CHAPTER 2

Methodology

This chapter provides a summary of the methodology and recent developments in the financial econometrics. This part of the study is divided into five subsections for using in Chapter 3, 4 and 5. Firstly, to determine the optimal portfolio weight, we present three different methods that are used to compute based on the risk and return framework. Secondly, we present the Demster-Shafer theory and linear belief function for combining the information from different events. Thirdly, we explain the GARCH model and the vector autoregressive model. Fourthly, we give some detail of the copula theory. Lastly, we provide the concept of extreme value theory for applying the financial data.

2.1 Portfolio Optimization Method

2.1.1 Markowitz method

In the conventional method, Markowitz (1952) introduced this method based on risk and return for portfolio selection problem. The return of a portfolio can be defined by the expected return μ_p and modeled as a weight sum of each stock's mean returns $\mu_1, \ldots, \mu_m$ from stock 1 to stock m

$$\mu_p = \sum_{i=1}^{m} p_i \mu_i \tag{2.1}$$

where p_i is the weight of each stock returns in portfolio. The risk of portfolio can be defined by the variance σ_p^2 and modeled as a weight sum of variance-covariance of each stock's returns.

$$\sigma_p^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} p_i p_j \sigma_{ij} \tag{2.2}$$

where σ_{ij} is variance-covariance of stock return in portfolio. We also write the portfolio variance in matrix notation as

$$\sigma_p^2 = p^T \Sigma p = (p_1, \dots, p_m) \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1,m}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^2 & \cdots & \sigma_m^2 \end{pmatrix} \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix} \tag{2.3}$$

To optimize the portfolio by Markowitz framework, we can minimize the variance of portfolio in equation 2.2 subject to the return of a portfolio in equation 2.1. This problem can be set up as:

$$\begin{aligned} &\text{Minimize} && \sigma_p^2 \\ &\text{Subject to} && \sum_{i=1}^{m} p_i \mu_i \geq \mu_o \\ & && \sum_{i=1}^{m} p_i = 1 \\ & && p_i \geq 0 \,, i = 1, \dots, m \end{aligned} \tag{2.4}$$

where $\sum_i^m p_i \mu_i = p^T \mu$ is the summation of return stocks, μ_o is the minimum level of portfolio return that the investors would like to receive from this portfolio. The sum of the weight of all stock returns invested must be equal to one and the last constraint is the non-negativity weight, which means no short sales in this portfolio. Although, this method has some drawback as the assumption of multivariate normality is in sample returns.

2.1.2 Entropy method

Alternatively, this study applied the entropy method to the portfolio selection method. Entropy in the information theory refers to the uncertainty related with

random variable. The basic concept of entropy was proposed by Shannon (1948).

Definition of the entropy: Let X be the random variable that is taking value in $\{x_1, \dots, x_m\}$ and with probabilities $p_1, \dots, p_m$ and $p_i \geq 0, i = 1, \dots, m$, $\sum_i^m p_i = 1$. To measure the uncertainty, let $s(p_i)$ be the information content that interprets the uncertainty that is associated in the event $X = x_i, i = 1, .., n$ (see Golan et al. (1996))

$$s(p_i) = log_2 \frac{1}{p_i} \tag{2.5}$$

The entropy $S(X)$ of a discrete random X depends on the probabilities of X and the entropy is defined by the expected informational content as

$$S(X) = \sum_{i=1}^{m} p_i h(p_i) = -\sum_{i=1}^{m} p_i \ln(p_i) \tag{2.6}$$

Then, Janynes (1992) applied the entropy to maximum entropy principle that is a technique to estimate the probabilities. The study used this method for a portfolio selection. The mean entropy method is the following:

$$\begin{aligned} \text{Maximize} \quad & -\sum_{i=1}^{m} p_i \ln(p_i) \\ \text{Subject to} \quad & \sum_{i=1}^{m} p_i \mu_i \geq \mu_o \\ & \sum_{i=1}^{m} p_i = 1 \\ & p_i \geq 0, i = 1, \dots, m \end{aligned} \tag{2.6}$$

Moreover, this study also emphasizes the entropy method for optimizing the portfolio. We applied the entropy into the objective function subject to the first (mean) and second (variance) moments in statistics. Let μ_o is the maximum level of portfolio return and σ_0^2 is the maximum level of portfolio variance that the investors would like to receive from this portfolio. We say this method that the mean-variance entropy method is the following:

$$\begin{aligned} \text{Maximize} \quad & -\sum_{i=1}^{m} p_i \ln(p_i) \\ \text{Subject to} \quad & \sum_{i=1}^{m} p_i \mu_i \geq \mu_o \\ & \sum_{i=1}^{m}\sum_{j=1}^{m} p_i p_j \sigma_{ij} \leq \sigma_0^2 \\ & \sum_{i=1}^{m} p_i = 1 \\ & p_i \geq 0, i = 1, \ldots, m \end{aligned} \tag{2.8}$$

The excess return per unit of risk is the simple ratio formula for comparing the portfolio performance that was proposed by Sharpe (1966). This study proposes a new method that used the idea of the Sharpe ratio for constructing the constraint of entropy method. We call it the Sharpe ratio entropy method.

$$\begin{aligned} \text{Maximize} \quad & -\sum_{i}^{m} p_i \ln(p_i) \\ s.t. \quad & \frac{\mu_p}{\sigma_p} \geq \frac{\mu_o}{\sigma_o} \\ & \sum_{i}^{m} p_i = 1 \\ & p_i \geq 0, i = 1, \ldots, m \end{aligned} \tag{2.9}$$

2.1.3 VaR and CVaR portfolio optimization

Previously, many investors use variance for the decisions involving risk measures. However, variance has some limitations as it is evaluated under the symmetry assumption that is not suitable for measuring the market risk. Alternatively, VaR is one of a risk measure that can specify only the scenario of downside (or upside) risk by defining the potential loss for a given confidence level.

Definition of VaR: Let r be the random returns of m assets, p is a set of feasible portfolios, $q \in (0,1)$ is the confidence level and the linear loss function of portfolio $f(p,r) = -[p_1 r_1 + \ldots + p_m r_m]$ not exceeding a given threshold γ. VaR of portfolio is defined as follows:

$$VaR_q(p) = min\{\gamma \in \mathbb{R}: P[f(p,r) \leq \gamma] \geq q\} \tag{2.10}$$

However, VaR have some weakness for portfolio analysis because it doesn't satisfy the properties of subadditivity and concavity in case of non-elliptical distributions. Rockafellar and Uryasev (2000) developed a risk measure called the conditional value at risk (or expected shortfall).

Definition of $CVaR$: Let r be the random variables and $p(r)$ is the density function of r, $CVaR$ is the expected loss exceeding the VaR in the confidence level at $1-q$, which is defined by

$$CVaR_q(p) = \frac{1}{(1-q)} \int_{f(x,y) \geq VaR_q(p)} f(p,r)p(r)dr \tag{2.11}$$

To calculate the optimal portfolio weight, this study set up the $CVaR$ optimization problem by the following this equation 2.12

$$\begin{aligned} &\text{Minimize} && CVaR_q(p) \\ &\text{Subject to} && p^T r \geq r_o \\ & && e^T p = 1 \\ & && p \geq 0, i = 1, \ldots, m \end{aligned} \tag{2.12}$$

In equation 2.12, r is a vector of expected return, r_o is the expected portfolio return and $e^T p$ is the sum of portfolio weights.

2.2 Belief Functions

The theory of belief functions was proposed by Dempster (1967) and Shafer (1976), it was also called the Dempster and Shafer theory (DST). This theory emphasized the process for combining the evidence from different sources. There are two important components applied in the mathematical probability to measure the degrees of belief for expressing the subjective judgments. The degrees of belief are combined by following Dempster's rule.

For the basic concept of this theory, let Ω be a frame of discernment, the mass function m (also called the basic probability assignment or belief mass assignment) is a mapping from subsets A of a frame of discernment into the interval between zero and one :$m : 2^{\Omega} \rightarrow [0,1]$, in which $m(\phi) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. $m(A) > 0$ is the degree of evidential support for focal element A. To determine the upper and lower bound of an interval by following the mass function, there are two important measures called belief and plausibility functions. Belief refers the lower bound for a given set A , which can be represented by $bel(A) = \sum_{B \subseteq A} m(B)$, $bel(\phi) = 0$ and $bel(\Omega) = 1$. Plausibility refers to the upper bound for a given set A , which is defined by

$Pl(A) = \sum_{B \cap A \neq \phi} m(B)$, $Pl(\phi) = 0$ and $Pl(\Omega) = 1$.

Definition of Dempster's rule: let m_1 and m_2 are two mass functions that are the evidences from two different sources on Ω , $\oplus$ is the operator of combination. The combination of two belief structures $m = m_1 \oplus m_2$ can be defined as follows:

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - k}, A \neq \phi, (m_1 \oplus m_2)(\phi) = 0, \qquad (2.13)$$

where $k = \sum_{B \cap C = \phi} m_1(B) m_2(C)$ represents the degree of conflict between m_1 and m_2. If k is large value, the conflict between m_1 and m_2 is large size. $1 - k$ is normalization.

Linear Belief Functions (LBF)

Liu (1996), Liu (1999) and Liu,Shenoy and Shenoy (2003) introduced the linear belief function (also called Gaussian belief function), which is the extension of Dempster Shafer theory in case of continuous variables. LBF also is special type of belief function in the sense of focal element as parallel sub-hyperplanes over a hyperplane. In general, belief function is defined by the belief mass assignment (bma) over a class of focal element, In case of LBF, bma is the Gaussian distribution across the sub-hyperplanes. LBF represents the logical and uncertainty (probabilistic) knowledge. Logical knowledge is identified by linear equation, a certainty hyperplan. Uncertainty knowledge is identified by the Gaussian distribution across all parallel focal elements over the hyperplanes. For understanding, Suppose that there is the representation $(\boldsymbol{C}, \boldsymbol{B}, \boldsymbol{L}, \boldsymbol{\pi}, \boldsymbol{E})$ for LBF in random variable space $(\boldsymbol{S})$. where $\boldsymbol{C}$ is the certainty space, $\boldsymbol{B}$ is the belief space, $\boldsymbol{L}$ is the linear space, $\boldsymbol{\pi}$ is the covariance and $\boldsymbol{E}$ is the expectation. $\boldsymbol{\pi}$ and $\boldsymbol{E}$ are defined by a variable in $\boldsymbol{B}$ under Gaussian distribution. $\boldsymbol{C} \subseteq \boldsymbol{B} \subseteq \boldsymbol{L} \subseteq \boldsymbol{S}$ where $\boldsymbol{C}, \boldsymbol{B}$ and $\boldsymbol{L}$ are in the subspace of $\boldsymbol{S}$. The belief space is the combination between $\boldsymbol{C}$ and $\boldsymbol{U}$ ($\boldsymbol{B} = \boldsymbol{C} \oplus \boldsymbol{U}$) where $\boldsymbol{U}$ is the uncertainty space. X in $\boldsymbol{C}$ is certain when the Gaussian distribution is zero variance for the variable in the certainty space. If nonzero variance in variable X, then we can

call that X is variable in the uncertainty space $\boldsymbol{U}$. Therefore, Liu (1999) explained that the key of LBF is to identify the mean and variance-covariance of variables into the moment metrics for transforming a piece of evidence as follows;

Example 1: Let r is a random variable of the normal distribution with mean (μ) and variance (Σ), which can be represented by the moment matrix as: $m = \begin{pmatrix} \mu \\ \Sigma \end{pmatrix}$ .If there is the case of no uncertainty, then the variance is $\Sigma = 0$. the moment matrix is $m = \begin{pmatrix} \mu \\ 0 \end{pmatrix}$

Example 2: Let r_1 and r_2 are two random variables under the normal distribution with the mean (μ_1, μ_2) and variance-covariance $(\Sigma_1, \Sigma_2, \Sigma_{1,2})$. The moment matrix can be written as:

$$m(X,Y) = \begin{pmatrix} \mu_1 & \mu_2 \\ \Sigma_1 & \Sigma_{2,1} \\ \Sigma_{1,2} & \Sigma_{2,} \end{pmatrix} \tag{2.14}$$

Example 3: Let $r_1 = a + Br_2$ is the liner equation. We can specify the linear equation in the term of partial moment matrix as:

$$m(r_1, \overrightarrow{r_2}) = \begin{pmatrix} 0 & \mathrm{a} \\ 0 & \mathrm{B} \\ \mathrm{B}^{\mathrm{T}} & 0 \end{pmatrix} \tag{2.15}$$

Example 4: Let $r_1 = a + Br_2 + \varepsilon$ is the liner regression with $\varepsilon \sim N(0, \Sigma)$.We can this regressions in the term of partial moment matrix as:

$$m(r_1, \overrightarrow{r_2}) = \begin{pmatrix} 0 & \mathrm{a} \\ 0 & \mathrm{B} \\ \mathrm{B}^{\mathrm{T}} & \Sigma \end{pmatrix} \tag{2.16}$$

Moreover, Liu (1996) showed that the combination rule in LBF of variable space was agreed with Demster (1990) for a combination in finite case that was replaced by Gaussian density function in LBF. Liu (1999) presented that matrix sweeping technique can be used alternatively for the Dempster's rule of combination. The sweeping technique is composed of forward sweep and reverse sweep. This technique is used for combining the means, variance, residual covariance and regression coefficients.

Suppose that $r_1, r_2, \ldots, r_n$ be a random return under the assumption of Gaussian distribution, $E(r_i) = \mu_i$ is expected mean, $Var(r_i) = \Sigma_{i,i}$ is variance and $Cov(r_i, r_j) = \Sigma_{i,j}$ is covariance between market returns i and j. $m(r_1, r_2) = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}$ is a moment matrix that presents the parameters in Gaussian distribution. Liu (1999) showed that the operation on moment matrices can be explained by the definitions below.

Definition of Marginalization: Liu (1999), suppose $m(r_1, r_2)$ is the moment matrix for the random variables r_1 and r_2 and $m^{\downarrow r_2}$ represents the marginal of $m(r_1, r_2)$ to r_1 that is implies to the conditional moment matrix of linear regression coefficient as follows:

$$m^{\downarrow r_2}(r_1, r_2) = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix} \tag{2.17}$$

Definition of Forward sweep: Liu (1999), suppose $m_k(r_1, r_2, \ldots, r_n)$ is the moment matrices for the random variables $r_1, r_2, \ldots, r_n$ from 1 to k matrices. To combine them, all moment matrices must be transformed to fully swept matrices $m_k(\overrightarrow{r_1}, \overrightarrow{r_2}, \ldots, \overrightarrow{r_n})$ by using forward sweeping. Forward sweeping of $m_k(r_1, r_2, \ldots, r_n)$ on r_s can be represented by $m_k(r_1, r_2, \ldots, \overrightarrow{r_s}, \ldots, r_n)$ as follows:

$$m_k(r_1, r_2, \ldots, \overrightarrow{r_s}, \ldots, r_n) = \begin{pmatrix} \mu_{j,s} \\ \Sigma_{ij,s} \end{pmatrix} \tag{2.18}$$

Where

$$\mu_{j,s} = \begin{cases} \mu_j - \mu_s \Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for j} \neq \text{s} \\ \mu_s \Sigma_{ss}^{-1} & ,\text{for j} = \text{s} \end{cases}$$

$$\Sigma_{ij,s} = \begin{cases} -\Sigma_{ss}^{-1} & ,\text{for i} = \text{s} = \text{j} \\ \Sigma_{is} \Sigma_{ss}^{-1} & ,\text{for j} = \text{s} \neq \text{i} \\ \Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for i} = \text{s} \neq \text{j} \\ \Sigma_{ij} - \Sigma_{is} \Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for otherwise} \end{cases}$$

Definition of Reverse sweep: Liu (1999), suppose $m_k(\overrightarrow{r_1}, \overrightarrow{r_2}, \ldots, \overrightarrow{r_n})$ is fully swept matrices for the random variables $r_1, r_2, \ldots, r_n$ from 1 to k matrices. Conversely, the fully swept matrices can be transformed to the moment matrices by using reverse sweeping. Reverse sweeping of $m_k(\overrightarrow{r_1}, \overrightarrow{r_2}, \ldots, \overrightarrow{r_n})$ on r_s can be represented by $m_k(\overrightarrow{r_1}, \overrightarrow{r_2}, \ldots, r_s, \ldots, \overrightarrow{r_n})$ as follows:

$$m_k(\overrightarrow{r_1}, \overrightarrow{r_2}, \ldots, r_s, \ldots, \overrightarrow{r_n}) = \begin{pmatrix} \mu_{j,s} \\ \Sigma_{ij,s} \end{pmatrix} \tag{2.19}$$

Where

$$\mu_{j,s} = \begin{cases} \mu_j - \mu_s \Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for j} \neq \text{s} \\ -\mu_s \Sigma_{ss}^{-1} & ,\text{for j} = \text{s} \end{cases}$$

$$\Sigma_{ij,s} = \begin{cases} -\Sigma_{ss}^{-1} & ,\text{for i} = \text{s} = \text{j} \\ -\Sigma_{is} \Sigma_{ss}^{-1} & ,\text{for j} = \text{s} \neq \text{i} \\ -\Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for i} = \text{s} \neq \text{j} \\ \Sigma_{ij} - \Sigma_{is} \Sigma_{ss}^{-1} \Sigma_{sj} & ,\text{for otherwise} \end{cases}$$

Definition of The combined linear belief function: Liu (1999), suppose $\vec{m}_1 = m_1(\vec{r_1}, \vec{r_2})$ and $\vec{m}_2 = m_2(\vec{r_1}, \vec{r_2})$ are fully swept matrices. The combination of $k = 2$ linear belief function is the direct sum of fully swept matrices as follows:

$$\vec{m} = \vec{m}_1 \oplus \vec{m}_2 = \begin{pmatrix} \vec{\mu}_1 + \vec{\mu}_2 \\ \vec{\Sigma}_1 + \vec{\Sigma}_2 \end{pmatrix} \tag{2.20}$$

2.3 Time Series Models

2.3.1 Vector Autoregressive Model

To consider the relationships between the variables in the case of multivariate time series, a vector autoregressive (VAR) model can be described for this situation. The reduced form VAR can be defined as

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + e_t \tag{2.21}$$

where Φ_0 is a $n \times 1$ vector of intercept parameters, $\Phi_1, \ldots, \Phi_p$ are the $n \times n$ coefficient matrices for lag $1, \ldots, p$ and the optimal lag length is conducted by AIC, SIC and HQ. y_t is $n \times 1$ a vector of market returns at time t , $y_{t-1}, \ldots, y_{t-p}$ are $n \times 1$ the vectors of market returns at time $t-1, \ldots, t-p$, and e_t is a $n \times 1$ vector of white noise process. Additionally, e_t is assumed to be $e_t \sim N(0, \Sigma)$ with following $E(e_t) = 0$, $E(e_t, e'_{t-1}) = 0$ (no serial correlation) and $E(e_t, e'_t) = \Sigma$ is the variance-covariance matrix of the white noise process. The simplest VAR can be determined by

$$Y = \boldsymbol{\Psi X} + E, \tag{2.22}$$

where $Y = [y_t]$, $\boldsymbol{X}= [1, y_{t-1}, \dots, y_{t-p}]$, $\boldsymbol{\Psi} = [\Phi_0, \dots, \Phi_p]$ and $E = [e_t]$. The VAR parameters can be estimated by multivariate least squares (MLS) method for $\boldsymbol{\Psi}$ yields

$$\hat{\boldsymbol{\Psi}} = \mathrm{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1} \tag{2.23}$$

2.3.2 Conditional Heteroscedastic Models

To estimate the volatility of a time series in financial data, Engle (1982) proposed a non-linear model that describes the risk under unstable condition namely the autoregressive conditional heteroskedasticity (or called ARCH) model. Bollerslev (1986) extended the ARCH model to allow the conditional volatility on the past value of itself namely generalized autoregressive conditional heteroskedasticity (or called GARCH) model. The basic GARCH (p,q) model can be specified by

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \dots + \alpha_p\varepsilon_{t-p}^2 + \beta_1\sigma_t^2 + \dots + \beta_q\sigma_{t-q}^2, \tag{2.24}$$

where $\varepsilon_t = \sigma_t z_t$ is the residuals, which are equal to the conditional variance (or volatility) and the standardized residuals. Generally, z_t is i.i.d (0,1) random variables. $\theta_g = [\alpha_i, \beta_j]$ is the parameter set of the GARCH model that requires the stationary condition as $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$. The positive and negative returns in the standard GARCH model are given the same volatility, but in a general situation it may not appear to be presented in such a format. Then, Glosten et al. (1993) extended the standard GARCH model that takes into account the asymmetries of volatility process namely the GJR-GARCH model.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\gamma_i\varepsilon_{t-i}^2 I_{t-i} + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \tag{2.25}$$

where $\gamma_i \varepsilon_{t-i}^2 I_{t-i}$ is leverage terms that represents the term of the asymmetric volatility, $I = 1$ if $\varepsilon < 0$ for negative shock and $I = 0$ for otherwise. $\theta_{gjr} = [\alpha_i, \beta_j, \gamma_i]$ is the parameter set of the GJR-GARCH model with the stationary condition as follows:

$$\alpha_0 > 0,\ \alpha_i \geq 0, \beta_j \geq 0,\ \alpha_i + \gamma_i \geq 0$$

and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j + \frac{1}{2}\sum_{i=1}^{p} \gamma_i < 1.$

In estimating all parameters of the GARCH models given the distribution $g(\cdot)$ of the standardized residuals, we can be fitted by using maximum log-likelihood (MLL).

$$\ell_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T} log\ \sigma_t + \sum_{t=1}^{T} log\left(g\left(\frac{r_t - \mu_t}{\sigma_t^2}\right)\right) \tag{2.26}$$

where $r_t - \mu_t = \varepsilon_t$ is mean residuals and θ are the parameter sets to be estimated.

2.4 Copula Approach

From the previous section, the dependence between two variables can be measured by Pearson's correlation coefficient; however there are some restrictions such as measuring the linear dependence between two variables under normal distribution, where the financial data have been tail-dependence between variables for some time. Alternatively, copulas are useful method for describing the dependence structures between elements of random vectors and it can provide a specific dependence structure to simulate the data. Copulas are the joint distribution functions into one dimensional marginal distribution function as

Definition of Copula: Let $\boldsymbol{x} = [x_1, \dots, x_n]'$ is set of random variables and a n-dimensional copula is a multivariate CDF of C with standard uniform marginal distribution$[0,1]$. The multivariate distribution of copulas is represented by $C(u) = C(u_1, \dots, u_n)$ that C is a mapping from: $[0,1]^n \rightarrow [0,1]$ from a unit hypercube into the unit interval.

Sklar provided the theorem for the dependence between the joint distributions for representing a copula.

Sklar's Theorem (1959): Consider $\boldsymbol{x} = [x_1, \dots, x_n]'$ be vector of random variables for $i = 1, \dots, n$ with a joint n- dimensional distribution function F and marginal distributions $F_1, \dots, F_n$. There is an unique copula $C \; \forall \; \boldsymbol{x} \in R^n$ that can be written by

$$C(u_1, \dots, u_n) = F\left(F_1^{-1}(u_1), \dots, F_n^{-2}(u_n)\right), \tag{2.27}$$

and equation 2.27 can be written

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \tag{2.28}$$

where $F_i^{-1}(u_i)$ is the inverse distribution function of the marginal x_i and $u_i \in [1,0]$. In equation 2.27, we can write the copula density function c by partial differentiation.

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1, \dots, \partial u_n}, \tag{2.29}$$

and we can write the joint density function f as follows:

$$f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1, \dots, \partial x_n}, \tag{2.30}$$

$$f(x_1,\dots,x_n)=\prod_{i=1}^{n} f(x_i)\frac{\partial^n C(u_1,\dots,u_n)}{\partial u_1,\dots,\partial u_n}, \tag{2.31}$$

$$f(x_1,\dots,x_n)=\prod_{i=1}^{n} f(x_i)\,c(u_1,\dots,u_n), \tag{2.32}$$

In equation 2.32 shows that the joint density function is composed of the copula density function and the marginal densities.

2.4.1 The Multivariate t -Copula

The most popular copulas based on the multivariate t-distribution, the standard t-copula is used to measure symmetric dependence and it can handle fatter tails dependence than Gaussian copula. The t-copula in n- dimensional distribution is defined by

$$C_t(u_1,\dots,u_n;\Re,v)=t_{v,\Re}(t_v^{-1}(u_1),\dots,t_v^{-1}(u_n)) \tag{2.33}$$

where $t_{v,\Re}$ denotes the CDF of the multivariate t-distribution with correlation matrix $\Re$ and degree of freedom v. t_v^{-1} denotes an CDF of inverse univariate t-distribution. Similarly, the multivariate t-copula can also be written as

$$C_t(u_1,\dots,u_n;\Re,v)=\int_{\infty}^{t_v^{-1}(u_1)}\cdots\int_{\infty}^{t_v^{-1}(u_n)}\frac{\Gamma\left(\frac{v+n}{2}\right)|\Re|^{-\frac{1}{2}}}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{\frac{n}{2}}}\left(1+\frac{1}{v}x'\Re^{-1}x\right)^{-\frac{v+n}{2}}d\boldsymbol{x} \tag{2.34}$$

For estimating the parameters in the t-copula, we used the density function in equation 2.34.

$$c_t(u_1,\dots,u_n;\Re,v)=\frac{\partial^n C_t(u_1,\dots,u_n;\Re,v)}{\partial u_1,\dots,\partial u_n} \tag{2.35}$$
$$=|\Re|^{-\frac{1}{2}}\frac{\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)}\left(\frac{\Gamma\left(\frac{v+n}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\right)^n\frac{\left(1+\frac{1}{v}\Lambda^{-1}\Re^{-1}\Lambda\right)^{-\frac{v+n}{2}}}{\prod_{j=1}^{n}\left(1+\frac{\Lambda_j^2}{v}\right)^{-\frac{v+n}{2}}}$$

where $\Gamma(\cdot)$ is the gamma function, $\Lambda=[t_v^{-1}(u_1),\dots,t_v^{-1}(u_n)]'$ is the vector of inverse univariate t-distribution. We use the density function in equation 2.35 to construct the log-likelihood function and maximize the log likelihood function with respect to v and $\Re$.

$$\ell_N(v,\Re)=\sum_{j=1}^{N}logc_{v,\Re}\left(\boldsymbol{u}^{(j)}\right) \tag{2.36}$$

2.4.2 Vine Copulas

In the previous section give the detail has been given about the multivariate copulas. This section illustrates the alternative the method for determining the dependence structures for the high dimensional variables (or more than two variables) based on pair copula constructions (PCC). Bedford and Cooke (2001, 2002) introduced PPC that the multivariate copulas are determined by various bivariate copulas and nested tree structure and it is also called the regular vine copulas. Subsequently, Aas et al. (2009) extended special structure of regular vine copulas, namely the canonical (C)-vine copula and the drawable (D)-vine copula. This thesis focuses on two vine copulas. Let $\boldsymbol{x}=[x_1,\dots,x_n]'$ is set of random variables for n-dimensions. There are the number of trees $n-1$ and bivariate copula densities $n(n-1)/2$ in the vine structures. $n!$ is the possible vines. The density function $f(\boldsymbol{x})$ of C-vine can be expressed as follows:

$$f(\boldsymbol{x})=\prod_{i=1}^{n}f_i(x_i)\cdot\prod_{j=1}^{n-1}\prod_{j=1}^{n-j}c_{j,j+k|1:(j-1)}\left(F(x_j|\boldsymbol{v_1}),F(x_{j+k}|\boldsymbol{v_1})|\delta_{i,i+j|1:(i-1)}\right), \tag{2.37}$$

where f_i is the marginal density of x_i, $c_{i,i+j|1:(i-1)}$ is the bivariate copula density with the parameters $\delta_{j,j+k|1:(j-1)}$ that also represent the pair-copulas, $F(x_j|\boldsymbol{v_1})$ and $F(x_{j+k}|\boldsymbol{v_1})$ are the conditional distribution function on $\boldsymbol{v_1} = x_1, \dots, x_{j-1}$. the density function $f(\boldsymbol{x})$ of D-vine can be expressed as follows:

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} f_i(x_i) \cdot \prod_{j=1}^{n-1} \prod_{k=1}^{n-j} c_{k,k+j|(k+1):(k+j-1)} \left(F(x_k|\boldsymbol{v_2}), F(x_{k+j}|\boldsymbol{v_2})|\delta_{k,k+j|(k+1):(k+j-1)}\right) \quad (2.38)$$

Where , $c_{k,k+j|(k+1):(k+j-1)}$ is the bivariate copula density with the parameters $\delta_{k,k+j|(k+1):(k+j-1)}$. $F(x_k|\boldsymbol{v_2})$ and $F(x_{k+j}|\boldsymbol{v_2})$ are the conditional distribution function on $\boldsymbol{v_2} = x_{k+1}, \dots, x_{k+j-1}$, j represents the trees and k represents the edge in each tree. Regarding the conditional distribution in equation 2.37 and 2.38, we can compute it according to Joe (1996) as

$$F(x|\boldsymbol{v}) = \frac{\partial C_{xv_j|\boldsymbol{v}_{-j}}\left(F(v_j|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\right)}{\partial F(v_j|\boldsymbol{v}_{-j})}, \quad (2.39)$$

where $C_{xv_j|\boldsymbol{v}_{-j}}$ be the bivariate copula distribution between x and v_j with conditional on $\boldsymbol{v}_{-j}$ and $\boldsymbol{v}_{-j}$ is the vector v_j that not obtain the component v_j. In figure 2.1 and 2.2, we give the examples of C and D-vines for 5 dimensions. There are $5! = 120$ possible vines for estimation. Hence, this study uses the maximum of the absolute empirical Kendel's tau values from all bivariate copulas to order the possible vines.

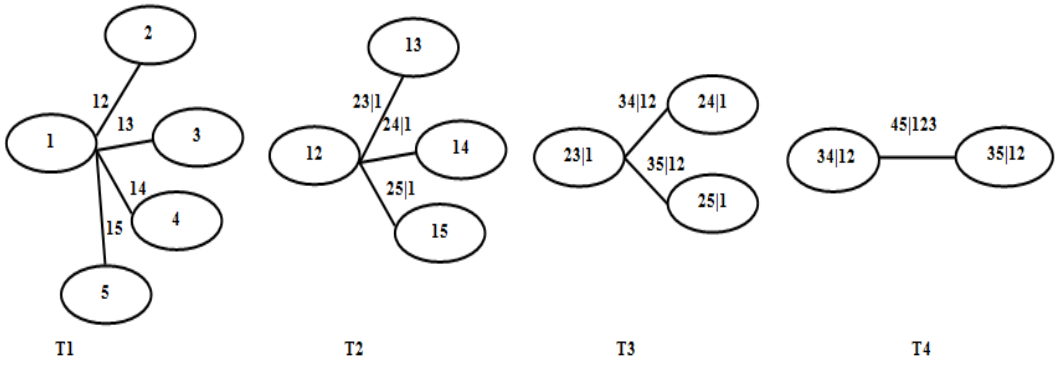

Figure 2.1 Five dimensions of C-vine structure

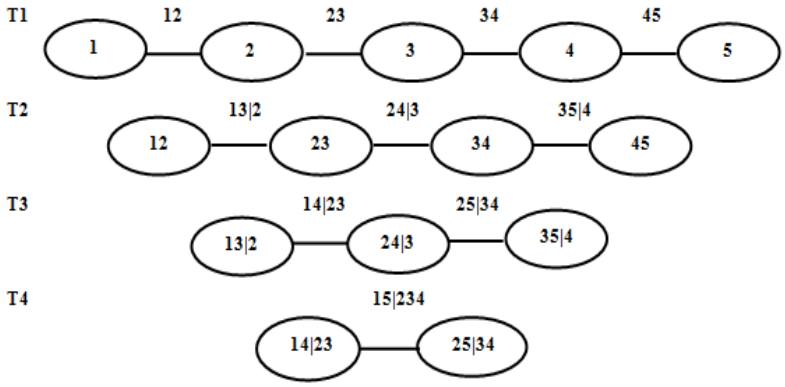

Figure 2.2 Five dimensions of D-vine structure

After we got the appropriate vine structure, we use the maximum likelihood method to estimate the parameters of copulas in the vine structures as follows:

$$\ell_C(\boldsymbol{\delta}|\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n-1}\sum_{k=1}^{n-j} log\left[c_{j,j+k|1:(j-1)}\left(F\left(x_j|v_1\right), F(x_{j+k}|v_1)|\delta_{i,i+j|1:(i-1)}\right)\right], \quad (2.40)$$

$$\ell_D(\boldsymbol{\delta}|\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n-1}\sum_{k=1}^{n-j} log\left[c_{k,k+j|(k+1):(k+j-1)}\left(F(x_k|v_2), F(x_{k+j}|v_2)|\delta_{k,k+j|(k+1):(k+j-1)}\right)\right] \quad (2.41)$$

where $c_{j,j+k|1:(j-1)}$ and $c_{k,k+j|(k+1):(k+j-1)}$ are the bivariate copula densities in equation 2.37 and equation 2.38.

2.5 Extreme Value Theory

Lastly, this methodology is the combination between the probability theory and mathematical statistics for analyzing about rare events. Normally, there are two important methods for identifying in extreme data, namely the peak over threshold method (POT) and block maxima method (BM). The first approach observes the extreme data that exceeds over a given threshold. The second approach selects the extreme data from maximum value of data per period. With regards to the study by Embrechts et al (1997), McNeil and Frey (2000), this thesis applies EVT with econometric model in high frequency data (financial data) that focuses on POT method. The popular one of POT method to approximate the excess tail events in financial context is the Generalized Pareto Distribution as below.

Definition of Generalized Pareto Distribution (GPD): Let $x_{1,\dots,}x_n$ are random data that satisfy the property of independent identically distributed. The Generalized Pareto distribution function can be defined as

$$F_{\eta,\vartheta}(x)\begin{cases}1-\left(1+\eta\left(\frac{x}{\vartheta}\right)\right)^{\frac{-1}{\eta}} & if\ \eta\neq 0\\ 1-\exp\left(-\frac{x}{\vartheta}\right) & if\ \eta=0\end{cases} \tag{2.42}$$

where $\vartheta > 0, x > 0$ if $\eta \geq 0$ and $0 \leq x \leq -1/\eta$ are the condition of GPD. ϑ and η are scale and shape parameters, $x_i = y_i - u$ is the exceedance over a given threshold u.

Definition of Generalized Pareto probability density function: it can be defined by the function $f_{\eta,\vartheta}(x)$ as

$$f_{\eta,\vartheta}(x)\begin{cases}\left(\frac{1}{\vartheta}\right)\left(1+\eta\left(\frac{x}{\vartheta}\right)\right)^{-1-\frac{1}{\eta}} & if\ \eta \neq 0\\ \left(\frac{1}{\vartheta}\right)-\exp\left(-\frac{x}{\vartheta}\right) & if\ \eta = 0\end{cases} \tag{2.43}$$

where $f(x=0)=\left(\frac{1}{\vartheta}\right)$. To find the optimal parameters in GPD, we maximize the log likelihood function of the n observations as follows:

$$\ell(\eta,\vartheta)=\sum_{j=1}^{n} log f_{\eta,\vartheta}(x_j)$$

$$\ell(\eta,\vartheta)=\begin{cases}-nlog\vartheta-\left(\frac{1+\eta}{\eta}\right)\sum_{j=1}^{n} log\left(1+\eta\left(\frac{x_j}{\vartheta}\right)\right) & if\ \eta \neq 0\\ -nlog\vartheta-\frac{1}{\vartheta}\sum_{j=1}^{n} x_j & if\ \eta = 0\end{cases} \tag{2.44}$$