Article

TOURISM ECONOMICS
The Business and Finance of Tourism and Recreation

Tourism Economics
2017, Vol. 23(7) 1419–1436

DOI: 10.1177/1354816617693964
journals.sagepub.com/home/teu
SAGE

Analysis of tourism demand serial dependence structure for forecasting

Liang Zhu
Nanyang Technological University, Singapore

Christine Lim
Nanyang Technological University, Singapore

Wenjun Xie
Nanyang Technological University, Singapore

Yuan Wu
Nanyang Technological University, Singapore

Abstract

This study aims to extend knowledge of serial dependence structure in tourism demand modelling and make a contribution to tourism forecasting with the use of copula method. Analysis of serial dependence can reveal the impact of current tourism demand on the future. This is important for tourism demand forecasting, as the prediction of future tourism demand relies highly on the historical demand information. However, serial dependence, especially its structure, has received very little attention in previous tourism research. The copula method is flexible as it provides various functions to specify different serial dependence structures and allows arbitrary distributions of tourism demand. We used five types of copulas to analyse two-dimensional serial dependence structure for 10 arrivals series to Singapore. The empirical findings show that serial dependence structures of arrivals can be non-linear. Additionally, the Student-*t* copula generates forecasts of tourism demand with higher accuracy than the autoregressive integrated moving average (ARIMA) and seasonal ARIMA models.

Keywords

copula method, forecasting performance, serial dependence structure, tourism demand, univariate time series

Corresponding author:

Liang Zhu, Division of Marketing and International Business, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore.

Email: lzhu003@e.ntu.edu.sg

Introduction

Destination governments and tourism-related businesses undertake tourism demand forecasting as it can guide policy decisions and strategic plans. Many researchers have applied or modified various techniques routinely used in business for computing tourism forecasts and increasing forecast accuracy. However, prior research clearly shows that different data frequencies, length of sample period and forecast horizon, origin countries or destinations used have some impact on the forecasting performance of tourism demand models (Li et al., 2005; Witt and Song, 2001). The variation in the performance of forecasting models under different situations has raised challenges to forecasters and policymakers to identify and overcome these drawbacks.

The basis of tourism demand forecasting is using the historical demand information to predict the future demand. Thus, it is very essential to understand the dependence of tourism demand (denoted by Y_t) on its historical information (or lagged variables, denoted by $Y_{t-1},\ Y_{t-2},\ \ldots,\ Y_{t-k}$). This relationship is known as serial dependence (or temporal association) in statistics. Specification of different serial dependence structures can reveal how the current tourism demand affects the future, and hence contributes to tourism demand forecasting. This is meaningful for both researchers and practitioners. However, the common practice in past research is to assume a linear serial dependence structure of tourism demand and to construct linear time series models for forecasting.

Among the different time series approaches, the autoregressive moving average (ARMA) model proposed by Box and Jenkins (1970) is most frequently applied because of its simplicity and generality. Many extensions of ARMA models have been proposed to accommodate the features of tourism demand time series, such as autoregressive integrated moving average (ARIMA) models (Cho, 2001; Papatheodorou and Song, 2005; Turner et al., 1995), seasonal autoregressive integrated moving average (SARIMA) models (Liang, 2014; Lim and McAleer, 2000; Kim and Ngo, 2001; Vergori, 2012), fractional integration ARMA or ARFIMA (Chu, 2008; Nowman and Van Dellen, 2012), among others. These extensions provide different alternatives to increase the performance of tourism time series forecasting. However, the models only capture linear serial dependence.

Using linear correlation as a measure of dependence requires the random variables to follow an elliptical joint distribution; however, most random variables do not meet this requirement (Embrechts et al., 2001). Similarly, there is no evidence that the current and lagged variables of tourism demand in a univariate time series, such as Y_t and Y_{t-1}, are jointly elliptically distributed. The frequent use of linear structure to model the dependence between Y_t and Y_{t-1} is problematic. According to Crane and Van Der Hoek (2008), if the association between Y_t and Y_{t-1} is not elliptical or is concentrated at one extreme, models based on linear association would not be able to capture the dependence structure adequately, leading to over- or underestimation problem.

Tourism research undertaken on dependence structure of a univariate time series has rarely gone beyond the use of linear models. Non-linear models such as neural network approaches are proposed to give better tourism forecasting performance (Álvarez-Díaz and Rosselló-Nadal, 2010; Sencheong and Turner, 2005). As explained in Peng et al. (2012), the neural network structure is typically based on having one input layer (explanatory variable/s), one output layer (dependent variable) and one-to-two hidden layers (which consists of explanatory and dependent variables with different weights). Among other limitations, the model suffers from difficulty in network structure determination (Kisi, 2011). Other non-linear models, such as sine wave approach (Chan, 1993) or sparse Gaussian process (Wu et al., 2012), also impose certain types of serial dependence structures on tourism time series, thus restricting the generality of these models.

Even though any type of association can be found between current and lagged observations in a univariate time series, modelling non-linear serial dependence structure has received very little attention in previous tourism demand forecasting research. As non-linear models have the potential to produce better forecasts, progress in the development of non-linear univariate time series analysis and forecasting merits continued research. The intention of this article is to propose the use of copula as an alternative method for modelling serial dependence structure in a univariate tourism time series and to provide a new direction in tourism forecasting. Ten tourist arrivals series are examined using this approach.

The copula method is flexible as it provides various functions to describe different dependence structures, whether they are linear or non-linear and symmetric or asymmetric (ranging from perfect negative to perfect positive) (de Melo Mendes and Aíube, 2011). Moreover, the copula method does not need any assumption for the distribution of tourism demand variable, while the existing models usually assume that tourism demand follows certain distribution, such as normal distribution in most cases. The benefits of copula have been demonstrated in the finance, insurance and economics literature (Balakrishnan and Lai, 2009; Embrechts et al., 2002; Frees and Valdez, 1998; Joe, 1997; Klugman et al., 2012; Nelsen, 2007). Copulas are also used to model the serial dependence structure in a univariate time series (Beare, 2010; Chen and Fan, 2006; Chen et al., 2009; Ibragimov, 2009; Smith et al., 2010).

The copula method is still a relatively new approach used in empirical tourism research. Related studies apply this method to analyse either the relationship between tourism demand and its determinants (Pérez-Rodríguez et al., 2015; Tang et al., 2016) or the co-movement of different tourism flows (Liu and Sriboonchitta, 2013; Liu et al., 2014; Puarattanaarunkorn and Sriboonchitta, 2014). In these studies, copulas are used for multivariate time series analysis. However, the application of copulas has not been extended to include demand forecasting in tourism research. This article intends to fill this gap.

Methodology

Sklar's theorem sets the foundation of copulas, which provides the link between marginal distributions and their joint distribution (Sklar, 1959). According to this theorem, if $H(.)$ is an n-dimensional joint distribution function for the random variables $Y_1, Y_2, \dots, Y_n$ with continuous marginal distribution functions $F_1, F_2, \dots, F_n$, $H(.)$ can be represented by a unique copula function $C(.)$ as

$$H(y_1,\ y_2,\ \dots,\ y_n) = C\{F_1(y_1),\ F_2(y_2),\ \dots,\ F_n(y_n);\ \theta\} \tag{1}$$

where y_i is the observation of Y_i and $i = 1, \dots, n$; parameter vector θ captures information about the dependence or association among Y_i. Sklar's theorem shows that any given marginal distributions can be combined with a copula function to obtain a valid joint distribution of the random variables.

In their study of copula-based Markov models, Chen and Fan (2006) argued that copulas can be used to analyse the serial dependence of stationary first-order Markov process. In light of this development, we will apply copulas to univariate time series analysis for tourism. In univariate copula models, $H(.)$ is the joint distribution function for $Y_t, Y_{t-1}, \dots, Y_{t-k}$ with continuous marginal distribution function F_m. $H(.)$ can be represented by a copula function $C(.)$ as

$$H(y_t,\ y_{t-1},\ \dots,\ y_{t-k}) = C\{F_m(y_t),\ F_m(y_{t-1}),\ \dots,\ F_m(y_{t-k});\ \theta\} \tag{2}$$

Equation (2) can be rewritten as follows:

$$H(y_t, y_{t-1}, \ \ldots, y_{t-k}) = C(u_t, u_{t-1}, \ \ldots, u_{t-k}; \theta) \tag{3}$$

where u_{t-i} is the cumulative density and $u_{t-i} = F_m(y_{t-i})$ for $i = 0, 1, \ldots, k$.

In the simplest case, random variables can be fitted using two-dimensional copulas, in which $k = 1$ and $H(.)$ is given as follows:

$$H(y_t, y_{t-1}) = C(u_t, u_{t-1}; \theta) \tag{4}$$

where θ is the dependence parameter between u_t and u_{t-1}. In this article, we analyse the serial dependence structure between Y_t and Y_{t-1} and construct two-dimensional copula models for forecasting purpose. The series has to be first-order Markov process.

A likelihood-based estimation method is used to estimate the dependence parameter θin equation (4). Let $c(.)$ denotes the probability density function corresponding to the copula distribution function $C(.)$. The log-likelihood (LL) function is as follows:

$$L = \sum_{t=1}^{T} \ln c(u_t, u_{t-1}; \theta) \tag{5}$$

where

$$c(.) = \frac{\partial^2 C}{\partial u_t \partial u_{t-1}} \tag{6}$$

As mentioned earlier, there is an eclectic mix of copula functions available for the specification of serial dependence structure. We apply five commonly used copulas belonging to two families, namely, elliptical and Archimedean families. The elliptical family is widely applied in statistics and econometrics (Frahm et al., 2003); the Gaussian and Student-*t* are the well-known copulas in this class. Hence, we employ these two copulas to model the serial dependence depicted in equation (4). The Archimedean family is also one of the most important copula classes which allows for a great variety of dependence structures (Embrechts et al., 2001). We select Frank, Clayton and Gumbel copulas from this family to complement the elliptical copulas, as the three can represent different kinds of tail dependence.

If the dependence structure is symmetric and elliptical, the Gaussian copula should be considered as follows:

$$H(y_t, y_{t-1}) = \Phi_G[\Phi^{-1}(u_t), \Phi^{-1}(u_{t-1}); \theta] \tag{7}$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution and Φ_G is the standard bivariate normal distribution with dependence parameter . The Student-*t* copula is also for symmetric and elliptical structure. Unlike the Gaussian copula, the Student-*t* copula has joint fat tails and allows for higher probability of joint extreme events (Kjersti, 2004). The joint distribution represented by this copula is written as follows:

$$H(y_t, y_{t-1}) = T_{\theta,\nu}[T_\nu^{-1}(u_t), T_\nu^{-1}(u_{t-1}); \theta, \nu] \tag{8}$$

where T_ν^{-1} is the inverse of the scalar standard Student's-*t* distribution with $\nu > 2$ degrees of freedom, $T_{\theta,\nu}$ is the bivariate Student's-*t* distribution with $\nu > 2$ degrees of freedom and dependence parameter θ.

If the dependence structure is symmetric and strongest in the centre of the joint distribution, the Frank copula can be used to specify these features as follows:

$$H(y_t,\ y_{t-1}) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_t} - 1)(e^{-\theta u_{t-1}} - 1)}{e^{-\theta} - 1}\right) \tag{9}$$

However, if the dependence is asymmetric and strongest in the negative (left) tail of the joint distribution, the characteristics are captured by the Clayton copula as follows:

$$H(y_t,\ y_{t-1}) = \left[\max\left((u_t)^{-\theta} + (u_{t-1})^{-\theta} - 1;\ 0\right)\right]^{-1/\theta} \tag{10}$$

Finally, if the dependence is asymmetric and strongest in the positive (right) tail of the joint distribution, the Gumbel copula should be considered as follows:

$$H(y_t,\ y_{t-1}) = \exp\left[-\left(\left(-\log(u_t)\right)^{\theta} + \left(-\log(u_{t-1})\right)^{\theta}\right)^{1/\theta}\right] \tag{11}$$

Empirical modelling

Data description

Tourist arrival is the continuous random variable used as the proxy of tourism demand. Initially, our data set comprised total and disaggregated monthly arrivals for Singapore from January 1995 to December 2014. The disaggregated data comprised arrivals from 5 continents and 15 major source countries obtained from the Singapore Tourism Board. Procedurally, the application of two-dimensional copula model requires the time series to be first-order Markov processes, in which Y_t is only associated with Y_{t-1}, but not the higher order lags. However, we have checked and found that the monthly data cannot meet this requirement, as the monthly seasonality indicates that Y_t is also associated with Y_{t-12}. The monthly data are thus seasonally adjusted with Census X-13 developed by the US Census Bureau. Finally, we analyse 10 arrivals series for serial dependence structure and its contribution to forecasting. These arrivals series are selected because they are the first-order Markov processes, and the latter will be discussed in the following section. The adjusted data include total tourist arrivals to Singapore; arrivals from four continents: America (AM), Asia (AS), Europe (EU) and Oceania (OC) and arrivals from Australia (AU), Japan (JP), South Korea (KO), Malaysia (MY) and the United States.

Figures 1 and 2 provide graphical illustrations of the various arrivals series. It can be seen from Figure 1 that AS is the most important continent for Singapore tourism. It actually accounts for 76.7% of tourist arrivals to Singapore in 2014. While there are strong growths in tourism flows from AS for the period under investigation, tourist arrivals from the other continents remain relatively stable. Figure 2 shows that Singapore receives more tourists from AU, MY and JP than from the United States and KO.

The data for each arrival series are split into the estimation period (from 1995M1 to 2010M12) and the validation period (from 2011M1 to 2014M12). The estimation sample is used to specify the copula and the benchmark models. Subsequently, the specified models are used to generate 1-, 2-, 3-, 6-, 12-, 18- and 24-month-ahead forecasts. The forecasts will be compared using the expanding windows (or recursive) forecasting technique with the data in the validation sample.

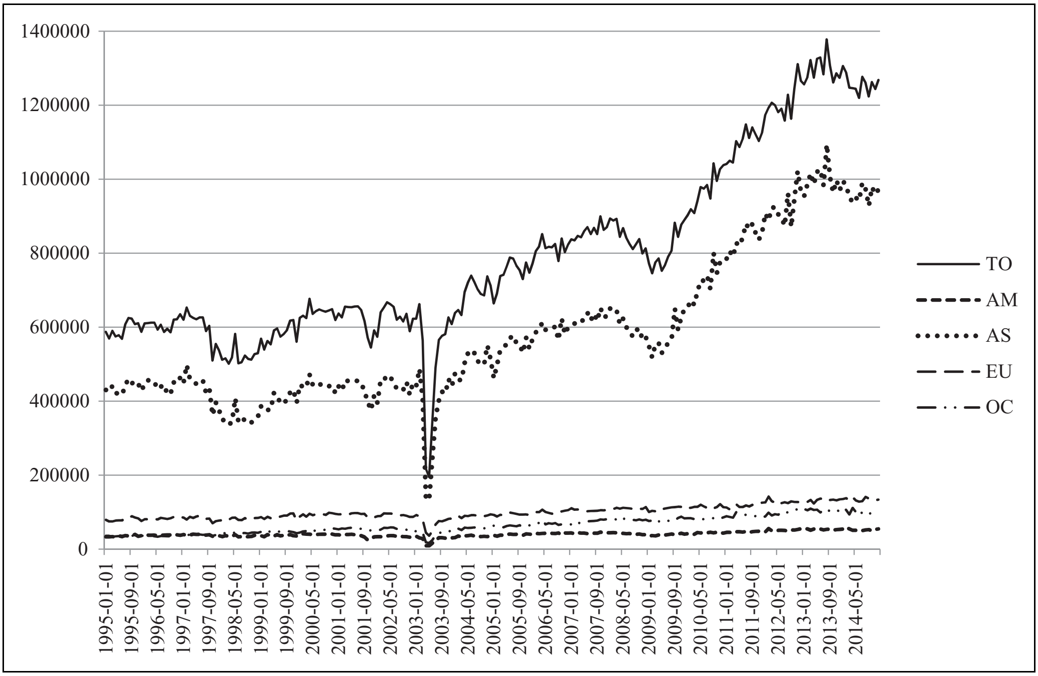

Figure 1. TO and arrivals from AM, AS, EU and OC.θTO: total tourist arrival; AM: America; AS: Asia; EU: Europe; OC: Oceania.

The aforementioned five copulas in equations (7) to (11) are used to specify serial dependence structure of tourism demand for Singapore. The appropriate specification of serial dependence is expected to contribute to tourism demand forecasting. We benchmarked the forecasting performance of the copula models against those of the ARIMA and SARIMA models.

Preliminary test

Before we carry out the empirical analysis of dependence structure, we have to examine if the individual arrival series (1) is stationary and (2) follows a specified distribution. Additionally, we test the Markov property of the series in this section.

We conduct the augmented Dickey–Fuller (ADF) tests to determine whether the monthly tourist arrivals series are stationary and the test results are presented in Table 1. All the arrivals series yield p-value of ADF statistics greater than 0.05, which indicate that these series have unit roots. However, taking first differences renders each series stationary. Therefore, we analyse arrivals' growth for all 10 series (in first differenced).

Each stationary arrival series has its own distribution, which is also called marginal distribution in order to distinguish it from the joint distribution. Modelling serial dependence relies on the marginal distribution of arrival series. The χ^2 test is used to determine if a particular tourist arrival series follows the normal distribution or alternative form, namely, Cauchy distribution. The null hypotheses for the χ^2 test are Y $\sim$ Normal or Y $\sim$ Cauchy (in which '$\sim$' denotes 'is distributed approximately as'). If the test statistic is greater than the critical value, the null hypothesis is

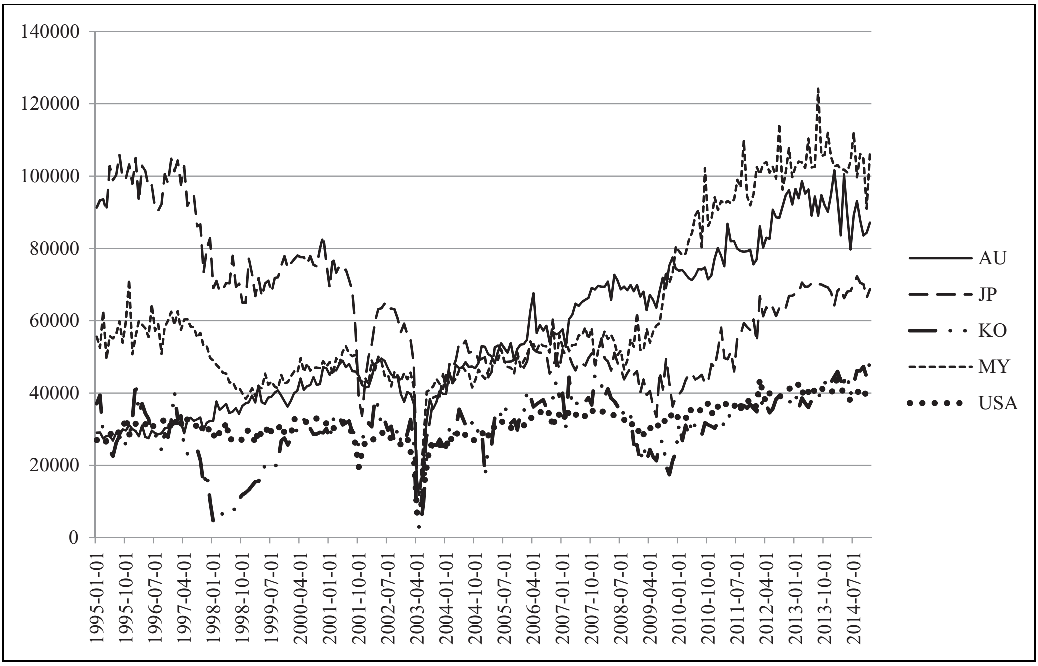

Figure 2. Tourist arrivals from AU, JP, KO, MY and the United States.θAU: Australia; JP: Japan; KO: South Korea' MY: Malaysia.

Table 1. Unit root tests for tourist arrivals to Singapore.

	ADF statistic	
Tourist arrivals series	Level	First difference
Total arrivals	−0.504	−13.061***
By continents		
America	−2.387	−15.405***
Asia	−0.308	−14.365***
Europe	−1.714	−13.614***
Oceania	−0.509	−14.641***
By source countries		
Australia	−1.055	−14.672***
Japan	−1.925	−14.757***
Malaysia	−0.653	−20.080***
South Korea	−2.573	−10.197***
USA	−2.378	−10.215***

Note: ADF: augmented Dickey–Fuller.
***Results are statistically significant at 0.001 level.

Table 2. χ^2 tests for distributions and LM test for serial correlation.

χ^2 test statistics				Serial correlation test	
	Specified distribution				
Tourist arrivals series	Normal	Cauchy	Critical value $\alpha = 0.01$	LM(12)*	p-Value
Total arrivals	34.753	*13.692*	21.670	16.024	0.190
America	27.437	*11.802*		16.736	0.160
Asia	28.835	*7.319*		13.610	0.326
Europe	21.777	*9.102*		20.036	0.066
Oceania	28.371	*7.108*		15.706	0.205
Australia	29.377	*8.672*		15.469	0.217
Japan	24.792	*11.977*		15.748	0.203
Malaysia	36.058	*17.839*		19.396	0.079
South Korea	20.540	*12.074*		17.357	0.137
USA	24.567	*12.691*		18.075	0.113

*Include 12 lags for monthly data.

rejected, implying that the arrival series does not follow the corresponding distribution. Table 2 presents the χ^2 test results. It can be seen that the test statistics for normal distribution are usually greater than the critical values at the 0.01 significance level. The only exception is growth in arrivals from KO. In contrast, Cauchy distribution can pass the test and possesses lower χ^2 statistic than the normal. Therefore, Cauchy distribution is selected for all growths in arrivals to generate the cumulative density u_t.

If a series follows an AR(1) specification with independent residuals, the series is a first-order Markov process (Deco and Schürmann, 2012). We can test the latter by conducting diagnostic checking on the residuals for serial correlation using the Breusch–Godfrey Lagrange Multiplier (LM) test. The null hypothesis of the LM test is that the residuals are not correlated. Table 2 also presents the results of the LM test; there is no serial correlation in the residuals of the AR(1) models for all the arrivals series since the p-values of LM statistics are greater than 0.05. Therefore, these series are the first-order Markov processes.

Analysis of serial dependence structure

We argue that it is problematic to use only linear structure to describe the serial dependence of tourism demand. To support this argument, we analyse the serial dependence structures of the 10 arrivals series. The most direct way is to check the serial association coefficients. Since we only consider the two-dimensional case in this study, we examine Pearson's coefficient (r), Spearman's rho coefficient (ρ) and Kendall's tau coefficient (τ) of Y_t and Y_{t-1}. These coefficients are the three most popular measures of association between random variables, among which Pearson's coefficient (r) only measures the strength of the linear relationship, while Spearman's ρ and Kendall's τ can measure either linear or non-linear associations (Hauke and Kossowski, 2011; Nelsen, 2007). Therefore, the comparison of Pearson's coefficient r and Spearman's ρ/Kendall's τ can be used to identify whether the linear structure is appropriate. For any given tourist arrivals series, we make the inference that the non-linear structure is a better choice for the specification of

Table 3. Associations between tourist arrivals and their first lags.

Tourist arrivals series	Pearson correlation coefficient	Kendall's τ	Spearman's ρ
Arrivals: Growth			
Total	0.054	−0.179**	−0.260**
America	−0.166	−0.236**	−0.339**
Asia	−0.035	−0.200**	−0.286**
Europe	−0.000	−0.138**	−0.195**
Oceania	−0.077	−0.181**	−0.253**
Australia	−0.072	−0.174**	−0.245**
Japan	−0.073	−0.168**	−0.237**
Malaysia	−0.369**	−0.281**	−0.391**
South Korea	−0.148*	−0.144**	−0.218**
USA	−0.119	−0.230**	−0.325**

*Results are statistically significant at 0.05 level.
**Results are statistically significant at 0.01 level.

serial dependence than the linear structure if either of the following conditions is met. First, Pearson's coefficient r of Y_t and Y_{t-1} is different from Spearman's ρ or Kendall's τ in terms of the coefficient sign (positive/negative). Second, Pearson's coefficient r is not significant, while Spearman's ρ or Kendall's τ is significant.

According to Table 3, the total arrivals' growth satisfies both conditions mentioned above. Pearson's coefficient r is positive but not significant, while Spearman's ρ and Kendall's τ are negatively significant. The serial dependence of total arrivals is more likely a non-linear negative association. As for arrivals from AM, AS, EU, OC, AU, JP and the United States, the Pearson's coefficients are not significant. Nevertheless, we cannot infer that there is no association between Y_t and Y_{t-1} for these cases, because Spearman's ρ and Kendall's τ for these series are all negatively significant. Since these seven cases satisfy the second condition, forecasting models with non-linear dependence structure should be considered. All the other series, namely growths in arrivals from MY and KO, do not meet the two conditions. For each of these two series, all the association coefficients are significant and have the same coefficient sign. Hence, we cannot conclude that the non-linear association exists. However, we can still determine the appropriate copula for these series, as the copula method encompasses both linear and non-linear dependence structures.

Further insight into the serial dependence structures of tourism demand can be obtained from the scatter plots of the cumulative densities generated for the tourist arrivals series under study. According to the scatter plots in Figure 3, the series which correspond to arrivals' growth seem to show symmetric dependence structures with heavy tails. The heavy tails are usually a sign of non-linear associations.

Table 3 and Figure 3 reveal that there is little support for using only one type of serial dependence specification, namely the linear structure, for tourism demand modelling and forecasting. The copula method is thus more appealing as it flexibly accounts for different serial dependence structures.

Copula model specification

The scatter plots show the serial dependence patterns of different tourist arrivals series. To determine the appropriate type of structure for each series and the copula to use, certain selection

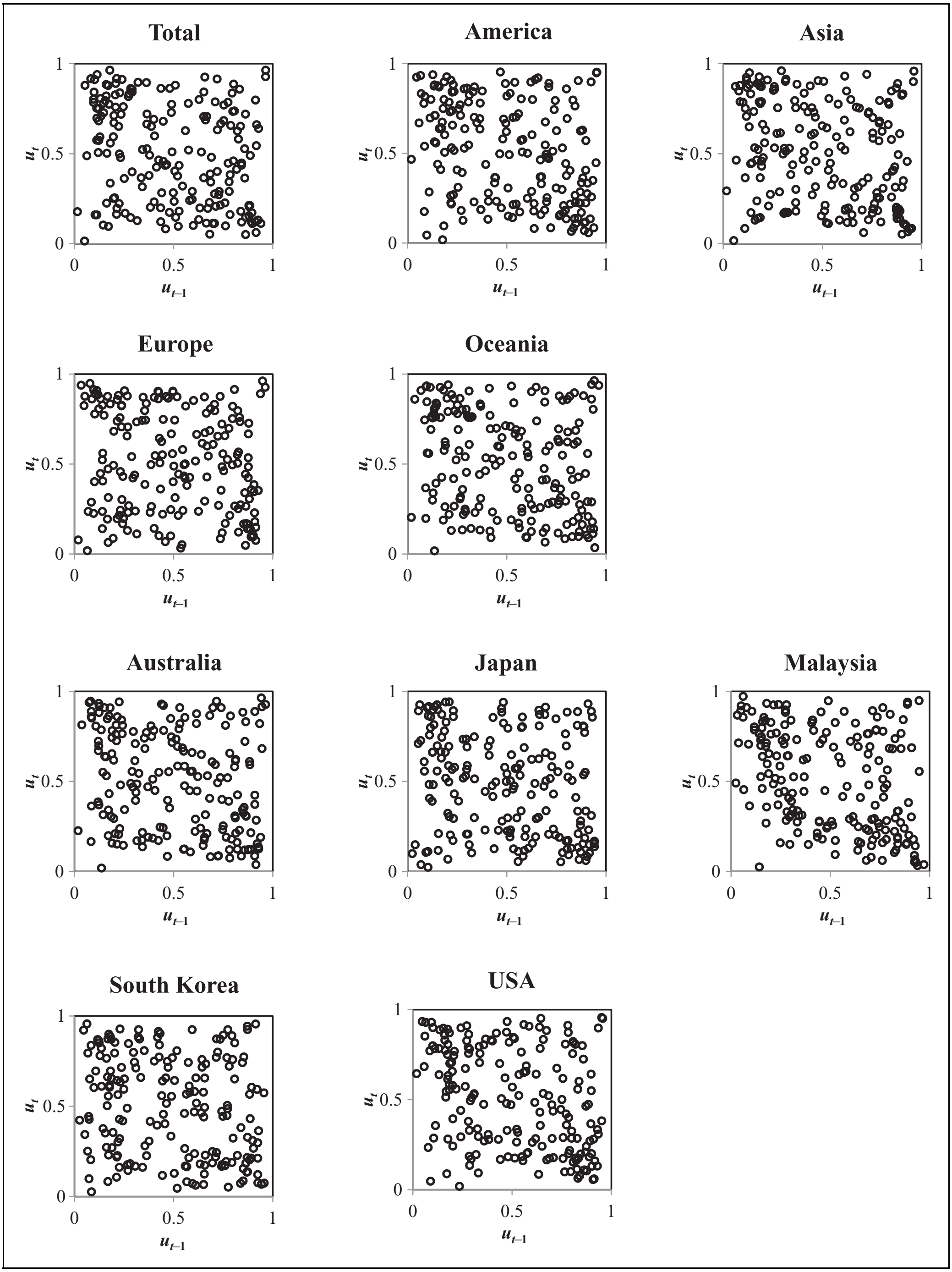

Figure 3. Scatter plots of cumulative densities for tourist arrivals to Singapore.

Table 4. Goodness-of-fit test for copula selection.

	p-Value of S_n^C statistics				
Tourist arrivals series	Gaussian	Student-*t*	Frank	Clayton	Gumbel
Total	0.400	0.463	0.481	0.000	0.002
America	0.327	0.413	0.385	0.000	0.000
Asia	0.469	0.485	0.457	0.001	0.001
Europe	0.496	0.518	0.473	0.024	0.037
Oceania	0.386	0.436	0.387	0.004	0.002
Australia	0.429	0.461	0.418	0.014	0.005
Japan	0.376	0.425	0.370	0.004	0.005
Malaysia	0.388	0.421	0.412	0.000	0.000
South Korea	0.480	0.533	0.543	0.034	0.037
USA	0.328	0.383	0.441	0.001	0.000

Table 5. Log-likelihood values comparison for optimal copula.

	Log-likelihood value		
Tourist arrivals series	Gaussian	Student-*t*	Frank
Total	5.711	*19.100*	6.796
America	11.150	*22.460*	12.670
Asia	7.368	*21.750*	8.069
Europe	3.195	*17.320*	3.334
Oceania	6.004	*20.620*	6.760
Australia	5.546	*17.760*	6.111
Japan	4.925	*18.890*	5.763
Malaysia	17.440	*26.580*	15.480
South Korea	4.830	*10.560*	4.715
USA	9.967	*20.950*	10.920

criterion is needed. Genest et al. (2009) summarized several goodness-of-fit testing for copula models, and we use the statistic $S_n^{(C)}$ recommended by them as the selection criterion for the copulas used in this study. If the p-value of $S_n^{(C)}$ is higher than the significance level, the corresponding copula can be used to specify the serial dependence. We use 0.05 significant level in our study, and the goodness-of-fit test results are reported in Table 4. According to the test results (given by the p-value of $S_n^{(C)}$), the Gaussian, Student-*t* and Frank copulas can be used to describe the serial dependence structures for all arrival series, while the Clayton and Gumbel copulas are not applicable.

To select the optimal copula for each series, we compare the LL values of the four models. The LL values of the copula models are reported in Table 5 and the copula with the largest LL value provides the best fit. It is apparent that the Student-*t* copula fits the data better than the other copulas for all the arrivals' growth series. This is consistent with the patterns shown in Figure 3.

Table 6 reports the association parameters of the optimal copulas selected. The parameters are all significant at 0.001 level and they provide further information about the serial dependence of

Table 6. Association parameters of optimal copula.

Tourist arrivals series	Optimal copula	Association parameter
Arrivals: Growth		
Total	Student-*t*	−0.305***
America	Student-*t*	−0.417***
Asia	Student-*t*	−0.311***
Europe	Student-*t*	−0.194***
Oceania	Student-*t*	−0.306***
Australia	Student-*t*	−0.294***
Japan	Student-*t*	−0.270***
Malaysia	Student-*t*	−0.415***
South Korea	Student-*t*	−0.261***
USA	Student-*t*	−0.386***

***Results are statistically significant at 0.001 level.

different arrivals series. Since the dependence structures of all the growth series are described by the Student-*t* copula, the negative parameters reveal negative tail dependence. This shows that the negative association between current and next period/future growths is stronger in the tails than the association in the centre. We can also infer that the current high growth in the right tail is associated with low growth in the future. The opposite is true for the left tail.

Forecasting comparison

The Student-*t* copula model is used to forecast the 10 tourist arrivals series. From the perspective of practical implications, it is more useful to forecast actual rather than seasonally adjusted arrivals. Thus, we transform the estimated forecasts to seasonalized/actual arrivals using monthly seasonal factors extracted by Census X-13.

To examine the contribution of well-specified serial dependence structures to tourism demand forecasting, we compare the predicted seasonalized arrivals of the Student-*t* copula with the forecasting results of ARIMA and SARIMA models. The optimal ARIMA and SARIMA models for each original series are determined by the Akaike information criterion and they are provided in both Tables 7 and 8. The common measures of forecast error, which are the mean absolute percentage error (MAPE) and root mean square error (RMSE), are used to evaluate tourism forecast performance.

According to Vergori (2012), MAPE values lower than 50% are acceptable and values lower than 20% correspond to good forecasts. The out-of-sample MAPE results in Table 7 demonstrate that all MAPEs are lower than 20%, indicating that all the three models perform well in forecasting tourist arrivals to Singapore. However, the Student-*t* copula outperforms both ARIMA and SARIMA models in forecasting. In general, the Student-*t* copula has lower average MAPEs than ARIMA and SARIMA models (see the last column of Table 7). We also make detailed comparison of forecasting performance for copula and ARIMA/SARIMA models at different forecasting horizons. The Student-*t* copula performs better than ARIMA model at different horizons for all arrivals series. It outperforms SARIMA model as well, with the only exception of total arrivals at 24-month-ahead forecast.

Table 7. MAPEs of copula, ARIMA and SARIMA models: Forecasting comparison.

Tourist arrivals series	Model	Horizon							
		$i=1$	$i=2$	$i=3$	$i=6$	$i=12$	$i=18$	$i=24$	Average
Total	Student-*t*	*2.505*	*2.225*	*2.212*	*2.217*	*2.165*	*2.211*	1.720	*2.179*
	ARIMA (10,1,9)	4.096	3.257	3.352	3.226	3.037	3.219	2.353	3.220
	SARIMA (1,1,2) × $(0,1,1)_{12}$	3.616	2.718	2.778	2.765	2.721	2.727	*1.708*	2.719
America	Student-*t*	*3.585*	*3.781*	*3.778*	*3.844*	*4.057*	*3.574*	*3.907*	*3.789*
	ARIMA (12,1,6)	5.990	5.727	5.703	5.604	5.891	6.687	6.869	6.067
	SARIMA (1,1,1) × $(0,1,1)_{12}$	4.153	4.251	4.405	4.224	4.262	4.609	5.146	4.436
Asia	Student-*t*	*2.724*	*2.975*	*3.555*	*3.664*	*3.828*	*4.068*	*3.493*	*3.472*
	ARIMA (11,1,9)	4.209	4.596	5.889	5.812	5.757	6.761	6.203	5.604
	SARIMA (1,1,1) × $(0,1,1)_{12}$	3.315	3.474	4.686	4.791	4.884	5.518	4.600	4.467
Europe	Student-*t*	*2.547*	*2.662*	*2.453*	*2.186*	*2.158*	*1.785*	*1.753*	*2.221*
	ARIMA (12,1,12)	3.814	3.932	3.780	3.555	3.598	3.080	2.686	3.492
	SARIMA (1,1,1) × $(1,1,1)_{12}$	3.658	3.823	3.796	3.575	3.640	2.624	2.434	3.364
Oceania	Student-*t*	*3.820*	*4.147*	*3.993*	*3.921*	*4.050*	*4.214*	*3.912*	*4.008*
	ARIMA (9,1,10)	4.968	5.809	6.177	6.125	6.533	6.338	5.901	5.979
	SARIMA (1,1,1) × $(1,1,1)_{12}$	4.972	5.416	5.490	5.760	5.709	4.689	4.070	5.158
Australia	Student-*t*	*3.975*	*4.315*	*4.165*	*4.162*	*4.316*	*4.448*	*4.140*	*4.217*
	ARIMA (12,1,12)	5.264	5.666	5.992	6.400	6.192	6.642	5.693	5.978
	SARIMA (1,1,1) × $(0,1,1)_{12}$	5.260	5.709	5.939	6.156	5.921	5.105	4.489	5.511
Japan	Student-*t*	*2.162*	*2.136*	*1.822*	*1.598*	*1.422*	*1.432*	*1.614*	*1.741*
	ARIMA (12,1,12)	4.836	4.857	4.624	4.375	3.459	4.294	3.789	4.319
	SARIMA (2,1,2) × $(0,1,1)_{12}$	4.837	5.238	4.440	4.637	3.922	2.157	1.864	3.871
Malaysia	Student-*t*	*3.852*	*4.322*	*4.188*	*4.421*	*3.869*	*3.861*	*2.819*	*3.905*
	ARIMA (12,1,1)	5.372	6.428	6.798	6.912	5.762	7.425	5.168	6.267
	SARIMA (2,1,2) × $(1,1,1)_{12}$	4.844	6.081	6.138	6.386	4.956	5.018	5.293	5.531
South Korea	Student-*t*	*4.430*	*4.588*	*4.582*	*4.517*	*5.062*	*4.658*	*5.130*	*4.710*
	ARIMA (12,1,12)	7.761	8.753	8.641	8.619	9.095	8.526	8.865	8.609
	SARIMA (2,1,2) × $(1,1,1)_{12}$	7.334	8.128	8.135	7.380	8.107	8.168	9.697	8.135
USA	Student-*t*	*3.593*	*3.691*	*3.730*	*3.733*	*3.846*	*3.109*	*3.457*	*3.594*
	ARIMA (12,1,9)	5.599	5.566	5.836	5.424	5.517	5.833	5.814	5.656
	SARIMA (1,1,1) × $(0,1,1)_{12}$	4.175	4.242	4.415	4.209	4.395	4.825	5.035	4.471

Note: MAPE: mean absolute percentage error; ARIMA: autoregressive integrated moving average; SARIMA: seasonal autoregressive integrated moving average. MAPEs are percentage values.

Table 8. RMSEs of copula, ARIMA and SARIMA models: Forecasting comparison.

Tourist arrivals series	Model	Horizon							
		$i=1$	$i=2$	$i=3$	$i=6$	$i=12$	$i=18$	$i=24$	Average
Total	Student-*t*	*38,044*	*38,298*	*38,195*	*38,663*	*39,946*	*41,992*	37,646	*38,969*
	ARIMA (10,1,9)	65,235	61,017	62047	62855	63,298	72,271	63,567	64,327
	SARIMA $(1,1,2)\times(0,1,1)_{12}$	56,675	52,698	53,296	53,342	55,120	51,099	*36,153*	51,198
America	Student-*t*	*2514*	*2652*	*2630*	*2650*	*2795*	*2368*	*2574*	*2598*
	ARIMA (12,1,6)	3912	3819	3869	3804	3947	4148	4303	3972
	SARIMA $(1,1,1)\times(0,1,1)_{12}$	2803	2848	2920	2830	2901	3309	3618	3033
Asia	Student-*t*	*37,787*	*40,941*	*43,019*	*44,321*	*46,659*	*49,345*	*44,086*	*43,737*
	ARIMA (11,1,9)	55,907	61,521	67,958	67,065	70,711	87,694	82,937	70,542
	SARIMA $(1,1,1)\times(0,1,1)_{12}$	46,984	50006	56293	57,249	58,826	62,093	53,943	55,056
Europe	Student-*t*	*5183*	*5451*	*4814*	*4182*	*4361*	*3896*	*4113*	*4571*
	ARIMA (12,1,12)	7247	7380	6946	6655	6850	6149	5799	6718
	SARIMA $(1,1,1)\times(1,1,1)_{12}$	6642	6928	6674	6411	6709	6415	6553	6619
Oceania	Student-*t*	*5433*	*5867*	*5768*	*5828*	*6014*	*6344*	*6361*	*5945*
	ARIMA (9,1,10)	6829	7923	8193	8236	8985	9221	10,221	8515
	SARIMA $(1,1,1)\times(1,1,1)_{12}$	6799	7508	7551	7809	7780	7184	6675	7330
Australia	Student-*t*	*5073*	*5451*	*5365*	*5453*	*5661*	*5947*	*5993*	*5563*
	ARIMA (12,1,12)	6526	6986	7381	7663	7562	8815	8357	7613
	SARIMA $(1,1,1)\times(0,1,1)_{12}$	6503	7156	7324	7480	7313	6835	6475	7012
Japan	Student-*t*	*2546*	*2639*	*1948*	*1764*	*1660*	*1628*	*1792*	*1997*
	ARIMA (12,1,12)	4530	4756	4484	4409	3953	4907	4598	4520
	SARIMA $(2,1,2)\times(0,1,1)_{12}$	4663	4973	4334	4689	4154	2310	2123	3892
Malaysia	Student-*t*	*6702*	*7754*	*7481*	*7865*	*7496*	*7703*	*6324*	*7332*
	ARIMA (12,1,1)	8564	10662	10,950	11,266	10,067	12,897	10,649	10,722
	SARIMA $(2,1,2)\times(1,1,1)_{12}$	7579	9968	10,042	10,455	8858	9998	10,836	9677
South Korea	Student-*t*	*2267*	*2389*	*2389*	*2430*	*2616*	*2412*	*2621*	*2446*
	ARIMA (12,1,12)	3737	4164	4172	4205	4451	4718	4989	4348
	SARIMA $(2,1,2)\times(1,1,1)_{12}$	3940	4330	4373	4139	4453	4533	5052	4403
USA	Student-*t*	*1970*	*2082*	*2069*	*2092*	*2199*	*1679*	*1829*	*1989*
	ARIMA (12,1,9)	2680	2672	2843	2725	2671	2759	2931	2754
	SARIMA $(1,1,1)\times(0,1,1)_{12}$	2182	2228	2305	2242	2337	2624	2801	2388

Note: RMSE: root mean square error; ARIMA: autoregressive integrated moving average; SARIMA: seasonal autoregressive integrated moving average.

The out-of-sample RMSE results displayed in Table 8 are consistent with the MAPEs. The Student-*t* copula generates better forecasts than ARIMA and SARIMA models on average and at different horizons for almost all the series. Taking both MAPEs and RMSEs into account, we can conclude that the well-specified copula-based serial dependence structure contributes to the increase of forecast accuracy.

Conclusion

This article initiates the application of copula functions in specifying the serial dependence structures of tourism demand for forecasting. In many past studies, the linear specification is employed to model the serial dependence structure of tourist arrivals, as proxy for tourism demand. The article draws attention to using a generalized approach as it is insufficient to use the linear structure to describe the serial dependence of tourism demand. The proposed copula method provides numerous functions to combine marginal distributions and approximate joint dependence. As a result, both linear and non-linear relations between current and past tourist arrivals can be estimated.

Our research on copula modelling makes a contribution to the tourism empirical literature as a propitious approach for serial dependence structure analysis and univariate time series forecasting. In addition, tourist arrivals are not necessarily assumed to be normally distributed in copula modelling. The comparison of forecasting performance between copula and the two benchmark models supports that well-specified serial dependence structure increases the accuracy of forecasts.

Besides methodological contributions, our study has policy and future research implications. Analysis of serial dependence structure provides some tourism insight to decision makers of the destination government, as well as various businesses supplying tourism facilities and/or services. Our findings show that negative tail dependence structure exists in the growth series. This means that a big decline (increase) in current arrivals is more likely to lead to a positive (negative) growth in the near future than a small decline (increase). Decision makers can implement informed operational strategies in preparation for positive (negative) future growth which comes after a strong negative (positive) current growth.

According to the World Travel and Tourism Council (WTTC), tourism capital investment in Singapore has exceeded SG$17 billion in 2014 and is expected to increase by 6.1% in 2015 (WTTC, 2015). Undoubtedly, the government has tourism planning in place to capitalize on the massive investment outlays. The potential to improve the forecasts of tourist arrivals can contribute to better promotional efforts by Singapore and the operations of tourism-related businesses in the country. Copula modelling can assist to reinvigorate the planning process and Destination Marketing Organization (DMO) to gain foresight on tourism demand, leading to more sensible decision-making on tourism promotion and destination marketing.

While recognizing the benefits of forecasting research to policy formulation, there is always a trade-off between forecasting accuracy and simplicity of technique/s used (Lim and McAleer, 2002). The linear model is still appealing in practice as it is parsimonious and easy to apply. For instance, the two linear benchmark models used in our study produce good forecasts with low MAPEs, although their performances are worse than the copula method. However, the appropriateness of using the copula technique for tourism practitioners should not be underestimated in forecasting tourist arrivals to Singapore. The copula method does not only serve as a point forecasting tool, as the tail dependence revealed by the selected copula can also provide decision makers with tourism insight into the future trend of tourism demand, making the method more informative than the linear ones.

Future research can extend the study in different directions. In this study, we analysed the first-order serial association of tourism demand. Future research can generalize the copula method to higher dimensional models with seasonal and MA terms, as well as more lags for the AR terms. The copula method can be applied to other destination countries to verify its forecasting performance.

Acknowledgements

The authors are grateful to the referees for helpful comments and suggestions. The first and second authors would like to thank the Ministry of Education Tier 1 and the Institute on Asian Consumer Insight.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was financially supported by the Ministry of Education Tier 1 and the Institute on Asian Consumer Insightd.

References

Álvarez-Díaz M and Rosselló-Nadal J (2010) Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics* 16(1): 153–168.

Balakrishnan N and Lai CD (2009) *Continuous Bivariate Distributions*. New York: Springer Science & Business Media.

Beare BK (2010) Copulas and temporal dependence. *Econometrica* 78(1): 395–410.

Box GE and Jenkins GM (1970) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Chan YM (1993) Forecasting tourism: A sine wave time series regression approach. *Journal of Travel Research* 32(2): 58–60.

Chen X and Fan Y (2006) Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130(2): 307–335.

Chen X, Wu WB and Yi Y (2009) Efficient estimation of copula-based semiparametric Markov models. Cowles Foundation Discussion Paper No. 1691. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1349768 (accessed 10 February 2017).

Cho V (2001) Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality & Tourism Research* 25(4): 399–420.

Chu FL (2008) A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management* 29(1): 79–88.

Crane GJ and Van Der Hoek J (2008) Conditional expectation formulae for copulas. *Australian & New Zealand Journal of Statistics* 50 (1): 53–67.

de Melo Mendes BV and Aíube C (2011) Copula based models for serial dependence. *International Journal of Managerial Finance* 7(1): 68–82.

Deco G and Schürmann B (2012) *Information Dynamics: Foundations and Applications*. New York: Springer Science & Business Media.

Embrechts P, Lindskog F and McNeil A (2001) *Modelling Dependence with Copulas. Rapport Technique, Départment de mathématiques*. Zurich: Institut Fédéral de Technologie de Zurich.

Embrechts P, McNeil A and Straumann D (2002) Correlation and dependence in risk management: Properties and pitfalls. In: Dempster MAH (ed.) *Risk Management: Value at Risk and Beyond*. New York: Cambridge University Press, pp. 176–223.

Frahm G, Junker M and Szimayer A (2003) Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters* 63(3): 275–286.
Frees EW and Valdez EA (1998) Understanding relationships using copulas. *North American Actuarial Journal* 2(1): 1–25.
Genest C, Rémillard B and Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44(2): 199–213.
Hauke J and Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 30(2): 87–93.
Ibragimov R (2009) Copula-based characterizations for higher order Markov processes. *Econometric Theory* 25(03): 819–846.
Joe H (1997) *Multivariate Models and Multivariate Dependence Concepts*. Florida: CRC Press.
Kim JH and Ngo T (2001) Modelling and forecasting monthly airline passenger flows among three major Australian cities. *Tourism Economics* 7(4): 397–412.
Kisi O (2011) Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resources Management* 25(2): 579–600.
Kjersti A (2004) *Modelling the Dependence Structure of Financial Assets: A Survey of Four Copulas*. Norway: Norwegian Computing Center.
Klugman SA, Panjer HH and Willmot GE (2012) *Loss Models: From Data to Decisions*. New Jersey: John Wiley & Sons.
Li G, Song H and Witt SF (2005) Recent developments in econometric modeling and forecasting. *Journal of Travel Research* 44(1): 82–99.
Liang YH (2014) Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering* 74: 111–119.
Lim C and McAleer M (2000) A seasonal analysis of Asian tourist arrivals to Australia. *Applied Economics* 32(4): 499–509.
Lim C and McAleer M (2002) Time series forecasts of international travel demand for Australia. *Tourism Management* 23(4): 389–396.
Liu J and Sriboonchitta S (2013) Analysis of volatility and dependence between the tourist arrivals from China to Thailand and Singapore: A copula-based GARCH approach. In: Huynh VN, Kreinovich V, Sriboonchitta S and Suriya K (eds) *Uncertainty Analysis in Econometrics with Applications*. Thailand: Springer, pp. 283–294.
Liu J, Sriboonchitta S, Nguyen HT, et al. (2014) Studying volatility and dependency of Chinese outbound tourism demand in Singapore, Malaysia, and Thailand: A vine copula approach. In: Huynh VN, Kreinovich V and Sriboonchitta S (eds) *Modeling Dependence in Econometrics*. Thailand: Springer, pp. 259–274.
Nelsen RB (2007) *An Introduction to Copulas*. New York: Springer Science & Business Media.
Nowman KB and Van Dellen S (2012) Forecasting overseas visitors to the UK using continuous time and autoregressive fractional integrated moving average models with discrete data. *Tourism Economics* 18(4): 835–844.
Papatheodorou A and Song H (2005) International tourism forecasts: Time-series analysis of world and regional data. *Tourism Economics* 11(1): 11–23.
Peng GB, Song H and Witt SF (2012) Demand modelling and forecasting. In: Dwyer L, Gill A and Seetaram N (eds) *Handbook of Research Methods in Tourism: Quantitative and Qualitative Approaches*. UK: Edward Elgar Publishing, pp. 71–90.
Pérez-Rodríguez JV, Ledesma-Rodríguez F and Santana-Gallego M (2015) Testing dependence between GDP and tourism's growth rates. *Tourism Management* 48: 268–282.
Puarattanaarunkorn O and Sriboonchitta S (2014) Copula based GARCH dependence model of Chinese and Korean tourist arrivals to Thailand: Implications for risk management. In: Huynh VN, Kreinovich V and Sriboonchitta S (eds) *Modeling Dependence in Econometrics*. Thailand: Springer, pp. 343–365.

Sencheong K and Turner LW (2005) Neural network forecasting of tourism demand. *Tourism Economics* 11(11): 301–328.

Sklar A (1959) *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Paris: Université Paris 8.

Smith M, Min A, Almeida C, et al. (2010) Modeling longitudinal data using a pair–copula decomposition of serial dependence. *Journal of the American Statistical Association* 105(492): 1467–1479.

Tang J, Sriboonchitta S, Ramos V, et al. (2016) Modelling dependence between tourism demand and exchange rate using the copula-based GARCH model. *Current Issues in Tourism* 19(9): 876–894.

Turner LW, Kulendran N and Pergat V (1995) Forecasting New Zealand tourism demand with disaggregated data. *Tourism Economics* 1(1): 51–69.

Vergori AS (2012) Forecasting tourism demand: The role of seasonality. *Tourism Economics* 18(5): 915–930.

Witt S and Song H (2001) Forecasting future tourism flows. In: Medlik S and Lockwood A (eds.) *Tourism and Hospitality in the 21st Century*. Oxford: Butterworth-Heinemann, pp. 106–118.

WTTC (2015) *The Economic Impact of Travel & Tourism, Singapore*. London: World Travel & Tourism Council.

Wu Q, Law R and Xu X (2012) A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications* 39(5): 4769–4774.