CHAPTER 3

Data Selection and Econometric Models

3.1 Data Selection

Among economic statistics on international tourism demand, international tourist arrivals is the key indicator required by the tourism industry, in particular for policy makers, marketers, and researchers. It is used in monitoring and assessing international tourism trends.

The number of tourist arrivals is the most common unit of measure used to quantify the volume of international tourism. Over the past few years, the international tourist arrivals variable is the most popular measure of tourism demand (Song and Li, 2008). Hence, this study use the number of tourist arrivals as the measure of tourism demand.

In the first case in my study, four variables were designated. There is the number of China's tourist arrivals to the following four destinations: Thailand, Singapore, South Korea, and Japan. The monthly data is from January 1993 to October 2011, yielding a total of 178 observations. The purpose of the first purpose is to estimate the volatility and dynamic dependent structure of tourism demand among the four destinations. The names of variables are listed in Table 3.1.

In the second case, we apply twelve variables, which are the number of China's inbound tourist arrivals from six top source countries and corroding exchange rate, namely South Korea, Japan, Russia, USA, Malaysia and Singapore, to examine the dependence between tourism demand and exchange rate. Monthly tourist arrivals and exchange rate are from January 1994 to December 2011, yielding a total of 216 observations. The names of variables are listed in Table 3.2.

Three variables are used in the third case, which are the number of Chinese tourist arrivals to three Southeast Asia destinations: Singapore, Thailand, and Malaysia, respectively. The sample period is from January 1998 to June 2012, which gives 174 observations for each destination. This case examines the volatility and co-movement between tourism demands in three China's outbound tourism markets. Table 3.3 shows the name of variable.

3.2 The copula function

A copula function is a statistical tool which can capture enables a flexible dependence structure between two (or more) random variables to be represented (Reboredo, 2011). In 1959, Sklar proved that a joint distribution can be separated into the marginal and a dependence function, which called a copula. According to the Sklar's (1959) theorem, the copula can be constructed as follow:

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) \quad (3.1)$$

where $F_{X,Y}(x, y)$s a joint distribution of two continuous random variables X and Y, with marginal functions $F_X(x)$ and $F_Y(y)$. The copula is a multivariate cumulative distribution function with uniform marginal U [0, 1] and V [0, 1], which relates the quantiles of the marginal distributions rather than the original variables. And it is also defined by

$$c(u, v) = Pr[U \leq u, V \leq v] \quad (3.2)$$

where u = $F_X(x)$ and v = $F_Y(y)$. Dependence (Eq. (3.2)) is invariant under strictly monotonic transformation of the variables u and v.

Subsequently, Patton (2006) presented the conditional copula function, which can be written as:

$$F_{XY|W}(x, y|w) = C(F_{X|W}(x|w), F_{Y|W}(y|w)|w) \quad (3.3)$$

where W is the conditioning variable, $F_{X|W}(x|w)$ is the conditional distribution of X|W = w, $F_{Y|W}(y|w)$is the conditional distribution of Y|W= w and $F_{XY|W}(x, y|w)$ is the joint conditional distribution of (X, Y)|W =w.

Differentiate Eq. (3.1) and Eq. (3.3), the corresponding unconditional and conditional joint densities are obtained:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)\frac{\partial^2 C(u,v)}{\partial u\,\partial v} = f_X(x)f_Y(y)c(u,v) \tag{3.4}$$

$$f_{XY|W}(x,y|w) = f_{X|W}(x|w)f_{Y|W}(y|w)\frac{\partial^2 C(u,v|w)}{\partial u\,\partial v}$$
$$= f_{X|W}(x|w)f_{Y|W}(y|w)c(u,v|w) \tag{3.5}$$

where $c(u,v)$ and $c(u,v|w)$ are the densities of unconditional and conditional copula , respectively. Hence, the conditional joint density of the two variables X and Y is represented by the product of the conditional copula density and the two conditional marginal densities, $f_{X|W}(x|w)$ and $f_{Y|W}(y|w)$.

There are two kinds of copula: tail independence and tail dependence. The tail dependence of copula can measure the dependence of the probability that two variables are in the lower or upper joint tails of bivariate distributions. We express the coefficient of right (upper) and left (low) tail dependence in terms of the copula between X and Y as:

$$\lambda_R = \lim_{u\to 1}\Pr[X \geq F_X^{-1}(u)|Y \geq F_Y^{-1}(u)] = \lim_{u\to 1}\frac{1-2u+c(u,u)}{1-u} \tag{3.6}$$

$$\lambda_L = \lim_{u\to 0}\Pr[X \geq F_X^{-1}(u)|Y \geq F_Y^{-1}(u)] = \lim_{u\to 0}\frac{1-2u+c(u,u)}{1-u} \tag{3.7}$$

where λ_R and λ_L belong to [0,1] and $F_X^{-1}(u)$ and $F_Y^{-1}(u)$ are the marginal quantile functions. If λ_R and λ_L are positive, then there are right and left tail dependence, otherwise there is right and left tail independence.

3.3 ARMAX-GARCH model

The ARMAX(γ, m)—GARCH (1,1) model can be described as follows:

$$r_{s,t} = c_0 + \sum_{i=1}^{r}\phi_{s,i}r_{s,t-i} + \sum_{j=1}^{m}\theta_{s,j}e_{s,t-j} + \sum_{n=1}^{12}\varphi_{s,n}D_n + e_{s,t} \tag{3.8}$$

$$e_{s,t} = \sqrt{h_{s,t}}z_{s,t} \tag{3.9}$$

$$h_{s,t} = \omega_s + \alpha_s {e^2}_{s,t-1} + \beta_s h_{s,t-1} \quad (3.10)$$

where $\phi_{s,i}$ is the autoregressive (AR) parameters, $\theta_{s,j}$ denotes moving average (MA) parameters, and γ and m are positive integers. And α_s and β_s are associated with the degree of innovation (ARCH effect) and volatility spillover effect (GARCH effect) from previous period, respectively. And the restrictions in the variance equation include $\omega_s > 0, \alpha_s, \beta_s \geq 0$, and $\alpha_s + \beta_s < 1$. D_n denotes the 12 month seasonal dummy variables and $\varphi_{s,n}$ is seasonal coefficient. $z_{s,t}$ is the standardized residual. It can be assumed to Gaussian distribution, Student-t distribution, GED, skewed-t distribution and so on.

3.4 The marginal distribution

According to the descriptive of the data, we choose the suitable distribution in different case. William (1908) discovered the student-t distribution, density of function is

$$\text{student} - t(z_{s,t}, \lambda_s) = f(z_{s,t}) = \frac{\Gamma\left(\frac{\lambda_s + 1}{2}\right)}{\sqrt{\lambda_s \pi}\,\Gamma\left(\frac{\lambda_s}{2}\right)} \left(1 + \frac{{z_s}^2}{\lambda_s}\right)^{-\frac{\lambda_s+1}{2}} \quad (3.11)$$

where λ is the number of degrees of freedom and $\Gamma$is the gamma function.

Hansen (1994) proposed a kind of skewed t distribution, which the density of function is

$$\text{skewed} - t(z_{s,t} | \eta_s, \lambda_s) = \begin{cases} nd\left(1 + \frac{1}{\eta_s - 2}\left(\frac{nx + m}{1 - \lambda_s}\right)^2\right)^{-\frac{\eta_s+1}{2}}, x < -\frac{m}{n} \\ nd\left(1 + \frac{1}{\eta_s - 2}\left(\frac{nx + m}{1 + \lambda_s}\right)^2\right)^{-\frac{\eta_s+1}{2}}, x \geq -\frac{m}{n} \end{cases} \quad (3.12)$$

The value of m,n and d are defined as: $m \equiv 4\lambda d \frac{\eta-2}{\eta-1}$, $n^2 \equiv 1 + 2\lambda^2 - n^2$ and $d \equiv \frac{\Gamma(\eta+1/2)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)}$, where λ and η are the asymmetry and kurtosis parameters, respectively. λ is restricted within (-1,1).

Fernandez and Steel (1998) proposed another kind of the skewed-t distribution, the skewed-t distribution λ_s degrees of freedom (df) has the following density:

$$\text{skewed}-t\left(z_{s,t}\middle|v_s,\lambda_s\right)=f\left(z_{s,t}\right)=\begin{cases}\dfrac{2}{v_s+\dfrac{1}{v_s}}f(v_s z_s), z_s<0,\\ \dfrac{2}{v_s+\dfrac{1}{v_s}}f\left(\dfrac{z_s}{v_s}\right), z_s\geq 0,\end{cases} \tag{3.13}$$

where $v_s$is the skewness parameter and λ_s is the degrees of freedom. Whenv_s is smaller (greater) than one, it is skewed to the left (right). If v_s =1, the skew-t distribution turn to be the usual Student's t distribution.

3.5 Static copula

We will use several copula models to measure the static dependence structure in China tourism market and capture the following patterns of dependence: static tail independence, symmetric and asymmetric tail dependence. In this subsection we will briefly describe these copula models and the statistical inference derived from them.

Followed by Jondeau and Rockinger (2002), the Gaussian Copula is

$$C_\rho^{Gau}(u,v|\rho)=\Phi_\rho\left(\phi^{-1}(u),\phi^{-1}(v)\right), \tag{3.14}$$

where Φ_ρ is the bivariate standardard normal cdf with the correlation ρ between u and v. And $\phi^{-1}(u)$ and $\phi^{-1}(v)$ are standard normal quantile functions. $\rho\in(-1,1)$ is the dependence between u and v.

The Student-t copula (Jondeau and Rockinger, 2002) is defined by

$$C_\rho^{Stu}(u,v|\rho,n)=T_{\rho,n}\left[t_n^{-1}(u),t_n^{-1}(v)\right], \tag{3.15}$$

where $T_{\rho,n}$ is the bivariate Student-t cdf with with a degree-of-freedom parameter n and correlation ρ. $t_n^{-1}(u)$ and $t_n^{-1}(v)$are the univariate Student-t quantile functions, with n as the df parameter. $\rho\in(-1,1)$, when $n\to\infty$, the Student-t copula converges to the Gaussian with zero dependence on the two side tails. Both Gaussian and Student-t copulas describe the symmetric dependence. But there is different feature. The feature of Gaussian copula does not have either left tail or right tail ($\lambda_L = \lambda_R = 0$). While the characteristic of the Student-t copula has non-zero dependence (or extreme-value dependence) in right and left tail, $\lambda_L = \lambda_R = 2t_{n+1}(-\sqrt{n+1}\sqrt{1-\rho}/\sqrt{1+\rho}) > 0$ (see Embrechts, Lindskog and McNeil, 2003).

The Gumbel and Clayton copulas reflect the asymmetric dependence. To capture the right (is also called upper tail, it means upper side tail not equal to zero $\lambda_R = 2 - 2^{1/\tau}$, while low side tail equal to zero $\lambda_L = 0$) tail dependence, the Gumbel copula (Gumbel, 1960) is defined as

$$C_\tau^{Gum}(u, v|\tau) = \exp\{-(\tilde{u}^\tau + \tilde{v}^\tau)^{1/\tau}\}, \tag{3.16}$$

where $\tilde{u} = -\ln(u)$ and $\tilde{v} = -\ln(v)$, and $\tau \in [1, +\infty)$ is the dependence between u and v. $\tau = 1$ shows no dependence and $\tau \rightarrow \infty$ represents a fully dependence relationship betweenu and v.

The Clayton copula captures the left tail dependence (is also called low tail, it means upper side tail equal to zero $\lambda_R = 0$, while low side tail not equal to zero $\lambda_L = 2^{-1/\tau}$), and we follow the Clayton (1978) and define as

$$C_\tau^{Clay}(u, v|\tau) = (u^{-\tau} + v^{-\tau} - 1)^{\ -1/\tau}, \tag{3.17}$$

where $\tau \in [0, +\infty)$ is the degree of dependence between u and v. $\tau = 0$ shows no dependence and the increase of the value of τ indciate the increase of the dependence between u and v.

The Plackett copula is followed by Nelsen (1999) and given by

$$C_\tau^{Pla}(u, v|\tau) = \frac{1}{2(\tau - 1)}\left(1 + (\tau - 1)(u + v)\right) - \sqrt{\left(1 + (\tau - 1)(u + v)\right)^2 - 4\tau(\tau - 1)uv}, \tag{3.18}$$

does not capture the tail dependence. The degree of dependence τ belongs to $[0, +\infty)$. If $\tau = 1$ means independence between u and v; if $\tau \rightarrow 0$ represents perfectly negative dependence, if $\tau \to \infty$ indicates perfectly positive dependence.

The Frank copula (Nelsen, 1999) is defined as

$$C_{\tau}^{Fra}(u, v|\tau) = -\frac{1}{\tau}\ln\left(1 + \frac{(e^{-\tau u} - 1)(e^{-\tau v} - 1)}{e^{-\tau} - 1}\right), \quad (3.19)$$

measures the symmetric dependence. $\tau \in [-\infty, +\infty)$, when $\tau = 0$, u and v are independent; when $\tau > 0$, they are positively dependent; and when $\tau < 0$, they are negatively dependent.

3.6 Dynamic copula

In the dynamic Gaussian and Student-t copulas, the Pearson correlation coefficient ρ_t is commonly used to describe the dependence structure. In this paper we assume that the dependence relies on the one lag dependence ρ_{t-1} and historical information $|(\mu_{t-1} - \mu_{t-2})(v_{t-1} - v_{t-2})|$ (Patton, 2006) in the dynamic Gaussian copula. Thus, the dynamic dependence process of the Gaussian follows

$$\rho_t = \Lambda(\alpha_c + \beta_c\rho_{t-1} + \gamma_c|(\mu_{t-1} - \mu_{t-2})(v_{t-1} - v_{t-2})|), \quad (3.20)$$

While in the dynamic Student-t copula, we follow Wu et al. (2012). The dynamic dependence process of the Student-t is

$$\rho_t = \Lambda(\alpha_c + \beta_c\rho_{t-1} + \gamma_c(\mu_{t-1} - 0.5)(v_{t-1} - 0.5)), \quad (3.21)$$

The conditional dependence, ρ_t determined from ρ_{t-1} captures the persistence effect. The products $(\mu_{t-1} - \mu_{t-2})(v_{t-1} - v_{t-2})$ and $(\mu_{t-1} - 0.5)(v_{t-1} - 0.5)$ capture historical information while $\Lambda = -\ln[(1 - x_t)/(1 + x_t)]$is the logistic transformation, which is used to ensure the dependence parameters fall within the interval (-1, 1).

The dynamic dependence process of the Gumbel (Wu et al., 2012) is

$$\tau_t = \Lambda(\alpha_c + \beta_c\tau_{t-1} + \gamma_c(u_{t-1} - 0.5)(v_{t-1} - 0.5)) \quad (3.22)$$

Followed Patton (2006), we change the historical information to $\frac{1}{10}\sum_{i=1}^{10}|a_{t-1} - b_{t-1}|$. We proposed dynamic dependence processes for Clayton copula as:

$$\tau_t = \Pi\left(\alpha_c + \beta_c\tau_{t-1} + \gamma_c \frac{1}{10}\sum_{i=1}^{10}|u_{t-1} - v_{t-1}|\right) \quad (3.23)$$

In the dynamic Gumbel and Clayton copula, the conditional dependence, τ_t determined from τ_{t-1} captures the persistence effect. In the dynamic Gumbel copula, $(u_{t-1} - 0.5)(v_{t-1} - 0.5)$ apply to capture historical information. In the Clayton dynamic copula, we use $\frac{1}{10}\sum_{i=1}^{10}|u_{t-1} - v_{t-1}|$ to capture historical information. And Π is also the logistic transformation, which is used to ensure the dependence parameters fall within the interval (-1, 1).

3.7 Estimation and calibration of the copula

The inference function for margins (IFM) method, full maximum likelihood estimation method (FML), Canonical Maximum Likelihood (CML) method and Non-parametric kernel density estimation-ML method are the methods to estimate and calibrate of the copula. The optimization procedure will confront problems in terms of extensive computation and estimate accuracy. To sum up, CML are superior to the other method. Therefore, we employ IFM and CML in this study.

Followed Jeo and Xu (1996), the two-step ML procedure (called inference function for marginal (IFM) method) is used in our study. In the procedure, the first step is to use the maximum likelihood (ML, Eq. (23)) to estimate the parameters of the marginal distribution while the second step is to use Eq. (24) to estimate the parameters of copula. The efficiency equations are

$$\hat{\theta}_{st} = \operatorname{argmax}\sum_{t=1}^{T} \ln f_{st}(z_{s,t}, \theta_{st}), s = 1,2 \quad (3.24)$$

and

$$\hat{\theta}_{IFM} = \operatorname{argmax}\sum_{t=1}^{T} \ln c_{st}(F_{st}(z_{1,t}), F_{2t}(z_{2,t}), \theta_{ct}, \hat{\theta}_{st}) \quad (3.25)$$

Standard maximum likelihood estimation CML (Canonical Maximum Likelihood method) is applied in our study. In the procedure, there are two steps. First step, after estimated the parameters of ARMA-GARCH model, the standardized innovations from the model of each variable are computed. Thus, we transform the standardized innovations series $z_{s,t}$ into uniform variate $\hat{u}_{i,t}$ and $\hat{v}_{j,t}$ $(i, j = s, \text{but } i \neq j)$ using the ECDF, which such that:

$$\hat{u} = \hat{F}_1(x_1) = \frac{1}{T+1}\sum_{s=1}^{T} I_{x_s \leq x_1} \tag{3.26}$$

$$\hat{v} = \hat{F}_2(x_2) = \frac{1}{T+1}\sum_{s=1}^{T} I_{x_s \leq x_2} \tag{3.27}$$

where T is the number of observations and I is the indicator function with I(expression) $= \begin{cases} 1 & \text{if expression is true} \\ 0 & \text{if expression is false} \end{cases}$.

Second, we apply the following equation to obtain the following CML estimate $\hat{\theta}_{CML}$, of the parameter, θ:

$$\hat{\theta}_{CML} = \arg\max \sum_{t=1}^{T} \ln c(\hat{u}, \hat{v}; \theta) \tag{3.28}$$

where θ is unknown parameter of the copula.

3.8 Uniform Distribution Test and Autocorrelation Test

Given the availability of the estimates of GARCH models, we turn to estimate copula functions for each pair of electricity price. For that, we transform the standardized residuals $z_{s,t}$ from GARCH model into the variates$\hat{u}_{i,t}$ and $\hat{v}_{j,t}$ $(i, j = s, \text{but } i \neq j)$, using the ECDF. Each variates $u_{i,t}(\text{or } \hat{v}_{j,t},\ ) i, j = s, \text{but } i \neq j$ should be uniform (0, 1), otherwise the copula model could be mis-specified. Following Patton (2006) and Reboredo (2011) this paper uses two steps to examine $u_{it}(\text{or } \hat{v}_{j,t},\ )$. The first test is to examine the serial correlation under the null hypothesis of serial independence, which is named a Ljung-Box (LB) test. Secondly, the Kolmogorow- Smirnov (KS) test is used to test the null hypothesis that the $u_{i,t}(\text{or } \hat{v}_{j,t},\ )$ are uniform (0, 1). The test statistic is defined as:

$$K_i = |A_i - O_i| \quad (3.29)$$

where Ai is the cumulative relative frequency for the theoretical distribution within each category, Oi is the corresponding value of the sample frequency

3.9 Goodness of Fit Test

The evaluations of the copula model have become a crucially important step. Therefore, Goodness-of-Fit (GOF) test was applied to the copula, based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. This paper used Genest, Remillard and Beaudoin's (2009) way to compute approximate P-values for statistics derived from this process consisting of using a parametric bootstrap procedure. The test statistic is the Cramer-von Mises functional:

$$S_n = \int_{[0,1]^d} C_n(x)^2 \, dC_n(x) = \sum_{i=1}^{n} \left\{ C_n(\hat{X}_i) - C_{\theta_n}(\hat{\hat{X}}_i) \right\} \quad (3.30)$$

which is from Genest, Remillard and Beaudoin (2009). And C_n is the empirical copula and C_{θ_n} is an estimator of C under the hypothesis that $H_0: C \in \{C_\theta\}$ holds. A p-value is less than 0.05, which indicate a rejection of the null hypothesis that the model is well specified.

Table 3.1 The name of variable in the first case

Variable	Name
South Korea	the number of China's tourist arrivals to South Korea
Japan	the number of China's tourist arrivals to Japan
Thailand	the number of China's tourist arrivals to Thailand
Singapore	the number of China's tourist arrivals to Singapore

Table 3.2 The name of variable in the third case

Variable	Name
South Korea	the number of South Korea's tourist arrivals to China
Japan	the number of Japan's tourist arrivals to China
Russia	the number of Russia's tourist arrivals to China
USA	the number of USA's tourist arrivals to China
Malaysia	the number of Malaysia's tourist arrivals to China

Table 3.3 The name of variable in the third case

Variable	Name
South Korea	the number of tourism arrivals to China from South Korea
Japan	the number of tourism arrivals to China from Japan
Russia	the number of tourism arrivals to China from Russia
USA	the number of tourism arrivals to China from USA
Malaysia	the number of tourism arrivals to China from Malaysia
Singapore	the number of tourism arrivals to China from Singapore
CNY/KRW	Exchange rate between CNY and KRW
CNY/JPY	Exchange rate between CNY and JPY
CNY/SUR	Exchange rate between CNY and SUR
CNY/USD	Exchange rate between CNY and USD
CNY/MYR	Exchange rate between CNY and MYR
CNY/SGD	Exchange rate between CNY and SGD