

Chapter 3

Data Selection and Econometric Models

In this chapter, we will discuss about the data selection and econometric models from fore proposes of this research which we introduce in chapter.

3.1 Data Selection

There are five levels of natural rubbers, from RSS1 to RSS5. The highest level is RSS1. However, RSS3 is primarily used in current and future world markets. Table 1 presents the numerous regions that import rubber from Thailand. The top six regions in terms of export volume are China, Malaysia, Japan, Europe, the United States, and South Korea. Rubber exports comprise approximately 90% of the total output of rubber in Thailand. For my first part of my study, seven variables were examined: one concerning the rubber price in Thailand and six concerning the exchange rate in the six regions mentioned above. Each variable involved 1577 observations. The first case aims to examine the relationship between the rubber price in Thailand and the exchange rates of countries importing Thai rubber. Table 1 lists the variable names.

Table 3.1 The first case of Introduce of Variable Names

Variables	Names
PRICE	Rubber price
BAHT	Exchange rate of Thailand Baht/US Dollar
CNY	Exchange rate of Chinese Yuan/US Dollar
EUR	Exchange rate of Euro Dollar/US Dollar
JPY	Exchange rate of Japanese Yen/US Dollar
KRW	Exchange of rate of Korea Won/US Dollar
MYR	Exchange rate Malaysia Riggit /US Dollar

For the second case of my study, given that Thailand trade depends highly on the USA and Japan, the exchange rate of Thai baht is a crucial factor. Other uncontrollable elements, such as tsunamis, floods and political environments, also have a direct effect on the exchange. Therefore, the US dollar/Thai baht index was chosen as the first variable.

The second variable chosen was the crude oil price. Two kinds of crude oil are traded in the futures market in Asia; these are traded exclusively by the Tokyo Commodity Exchange (TOCOM).

This study used daily data from May 28, 2004 to Dec 31, 2010, i.e., a total of 1,581 observations to match the first trading day of AFET for Thai rubber. This study aims to determine the relationships among exchange rate, crude oil, and Thai rubber. Therefore, crude oil traded in TOCOM was selected.

Finally, the growth in rubber output is closely related to seasonality. Due to the fact that temperature and precipitation are important factors in natural rubber output

(as mentioned in the introduction), the variables representing the production environment of rubber were included. Thus, climatic data from 25 locations with high rubber outputs were chosen.

The complete set of variables assumed to be related to volatility in Thai rubber prices used in the study is presented in Table 2.

Table 3.2 The second case of Introduce of Variable Names

Variables	Unit	Names
Rubber price	Baht	Export Price of Thai natural rubber
US dollar	Index number	US dollar index at close – Trade weighted
Crude oil	Index number	Crude oil index in TOCOM
TempD	Celcius	Difference between today's temperature from yesterday, 1581 observations, which is made up of average temperatures by top 25 rubber producing areas in Thailand. These are: Burirum, Chanthaburi, Chon buri, Chumphon, Krabi, Nakhon Si Thammarat, Narathiwat, Nong Khai, Pattani, Phangnga, Phattaluang, Phetchabun, Phitsanulok, Ranong, Rayong, Sakon Nakhon, Satun, Si Sa Ket, Songkhla, Surat Thani, Trad, Trang, Ubon Ratchathuni, Udon Thani and Yala.
Rainfall	mm	Average precipitation per day, 1581 observations, where the average precipitation is from the top 25 rubber producing areas named above.

Compared with the Agricultural Futures Exchange of Thailand (AFET), both the TOCOM and Singapore Commodity Exchange Limited (SICOM) possess higher volumes in the world. The volatility of oil price can affect the price of natural rubber as well as the price of synthetic rubber, which is the substitute of natural rubber. In Asia, two kinds of oil are traded in the futures market. TOCOM only trades crude oil and gas oil. Regarding this, five variables are concerned with futures products from the three futures markets mentioned above. For each variable, 1,609 observations are required from May 28, 2004 (the first trading day of AFET) to December 31, 2010. The third case aims to identify the relationships between the futures prices of rubber and oil. The variable names are introduced in Table 3.3.

Table 3.3 The Third case study of Introduce of Variable Names

Variables	Names
AFET	Rubber Futures in AFET
SICOM	Rubber Futures in SICOM
TOCOM	Rubber Futures in TOCOM
Crude	Crude Oil Futures in TOCOM
Gas	Gas Oil Futures in TOCOM

3.2 The Statistics of the Variables

The returns of asset i at time t are calculated as following:

$$R_{i,t} = \log \left(\frac{P_{i,t}}{P_{i,t-1}} \right) \quad (6)$$

were $P_{i,t}$ and $P_{i,t-1}$ are the closing prices of asset i for days t and $t-1$, separately.

All series data are stationary and tested by using the Augmented Dickey-Fuller (ADF) test, which is given as following:

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \varepsilon_t \quad (7)$$

The null hypothesis is $\theta = 0$ which, if rejected, then means that the series y_t is stationary, or it is not stationary.

3.3 VARMA-GARCH and VARMA-AGARCH model

We apply VARMA-GARCH model to analyze the data proposed by Ling and McAleer (2003) and VARMA-AGARCH model proposed by McAleer et al., (2009).

The effect of fluctuation cannot be distinguished individually very clearly in the traditional multivariate GARCH model.

The VARMA-GARCH model is expressed as:

$$Y_t = E(Y_t | F_{t-1}) + \varepsilon_t \quad (8)$$

$$\varepsilon_t = D_t \eta_t \quad (9)$$

$$H_t = \omega + \sum_{j=1}^r \alpha_{ij} \varepsilon_{i,t-j} + \sum_{j=1}^s \beta_{ij} H_{i,t-j} \quad (10)$$

And VARMA-AGARCH model is in following:

$$H_t = \omega + \sum_{j=1}^r \alpha_{ij} \varepsilon_{i,t-j} + \sum_{j=1}^r C_{ij} I_{ij} \varepsilon_{i,t-j} + \sum_{j=1}^s \beta_{ij} H_{i,t-j} \quad (11)$$

Where $H_t = (h_{1t}, h_{2t}, \dots, h_{mt})$, $\eta_t = (\eta_{1t}, \eta_{2t}, \dots, \eta_{mt})$,

$$D_t = \text{diag}(h_{1t}^{1/2}, h_{2t}^{1/2}, \dots, h_{mt}^{1/2})$$

For the first case study, the full model is in following:

$$A_t = \gamma_{A0} + \gamma_{A1} P_{t-1} + \gamma_{A2} B_{t-1} + \gamma_{A3} C_{t-1} + \gamma_{A4} E_{t-1} + \gamma_{A5} J_{t-1} + \gamma_{A6} K_{t-1} + \gamma_{A7} M_{t-1} + \varepsilon_{A,t} \quad (12)$$

Where P is PRICE, B is BAHT, C is CNY, E is EUR, J is JPY, K is KRW, M is MYR and ε is error term.

For the second case study, the full model is in following:

$$A_t = \gamma_{A0} + \gamma_{A1}A_{t-1} + \gamma_{A2}B_{t-1} + \gamma_{A3}C_{t-1} + \gamma_{A4}D_{t-1} + \gamma_{A5}E_{t-1} + \varepsilon_{A,t}$$

$$\begin{bmatrix} \varepsilon_{A,t} \\ \varepsilon_{E,t} \end{bmatrix} | \Omega_{t-1} \sim N(0, H_t)$$

Where A is the export price of natural rubber in Thailand, B is the futures price of crude oil in TOCOM, C is the dollar index, D is the difference of average temperature with yesterday, E is the average precipitation and ε is error term.

We use normal distribution and MLE (Maximization Likelihood Estimation) procedure to estimate the parameters of this model.

$$\hat{\theta} = \operatorname{argmin}_{\theta} \frac{1}{2} \sum_{t=1}^n (\log |Q_t| + \varepsilon_t' Q_t^{-1} \varepsilon_t) \quad (15)$$

Where θ is the vector of parameters to be estimated on the conditional log-likelihood function, and $|Q_t|$ is the determinant of Q_t , the conditional covariance matrix.

3.4 GARCH model

Bollerslev (1986) proposed the GARCH model which put conditional variance of lags in to ARCH model and make it general. The GARCH model is given by:

$$R_t = \mu_{t-1} + \varepsilon_t \quad (16)$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, h_t) \quad (17)$$

$$H_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (18)$$

When $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$, the GARCH model is stable.

GARCH (p,q) model can be describes as follows:

$$R_t = \mu_{t-1} + \varepsilon_t \quad (19)$$

$$E_t[\Omega_{t-1} \sim N(0, h_t)] \quad (20)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^p \beta_j h_{t-1} \quad (21)$$

Where $i=1,2,\dots,q$, $J=1,2,\dots,p$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$.

In this model, μ_{t-1} is the conditional mean of R_t at time t , h_t is the conditional variance and Ω_{t-1} is the all useful information set at time $t-1$.

The GARCH (1,1) model can be described as follows:

$$r_{i,t} = c_0 + c_1 r_{i,t-1} + e_{i,t}, e_{i,t} | \Psi_{t-1} = h_{i,t}, z_{i,t}, z_{i,t} \sim \text{skewed-t}(z_i | \eta_i, \lambda_i) \quad (22)$$

$$e_{i,t} \sim N(0, h_{i,t}^2) \quad (23)$$

$$h_{i,t}^2 = \omega_{i,t} + \alpha e_{i,t-1}^2 + \beta h_{i,t-1} \quad (24)$$

Where $\omega_{i,t} > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$.

The error term $e_{i,t}$ is assumed to be skewed-t distribution which can be used to describe the possibly asymmetric and heavy tail characteristics of each variable.

Following Hansen (1994), the density function is

$$\text{skewed-t}(z|\eta, \lambda) = \begin{cases} BC \left(1 + \frac{1}{\eta-2} \left(\frac{Bz+A}{1-\lambda} \right)^2 \right)^{-(\eta+1)/2}, & z < -\frac{A}{B} \\ BC \left(1 + \frac{1}{\eta-2} \left(\frac{Bz+A}{1-\lambda} \right)^2 \right)^{-(\eta+1)/2}, & z \geq -\frac{A}{B} \end{cases} \quad (25)$$

The value of A, B and C are defined in following:

$$A \equiv 4\lambda \frac{\eta-2}{C^{\eta-1}}, B \equiv 1 + 2\lambda^2 - A^2 \text{ and } C \equiv \frac{\Gamma(\eta+1/2)}{\sqrt{\pi(\eta-2)\Gamma(\eta/2)}} \quad (26)$$

Where λ and η are the asymmetry and kurtosis parameters, separately. Those are restricted to be $-1 < \lambda < 1$ and $2 < \eta < \infty$. When $\lambda = 0$, it will turn to the Student $-t$ distribution. If $\lambda = 0$ and η diverge to infinite, it will be the normal distribution.

3.5 Elliptical Copula

The copula function is used in discussing problems between many variables, and is also called the dependence function (Deheuvels, 1978). Sklar (1959) advances the copula theory, pointing out that one unit distribution can be analyzed to n marginal distribution and one copula function. Given that the number of parameters can be large, two-step methods are generally employed. Thus, in this paper, the marginal parameters were first estimated by optimizing the marginal log likelihoods independently of each other. Second, the copula parameters were estimated by optimizing the corresponding copula log likelihood at the second step.

The marginal log likelihoods function:

$$m\mathcal{L}(\theta; \mathbf{x}) = \sum_{i=1}^p \sum_{j=1}^T \log (F_i(x_{1,t}; \theta_i)) \quad (27)$$

The copula log-likelihood function:

$$c\mathcal{L}(\theta; \mathbf{u}, \emptyset) = \log (c(F_1(x_{1,t}), \dots, F_p(x_{p,t}); \theta)) \quad (28)$$

Therefore, the log likelihoods of two elliptical copula, the Gaussian and Student- t copula are given by:

$$\mathcal{L}_G(\mathbf{R}; \mathbf{u}_t) = -\frac{1}{2} \sum_{t=1}^T (\log |R| + \epsilon_t'(R^{-1} - I)\epsilon_t) \quad (29)$$

$$\begin{aligned}
\mathcal{L}_{St}(R, d, u_t) = & \\
& -T \log \frac{\Gamma(\frac{d+p}{2})}{\Gamma(\frac{d}{2})} - pT \log \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} - \frac{d+p}{2} \sum_{t=1}^T \log \left(1 + \frac{\epsilon_t' R^{-1} \epsilon_t}{d} \right) - \sum_{t=1}^T \log |R| + \\
& \frac{d+1}{2} \sum_{t=1}^T \sum_{i=1}^p \log \left(1 + \frac{\epsilon_{it}^2}{d} \right)
\end{aligned} \tag{30}$$

where the vector ϵ_t is the vector of the transformed standardized residuals which depends on the copula specification. For the Gaussian copula, the vector ϵ_t is defined as: $\epsilon_t = (\Phi^{-1}(u_{1,t}), \dots, \Phi^{-1}(u_{p,t}))$, which Φ^{-1} is the inverse univariate standard normal distribution. For the Student-t copula, it defined analogously as: $\epsilon_t = (t_d^{-1}(u_{1,t}), \dots, t_d^{-1}(u_{p,t}))$, which t_d^{-1} is the inverse student's t distribution with d degrees of freedom. In both of likelihoods R denotes the correlation matrix of ϵ_t .

The DCC (1.1) model of Engle (2002) defined that the degree of freedom parameter is static for the Student-t copula and the correlation R^t evolves through time.

$$Q_t = (1 - \alpha - \beta) \cdot \bar{Q} + \alpha \epsilon_{t-1} \cdot \epsilon_{t-1}' + \beta \cdot Q_{t-1} \tag{31}$$

$$R_t = \bar{Q}_t^{-1} Q_t \bar{Q}_t^{-1} \tag{32}$$

Where \bar{Q} is sample covariance of ϵ_t , \bar{Q}_t is a square $p \times p$ matrix with zeros as off-diagonal elements and diagonal element the square root of those of Q_t . The parameter constraints for the DCC are the same as for the univariate GARCH (1,1) models.

$$\alpha + \beta < 1, \alpha, \beta \in (0,1) \tag{33}$$

3.6 Archimedean Copula

In general, the Copulas of Gumbel and Clayton are commonly employed in the financial studies. Because the Gumbel copula is limited to the description of a positive dependence structure, this paper will use the Clayton copula, which possesses similar properties with the Gumbel copula, but it does not have a positive dependence restriction. The log likelihood of Clayton Copula is given by:

$$\mathcal{L}_{\text{Clayton}}(d; \mathbf{u}_t) = \sum_{t=1}^T \log \left((1+d)(u_{1t} \cdot u_{2t})^{-1-d} (u_{1t}^{-d} + u_{2t}^{-d} - 1)^{-2-\frac{1}{d}} \right) \quad (34)$$

With $d = \frac{2\mathcal{J}}{1-\mathcal{J}}$, where \mathcal{J} is the Kendall's tau and $\mathbf{u}_t = (u_{1t} \cdot u_{2t})$.

The equation for the Clayton Copula in Patton (2006) defined the evolution of dependency parameter as following:

$$\mathcal{J}_t = \Lambda(w + \beta\mathcal{J}_{t-1} + \alpha \cdot |u_{1,t-i} - u_{2,t-i}|) \quad (35)$$

Where: Λ denotes the logistic transformation: $\Lambda(x) = (1 + e^{-x})^{-1}$ in order to keep the parameters of Clayton copula in (0,1).

For the multivariate Copula model, it should be defined the model that when it is be fitted to a multivariate ($p > 2$) data set that consist of iid uniform $U(0,1)$ margins. Therefore, a filtration of the data set often through a GARCH model will extract the iid residuals from the raw data and transform them to uniform. The multivariate Copula is the decomposition to a product of bivariate Copula in a form of a nested set of $p-1$ trees. The first tree consists of $p-1$ bivariate Copula and the second tree consists of $p-2$ Copula and so on. The corresponding multivariate Copula log likelihood is the sum of these $(p(p-1))/2$ bivariate log likelihoods. This paper will follow the suggestion from Aas et al., (2009) that one should estimate the copula

parameters with the multi step method first and then use these estimates as starting values for the one step method. The multivariate Copula tree is written as following:

$$\begin{aligned}
 C(F(X1), \dots, F(X5)) = & \\
 C12 \cdot C13 \cdot C14 \cdot C15 & \quad \text{Tree1} \\
 C23|1 \cdot C24|1 \cdot C25|1 & \quad \text{Tree2} \\
 C35|12 \cdot C34|12 & \quad \text{Tree3} \\
 C45|123 & \quad \text{Tree4}
 \end{aligned} \tag{36}$$

For parameter estimation method, it uses the inference functions for margins (IFM) or calls two-stage estimation method to estimate parameter of copula-based GARCH modes. Joe (1997) showed that this estimator was closed to the maximum likelihood estimator and asymptotic efficiency to it. Therefore, the IFM will employ to compute the estimator efficiently without losing the realist information.

For the case study two, θ_p be the parameter of marginal distribution of spot price of rubber, θ_v be the parameters of marginal distributions of four variables, which are Oil, Dollar, TempD and Rain, separately, and θ_c be the parameters be the parameters in the Copula function Ct .

For the case study third, θ_p be the parameter of marginal distribution of futures price of rubber in AFET, θ_v be the parameters of marginal distributions of four variables, which are SICOM, TOCOM, Crude and Gas, separately, and θ_c be the parameters be the parameters in the Copula function Ct .

The likelihood function of $\theta = (\theta_p, \theta_v, \theta_c)$ is as following:

$$L_{p,v}(\theta) = L_p(\theta_p) + L_v(\theta_v) + L_c(\theta_c) \tag{37}$$

For the first stage, it will estimate the parameter of marginal distributions by the maximum likelihood method, separately,

$$\tilde{\theta}_p \equiv \arg \max \sum_{t=1}^T \log g_{p,t}(r_{p,t} | \Psi_{t-1}; \theta_p) \quad (38)$$

$$\tilde{\theta}_v \equiv \arg \max \sum_{t=1}^T \log g_{v,t}(r_{v,t} | \Psi_{t-1}; \theta_v) \quad (39)$$

For the second stage, given the marginal estimates obtained above, the dependence parameters are estimated by

$$\tilde{\theta}_c \equiv \arg \max \sum_{t=1}^T \log C_t(u_{p,t}, u_{v,t}, \tilde{\theta}_p, \tilde{\theta}_v; \theta_c) \quad (40)$$

Where: $u_{p,t} = (r_{p,t} | \Psi_{t-1}, \tilde{\theta}_p)$ and $u_{v,t} = (r_{v,t} | \Psi_{t-1}, \tilde{\theta}_v)$.