sloth Pedersen (2010)	Extreme Value Theory with Applications in Quantitative Risk Management	Danish OMX C20 Index consist of 1) Novo Nordisk 'B' (NOVO B) 2) Carlsberg 'B' (CARLS B) 3) Danske Bank (DANSKE)	Univariate Methods 1) Historical Simulation (HS) 2) HS with a GARCH-type model (HS-GARCH and HS-GARCH-t) 3) Filtered Historical Simulation (FHS) 4) HS with Conditional EVT (HSCONDEVT) Multivariate Methods 1) Variance-Covariance (VC) 2) Variance-Covariance with EWMA (VC-EWMA) 3) Constant Conditional Correlation Model (CCC-GARCH), 4) Dynamic Conditional Correlation Model (DCC-GARCH) 5) Multivariate Conditional EVT
Shawkat Hammoudeh, Farooq Malik, and Michael McAleer (2011)	Risk Management of Precious Metals	All precious metals are traded at COMEX in New York	1) Value-at-Risk 2) Risk Metrics 3) GARCH with t distribution 4) GARCH - Filtered Historical Simulation (FHS)

Table 2.1 (Cont.)

Author	Topic	Variables used in the study	Methodology
Bc. Viktoria Tesarov (2012)	Value at Risk GARCH vs. Stochastic Volatility Models: Empirical Study	1) S&P 500 (GSPC) 2) NASDAQ Composite (IXIC) 3) FTSE 100 (FTSE) 4) CAC 40 (FCHI) 5) DAX 30 (GDAXI)	1) Value at Risk 2) GARCH model 3) Stochastic Volatility Models 4) Backtesting
Ngoga Kirabo Bob (2013)	Value at Risk estimation A GARCH-EVT-Copula Approach	1) DAX 2) IBEX 35 3) FTSE MIB 4) CAC 40	1) Value at Risk 2) Copula functions 3) Extreme Value Theory (EVT) 4) GARCH models
Jiechen Tang et al. (2014)	Estimating Risk of Natural Gas Portfolios by Using GARCH-EVT-Copula Model	Spot and futures prices at the Title Transfer Facility (TTF) Hub	1) GARCH model 2) Extreme Value Theory (EVT) 3) Copula 4) Portfolio Risk Analysis consist of VaR and CVaR and Optimal Portfolio with Minimum Risk