Thai **J**ournal of **M**athematics : (2016) 215–229
Special Issue on Applied Mathematics : Bayesian Econometrics

http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Forecasting Chinese International Outbound Tourists: Copula Kink AR-GARCH Model

Pimonpun Boonyasana† **and Warattaya Chinnakum**‡,1

†Faculty of Economics, Chiang Mai University
e-mail : pui.econ@gmail.com
‡Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail : warattaya_chin@hotmail.co.th

Abstract : The aim of this paper is to model and forecast Chinese tourist arrivals to Thailand, Singapore, and Malaysia. Monthly tourist arrivals from 1999 to 2014 are used in the analysis. In this paper, we propose Kink AR(m)-GARCH(p,q) model that combine the classical GARCH model of Bollerslev (1986)[1] with the Kink model of Chan and Tsay (1998). In additional, we assume that there are dependence between growth rate of tourist arrivals to Thailand, Singapore, and Malysia, from China. Copula approach was employed to capture these dependency. Therefore, Copula-base Kink AR(m)-GARCH(p,q) was used in this study. According to minimizing The results show that T-Copula 2-regimes Kink AR(1)-GARCH(1,1) model with normal, student-t, and skewed student-t error distributions, delivers the most accurate predictions.

Keywords : forecast; chinese tourist; copula; Kink AR-GARCH.
2010 Mathematics Subject Classification : 62M10; 62P20; 91B42.

1 Introduction

China is becoming the largest source market for international travel. According to World Bank [2], Chinese International outbound tourists increased from 4.5 million in 1955 to 31.0 million in 2005 and 120 million in 2015. Word Bank [2] also reported that expenditures of Chinese international outbound visitors in other countries is 104.5 billion US dollars (12 percent of world's tourism expenditure).

[1] Corresponding author.

The main driving forces for the increases included rising of China's GDP, favorable policies, and appreciation of Renminbi (RMB).

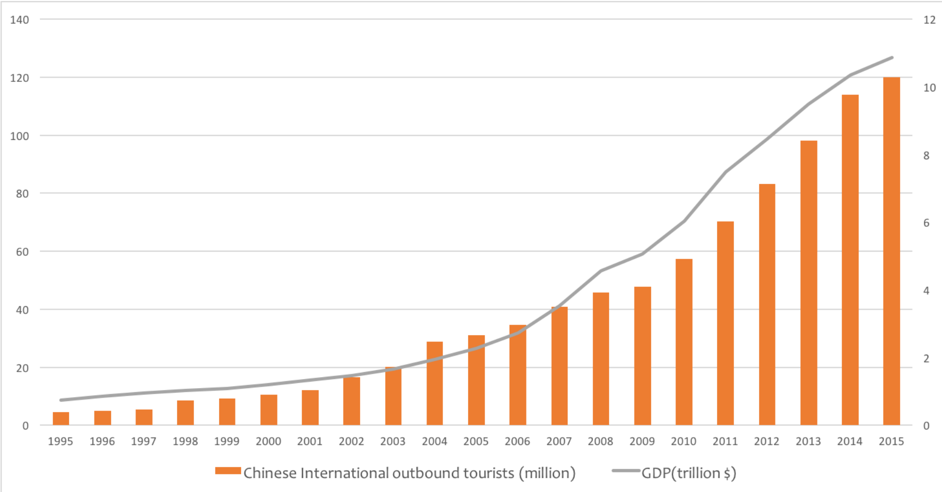

Figure 1: Chinese International outbound tourists compare with China's GDP (World Bank) [2]

Analyzing from travel destinations based on statistics collected by Euromoney Institution Investor Company [3], Asia especially ASEAN[2] country still dominated Chinas outbound tourism market. ASEAN countries namely, Thailand, Singapore, and Malaysia were among the top 10 destinations selected by Chinese travelers. The number of Chinese arrivals in ASEAN was 9.3 million, which accounted for about 10.5% of international tourist arrivals in ASEAN, and 11.1% of Chinese outbound tourists. From Figure 2 Thailand and Singapore have the highest market shares, 35.3% and 22.1%, respectively, followed by Malaysia and Vietnam, which account for 12.0% [3]. Untong and Kaosa-ard [4] assessed the competitive advantage in tourism of ASEAN countries in Chinese tourists market (except Brunei) and analyzed the growth of Chinese tourist arrivals for each ASEAN country. Their results show that Thailand has become the "Rising star". It has an increase in competitive advantage and can attract Chinese tourists more than ASEAN on average. Vietnam and Singapore ranked 2^{nd} and 3^{rd} in competitive advantage but their market grew more slowly than ASEAN, hence these countries were under "opportunity".

[2] ASEAN or the Association of Southeast Asian Nations has 10 member states that are Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam

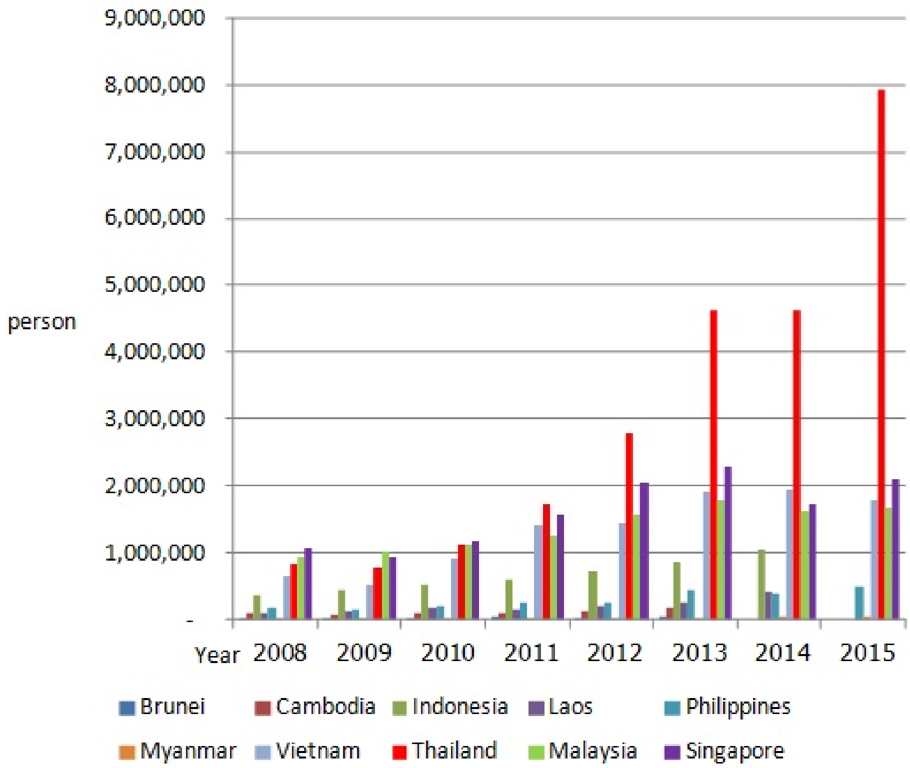

Figure 2: Chinese tourist arrivals to ASEAN (Euromoney Institutional Investor Company) [3]

Since Thailand, Singapore, and Malaysia, become favorite destinations for Chinese tourists, forecasting the Chinese tourist arrivals to Thailand, Singapore, and Malaysia is important for both public and private sectors. In the last few decades, numerous researchers have studied international tourism demand and a wide range of the available forecasting techniques have been tested. In this paper, we propose an alternative: Copula Kink AR(m)-GARCH(p,q) model to capture mean and volatility asymmetries in Chinese tourist outbound. The objective of this paper is to analyze the nonlinear behavior of Chinese tourist arrivals to Thailand, Singapore, and Malaysia and assess the forecasting performance of the model when applied to tourist data.

The remainder of the paper is organized as follows. Section 2 contains a brief literature review. In Section 3 we present a rigorous description of the methodology used in the analysis. Section 4 describes the data and presents the results of preliminary data analysis. The estimated models and empirical results for the Copula Kink AR(m)-GARCH(p,q) model are discussed in Section 5. Finally, Section 6 discusses the findings and draws conclusions.

2 Literature Review

The existing literature on forecasting tourism demand is wide ranging both in terms of the different techniques employed and in terms of the different countries covered. The selection of the most accurate forecasting model for a particular destination is often based on the out-of-sample forecasting performance. The mean absolute percentage error (MAPE) or root mean squared percentage error (RMSPE) are computed and compared.

Song and Li [5] reviewed the published studies on tourism demand modeling and forecasting since 2000. They found that there is no single model that consistently outperforms other models in all situations. Authors differ on the best method for tourism forecasting. For example, whereas Untong [6] analyzed Chinese tourist arrivals to Thailand during the last 25 years using the Ordinary Least Squares (OLS) technique, Sookmark [7] applied Seasonal Autoregressive Integrated Moving Average (SARIMA) and the Ordinary Least Squares (OLS) techniques, Chaitip et al. [8] provided non-linear forecasting model which is Markov Switching Vector Autoregressive model (MS-VAR model) and Min et al. [9] applied the belief function approach to statistical forecasting of tourist arrivals to Thailand.

3 Methodology

In this study, our propose is to model Chinese tourist arrivals to Thailand, Singapore, and Malaysia by using Kink AR(m)-GARCH(p,q). The behavior of Chinese tourists to three countries are assumed to have a correlation. The copula is an effective tool that is used to measure the dependence structure in many studies. Therefore, the Copula based AR(m)-GARCH(p,q). model was used in this study.

3.1 Copula Based AR(m)-GARCH(p,q) Model

First, we will introduce trivarite Kink AR-GARCH model and then we employ copula approach to capture the dependence of the Chinese tourist outbound to Thailand, Singapore, and Malaysia.

3.1.1 Trivarite Kink AR-GARCH Model

The Autoregressive (AR) process of order m is defined as

$$y_t^{(1)} = \alpha^{(1)} + \sum_{i=1}^{m} \beta_{1i}^{(1)} y_{t-i}^{(1)} I\left(y_{y-d}^{(1)} \le r^{(1)}\right) + \sum_{i=1}^{m} \beta_{2i}^{(1)} y_{t-i}^{(1)} I\left(y_{t-d}^{(1)} > r^{(1)}\right) + \varepsilon_{1,t}$$

$$y_t^{(2)} = \alpha^{(2)} + \sum_{i=1}^{m} \beta_{1i}^{(2)} y_{t-i}^{(2)} I\left(y_{y-d}^{(2)} \le r^{(2)}\right) + \sum_{i=1}^{m} \beta_{2i}^{(2)} y_{t-i}^{(2)} I\left(y_{t-d}^{(2)} > r^{(2)}\right) + \varepsilon_{2,t}$$

$$y_t^{(3)} = \alpha^{(3)} + \sum_{i=1}^{m} \beta_{1i}^{(3)} y_{t-i}^{(3)} I\left(y_{y-d}^{(3)} \leq r^{(3)}\right) + \sum_{i=1}^{m} \beta_{2i}^{(3)} y_{t-i}^{(3)} I\left(y_{t-d}^{(3)} > r^{(3)}\right) + \varepsilon_{3,t}, \tag{3.1}$$

where $y_t^{(1)}$ is growth rate of Chinese tourist outbound to Thailand at time t, $y_t^{(2)}$ is growth rate of Chinese tourist outbound to Singapore at time t, $y_t^{(3)}$ is growth rate of Chinese tourist outbound to Malaysia at time t. $\alpha^{(1)}$, $\alpha^{(2)}$, and $\alpha^{(3)}$ are mean, $\beta_{1i}^{(1)}$, $\beta_{1i}^{(2)}$, and $\beta_{1i}^{(3)}$, is lower regime autoregressive coefficients, and $beta_{2i}^{(1)}$, $beta_{2i}^{(2)}$, and $beta_{2i}^{(3)}$ are upper regime autoregressive coefficients. I is indicator variable with y_{t-d} is an observed variable determining the switching point and r is the threshold parameter or Kink point values defining the regime for both mean and variance equations through indicator function $I\left(y_{t-d} \leq r\right)$ for lower regime and $I\left(y_{t-d} > r\right)$ for upper regime. The $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$, term in the Kink-AR mean equation 3.1 are the innovations of the time series process. Engle [10] defined them as an autoregressive conditional heteroscedastic process where all ε_t are of the form

$$\begin{aligned} \varepsilon_{1,t} &= h_{1,t} u_{1,t} \\ \varepsilon_{2,t} &= h_{2,t} u_{2,t} \\ \varepsilon_{3,t} &= h_{3,t} u_{3,t}, \end{aligned} \tag{3.2}$$

where $u_{i,t}$ is an *iid* process with zero mean and unit variance. Although $\varepsilon_{i,t}$ is serially uncorrelated by definition its conditional variance equals $h_{i,t}^2$ and, therefore, may change over time.

The variance equation of the Kink-GARCH(p,q) model can be expressed as

$$\begin{aligned} h_{1,t}^2 =& \delta^{(1)} + \sum_{i=1}^{p} \zeta_{1i}^{(1)} \varepsilon_{1,t-i}^2 I\left(y_{t-d}^{(1)} \leq r^{(1)}\right) + \sum_{i=1}^{p} \zeta_{2i}^{(1)} \varepsilon_{1,t-i}^2 I\left(y_{t-d}^{(1)} > r^{(1)}\right) \\ &+ \sum_{j=1}^{q} \theta_{1j}^{(1)} h_{1,t-j}^2 I\left(y_{t-d}^{(1)} \leq r^{(1)}\right) + \sum_{j=1}^{q} \theta_{2j}^{(1)} h_{1,t-j}^2 I\left(y_{t-d}^{(1)} > r^{(1)}\right) \\ h_{2,t}^2 =& \delta^{(2)} + \sum_{i=1}^{p} \zeta_{1i}^{(2)} \varepsilon_{2,t-i}^2 I\left(y_{t-d}^{(2)} \leq r^{(2)}\right) + \sum_{i=1}^{p} \zeta_{2i}^{(2)} \varepsilon_{2,t-i}^2 I\left(y_{t-d}^{(2)} > r^{(2)}\right) \\ &+ \sum_{j=1}^{q} \theta_{1j}^{(2)} h_{2,t-j}^2 I\left(y_{t-d}^{(2)} \leq r^{(2)}\right) + \sum_{j=1}^{q} \theta_{2j}^{(2)} h_{2,t-j}^2 I\left(y_{t-d}^{(2)} > r^{(2)}\right) \\ h_{3,t}^2 =& \delta^{(3)} + \sum_{i=1}^{p} \zeta_{1i}^{(3)} \varepsilon_{3,t-i}^2 I\left(y_{t-d}^{(3)} \leq r^{(3)}\right) + \sum_{i=1}^{p} \zeta_{2i}^{(3)} \varepsilon_{3,t-i}^2 I\left(y_{t-d}^{(3)} > r^{(3)}\right) \\ &+ \sum_{j=1}^{q} \theta_{1j}^{(3)} h_{3,t-j}^2 I\left(y_{t-d}^{(3)} \leq r^{(3)}\right) + \sum_{j=1}^{q} \theta_{2j}^{(3)} h_{3,t-j}^2 I\left(y_{t-d}^{(3)} > r^{(3)}\right). \end{aligned} \tag{3.3}$$

From equation 3.1 and 3.3, $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$, are assumed to have a correlation. Therefore, the dependence $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{2,t}$ between is modeled through copulas.

3.1.2 Copula

For the benefit of our readers, we recall necessary background information on copulas in this subsection.
The notion of a copula was introduced by Sklar [11] who was answering a question raised by M. Frèchet about the relationship between a multidimensional probability function and its lower-dimensional marginals.

Consider a random vector $(Y_1, Y_2, \ldots, Y_n)$. Suppose its margins $F_1, F_2, \ldots, F_n$ are continuous functions. By applying the probability integral transform to each component, the random vector

$$(U_1, U_2, \ldots, U_n) = (F_1(Y_1), F_2(Y_2), \ldots, F_n(Y_n)) \tag{3.4}$$

has uniform margins.

The copula of $(Y_1, Y_2, \ldots, Y_n)$ is defined as the joint cumulative distribution function of $(U_1, U_2, \ldots, U_n)$:

$$C(u_1, u_2, \ldots, u_n) = P(U_1 \le u_1, U_2 \le u_2, \ldots, U_n \le u_n) \tag{3.5}$$

The copula C contains all information on the dependence structure between the components of $(Y_1, Y_2, \ldots, Y_n)$ whereas the marginal cumulative distribution functions F_i contain all information on the marginal distributions. The above formula for the copula function can be rewritten to correspond to this as:

$$C(u_1, u_2, \ldots, u_n) = P\left(Y_1 \le F_1^{-1}(u_1), Y_2 \le F_2^{-1}(u_2), \ldots, Y_n \le F_n^{-1}(u_n)\right) \tag{3.6}$$

Sklar's theorem provides the theoretical foundation for the application of copulas. Sklars theorem states that a multivariate cumulative distribution function $H(Y_1, Y_2, \ldots, Y_n)$ of a random vector $(Y_1, Y_2, \ldots, Y_n)$ with marginals F_i can be written as

$$H(Y_1, Y_2, \ldots, Y_n) = C(F_1(Y_1), F_2(Y_2), \ldots, F_n(Y_n)), \tag{3.7}$$

where C is a copula. Extensive theoretical discussion on copulas can be found in Joe [12] and Nelsen [13].

In this study, we consider to model the dependence between $\varepsilon_{i,t}$ through the Gaussian and Student-t copula functions, which is the most commonly use in literature. We conduct a Bayesian estimation as the estimator. Bayesian requires the specification of a likelihood and prior distributions on the model parameters. For the copula parameter, we select prior distributions that are uninformative. Following Smith [14], Smith, Gan and Kohn [15], and Wichitaksorn and Choy [16], we can impose the prior distribution to copula density to obtain the posterior densities of the Elliptical copulas as in the following:

1. Posterior gaussian copulas density,

for n-dimension, the α^{th} conditional quantile of posterior gaussian or normal copula can be written as:

$$f\left(R_{\alpha,G},\Phi_{\alpha,i}\left(u_{\alpha,i}\right)\right)=p\left(R_{\alpha,G}\right)\cdot\left(\sqrt{detR_{\alpha,G}}\right)^{-1}\ll$$
$$\left[\frac{1}{2}\left(\Phi_{\alpha,1}^{-1}\left(u_{\alpha,1}\right),\Phi_{\alpha,2}^{-1}\left(u_{\alpha,2}\right),\ldots,\Phi_{\alpha,n}^{-1}\left(u_{\alpha,n}\right)\right)\cdot\left(R_{\alpha,G}^{-1}-I\right)\right.$$
$$\left.\cdot\left(\Phi_{\alpha,1}^{-1}\left(u_{\alpha,1}\right),\Phi_{\alpha,2}^{-1}\left(u_{\alpha,2}\right),\ldots,\Phi_{\alpha,n}^{-1}\left(u_{\alpha,n}\right)\right)^{T}\right], \quad (3.8)$$

where $\Phi_{\alpha,i}$ is n-dimensional standard normal cumulative distribution at quantile α, $R_{\alpha,G}$ is dependence matrix of gaussian copula at quantile α, and $p\left(R_{\alpha,G}\right)$ is the prior density at quantile α which is applied to be uniform distribution $[0,1]$. To draw the updated copula dependence parameters, we random these parameters from the proposal trancated normal distribution $[-1,1]$ interval.

2. Posterior student-t copulas density,

   in the case of n-dimension, the posterior student-t copula density is evaluated by

$$f\left(R_{\alpha,T},t_{\alpha,v}\left(u_{\alpha,i}\right)\right)=p\left(R_{\alpha,T}\right)\cdot p\left(v_{\alpha}\right)\cdot\frac{f_{\alpha,v,R}\left(t_{\alpha,v}^{-1}\left(u_{\alpha,1}\right),\ldots t_{\alpha,v}^{-1}\left(u_{\alpha,n}\right)\right)]}{\prod_{i=1}^{n}f_{\alpha,v}\left(t_{\alpha,v}^{-1}\left(u_{\alpha,i}\right)\right)}, \quad (3.9)$$

where $f_{\alpha,v,R}$ is the joint density of a $t_{\alpha,v}\left(u_{\alpha,i}\right)$-distributed random vector where $R_{\alpha,T}$ isthe correlation matrix. The prior density for v_{α} and $R_{\alpha,T}$ are exponential distribution and uniform $[-1,1]$, respectively. To draw the updated dependence parameters and degree of freedom (v_{α}), we random v_{α} and $R_{\alpha,T}$ from the proposal trancate normal distribution interval $[-1,1]$ and distribution interval $[0,\infty]$.

3.2 Priors and Likelihood for Multivariate Copula Kink AR(m)-GARCH(p,q) Model

Let the full parameter vector be denoted as

$$\Theta=\left(\alpha^{(i)},\beta_{1i}^{(i)},\beta_{2i}^{(i)},\delta^{(i)},\zeta_{1i}^{(i)},\zeta_{2i}^{(i)},\theta_{1j}^{(i)},\theta_{2j}^{(i)},r_{i},R_{\alpha}\right).$$

A Bayesian statistical model, it is required a parametric statistical model, $f\left(y|\Theta\right)$, and a prior distribution on the parameters, $p\left(\Theta\right)$. Here we will choose conjugate priors wherever possible and normal priors. Since we work on large data sets, we do not consider the choice of the prior distribution a critical issue and rely on the asymptotic efficiency of the Bayes estimator [17]. Thus, in this study, we can write the conditional likelihood function and prior of the trivariate Gaussian copula-based two-regime Kink AR(m)-GARCH(p,q) model as

$$f\left(y|\Theta\right) = [\prod_{i=1}^{3} f\left(y|\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i\right) \cdot f\left(R_{\alpha,G}, \Phi_{\alpha,1}\left(u_{\alpha,1}\right), \Phi_{\alpha,2}\left(u_{\alpha,2}\right), \Phi_{\alpha,3}\left(u_{\alpha,3}\right)\right)] \cdot p\left(\Theta\right), \tag{3.10}$$

and trivariate T copula based two-regime Kink AR(m)-GARCH(p,q) model as

$$f\left(y|\Theta\right) = [\prod_{i=1}^{3} f\left(y|\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i\right) \cdot f\left(R_{\alpha,G}, t_{\alpha,1}\left(u_{\alpha,1}\right), t_{\alpha,2}\left(u_{\alpha,2}\right), t_{\alpha,3}\left(u_{\alpha,3}\right)\right)] \cdot p\left(\Theta\right), \tag{3.11}$$

where $f\left(y|\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i\right)$ is a likelihood function of which can be normal, student-t or skew-t distribution. Each likelihood function can be written as follows:

1. Normal likelihood

$$L\left(\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i|y\right) = \prod_t \frac{1}{\sqrt{2\pi h_t^2}} exp\left\{\frac{1}{2h_t^2}u_t^2\right\}. \tag{3.12}$$

2. Student-t likelihood

$$L\left(\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i|y\right) = \prod_t \frac{\tau\left(\frac{1}{2}\left(v+1\right)\right)}{\pi^{\frac{1}{2}}\tau\left(\frac{1}{2}v\right)} \left[\left(v-2\right)h_t^2\right]^{-\frac{1}{2}} \left[1 + \frac{\varepsilon_t^2}{\left(v-2\right)h_t^2}\right]^{-\frac{1}{2}\left(v+1\right)}, \tag{3.13}$$

where v is the degree of freedom.

3. Skewed-t likelihood

$$L\left(\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i|y\right) = \prod_{t=1}^{n} \left[\frac{2}{\xi = 1/\xi} f\left(\xi y\right)\right], \text{ for } y < 0$$
$$L\left(\alpha^{(i)}, \beta_{1i}^{(i)}, \beta_{2i}^{(i)}, \delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}, \theta_{2j}^{(i)}, r_i|y\right) = \prod_{t=1}^{n} \left[\frac{2}{\xi = 1/\xi} f\left(\frac{y}{\xi}\right)\right], \text{ for } y \geq 0 \tag{3.14}$$

where ξ is a skew parameter and $f\left(\xi y\right)$ and $f\left(\frac{y}{\xi}\right)$ are a density of student-t distribution.

Our prior settings are similar to those of Henneke et al. [17] and Chen et al. [18]. We set the normal prior for $\alpha^{(i)}, \beta_{1i}^{(i)}$, and $\beta_{2i}^{(i)}$, truncated normal priors on the GARCH parameters $\delta^{(i)}, \zeta_{1i}^{(i)}, \zeta_{2i}^{(i)}, \theta_{1j}^{(i)}$, and $\theta_{2j}^{(i)}$, translated exponential prior for degree of freedom v which is restricted to be larger than 2 to ensure the GARCH parameters to be positive and uniform prior for r_i. To draw the Markov Chain Monte Carlo (MCMC) iterates for the set of estimated parameters, we conducted the Metropolis-Hastings (MH) algorithm, where the proposal parameters are randomed from the normal distribution with mean Θ_{t-1} and variances are obtained from the standard error of each parameter from the Maximum likelihood estimator. The parameters are updated by block consisting of mean equation block (AR), variance equation block (GARCH), and threshold or Kink point block while the degree of freedom parameter is sampled using an optimized rejection technique from a translated exponential source density [19].

In this study, we set the MCMC sample size with 10,000 iterations, discarding the 2,000 burn-in iterates and keep the last 8,000 iterates for inference.

4 Data

In this study of Chinese outbound tourism to three major ASEAN countries comprising Thailand, Singapore, and Malaysia, the number of Chinese tourist arrivals to these destinations is used. The data are monthly time series data for the period from January 1999 to December 2014, collected from CEIC database [3]. Additionally, we transform these variables into growth rate before estimating.

From Table 1, 1^{st} column is describtive statistics of growth rate of Chinese outbound tourism to Thailand, 2^{nd} column is describtive statistics of growth rate of Chinese outbound tourism to Singapore, and 3^{rd} column is describtive statistics of growth rate of Chinese outbound tourism to Malaysia.

As can be seen above, Table 1 gives some standard summary statistics along with the Jarque-Bera(JB) test for normality. Under the null that the data are *iid* normal, JB is asymptotically distributed as chi-square with 2 degrees of freedom. The distribution of the growth rate of Chinese outbound tourism to Thailand, Singapore, and Malaysia, is clearly normal.

Table 1: Descriptive statistics

	Thailand	Singapore	Malaysia
Mean	0.174	0.052	0.059
Median	0.033	0.025	0.036
Maximum	13.643	1.518	1.452
Minimum	−0.884	−0.826	−0.722
Std. Dev.	1.107	0.305	0.308
Skewness	8.808	1.120	0.855
Kurtosis	97.857	6.704	5.720
Jarque-Bera	102,001.300	205.293	113.088
Probability	0.000	0.000	0.000

Source:Calculation

In this study, we employ the Augmented Dickey-Fuller (ADF) [20] test statistic to analyze the order of integration of our variables. The null hypothesis tested is that the variable under investigation has a unit root against the alternative that it does not. The results of the ADF test indicate that all data are statistically significant with stationarity at the level I(0).

5 Empirical Results

5.1 Lag Length Criteria and Model Selection

First of all, we have to identify the order of AR. Bayesian information criterion (BIC) [21] is a standard commonly used for selecting statistical model. The autoregressive (AR) order of mean equation is determined by way of minimizing BIC. According to Table 2, for all countries, AR(1) is identified.

Table 2: Lag length criteria

	BIC				
	lag1	lag2	lag3	lag4	lag5
Thailand	**−615.456**	−614.875	−609.983	−613.818	−609.495
Singapore	**−619.259**	−610.839	−616.028	−618.738	−617.178
Malaysia	**59.200**	64.495	68.108	72.586	72.842

Source:Calculation
Note:smallest BIC in **bold**

Table 3: Number of regime selection

Thailand	BIC
1 regime	135.668
2 regimes	**134.553**
3 regimes	263.980
Singapore	BIC
1 regime	134.307
2 regimes	**115.167**
3 regimes	273.985
Malaysia	BIC
1 regime	633.644
2 regimes	**580.563**
3 regimes	669.818

Source: Calculation
Note: smallest BIC in **bold**

5.2 Selecting the Number of Regime

Before we estimate the model, the Bayesian information criterion (BIC) is employed to select the number of regime. According to minimizing BIC from Table 3, we then setup the estimated model for all countries as two-regime Kink AR(1)-GARCH(1,1). The study employs a GARCH(1,1) since it is the most prevalently used for empirical applications in many studies and it is able to reproduce the volatility dynamics of stochastic time series data [19].

Table 4: Model Selection

		BIC		
Marginals1	Marginals2	Marginals3	Gaussian copula	T-Copula
normal	normal	normal	−6.861	−1.971
normal	normal	student-t	−5.346	−134.854
normal	normal	skewed student-t	−7.801	−133.716
normal	student-t	normal	−110.599	122.125
normal	student-t	student-t	−40.132	−12.548
normal	student-t	skewed student-t	−125.804	−133.549
normal	skewed student-t	normal	−223.043	−133.716
normal	skewed student-t	student-t	−224.993	−237.753
normal	skewed student-t	skewed student-t	−148.510	−297.632
student-t	normal	normal	−8.178	−237.753
student-t	normal	student-t	−39.677	−13.281
student-t	normal	skewed student-t	−65.579	−121.592
student-t	student-t	normal	5.483	−26.534
student-t	student-t	student-t	−57.350	−19.982
student-t	student-t	skewed student-t	−37.663	−25.975
student-t	skewed student-t	normal	−59.758	−188.983
student-t	skewed student-t	student-t	−43.727	−84.603
student-t	skewed student-t	skewed student-t	−201.854	−214.142
skewed student-t	normal	normal	−26.304	−93.523
skewed student-t	normal	student-t	199.056	−133.907
skewed student-t	normal	skewed student-t	−265.792	−136.074
skewed student-t	student-t	normal	−190.382	−136.074
skewed student-t	student-t	student-t	−150.088	−81.526
skewed student-t	student-t	skewed student-t	−205.117	−214.079
skewed student-t	skewed student-t	normal	−269.055	−179.451
skewed student-t	skewed student-t	student-t	−209.668	**−314.707**
skewed student-t	skewed student-t	skewed student-t	−209.852	−188.983

Source:Calculation
Note:smallest BIC in **bold**

Before we estimate model of copula Kink AR(1)-GARCH(1,1) type, we have to know the joint distribution between growth rate of Chinese outbound tourism to Thailand, Singapore and Malaysia. From Table 4, 1^{st} column is marginal distribution of growth rate of Chinese outbound tourism to Thailand, 2^{nd} column is marginal distribution of growth rate of Chinese outbound tourism to Singapore, and 3^{rd} column is marginal distribution of growth rate of Chinese outbound tourism to Malaysia. We select the marginals of each model by using the smallest

Bayesian information criterion (BIC) and multivariate T-copula based two-regime Kink AR(1)-GARCH(1,1) with skewed student-t, skewed student-t, and student-t margins are suggested.

5.3 Forecasting Performance

After modeling and forecasting of Chinese tourism outbound to three major ASEAN countries comprising Thailand, Singapore, and Malaysia, the best fit model obtained from the model selection was investigated for its forecasting performance. The in-sample mean absolute error (MAE), root mean squared error (RMSE) and the out-of-sample mean absolute percentage error (MAPE) of T-copula two-regime Kink AR(1)-GARCH(1) were calculated and were shown in Table 5.

Table 5: In-sample forecasting performance

	Kink AR(1)-GARCH(1,1) 2 regimes		
	Thailand	Singapore	Malaysia
MAE	0.359	0.377	0.561
RMSE	0.460	0.472	1.699
MAPE	6.211%	13.474%	6.813%

Source: Calculation.

Figure 3, Figure 4, and Figure 5, shows clearly the forecasted ChineseI Tourist Arrivals to Thailand, Singapore and Malaysia, using T-Copula Kink AR(1)-GARCH (1,1) model.

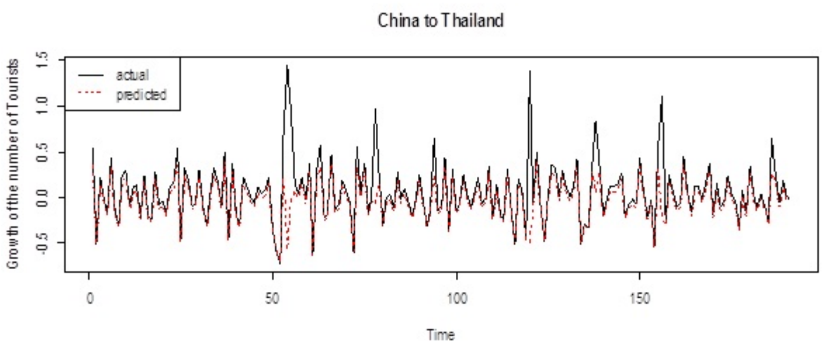

Figure 3: Forecasted ChineseI Tourist Arrivals to Thailand with 2 x E.S from the T-Copula Kink AR(1)-GARCH(1,1) Model

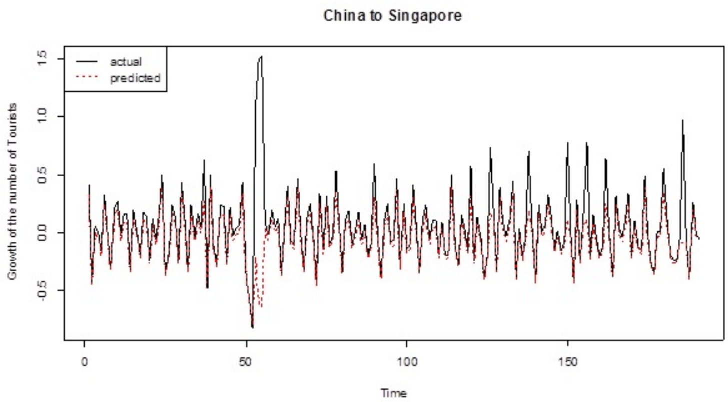

Figure 4: Forecasted ChineseI Tourist Arrivals to Singapore with 2 x E.S from the T-Copula Kink AR(1)-GARCH(1,1) Model

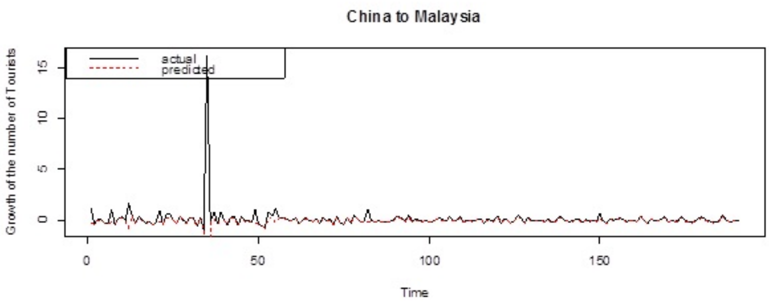

Figure 5: Forecasted ChineseI Tourist Arrivals to Singapore with 2 x E.S from the T-Copula Kink AR(1)-GARCH(1,1) Model

6 Conclusion

China has experienced phenomenal economic growth over the past decade with GDP per capita outperforming other large emerging markets. Because of the increase in standard of living, the China's outbound tourism grows by leaps and bounds. Among many destinations, Chinese tourists choose to travel to ASEAN especially, Thailand, Singapore, and Malaysia. Therefore, the aim of this paper is to model and forecast Chinese tourist arrivals to Thailand, Singapore, and Malaysia.

Considering the fact that modelling and forecasting of tourism demand is a challenging topic, that the adequacy and accuracy of a forecasting model is valued according to its in-sample and out-of-sample forecasts, and that is still difficult to indicate which modelling techniques are the most adequate for tourism demand modelling. There are several models for forecasting tourist arrivals. Song and Li [5] found that there is no single model that consistently outperforms other models in all situations. Authors differ on the best method for tourism forecasting. In this paper, we propose a Kink AR-GARCH model to capture mean and volatility asymmetries in Chinese tourist outbound. We also assume that there are dependence among behavior of tourist arrivals from China to each countries. Therefore, we apply copula approch to capture these dependence. Monthly tourist arrivals from 1999 to 2014 are used in the analysis. Empirical result shows that, for all countries, T-Copula 2-regime Kink AR(1)-GARCH(1,1) with skewed student-t, skewed student-t, and student-t margins are suggested.

According to this research, the future attention should be paid to further developments of forecasting techniques (especially regression models and its different functional forms), or to combination of different forecasts. Further research should focus to impact of other factors such as income and change in currency on tourism demand which will improve forecasting accuracy.

Acknowledgements : We would like to thank the referee(s) for his comments and suggestions on the manuscript. we also gratefully thank Professors Hung T. Nguyen and Songsak sriboonchitta for their relevant and helpful comments, which have helped us considerably improve the paper. This work was supported by Puay Ungpakorn Centre of Excellence in Econometrics, Faculty of Economics, Chiang Mai University.

References

[1] T. Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics 31 (1986) 307-327.

[2] World Bank, World Development Indicators. Retrieved April 8, 2016, from http://data.worldbank.org/data-catalog/world-development-indicators.

[3] Euromoney Institutional Investor Company, CEIC's database [computer software], Bangkok, Thailand, (2016)

[4] A. Untong, M. Kaosa-Ard, The Competitive Advantage in Tourism of Asean Countries in Chinese Tourists Market. Retrieved from http://www.Vijaichina.Com/Sites/Default/files/2_Dr.Akarapong_Pdf.Pdf.

[5] H. Song, G. Li, Tourism Demand Modelling and Forecasting-a Review of Recent Research, Tourism Management 29 (2008) 203-220.

[6] A. Untong, ASEAN Long-Run Tourism Elasticity Demand in Thailand, Applied Economics Journal 22 (2015) 77-101.

[7] S. Sookmark, An Analysis of International Tourism Demand in Thailand (Doctoral Dissertation), National Institute of Development Administration, Bangkok, 2011.

[8] P. Chaitip, C. Chaiboonsri, International Tourists Arrival to Thailand: Forecasting by Non-linear Model, Procedia Economics and Finance 14 (2014) 100-109.

[9] N. Min, J. Sirisrisakulchai, S. Sriboonchitta, Forecasting Tourist Arrivals to Thailand Using Belief Functions, Studies in Computational Intelligence 583 (2015) 343-357.

[10] R.F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica 50 (1982) 987-1007.

[11] A. Sklar, Fonctions de r*é*partition*à* n-dimensions et leurs marges, Publications de l'Institut de Statistique de l'Universit*é* de Paris, Paris, 1959.

[12] H. Joe, Multivariate models and dependence concepts, Chapman & Hall, London, 1997.

[13] R.B. Nelson, An Introduction to Copulas, Springer-Verlag, New York, 2006.

[14] M.S. Smith, Bayesian Approaches to Copula Modelling, Available at SSRN1974297, 2011.

[15] M.S. Smith, Q. Gan, R.J. Kohn, Modelling Dependence using Skew T Copulas: Bayesian Inference and Applications, Journal of Applied Econometrics 27 (2012) 500-522.

[16] N. Wichitaksorn, S.B. Choy, Modeling Dependence of Seemingly Unrelated Tobit Model Through Copula: a Bayesian Analysis, Thailand Econometrics Society 3 (2012) 6-19.

[17] J.S. Henneke, S.T. Rachev, F.J. Fabozzi, M. Nikolov, MCMC-Based Estimation of Markov Switching ARMAGARCH Models, Applied Economics 43 (2011) 259-271.

[18] C.W. Chen, M. Chen, S.Y. Chen, Pairs Trading via Three-Regime Threshold Autoregressive GARCH Models, Modeling Dependence in Econometrics, Springer International Publishing (2014) 127-140.

[19] D. Ardia, Financial risk management with bayesian estimation of GARCH Models: theory and applications, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, Germany, (2008) ISBN 978-3-540-78656-6.

[20] D.A. Dickey, W. A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74 (1979) 427-431.

[21] G. Schwarz, Estimating the dimension of a model, Ann. Statist 6 (1978) 461-464.

(Received 11 August 2016)
(Accepted 19 October 2016)