Chapter 3

Methodology

3.1 Data

The data have been gathered from the Chicago Ethanol Spot data (USD per barrel) and the North Sea (Forties) spot Crude Oil (USD per barrel). The data are collected from EcoWin. The data span is from November 4, 2005, to December 26, 2013, at a daily frequency, which amounts to a total of 1,188 observations. The daily return was computed as $R_{i,t} = \ln(p_{i,t} / p_{i,t-1})$, where $P_{i,t}$ and $P_{i,t-1}$ are the daily spot prices for days "t" and "t-1" for market i.

3.2 Conceptual Framework

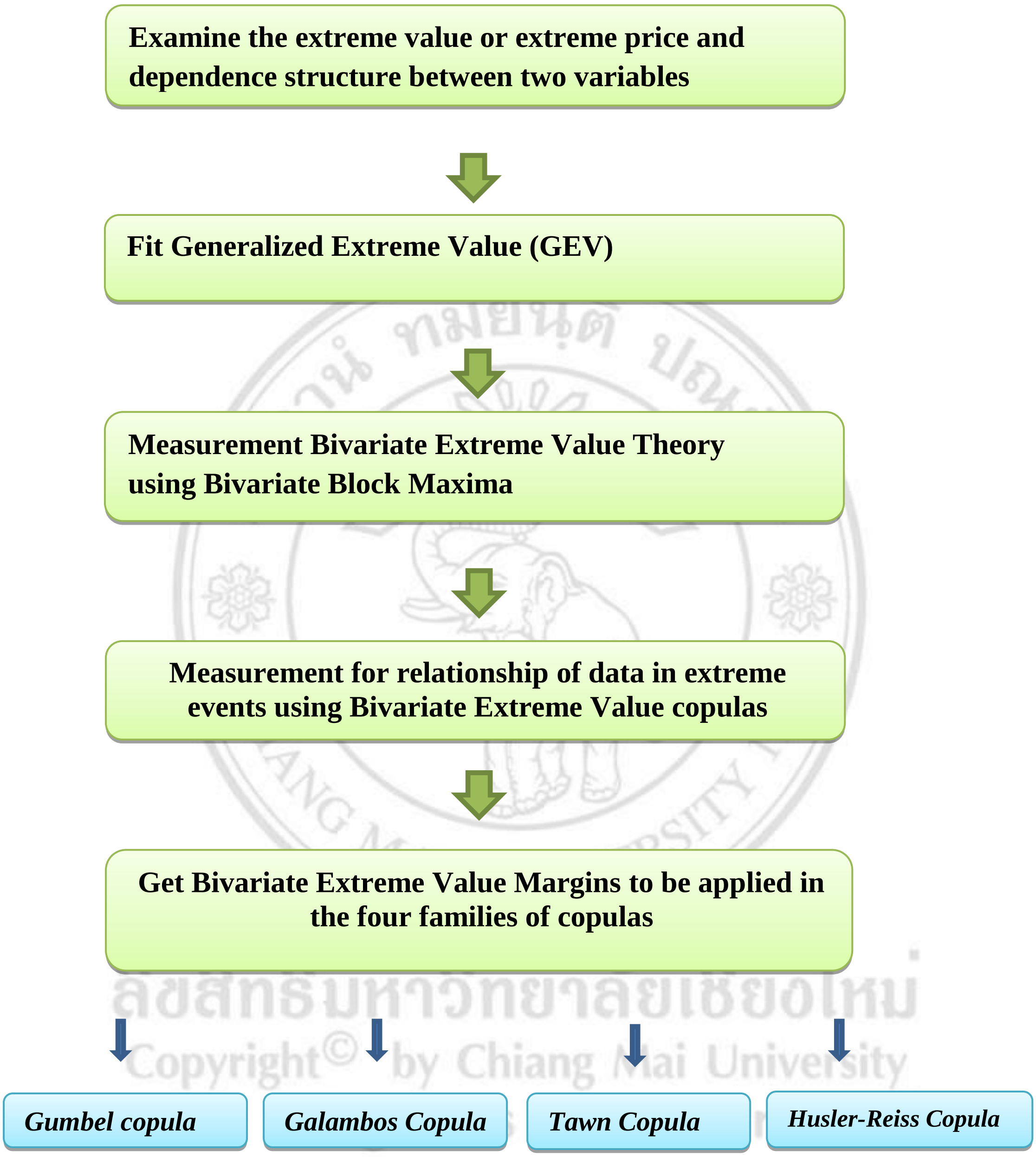

Figure 3.1 The conceptual framework

3.3 Research Methodology

3.3.1 Generalized Extreme Value (GEV) distribution

For a single margin, M_i is the maxima sequence, which is the same as defined before, and "i" is the number of blocks. F is the general price distribution, and G is the asymptotic extreme value distribution. The EVT shows that by founding a series of a_n and b_n, the maxima can be converted to the general extreme value distribution (GEV) G (Coles, 2001; Beirlant, 2004), given as follows:

$$G(x;b,a,\xi) = exp\left\{-\left[1+\xi\left(\frac{x-b}{a}\right)\right]^{-1/\xi}\right\}. \quad (3.1)$$

Where ξ is the shape parameter explaining the behavior of the tail of the distribution. When $\xi < 0$ the distribution is the Weibull, $\xi > 0$ the Fréchet, and $\xi = 0$ the Gumbel.

3.3.2 Bivariate block maxima

The bivariate block maxima model is investigated with non-parametric and parametric cases. The parametric models that can summarize the bivariate BM are provided below (Chuangchid et al, 2012):

$$G(x,y) = exp\left\{\log\left(G_1(x)G_2(y)\right) A\left(\frac{\log\left(G_2(y)\right)}{\log\left(G_1(x)G_2(y)\right)}\right)\right\}, \quad (3.2)$$

where x = ethanol price, y = crude oil price

G_1 = margin of ethanol price

G_2 = margin of crude oil price

A(t) = dependence structure between the margins of ethanol price and crude oil price, which is as follows:

1) A(t) is convex:

2) $max\{(1-t),t\} \leq A(t) \leq t$; and

3) A(0) = A(1) = 1.

The second property shows that the lower bound corresponds to the complete dependence $G(x,y) = \min\big(G_1(x)G_2(y)\big)$, whereas the upper bound corresponds to (complete) independence $G(x,y) = \big(G_1(x)G_2(y)\big)$.

For A(t), this paper chose one parametric model from nine models, using minimum Akaike Information Criterion (AIC) in the bivariate block maxima case. The nine parametric models are given in the following discussion, where "x" is the ethanol price, "y" is crude oil price, and "r" is the parameter of dependence between the ethanol price and the crude oil price (Stephenson, 2011).

1. Model Number 1 = "log" (Gumbel, 1960).

$$G(x,y) = \exp\left[-\left(x^{\frac{1}{r}} + y^{\frac{1}{r}}\right)^{r}\right], \tag{3.3}$$

where $0 < r \leq 1$. Complete independence is obtained when r = 1. Complete dependence is when $r \rightarrow 0$.

2. Model Number 2 = "alog" (Tawn, 1988)

$$G(x,y) = \exp\{-(1-t_1)x - (1-t_2)y - [(t_1x)^{1/r} + (t_2y)^{1/r}]^r\}, \tag{3.4}$$

where $0 < r < 1$ and $0 \leq t_1, t_2 \leq 1$. When $t_1 = t_2 = 1$, the asymmetric logistic model becomes equivalent to the logistic model. When r = 1, and either $t_1 = 0$ or $t_2 = 0$, there is said to be complete independence. When $t_1 = t_2 = 1$ and, $r \rightarrow 0$, there is said to be complete dependence.

3. Model Number 3 = "hr" (Husler and Reiss, 1989)

$$G(x,y) = \exp\left(-y_1\Phi\left\{r^{-1} + \frac{1}{2}r[log(x/y)]\right\} - y_2\Phi\left\{r^{-1} + \frac{1}{2}r[log(y/x)]\right\}\right) \tag{3.5}$$

$\Phi(\cdot)$ is the standard normal distribution function and r > 0.

When r reaches zero ($r \to 0$), there is said to be independence.

When r moves to infinity, it can be said that there is complete dependence.

4. Model Number 4 = "neglog" (Galambos, 1975)

$$G(x, y) = \exp\left[-(x - y + [x^{-r} + y^{-r}])^{-1/r}\right], \tag{3.6}$$

where r > 0. In the limit, as $r \to 0$, it denotes complete independence.

When r moves to infinity, there is said to be complete dependence.

5. Model Number 5 = "aneglog" (Joe, 1990)

$$G(x, y) = \exp\left[-x - y + [(t_1 x)^{-r} + (t_2 y)^{-r}]^{-1/r}\right] \tag{3.7}$$

where r > 0 and $0 < t_1, t_2 \leq 1$. When either t_1= 0 or t_2 =0, and r reaches 1, there is said to be complete independence. When $t_1 = t_2 = 1$ and r moves to infinity, it can be said that there is complete dependence. When t_1 and t_2 = are fixed, and r moves to infinity, then these are different limits.

6. Model Number 6 = "bilog" (Smith, 1990)

The equation foe the parameters α and β is

$$G(x, y) = \exp\{-xq^{1-\alpha} - y(1-q)^{1-\beta}\}, \tag{3.8}$$

where$q = q(x, y;\ \alpha, \beta)$ is the root of the equation

$$(1-\alpha)x(1-q)^{\beta} - (1-\beta)yq^{\alpha} = 0, \tag{3.9}$$

and $0 < \alpha, \beta < 1$. When $\alpha = \beta$, the bilogistic model equals the logistic model with the dependence parameter, dependence $= \alpha = \beta$. When $\alpha = \beta$ reaches 0, there is said to be complete dependence. When one of α, β is fixed and the other reaches 1, difference limits are said to happen. Independence is obtained as $\alpha = \beta$ approaches 1.

7. Model Number 7 = "negbilog" (Coles and Tawn, 1994)

The equation for α and β is

$$G(x, y) = \exp\{-x - y + xq^{1+\alpha} + y(1-q)^{1+\beta}\}, \quad (3.10)$$

where $q = q(x, y;\ \alpha, \beta)$ is the root of the equation

$$(1+\alpha)xq^{\alpha} - (1+\beta)y(1-q)^{\beta} = 0, \quad (3.11)$$

and $\alpha > 0$ and $\beta > 0$. When $\alpha = \beta$, the negative bilogistic model equals the negative logistic model with dependence parameter $= 1/\alpha = 1/\beta$. In the limit, as $\alpha = \beta$ reaches 0, there is said to complete dependence. When $\alpha = \beta$ moves to infinity, and when one of α, β is fixed and the other moves to infinity, it can be said that there is independence.

8. Model Number 8 = "ct" (Coles and Tawn, 1994)

Let the parameters be $\alpha > 0$ and $\beta > 0$

$$G(x, y) = \exp\left[-x[1 - Be(q; \alpha+1, \beta)] - y[1 - Be(q; \alpha+\beta, 1)]\right], \quad (3.12)$$

where $q = \alpha y/(\alpha y + \beta x)$ and $Be(q; \alpha, \beta)$ is the beta distribution function evaluated at q with shape 1 = α and shape 2 = β. In the limit, as $\alpha = \beta$ moves to infinity, there is said to be complete dependence. When $\alpha = \beta$ reaches zero, and when one of α, β is fixed and the other reaches 0, it can be said that there is independence.

9. Model Number 9 = "amix" (Tawn, 1988)

Let the parameters α and β have a dependence function in the following cubic polynomial form:

$$A(t) = 1 - (\alpha + \beta)t + \alpha t^2 + \beta t^3, \tag{3.13}$$

where $\alpha \geq 0$ and $(\alpha + 3\beta) \geq 0, (\alpha + \beta) \leq 1$ and $(\alpha + 2\beta) \leq 1$. The beta then lies in the interval [-0.5, 0.5] and the alpha in [0, 1.5]. The alpha could be larger than 1 if $\beta < 0$. When both the parameters are zero, it can be said that there is independence.

3.3.3 Parametric models of copulas

3.3.3.1 Gumbel copula (logistic copula)

Invented by Gumbel (1960), the Gumbel or logistic, copula is the oldest of the EVC models. It belongs to both the Archimedean and the extreme value copulas. The dependence function A(w) is given as follows:

$$A(w) = [(1-w)^r + w^r]^{1/r}, \tag{3.14}$$

where $r \geq 1$. The parameter r measures the degree of dependence, ranging from complete independence ($r = 1$) to complete dependence ($r = \infty$). Therefore, the Gumbel extreme value copula is given as

$$C(u_1, u_2) = \exp\left\{-[(-\ln u_1)^r + (-\ln u_2)^r)^{1/r}\right\}. \tag{3.15}$$

3.3.3.2 Galambos copula (negative Logistic Model)

Let, $\hat{C}_\phi$ be the distribution of the (1-U1,…, 1-Ud) random vector. The tail dependence function could be written as follow:

$$C_*(u_1, \dots, u_k) = exp\left[- \sum_{\substack{J \subset \{1,\dots,k\} \\ |J| \geq 2}} (-1)^{|J|} \left\{ \sum_{j \in J} (-log u_j)^{-\alpha} \right\}^{-1/\alpha} \right] \prod_{j=1}^{k} u_{j,} \quad \alpha > 0. \tag{3.16}$$

3.3.3.3 Tawn copula (asymmetric logistic copula)

The Tawn copula, or (asymmetric logistic copula), is much more flexible and combine several existing models such as the logistic model ($\emptyset = \theta = 1$), and a mixture of logistic and independence models. Complete dependence corresponds to $\emptyset = \theta = 1$ and $r = \infty$, whereas complete independence corresponds to $\emptyset = 0$ or $\theta = 0$ or r=1. The dependence function is as follows:

$$A(w) = \ [\theta^r (1-w)^r + \emptyset^r w^r\ ]^{1/r} + (0 - \emptyset)w + 1 - \theta, \tag{3.17}$$

with $\emptyset \leq 1$ or $\theta \geq 0$ and $r \geq 1$, and the copula function

$$C(u_1, u_2) = \exp\left\{\ln u_1^{1-\theta} + \ln u_2^{1-\phi} - [(-\theta \ln u_1)^r + (-\phi \ln u_2)^r\ ]^{1/r}\right\}. \tag{3.18}$$

3.3.3.4 Husler-Reiss (HR) copula

The drawbacks of the logistic and the negative logistic copulas are that they are too limited for large dimensional problems since the dependence is described only by a single parameter θ. However, the HR copula does not have this problem; we give the corresponding distribution of the bivariate case:

$$C_*(u_1, u_2) = exp\left[\Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{log u_2}{log u_1}\right)\right\}\log u_1 + \Phi\left\{\frac{a}{2} + \frac{1}{a}\log\left(\frac{\log u_1}{\log u_2}\right)\right\}\log u_2\right], \tag{3.19}$$

where Φ is the standard normal cumulative distribution function.

In our case, specifically, let u1 be the ethanol price return marginal and "v" be the crude oil price marginal. We apply from the above mentioned discussion the four EV copulas to calculate the dependence of the two energy prices.

3.3.4 Kendall tau Dependence Measure

The Kendall tau can be expressed uniquely in terms of the copula; it is in the range [-1, 1].

$$\tau = 4\int_0^1\int_0^1 C(u,v)\,dC(u,v) - 1. \qquad (3.20)$$

Especially, in terms of the dependence function A, the particular Kendall tau is given as follows:

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} A''(t)dt. \qquad (3.21)$$

3.3.5 Extreme value copulas

Extreme value copulas could be analyzed to find suitable models to obtain the dependence structure of the extreme values, with the presence of the component wise maxima. Here, we consider the bivariate case for our specific problem. Let $X_i = (X_{i1}, X_{i2}), i \in \{1,\dots,n\}$ be an i.i.d. sample random vectors with general distribution function F, margins F_1, F_2, and copula C_F. F is assumed to be continuous. Consider the vector of the component wise maxima:

$$M_n = (M_{n,1}, M_{n,2}), \qquad where\ M_{n,j} = \bigvee_{i=1}^{n} X_{ij}, \qquad (3.22)$$

Because the joint functions of M_n are given by F^n, and the marginal distributions are expressed by F_1^n and F_2^n, the copula is C_n of M_n:

$$C_n(u_1,u_2) = C_F(u_1^{1/n}, u_2^{1/n})^n. \qquad (3.23)$$

It is clear that the extreme value copula is the same as the Generalized Extreme Value (GEV) distribution, which shares the max-stable property (Gudendorf and Segers, 2009). Therefore, the simple of extreme- value copulas could be obtained by employing the max-stability. Also, we can see from the literature studies that copula is max-stable if and only if it is an extreme- value copula. The understanding of extreme value copula is when we know the maxima distribution; here, we know the joint maxima distribution. This is the point at which the extreme value copula is different from other copulas, and also gives the evidence to use the GEV as the margin.