Dependence issues in tourism demand modeling and forecasting : within and between dependence

Zhu, Liang

2017

Zhu, L. (2017). Dependence issues in tourism demand modeling and forecasting : within and between dependence. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/70340>

<https://doi.org/10.32657/10356/70340>

**DEPENDENCE ISSUES IN TOURISM DEMAND
MODELING AND FORECASTING: WITHIN AND
BETWEEN DEPENDENCE**

ZHU LIANG

NANYANG BUSINESS SCHOOL

2017

**DEPENDENCE ISSUES IN TOURISM DEMAND
MODELING AND FORECASTING: WITHIN AND
BETWEEN DEPENDENCE**

ZHU LIANG

NANYANG BUSINESS SCHOOL

A thesis submitted to the Nanyang Technological University

in fulfillment of the requirement for the degree of

Doctor of Philosophy

2017

ACKNOWLEDGEMENTS

PhD study requires extraordinary persistence, determination, and commitment. It is a tough but enjoyable journey for me with many people who have whole-heartedly supported me. Without their encouragement and assistance, it would have been impossible for me to complete this thesis. Every one of them has my sincere gratitude and appreciation.

I am especially grateful to my supervisors, Professor Christine Lim Mei Lai and Professor Wu Yuan, for their great enthusiasm, excellent guidance and kind encouragement. I would like to thank Professor Christine for the time and energy she spent in reviewing my work. Her exceptional knowledge in the tourism discipline and rigorous attitude in research are essential not only for this thesis, but also for the long-term development of my research career. Her patience and kindness have helped me tremendously when I experienced difficulties in research and life. I would also like to thank Professor Wu for his enlightening and valued insights, in particular on the methodological aspects of my thesis. His willingness to share his wisdom in copulas and time series analyses is a strong support for starting this research work. I would have never completed this thesis without the generous inputs of Professors Christine and Wu.

I wish to express my gratitude to my Thesis Advisory Committee, Proposal Defense Committee and Final Oral Defence Committee. I have very much benefited from the efficient instruction of Professors Wee Chow Hou, Chen Chien-Ming, Feng Qu, Goh Kim Huat and Joan Henderson. They are always very helpful. I really appreciate their insightful and constructive suggestions. In addition, I want to acknowledge the financial support of the Nanyang Technological University Research Scholarship for 2012-2016.

Many thanks are due to Professor Zhang Jianping who supervised my Master program at Beijing Jiaotong University. She has given much support in my study and life. My thanks are also extended to all my friends who laugh with me when I am happy and talk to me when I feel sad.

Special thanks go to my family members. In particular, I am forever grateful to my mother, Ms. Zhang Yun. I have learned how to love my family and friends, and how to love my life, from this kindhearted lady. She remains strong in spite of sickness and difficulties in life. Because of her, I am stronger. I rarely speak out my love to my mother, but I will let her know I love her forever. I wish to dedicate this thesis to my mother as a mark of respect and affection.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
TABLE OF CONTENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
ABSTRACT	vii
CHAPTER 1. INTRODUCTION	1
Economic Contributions of Tourism	1
Importance of Tourism Demand Forecasting	3
Research Motivations	6
Organization of Thesis	8
CHAPTER 2. TOURISM IN SINGAPORE	10
History of Singapore Tourism	11
Contribution of Tourism to Singapore	13
Tourism Investment and Forecasting	15
CHAPTER 3. LITERATURE REVIEW	18
Tourism Demand Forecasting Methods – Linear Models	18
Univariate Time Series Models	19
Econometric/Structural models	21
Tourism Demand Forecasting Methods – Nonlinear Models	22
Dependence between Tourism Flows	24
Concluding Remarks	25
CHAPTER 4. DEPENDENCE & COPULAS	27
Sklar’s Theorem for Joint Distribution & Copula	27
Copula Functions for Different Dependence Structures	28
Applications and Usefulness of Copula	34
CHAPTER 5. WITHIN DEPENDENCE AND UNIVARIATE TOURISM DEMAND FORECASTING	38
Two-Dimensional Copula Model for Within Dependence (Study 1)	39
Methodology	39
Data Description	41
Preliminary Test	42
Analysis of Within Dependence Structure	45
Copula Model Specification	49
Forecasting Results	53

High-Dimensional Copula Model for Within Dependence (Study 2).....	58
Methodology	59
Data Description	62
Preliminary Test.....	62
Forecasting Results	65
CHAPTER 6. BETWEEN DEPENDENCE AND MULTIPLE TOURISM FLOWS	
FORECASTING	72
Copula-based Approach for Between Dependence of Tourism Flows	72
Empirical Analysis.....	75
Data and Variables	75
Preliminary Test.....	77
Model Estimation.....	81
Forecasting Results	83
CHAPTER 7. WITHIN-BETWEEN DEPENDENCE	87
Copula-WB framework.....	87
Empirical Analysis.....	89
Data Description	89
Preliminary Test.....	89
Copula Model Specification.....	90
Forecasting Results	92
CHAPTER 8. CONCLUSION & RECOMMENDATIONS	98
Overall Summary of the Findings	98
Contributions of the Research.....	99
Limitations of the Research & Future Studies	102
REFERENCES	106
APPENDICES	115
Appendix A. Total Tourist Arrivals and Arrivals from Different Source Countries (1995M01 to 2013M12).....	115
Appendix B. Probability Density Functions of Five Copulas	117
Appendix C. Binary Search Algorithm.....	118
Appendix D. Robustness Check of Forecasting Results	119
Appendix E. Log-likelihood Values of Copula-WB Framework (Full Version).....	127

LIST OF TABLES

Table 4.1 Comparison among Copula, AR SARIMA & Structural Models	36
Table 5.1: Unit Root Tests for Tourist Arrivals to Singapore	43
Table 5.2: Chi-Square Tests for Distributions and LM Test for Serial Correlation	45
Table 5.3: Associations between Tourist Arrivals and Their First Lags	47
Table 5.4: Goodness-of-fit Test for Copula Selection	50
Table 5.5: Log-likelihood Values Comparison for Optimal Copula	51
Table 5.6: Association Parameters of Optimal Copula	52
Table 5.7: MAPEs of Copula and AR(1) Models over Five Time Horizons	57
Table 5.8: RMSEs of Copula and AR(1) Models over Five Time Horizons	58
Table 5.9: Unit Root Tests for Tourist Arrivals to Singapore	63
Table 5.10: Chi-Square Test for Distribution of Tourism Demand Series	64
Table 5.11: Breusch–Godfrey LM Test for Serial Correlation.....	65
Table 5.12: SARIMA Model Parameters for Arrivals Series	66
Table 5.13: MAPEs of Copula and Benchmark Models over Five Time Horizons.....	70
Table 5.14: RMSEs of Copula and Benchmark Models over Five Time Horizons.....	71
Table 6.1: Unit Root Tests for Related Variables	78
Table 6.2: HEGY Test for Tourist Arrivals Series^a	79
Table 6.3: Bounds Test for Cointegration and ARDL Specification	80
Table 6.4: Johansen Cointegration Test.....	80
Table 6.5: Goodness-of-fit of Copula Models	82
Table 6.6: Estimation Results of Dependence Structure of Copula Models	83
Table 6.7: MAPE & RMSE of Different Forecasting Models.....	84
Table 7.1: Log-likelihood Values of Copula-WB Framework	91
Table 7.2: ARIMA Model Parameters for Arrivals Series.....	92
Table 7.3: MAPEs of Copula and Benchmark Models over Five Time Horizons.....	96
Table 7.4: RMSEs of Copula and Benchmark Models over Five Time Horizons	96

LIST OF FIGURES

Figure 2.1: Tourist Arrivals to Singapore and Annual Percentage Change.....	10
Figure 2.2: Tourism Receipts, Singapore.....	14
Figure 4.1: Scatter Plots of Two-dimensional Gaussian Copula	30
Figure 4.2: Scatter Plots of Two-dimensional Student-t Copula	31
Figure 4.3: Scatter Plots of Two-dimensional Frank Copula.....	32
Figure 4.4: Scatter Plots of Two-dimensional Clayton Copula.....	33
Figure 4.5: Scatter Plots of Two-dimensional Gumbel Copula	34
Figure 5.1: Scatter Plots of Cumulative Densities for Tourist Arrivals to Singapore	48
Figure 5.2 Actual Total Arrivals and Forecasts of AR(1) and Copula Models (Horizon $i=1$)	54
Figure 5.3: Actual Total Arrivals and Forecasts of AR(1) and Copula Models (Horizon $i=24$)	54
Figure 5.4: Actual Total Arrivals and Forecasts of AR(13) and Copula Models (Horizon $i=1$)	67
Figure 5.5: Actual Total Arrivals and Forecasts of AR(13) and Copula Models (Horizon $i=24$)	67
Figure 5.6: Histograms of MAPE & RMSE for Total Tourist Arrivals by Forecast Horizons	69
Figure 6.1: Scatter Plots of Percentile Ranks of Residuals	81
Figure 6.2: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 1)	85
Figure 6.3: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 2)	85
Figure 6.4: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 3)	86
Figure 7.1: Actual Arrivals of Pair 1 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=1$).....	93
Figure 7.2: Actual Arrivals of Pair 1 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=24$).....	94
Figure 7.3: Actual Arrivals of Pair 2 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=1$).....	94
Figure 7.4: Actual Arrivals of Pair 2 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=24$).....	95

ABSTRACT

Tourism forecasting has long been an attractive topic in the tourism demand literature because of the undoubted significance of such studies for policy purposes. Many researchers have applied or modified various techniques routinely used in business for computing tourism forecasts and increasing forecast accuracy. Among these techniques, linear models are popular as they are computationally straightforward and can be easily applied in practice. However, a linear relationship is just one of the numerous dependence structures, and the misspecification of dependence structures can cause over- or under-estimation problem.

This thesis proposes the copula technique to address this issue. The research focuses on international tourism demand for Singapore and applies the copula method to examine tourism flows/arrivals to Singapore from 1995 to 2013. The empirical analysis of this thesis covers not only the serial dependence (labeled as Within Dependence) of individual tourism flows to Singapore, but also the dependence between different tourism flows (labeled as Between Dependence). Specifically, I conduct four studies to advance knowledge of dependence structure in tourism demand modeling and make a contribution to tourism forecasting with the use of the copula method. Study 1 and 2 analyze Within Dependence. The simple two-dimensional model is initially used to examine the Within Dependence of tourism demand and its first lag in Study 1. Study 2 investigates the sophisticated high-dimensional cases and includes higher order lags of tourism demand. The third study examines whether the dependence between tourism flows exists, to shed light on the Between Dependence and contribute to the joint forecast of tourism flows. In

the last study, I take both the Within and Between Dependences into account, and develop a comprehensive copula-WB framework for forecasting.

By analyzing the tourism flows to Singapore, the empirical results of the studies reveal the different serial dependence structures of tourism flows per se and the dependence structures between flows. Additionally, the studies show that in contrast to the AR, ARIMA, SARIMA, ARDL and ECM models, the copula method is flexible in specifying dependence structure and can improve tourism demand forecasting.

CHAPTER 1. INTRODUCTION

The past three decades have witnessed the dramatic growth of global tourism demand. According to the United Nation World Tourism Organization (UNWTO 2014), international tourist arrivals increased by 5% in 2013, leading to a record of 1087 million arrivals all over the world. International tourism is regarded as an engine of economic recovery, and is reported to have reached a growth rate of 4.7% in 2014 (UNWTO 2015a). This rapid development of the tourism industry stimulates the rising interest in tourism research. As one of the most important areas in tourism study, tourism demand modeling and forecasting has attracted considerable attention, because of its great impact on decision making of government and tourism related businesses (Song and Witt 2006).

In this introductory chapter, I will discuss the economic contribution of tourism and the importance of tourism demand forecasting. Subsequently, I will elaborate why dependence analysis should be considered in tourism demand forecasting and present the research motivations. The organization of this thesis will be provided at the end of the chapter.

Economic Contributions of Tourism

Li (2004) pointed out that the tourism sector “does not only draw upon the resources of the national economy and contributes directly to economic growth, but also stimulates the growth of other economic sectors via spill-over effects” (p. 4). These arguments have been supported by research on different destination countries, including developing countries such as Algeria (Poirier 2001), China (Oosterhaven and Fan 2006), Egypt (Var and Imam 2001), Fiji (Narayan 2004) and Turkey (Tosun et al. 2003), and developed

countries such as Australia (Dwyer et al. 2003; Dwyer and Forsyth 1996), Singapore (Khan et al. 1990), South Korea (Lee and Kwon 1995) and the United States (Blake et al. 2001) among others.

In terms of GDP, researchers have shown that tourism receipts significantly affect income and government revenue for many countries or regions (Fletcher 1989; Archer and Fletcher 1996; Daniels et al. 2004; Archer 1995; Dwyer et al. 2000). International tourism receipts exceeded \$50 billion for each of the following countries: United States, Spain, China, France and Macau in 2013 (UNWTO 2015b). For others such as Fiji, Greece and Thailand, tourism receipts accounted for 17.1%, 6.7% and 9.9% of the GDP in 2013, respectively. The tourism sector also increases GDP by attracting foreign investment. Studies by Barrowclough (2007), Endo (2006), Ivanovic et al. (2011), Khoshnevis Yazdi et al. (2015) and Peric and Niksic Radic (2011) have shown that foreign investment in tourism has been increasing in different countries. Moreover, tourism influences imports and exports of goods, both of which are also important components of GDP. Keum (2011) argued that the international tourism flows affect the trade flows between South Korea and its 21 trading partners worldwide, and the trade flows influence the tourism flows in turn. His empirical findings concur with other researchers' like Gray (1970), Kulendran and Wilson (2000) and Shan and Wilson (2001).

The contribution of tourism to national economies is also evident in employment generation. As a labour intensive industry, tourism creates many job opportunities for destination residents. According to Turner and Sears (2013), the global tourism industry has hired more than 98 million people directly, which is equivalent to over 3 percent of worldwide employment. If indirect and induced impacts are included, the tourism industry

contributes to around one in every eleven jobs worldwide in 2013. Furthermore, tourism receipts have a positive effect on employment in other industries, which can be measured by the employment multiplier. Empirical studies about the impact of tourism on employment are conducted by Dwyer and Forsyth (1998) on Australia, Witt et al. (2004) on Denmark, Narayan (2004) on Fiji, Huse et al. (1998) on Norway, and Lee and Kwon (1995) on South Korea.

Another significant role of tourism is its contribution to the foreign exchange reserves of tourist-receiving countries (Dwyer and Forsyth 1993; Balaguer and Cantavella-Jorda 2002; Tisdell and Wen 1991). International tourism receipts constitute an important part of foreign exchange earnings for many countries and their balance of payments (Narayan 2004; Li 2004). In addition, the foreign exchange earnings can be used to import capital goods for production, which are beneficial for economic growth (Seetanah 2011).

Importance of Tourism Demand Forecasting

Tourism demand refers to the amount of tourism products that a tourist is willing and able to buy (Croes and Vanegas Sr 2005). Turtureanu (2010) describes tourism as a composite product involving facilities and services like attractions, transportation, accommodation, entertainment, travel agencies and associated services. The rapid development of the tourism sector in many countries has caused tourism demand forecasting to become increasingly essential. Foresight of demand is acutely needed for the following reasons (Douglas 2001).

Firstly, most tourism products are perishable. The inability to stockpile tourism products such as empty airline, ferry and restaurant seats or hotel rooms will result in the loss of revenue for the respective businesses (Song and Witt 2006; Du Preez and Witt 2003). Foresight of tourism demand can help these businesses to make better decisions and plans on supply. Secondly, the production and consumption of most tourism products are inseparable and take place simultaneously at the destination. Demand foresight can help tourism-related firms to make better decisions with operational matters and manpower planning. Thirdly, tourism supply requires large, long lead-time investments in equipment and infrastructures. Long-term financial supports are essential for the constructions of these facilities and the expected net returns on the investments would not be achieved if tourism demand is insufficient to fully utilize the designed capacities of the facilities (Song and Witt 2006). Demand forecasting is thus used to examine the feasibility of investment spending on the supply of tourism, in order to avoid wastage of financial resources and investment opportunity foregone.

The practice of tourism demand forecasting can be extended beyond the travel component of a trip to include destination accommodation, attractions, among others (Witt and Witt 1995). The importance of tourism demand forecasting for related businesses has actually been discussed in the literature. Forecasting has a long track record in the airline industry and accurate forecasts are crucial for its yield management. Lee (1990) found that a 10% increase in forecast accuracy in the airline industry increased revenue by 0.5-3.0% on high demand flights. Tourism demand forecasting contributes to the revenue increase by assisting an airline in its fleet planning, route network development and annual operating plan (Srisaeng et al. 2015), and in reducing its risk

through an objective evaluation of the demand side of the airline business (Ba-Fail et al. 2000).

Accurate reservation forecasting is crucial to the hotel industry as well. If information about future demand for hotel rooms is inaccurate, either over-booking or under-booking will occur, and both are very costly to the hotel business (Law 2004). With accurate forecasts, decision-makers and managers in the hospitality industries can improve strategic planning and decision making (Claveria et al. 2015). Anticipating demand enables them to engage in inventory, workforce, supplies, financial budgeting and pricing management appropriately, which lead to the maximization of revenue and minimization of costs (Caicedo-Torres and Payares 2016). Schwartz and Hiemstra (1997) argued that a long forecasting horizon is required, as daily predictions of hotel demand are often used many days before the actual date of guest arrival (often as early as 90 days ahead). However, Hayes and Miller (2011) pointed out that in recent years, hoteliers have stressed the importance of short-term forecasting for the purpose of staying agile in a fiercely competitive marketplace.

Apart from the airline and hospitality industries, demand forecasting can also benefit other tourism related businesses. For instance, reliable forecasts can help a travel agency to design different tourism packages comprising hotel rooms, flights, tickets to scenic spots, among others. Huang (2013) pointed out that accurate forecasts of tourism demand can decrease the cost pressure caused by overstocking or lost orders due to shortage. It has also been shown that the US national parks provide visitation forecasts for approximately 350 parks, and understanding future demand is critical to the provision of

services, appropriate application of funds, and budgeting concerns (Wilmot and McIntosh 2014).

Prior research shows that the forecasting performance of tourism demand models could be diverse due to different data frequencies, length of forecast periods, origin countries or destinations used (Witt and Song, 2001, Li et al., 2005). As such, there is no universally best model for tourism demand forecasting. Although forecasting is a challenging task, continuous efforts should be made to improve the forecasting performance of tourism demand models.

Research Motivations

Forecast endeavours should be underpinned by knowledge of dependence structure; however, discussion of the latter has been insufficient in the tourism forecasting literature as it is a common practice to assume the dependence structure to be linear. Yu et al. (2007) have found that the accuracy of forecasting models are influenced by tourist arrival data patterns in a specific and systematic way. Using a large tourist arrival data set, they have identified four distinct data patterns: (1) a stable linear trend, (2) a nonlinear trend, (3) a wave-shaped trend, and (4) an abrupt change pattern. The last three reflect nonlinear patterns in tourist arrivals, and are usually caused by structure break, seasonality, holiday fluctuation or one-off event. Using linear models to represent tourist arrivals series with nonlinear patterns can be problematic. Thus, it is necessary to analyze the dependence structures of tourism demand series before we can use appropriate linear or nonlinear models for forecasting.

The purpose of this thesis is to investigate tourism flow dependence structure using the copula method. The appealing feature of the copula method is that it captures both the linear and non-linear forms of dependence in a time series, and there are no restrictions imposed on the distribution of the dependent variable. Specifically, I apply the copula method to examine the Within Dependence (aka serial dependence) and it refers to the dependence of tourism demand on its historical information. It is essential to understand the Within Dependence, as the basis of tourism demand forecasting is to use the historical information to predict future demand.

Examining the Between Dependence, which refers to the dependence between tourism flows from different source markets for the same destination, is also important, especially for the joint forecast of tourism flows. According to Du Preez and Witt (2003), the cross-correlation between tourism flows from origin countries in the same region can be included in multivariate time series models for forecasting. Many studies have invariably assumed independence or no association between tourism flows. However, if tourism flows are not independent, the misspecification of dependence structure will distort the forecasting performance of the tourism demand models. The copula technique can display both linear and nonlinear dependences, and is thus useful for modeling Between Dependence structure.

This thesis proposes the use of the copula technique as an alternative approach to take the dependence structure of tourism flows into account, and offers a new perspective on tourism demand modeling and forecasting. When the dependence structures are linear, the existing linear models are still preferable, as they are parsimonious and can achieve similar forecasting accuracy as the copula method does. However, if the dependence

structures are nonlinear, appropriate specifications of dependence structures with copulas give us more accurate forecasts of tourism demand than the linear models. Four studies are conducted on tourism demand for Singapore using copulas to analyze both Within and Between Dependence. The source countries under investigation are Australia, Bangladesh, Canada, Germany, Indonesia, Japan, Malaysia, Myanmar, New Zealand, South Korea, Switzerland, Thailand and the United States. Total tourist arrivals to Singapore are also examined. The nonlinear patterns in these arrivals series can be seen in Appendix A. The research contributions of the thesis are to extend the current discussion of tourism modeling and forecasting to a more generalized framework, and to advance knowledge of dependence structure in tourism demand.

Organization of Thesis

The outline of this thesis is as follows. Tourism development in Singapore is discussed in Chapter 2. Chapter 3 reviews previous studies on tourism demand forecasting, in which the commonly used forecasting models are evaluated. Chapter 4 describes the basic principle of the copula method and the commonly used copula functions. The application of copulas in different disciplines is presented in this chapter as well. Chapter 5 includes 2 studies which examine the Within Dependence structure in univariate tourism demand forecasting. The first study uses a simple two-dimensional method to analyze the Within Dependence of tourism demand and its first lag, while the second discusses a sophisticated high-dimensional case and incorporates higher-order lags. In Chapter 6, a study on the Between Dependence of multiple tourism flows and joint forecasting are investigated. In this study, tourism flows from six origin countries located in three regions

are classified into three country-pairs for analysis. The fourth study, which analyzes the Within-Between Dependence for two country-pairs to generate joint forecasts of tourism demand, is included in Chapter 7. The conclusion of this thesis is provided in Chapter 8. This chapter also presents the contributions and limitations of the thesis, as well as some suggestions for future research.

CHAPTER 2. TOURISM IN SINGAPORE

Singapore is an island country located at the southern tip of Malaysia in Southeast Asia. Famous for its green cityscape, multicultural environment and low crime rate, Singapore attracts tourists from all over the world with the additional advantage that English is widely spoken. Inbound tourism to Singapore has experienced a rapid increase with over 15 million visitor arrivals in 2013 (almost three times the population of Singapore). Except for the negative impacts of the Asian Financial Crisis in 1997, SARS in 2003 and the Global Financial Crisis in 2008-2009, tourism demand for Singapore increased at an average rate of 6.34% in the recent decades leading to 2013 (see Figure 2.1).

Figure 2.1: Tourist Arrivals to Singapore and Annual Percentage Change

In this chapter, the history of Singapore tourism, contribution of tourism and investment in tourism by the Singapore government are evaluated.

History of Singapore Tourism

Singapore has been promoted as a tourist destination for more than 50 years after the establishment of the Singapore Tourist Promotion Board (STPB) in 1964. The country has experienced great development in tourism, and tourist arrivals to this destination country increased to 15.6 million in 2013 from only 98,481 in 1965 (Henderson 2015).

The STPB, which was renamed Singapore Tourism Board (STB) in 1997, played a leading role and cooperated with various government agencies and industry stakeholders to enhance the tourism attractions in Singapore. The first remarkable extension happened in the 1970s. Various constructions, such as the Raffles Landing Tourist Project, the Merlion Project, the Jurong Bird Park and the Singapore Zoological Gardens, were undertaken, together with the development of Sentosa as a resort island. The follow-up projects in the 1980s included the redevelopment of Boat Quay and Clarke Quay. Since 2004, several new projects have also been launched to further improve the quality of Singapore tourism, including the Integrated Resorts (IRs), Singapore Flyer, Gardens by the Bay, among others. Most of these man-made attractions are successful in bringing tourists to Singapore, though some like the Singapore Flyer has struggled to meet its visitor target in the early few years after its opening and its ownership has been changed in 2014.

Apart from the construction of the attractions, a lot of events have been hosted to enrich the tourists' experience in Singapore. The various regular events include the annual

Singapore Food Festival, the Great Singapore Sale, Singapore Biennale, Art Stage Singapore and Laneway Music Festival. Mega events such as the Formula One™ have also been introduced. Since 2008, the Formula One™ has contributed an average of S\$150 million to Singapore tourism each year, except for 2009 (STB 2014).

STB which oversees all aspects of tourism in the country has developed several marketing strategies to publicise the destination as a leading leisure and services centre in Asia (STB 2014). During the 50 years of tourism development, the destination image of Singapore has undergone great changes. The initial destination image of Singapore showcased the combination of Chinese, Malay and Indian culture which was highlighted by the branding “Instant Asia” in 1970 mainly to attract western tourists (Chong 2010). New elements, such as the modernization of the city state and the fusion of Eastern and Western cultures, have been gradually incorporated to reposition Singapore tourism. Destination rebranding in the ensuing years includes “Surprising Singapore”, “New Asia – Singapore” and “Uniquely Singapore”.

As the tastes and preferences of tourists keep changing and become more diverse, STB introduced “YourSingapore” in 2010. The new ‘YourSingapore’ brand emphasizes the uniqueness of Singapore and captures a personalized vacation experience which tourists are looking for (Popescu and Corboş 2011). Additionally, STB has launched customised integrated marketing strategies and activities in Australia, China, India, Indonesia, Malaysia, the Philippines, Thailand and Vietnam. Meanwhile, new digital and social media initiatives have also been promoted by STB to increase interaction with potential tourists from key markets and encourage exchange of information and ideas online (STB 2016a). Lately, the marketing focus has gradually expanded to potential short

haul markets such as the Tier 2 cities of China, Indonesia and India and long haul markets such as United States, United Kingdom, Germany, France, Russia and Switzerland (STB 2016b).

To globally promote Singapore as a compelling leisure and business destination, STB has established 20 regional offices worldwide in 7 regions, namely Americas, Europe, Greater China, North Asia, Oceania, Southeast Asia, and South Asia, Middle East & Africa (STB 2015). These offices serve to enhance Singapore's destination awareness and image through marketing, media and trade initiatives.

The efforts of STB in partnership with public and private sectors have been successful and this is reflected in the numerous tourism awards and accolades received by Singapore. The city state has been considered as the top destination in Asia and even in the world. For example, it was ranked 1st in Asia and 10th globally by the World Economics Forums' 2009 Travel & Tourism Competitiveness Report, and won the "Destination of the Year" of Travel Trade Gazette (TTG) Asia Awards in 2010 and 2011. Moreover, it was selected as one of the top 25 destinations in Asia by TripAdvisor Travelers' Choice Awards in 2010 and 2013. It was continuously rated the best family destination, best business city, favourite leisure city, and top destination for meetings, incentives, conferences and exhibitions (MICE) by its key markets.

Contribution of Tourism to Singapore

Tourism has become a fast-growing industry and it is one of the most important sectors in Singapore (STB 2014). This industry has contributed enormous foreign exchange earnings through tourism receipts, which have grown rapidly in recent years.

Figure 2.2 shows the pattern of tourism receipts from 1995 to 2013 in Singapore which has increased dramatically after the devastating impact of Severe Acute Respiratory Syndrome (SARS) on Singapore tourism in 2003 and the 2009 Global Financial Crisis. Indeed, tourism has become one of Singapore's key economic contributors and earned 4% of Singapore's GDP with tourism receipts of S\$23.5 billion (approximately 19.30 billion USD) in 2013.

Figure 2.2: Tourism Receipts, Singapore

In addition to generating foreign exchange, tourism plays an important role in employment creation in Singapore. According to the World Travel & Tourism Council (WTTC 2015b), more than 140,000 jobs were directly generated by tourism in 2013 (which is equivalent to over 4% of total employment), and the number exceeded 150,000 in 2014 (4.3% of total employment). The total contribution (direct and indirect) of tourism

to employment was reported to reach 303,000 jobs in 2014 (or 8.5% of total employment). WTTC predicted that tourism will create 180,000 (323,000) direct (total) employment by 2025 (WTTC 2015b). In the Singapore Benchmarking Report, WTTC (2015a) further revealed that tourism directly employed more people than every other sector in 2014 and sustained a total of 4.4 million direct, indirect, and induced jobs.

Tourism Investment and Forecasting

Given the important contributions of tourism, the government has provided considerable investments in this industry to realize its economic benefits. In 2004, the Tourism 2015 Plan was launched by the Singapore Government, in which a S\$2 billion Tourism Development Fund was established to further support tourism development, sustainability and growth trajectory (STB 2014). Together with private-sector investment, the development of tourism-related infrastructure projects has enabled Singapore to host many events (MICE, business, cultural, sporting, among others). Mega tourism projects that have come on-stream include the two Integrated Resorts, Singapore Flyer, Gardens by the Bay, a range of attractions, Sports Hub, cruise terminal, and the construction of two new terminals at Changi Airport, Jewel Changi Airport Complex and an extensive land transport network are in the pipeline (BMI 2015; STB 2014). These and other infrastructures have transformed Singapore's tourism landscape and enhanced the destination attractiveness.

At the 2016 Singapore Tourism Industry Conference, STB disclosed a new five-year Tourism Development Fund (STB 2016b) for tourism product, event and capability developments. The priority of the investments is to support event and cruise development,

and hotel redesign. The fund is also to be used for business quality standards improvement, training of tourism industry professionals to enhance tourism capability, and improving overall visitor experience and satisfaction.

Given the enormous investment in tourism, the Singapore government should be very cautious to distribute the money appropriately to avoid over-supplies. According to Khan et al. (1990), the country actually experienced an over-supply of hotel rooms in the 1980s. The steady growth of tourism in the 1960s and 1970s resulted in widespread optimism, leading to the over-construction of new hotels and the expansion of the existing ones. Apparently, hotel occupancy rates fell from 86% in 1980 to about 65% in 1986. This situation could be averted if the government had a better understanding of future tourism demand.

STB releases information about tourism demand forecasts in different reports, news and other publications. These forecasts are useful for STB's planning, although there is no information on the source of the forecasts and the forecasting methodology used. Business Monitor International (BMI) Research provides quarterly reports on Singapore tourism, which include detailed industry forecasts and other expert analysis. Ordinary Least Squares (OLS) estimators are mainly used by BMI Research, and the forecasting tools include different linear models such as autoregressive moving average model (ARMA), vector autoregressions, traditional decomposition methods and smoothing methods (BMI 2015).

The forecasting tools in BMI Research reports have been used for decades. Advanced forecasting methods could be more informative. Researchers have proposed different methods to forecast tourism demand for Singapore. These methods include the

revised versions of the ARMA model (Chu 2008, 2009), as well as the complex nonlinear models (Chu 2004; Kon and Turner 2005; Chan 1993). Details of these forecasting methods will be reviewed in Chapter 3.

CHAPTER 3. LITERATURE REVIEW

Tourism demand modeling and forecasting relies heavily on dependence analysis. The dependence (also called association or relationship) among different variables is the basic issue to be addressed in statistical modeling (Frees and Valdez 1998). Actually, most empirical studies in economics and other social sciences conduct dependence analysis and determine if a change in a response variable labeled Y is caused by an explanatory variable labeled X (or a set of variables labeled $X_1, X_2, \dots, X_k$) (Wooldridge 2010). When it comes to time series analysis, the foundation is serial dependence in which the focal relationship is among the response variable Y_t and its lagged variables $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$. For tourism demand modeling and forecasting, the main focus is always on how to apply the dependence between tourism demand and the explanatory variables or serial dependence to construct the forecasting model.

In this chapter, studies on tourism demand forecasting are reviewed. Specifically, I will discuss major tourism demand forecasting models used in prior studies, including both linear models (univariate time series and econometric structural models) and nonlinear models. The limited discussion on interdependence between tourism flows will also be presented.

Tourism Demand Forecasting Methods – Linear Models

International tourism demand is often measured in terms of the number of tourist arrivals from an origin country to a foreign destination. During the past decades, researchers have implemented various approaches to tourism demand modeling and forecasting of tourist arrivals from either a single or multiple origin countries. Two types

of quantitative methods used for tourism demand analysis and forecasting are univariate time series models and econometric (or structural) models. The time series models involve the statistical analysis of tourist arrivals data only and attempt to predict future arrivals from current and past values. Econometric models are based on the statistical analysis of data for other related (explanatory) variables, and the use of these variables to forecast tourist arrivals.

Univariate Time Series Models

A review of studies on quantitative tourism forecasting methods by Peng, Song, and Crouch (2014) shows that time series models are the dominant methods used since the 1980s. The models forecast tourism demand based on one variable (usually tourist arrivals) by extrapolating its past values. Univariate time series methods rely on single-variable time-series data collected over time and project data points into the future using the historical information (Hu 2002). They include basic and advanced approaches, of which the basic form of the Box and Jenkins (1970) autoregressive moving average (ARMA) model is most frequently used because of its generality. This model linearly combines the historical values of tourism demand with the white noise processes (white noise residuals as the proxies) to construct the forecasting model for stationary series. Extensions of ARMA models have been proposed which include ARIMA (where 'I' stands for integrated) models (Cho 2001; Papatheodorou and Song 2005) for non-stationary series, seasonal ARIMA (or SARIMA) models (Liang 2014; Goh and Law 2002; Kim and Ngo 2001; Lim and McAleer 2000) for seasonality, and fractional integration ARMA or ARFIMA (Chu 2008) for non-integer values of the differencing parameter, among others.

Invariably, the family of ARMA models are popular applications in tourism demand forecasting studies. Some other univariate time series models, including Naïve models and exponential smoothing models, have also been used in tourism demand forecasting literature. The naïve forecasting method is the simplest since it assumes no change of tourism demand in the future and uses either the value of last period (y_{t-1}) or the long-term average of the historical values ($\bar{y} = \frac{1}{k} \sum_{i=1}^k y_{t-i}$) as the forecast of current tourism demand (y_t). The naïve models are usually applied with other approaches in forecasting tourism demand for comparison purposes (Gunter and Önder 2015; Chu 2014; Lin et al. 2014).

Exponential smoothing is another univariate modeling technique that relies on linear serial dependence for forecasting. It imposes a geometrically declining weighting scheme on the lagged values of a series (Brooks 2014). Several researchers initiated the use of this approach in the late 1950s (Brown 1959; Holt 1957; Winters 1960), which includes a variety of forecasting models such as the single exponential smoothing, the double exponential smoothing and the Holt-Winter's models (no seasonal, additive or multiplicative). Similar to the naïve models, most of the exponential smoothing models have been used and compared with other alternatives in forecasting tourism demand (Chen et al. 2007; Geurts and Ibrahim 1975; Athanasopoulos and Hyndman 2008). Exponential smoothing has been extended to an Error-Trend-Seasonal or Exponential Smoothing (ETS) framework by Hyndman et al. (2002), which encompasses the exponential smoothing models just mentioned. The application of the ETS framework can also be found in tourism demand forecasting research (Gunter and Önder 2015).

Econometric/Structural models

Researchers have also developed structural econometric models for tourism demand forecasting. Numerous statistical techniques have been proposed to analyze the multivariate relationships between different random variables in tourism demand. The autoregressive distributed lag (ARDL) model (or ADLM), the equilibrium correction (or error correction) model (ECM), the vector autoregressive (VAR) model and the time varying parameter (TVP) models have emerged as the main multivariate normal distribution methods used. For example, Kulendran and Witt (2003) used the ECM model and six other alternatives to forecast the demand for international business tourism and found that their forecasting performance depends highly on the length of forecasting horizon. Song et al. (2003) evaluated the forecasting accuracy of ADLM, ECM, TVP and other alternatives using expenditure-weighted nights spent by tourists in Denmark. They also found that rankings of these forecasting models vary for different lengths of horizons. Moreover, these structural models can be combined together to develop more advanced models. For example, Li et al. (2006) develop the TVP-ECM model and found that it outperformed the alternative models in forecasting, especially in forecasting the growth rate of tourism demand.

Research using econometric models shows that income of origin country, relative or tourism prices, exchange rates and transportation costs are the key factors (independent variables) that affect tourism demand (Lim 1997; Song and Li 2008). Other determinants include trends, seasonal factors, dynamics, marketing expenditures (Lim 1997), number of travel agencies (Icoz et al. 1998), hotels' age (Hanly and Wade 2007), accommodation

capacity (Ouerfelli 2008), climate index (Goh 2012) and qualitative factors, among others, which influence people's decision to demand international travel.

Since the analysis involves more than one variable, the multivariate model specification assumes there is some linear relationship between tourism demand and its determinants. The prediction of the explanatory variables is necessary to precede tourism demand forecasting. As many of the empirical tourism demand studies apply the ordinary least squares (OLS) estimation technique (Goh and Law 2011), the use of linear regression models for purposes of inference or prediction are also premised on the classical assumptions of the error terms.

Tourism Demand Forecasting Methods – Nonlinear Models

Using the 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). If the association between the variables is not elliptical or is concentrated at one extreme, models based on the linear association would not be able to capture the dependence structure adequately, leading to over- or under-estimation problem (Crane and van der Hoeek 2008).

As stated by Huarng et al. (2011), tourism demand series change drastically and their trends tend to be nonlinear. Prior empirical studies have shown evidence of nonlinearity in tourism demand series, including holiday tourism flow to Mountain Huashan in China (Chen et al. 2015), pilgrim visitors to a well-known shrine at Katra in India (Maheswaran and Khosa 2015), tourist arrivals to Taiwan (Chen et al. 2015; Zins et

al. 2012), as well as tourist arrivals to Singapore (Chu 2004). Therefore, using linear models could be problematic.

Different nonlinear models have been proposed to accommodate the nonlinear patterns. The Artificial Intelligence (AI) methods such as artificial neural networks (ANN) are the commonly used nonlinear methods in tourism demand forecasting. The use of ANN method to tourism forecasting was initiated in the late 1990s and researchers continued to introduce some improved versions in their studies (Song and Li 2008). These ANN methods have been empirically shown to outperform the linear models in tourism forecasting. For instance, Kon and Turner (2005) compared the ANN with the basic structural method (BSM) and simpler models such as the naïve and Holt-Winters methods. They found that ANN can outperform BSM and the simpler methods for short horizons when forecasting Singapore tourism demand. Chen et al. (2012) believed that incorporating empirical mode decomposition (EMD) in the ANN can contribute to tourism demand forecasting. The superiority of the model is evident when they used it to analyze international arrivals to Taiwan. Other AI methods, such as the fuzzy time series method (Zins et al. 2012), the genetic algorithms (Hernández-López and Cáceres-Hernández 2007) and the support vector machine (Chen et al. 2015), have also been proposed.

Apart from the AI methods, other nonlinear models can be found in tourism demand forecasting literature as well. The sine wave time series regression approach used to specify the cyclic trend, performed better in forecasting tourism demand to Singapore (Chan 1993). However, Chu (1998) showed that the sine wave model did not outperform an improved ARIMA model with the seasonal and nonseasonal elements. Recently, the

Singular Spectrum Analysis (SSA) emerges in the tourism forecasting research. This method decomposes a time series into a group of independent components prior to forecasting. It has been shown that this method outperformed the linear alternatives in forecasting UK tourism receipts (Beneki et al. 2012) and US tourist arrivals (Hassani et al. 2015).

While various complicated models have been proposed to fit the nonlinear data of tourism demand, some of them are rarely used in tourism demand forecasting studies. These nonlinear models do not emphasize the dependence structures in tourism demand or provide information about the dependence structures. Instead, they tend to impose certain types of dependence structures on tourism demand, and thus restricting the models from having general application.

Dependence between Tourism Flows

Another noteworthy issue in tourism demand forecasting is the dependence between different tourism flows. While it is a common practice by researchers to examine multiple tourism flows to validate the generalization of the forecasting models they propose, they seldom consider whether these flows are associated with each other. For example, Hassani et al. (2015) promoted the use of optimal SSA to forecast tourist arrivals from various source markets to the United States. Gounopoulos et al. (2012) focused on the total tourist arrivals, and arrivals from France, Germany, the United Kingdom, Italy, Netherlands and the United States to Greece. Athanasopoulos and De Silva (2012) used multivariate exponential smoothing method to forecast international tourist arrivals from 11 source countries to Australia and New Zealand. However, all these

studies ignore the dependence structure of tourism flows from various source markets. The implicit assumption in these studies is independence relations (no association) among tourism flows. If tourism flows are correlated, the misspecification of dependence structure will distort the forecasting performance of the tourism demand models.

The number of studies which investigate tourism flow dependence is very limited. By taking into account the correlation between tourist arrivals to Australia and New Zealand, Athanasopoulos and De Silva (2012) showed that it can contribute to forecasting accuracy. Similarly, Du Preez and Witt (2003) examined tourist arrivals from four European countries (France, Germany, Italy and UK) to the Seychelles and they argue that the multivariate ARMA time series models (which incorporate cross-correlations of the arrival series) can generate more accurate forecasts than univariate models when the arrival series are highly correlated with each other. Chan et al. (2005) also reported the presence of linear interdependence effect between tourism flows from the four leading source countries to Australia; however, forecasting is not performed by the authors.

Concluding Remarks

Although substantial efforts have been made to forecast tourism demand/flow, there are still some issues which deserve further discussion. Firstly, linear models are still the commonly used methods in tourism demand forecasting. However, the linear models are very limited and cannot exhibit the Within Dependence structure of tourism demand. Some researchers have analyzed nonlinear serial associations, but the models used have imposed certain types of serial dependence structures on tourism time series, thus restricting the generality of these models. Secondly, analysis of the Between Dependence

of different tourism flows also deserves more attention. It makes sense to examine the interdependence between tourism flows instead of assuming that the different tourism flows are independent of each other. Since both Within Dependence and Between Dependence have not been sufficiently discussed in previous studies, this thesis aims to apply the copula method to examine dependence structure for tourism forecasting.

CHAPTER 4. DEPENDENCE & COPULAS

In the empirical tourism literature, the discussion of dependence structure and its contribution to tourism demand forecasting has received little attention. The copula method proposed is an appropriate technique to address this shortcoming. In this chapter, Sklar Theorem which is the basis of the copula method is presented, as well as several commonly used copula functions. The application and usefulness of this method are also discussed at the end of the chapter.

Sklar's Theorem for Joint Distribution & Copula

Trivedi and Zimmer (2007) defined copulas as parametrically specified joint distribution generated from given marginals; hence, the properties of copulas are analogous to the properties of joint distributions. To obtain a valid joint distribution of the random variables, marginal distributions, which can be any set of disparate distributions, are used as inputs to the corresponding copula function. That is to say, copulas allow arbitrary continuous marginal distributions to be combined for describing the dependence structures. With the various copula functions and arbitrary continuous marginal distributions, the copula technique can well specify different nonlinear dependence structures. Moreover, this technique can also specify the linear structure by using the elliptical copula function and normal marginal distribution. The optimal dependence structures estimated by copulas are more robust and provide more accurate statistical inference of the variables under study.

Sklar's Theorem sets the foundation of copulas, which provides the link between marginal cumulative densities and their joint densities (Sklar 1959). If $H(\cdot)$ is an n -

dimensional joint distribution function for the random variables $Y_1, Y_2, \dots, Y_n$ with continuous marginal distribution functions $F_1, F_2, \dots, F_n$, according to Sklar's Theorem, $H(\cdot)$ can be represented by a unique copula function $C(\cdot)$ as:

$$H(y_1, y_2, \dots, y_n) = C(u_1, u_2, \dots, u_n; \theta) \quad (4.1)$$

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 ; u_i is the cumulative density and $u_i = F_i(y_i)$ for $i = 1, \dots, n$. 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 the simple case, two random variables can be fitted using two-dimensional copulas, in which $n = 2$ and $H(\cdot)$ is given by:

$$H(y_1, y_2) = C(u_1, u_2; \theta), \quad (4.2)$$

where θ is the dependence parameter between u_1 and u_2 . The two-dimensional case is initially used in the illustration of different copula functions. Extension of the two-dimensional case to the high-dimension is subsequently discussed.

Copula Functions for Different Dependence Structures

There is an eclectic mix of copula functions available for the specification of dependence structure. In terms of independence which indicates no association between random variables, the dependence structure can be represented by the Product copula. When the random variables are associated with each other, two families of copulas, namely elliptical and Archimedean families, are commonly used in the empirical copula

literature. The elliptical family is widely applied in statistics and econometrics (Frahm et al. 2003) with the Gaussian and Student-t being the well-known copulas in this class. 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). The popular representatives of this family are the Frank, Clayton and Gumbel copulas, as the three can exhibit different kinds of tail dependence.

The Product copula, or independence copula, is used to describe the independence structure among the variables. For example, if Y_1 and Y_2 are independent from each other, the Product copula function describing this relationship can be written as:

$$C_{\text{Product}}(u_1, u_2) = u_1 \times u_2, \quad (4.3)$$

The Gaussian copula, also called normal copula, uses only pairwise association among the variables to encode dependence. It allows for equal degrees of positive and negative dependence, and represents a symmetric and elliptical dependence structure. This copula takes the form of:

$$C_{\text{Gaussian}}(u_1, u_2) = \Phi_G[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta], \quad (4.4)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution and Φ_G is the standard bivariate normal distribution with dependence parameter θ ($\theta \in [-1, 1]$).

Figure 4.1 shows three simulated two-dimensional scatter plots of the Gaussian copula with $\theta = -0.5, 0$ and 0.5 . When $\theta = 0$, it is the same as the Product copula and there is no systematic pattern. When $\theta \neq 0$, the higher absolute value means the higher degree of association between the two variables. Moreover, the degree of association at the right tail is the same as the degree of association at the left tail.

Figure 4.1: Scatter Plots of Two-dimensional Gaussian Copula

The Student-t copula is also developed for symmetric and elliptical structure. Unlike the Gaussian copula, the Student-t copula has joint fat tails and allows for a higher probability of joint extreme events (Kjersti 2004). The joint distribution represented by this copula is written as:

$$C_{\text{Student-t}}(u_1, u_2) = T_{\theta, \nu}[T_{\nu}^{-1}(u_1), T_{\nu}^{-1}(u_2); \theta, \nu], \quad (4.5)$$

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 ($\theta \in [-1, 1]$). Figure 4.2 shows three simulated two-dimensional scatter plots of the Student-t copula with $\theta = -0.5, 0$ and 0.5 . Again, when $\theta = 0$, it is the same as the Product copula and there is no systematic pattern. Compared to the Gaussian copula, the Student-t copula displays stronger tail dependence in both the left and right tails.

Figure 4.2: Scatter Plots of Two-dimensional Student-t Copula

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

$$C_{\text{Frank}}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right), \quad (4.6)$$

where $\theta \in (-\infty, +\infty) \setminus \{0\}$. The Frank copula is also very popular, because it allows for both positive and negative dependence among the marginal distributions, and the dependence it exhibits is symmetric in both tails. Since the dependence of the Frank copula is strongest in the middle of the distribution, this copula is useful for data with weak tail dependence (Trivedi and Zimmer 2007). Figure 4.3 shows three scatter plots of two-dimensional Frank copula with $\theta = -10, 1$ and 10 . When the absolute value of θ becomes larger, the association between the two variables is stronger. Moreover, as the association parameter approaches 0 (such as when $\theta = 1$), the dependence pattern becomes very unclear.

Figure 4.3: Scatter Plots of Two-dimensional Frank Copula

The Clayton copula is an asymmetric Archimedean copula which demonstrates tail dependence. This copula function is given by:

$$C_{\text{Clayton}}(u_1, u_2) = ((u_1)^{-\theta} + (u_2)^{-\theta} - 1)^{-1/\theta}, \quad (4.7)$$

where $\theta \in (0, +\infty)$. The Clayton copula exhibits strong left tail dependence and relatively weak right tail dependence. Hence, when the association among different events is strongest in the left tail of the joint distribution, it will be better to select the Clayton copula for model construction (Trivedi and Zimmer 2007). Figure 4.4 shows three scatter plots of two-dimensional Clayton copula with parameter $\theta = 1, 5$ and 10 , respectively. It is clear that there is strong left tail dependence between the two variables, and the association between the two variables is stronger when θ becomes larger.

Figure 4.4: Scatter Plots of Two-dimensional Clayton Copula

The Gumbel copula is also an asymmetric Archimedean copula with strong tail dependence. This copula function is given by:

$$C_{\text{Gumbel}}(u_1, u_2) = \exp [-((-\log(u_1))^\theta + (-\log(u_2))^\theta)^{1/\theta}], \quad (4.8)$$

where $\theta \in [1, +\infty)$. The Gumbel copula exhibits strong right tail dependence and relatively weak left tail dependence. Hence, when the association across the distribution is strongest in the positive tail, the Gumbel copula should be considered. Figure 4.5 shows three scatter plots of two-dimensional Gumbel copula with parameter $\theta = 1, 5$ and 10 . When $\theta = 1$, the parameter is at the boundary and the Gumbel copula equals to the Product copula. When $\theta > 1$, it is clear that there is strong right tail dependence between the two variables, and the association between the two variables is stronger when θ becomes larger.

Figure 4.5: Scatter Plots of Two-dimensional Gumbel Copula

Applications and Usefulness of Copula

The copula method is flexible as it provides various functions to describe different dependence structures, whether they are linear or nonlinear, symmetric or asymmetric, and ranging from perfect negative to perfect positive (Vaz 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 a certain distribution, such as the normal distribution in most cases. The benefits of using copulas have been demonstrated in the finance, insurance and economics literature (Da Costa Dias 2004; Pfeifer and Nešlehová 2003; Ferreira 2008; Liew and Wu 2013; Frees and Valdez 2008; Reboredo 2011). Copulas are also used to model the serial dependence structure in a univariate time series (Chen and Fan 2006; Chen et al. 2009; Ibragimov 2009; Beare 2010; Domma et al. 2009).

In the following chapters, the copula method is used to analyze the dependence structures within and between tourist arrivals, and its performance is compared with the AR, SARIMA and structural models in forecasting. The advantages and disadvantages of

these methods are listed in Table 4.1. It can be seen that the copula method is attractive as it allows for an arbitrary distribution of the dependent variable. On the contrary, the AR and SARIMA models assume that the dependent variable is normally distributed. The structural models usually assume the dependent variable follows predetermined (mostly normal) distributions.

Another benefit of using copulas is that this method is flexible in specifying various types of dependence structures. This is particularly useful for tourism demand modelling, as it has been shown that tourism demand data such as tourist arrivals, have different serial dependence patterns. The other three methods account for linear dependence structure. The linear models are computationally simpler than the copula method. However, they may have lower accuracy in estimation and forecasting when the dependence structure of the tourism data is nonlinear.

Seasonality is an issue we should pay attention to when we analyze and forecast tourism demand. Seasonality may be caused by various factors, such as weather, school vacations and holidays, and it usually consists of periodic, repetitive, and generally regular and predictable patterns in the levels of a time series (Zdravevski et al. 2015). The SARIMA model is designed for seasonality in time series, while the other three need special techniques to address data with seasonality. These techniques include the use of seasonally adjusted data, seasonal dummy variables and high order lagged variables.

Table 4.1 Comparison among Copula, AR SARIMA & Structural Models

	Copula	AR	SARIMA	Structural
Distribution Assumption of Dependent Variable	No requirement	Normal distribution	Normal distribution	Normal or predetermined distribution
Dependence Structure	Various types of structures	Linear structure	Linear structure	Linear structure
Simplicity	Complex	Simplest	Simple	Simple
Seasonality	Requiring other adjustment to address data with seasonality	Requiring other adjustment to address data with seasonality	Specially designed to address data with seasonality	Requiring other adjustment to address data with seasonality

The copula technique is still a relatively new approach in empirical tourism research. Related studies apply this technique to analyze the relationship between tourism demand and its determinants. For instance, Tang et al. (2014) examined the relationship between tourism demand to China and exchange rate using copula-GARCH models. This copula-based GARCH approach has also been used to investigate the dependence between GDP and tourism's growth rates (Pérez-Rodríguez et al. 2015). Another application of the copula method in tourism research is to map the co-movement of different tourism flows, such as the dependence of China outbound tourism to any two destinations, be it Thailand, Singapore, South Korea or Japan (Liu and Sriboonchitta 2013; Tang et al. 2013), or dependence between Chinese and Korean tourist arrivals to Thailand (Puarattanaarunkorn and Sriboonchitta 2014a). The discussion can be extended to a three dimensional case by

using vine copulas (Liu et al. 2014; Puarattanaarunkorn and Sriboonchitta 2014b). The research focuses are still on the dependence of multivariate time series and the methods used in these studies have not gone beyond copula-GARCH models.

There is scope for the application of the copula method in empirical tourism research, especially to forecasting. The following chapters include four empirical studies in which the various copula functions introduced in this chapter, are applied to analyze the Within and Between Dependence structures of tourism demand for Singapore, prior to forecasting.

CHAPTER 5. WITHIN DEPENDENCE AND UNIVARIATE TOURISM DEMAND FORECASTING

Prior research has shown that the copula method can be used to model serial dependence (or Within Dependence) of stationary Markov processes (Chen and Fan 2006; Chen et al. 2009; Ibragimov 2009). This allows the use of the copula method in univariate tourism demand forecasting. The current chapter focuses on univariate tourism time series and applies the copula method to specify Within Dependence in two empirical studies. Monthly data used in these studies require the extraction of the seasonality which refers to the systematic intra-year movement in a time series (Hylleberg 1992). Seasonal adjustment is one way to remove the seasonality so that the underlying trend patterns of the series can be examined prior to forecasting (Newbold and Bos 1994). Deseasonalized tourism time series is used in Study 1 in which the simple two-dimensional copulas are employed to specify the Within Dependence structure between tourism demand and its first lag. The forecasting results of the copula models will be compared with AR(1) which corresponds to linear Within Dependence structure. The sophisticated analysis of the Within Dependence in the high-dimensional case is undertaken in Study 2, and seasonal differencing is used to address seasonality. Specifically, I employ a four-dimensional Gaussian copula model to forecast tourism demand to Singapore and compare the results with those of the ARMA family.

Two-Dimensional Copula Model for Within Dependence (Study 1)

Methodology

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

$$H(y_t, y_{t-1}, \dots, y_{t-k}) = C\{F_m(y_t), F_m(y_{t-1}), \dots, F_m(y_{t-k}); \theta\} \quad (5.1)$$

Equation (5.1) can be rewritten as:

$$H(y_t, y_{t-1}, \dots, y_{t-k}) = C(u_t, u_{t-1}, \dots, u_{t-k}; \theta) \quad (5.2)$$

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

I will analyze the serial dependence structure between Y_t and Y_{t-1} and construct two-dimensional copula models for forecasting purpose. Tourist arrival is the continuous random variable used as the proxy of tourism demand. The Within Dependence of tourism demand using two-dimensional copulas is given by:

$$H(y_t, y_{t-1}) = C(u_t, u_{t-1}; \theta) \quad (5.3)$$

where θ is the dependence parameter between u_t and u_{t-1} .

A likelihood based estimation method is used to estimate the dependence parameter θ in Equation (5.3). Let $c(\cdot)$ denote the probability density function corresponding to the copula distribution function $C(\cdot)$. The log-likelihood function is:

$$L = \sum_{t=1}^T \ln c(u_t, u_{t-1}; \theta) \quad (5.4)$$

where

$$c(.) = \frac{\partial^2 C}{\partial u_t \partial u_{t-1}} \quad (5.5)$$

The specific probability density functions of the five copulas, including the Gaussian, Student-t, Frank, Clayton and Gumbel copulas, can be seen in Appendix B. After estimating the dependence parameter θ , we can use θ and u_{t-1} to calculate the mean, or the conditional expectation of u_t as the forecasted value $\hat{u}_t$. Since it is difficult to calculate the mean, the binary search algorithm is applied to search the median of u_t as the approximation of the mean. The details of the binary search algorithm are shown in Appendix C.

In an AR model, the current value of tourism demand Y_t depends only upon its historical values. An AR(p) model, also known as autoregressive model of order p, is given by:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (5.6)$$

Equation (5.6) shows that the relationship between tourism demand Y_t and its lagged variables is linear. In other words, the AR(p) model exhibits a linear Within Dependence structure. We can use the ordinary least square (OLS) method to fit the AR(p) model and estimate the coefficient ϕ_i , in order to generate the predicted values of Y_t .

The two-dimensional copula method and the AR(1) model both address the Within Dependence between tourism demand and its first lag. However, there are numerous copula functions for different dependence structures, while the AR(1) model only represents the linear structure. Since the AR(1) model is used as a benchmark for comparing its forecast performance with that of copula, we expect the copula method to outperform the AR(1) in forecasting if the dependence structure is nonlinear.

Data Description

Initially, the dataset comprised total and disaggregated monthly arrivals from numerous source countries to Singapore (see Chapter 1 for details) from January 1995 to December 2013. The monthly data obtained from the Singapore Tourism Board (STB) are seasonally adjusted with Census X-13 developed by US Census Bureau. Only 12 arrivals series are selected because they are first-order Markov processes (test for Markov property will be shown in the next section). They include total tourist arrivals (TO) to Singapore and arrivals from Australia, Bangladesh, Canada, German, Japan, Malaysia, Myanmar, New Zealand, South Korea, Switzerland and the United States. These arrivals series will be analyzed for serial dependence in the following discussion.

The data for each arrival series are split into the estimation period (from 1995M01 to 2010M12) and the validation period (from 2011M01 to 2013M12). The estimation sample is used to specify the copula and the benchmark models. Subsequently, the specified models are used to generate 1-, 6-, 12-, 18- and 24-month-ahead forecasts. The forecasts are compared using the expanding windows (or recursive) forecasting technique with the data in the validation sample.

The five copulas introduced in Chapter 4 (in Equations (4.4) to (4.8)), namely the Gaussian, Student-t, Frank, Clayton and Gumbel copulas, are considered in this study. I intend to identify the optimal model from these five copulas to specify the serial dependence structure of each arrivals series. Subsequently, the forecasting results of the selected copula models and that of the AR(1) model will be compared.

Preliminary Test

Before carrying out the empirical analysis of dependence structure, it is necessary to examine if the individual arrival series: 1) is stationary; and 2) follows a specified distribution. Additionally, the Markov property of the series is tested in this section.

I 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 5.1. It can be seen that arrivals from Canada, Germany, South Korea and the United States yield p-value of ADF statistics smaller than 0.05, which indicate that these series are stationary in levels. The other series have unit roots and taking first difference renders each series stationary. Therefore, I analyze actual arrivals (untransformed series) from Canada, Germany, South Korea and the United States, and arrivals' growth for the other 8 series (in first differenced).

Table 5.1: Unit Root Tests for Tourist Arrivals to Singapore

Tourist Arrivals Series	ADF Statistic	
	Level	1 st Difference
Total Arrivals	-1.253	-13.681***
By Source Countries:		
Australia	-0.996	-15.312***
Bangladesh	-0.271	-21.139***
Canada	-3.917**	-----
Germany	-3.947**	-----
Japan	-1.949	-16.549***
Malaysia	-0.269	-19.440***
Myanmar	0.499	-19.957***
New Zealand	-2.490	-18.835***
South Korea	-3.098*	-----
Switzerland	-1.749	-12.761***
USA	-3.685**	-----

* indicates results are statistically significant at 0.05 level;

** indicates results are statistically significant at 0.01 level;

*** indicates results are statistically significant at 0.001 level.

Each stationary arrival series has its own distribution, which is also called marginal distribution in order to distinguish it from the joint distribution. Modeling serial dependence relies on the marginal distribution of arrival series. The chi-square test is used to determine if the tourist arrival series follow the normal distribution or alternative forms, namely Cauchy or Logistic distribution. The null hypotheses for the chi-square test are: $Y \sim \text{Normal}$, $Y \sim \text{Cauchy}$ or $Y \sim \text{Logistic}$ (in which ‘ $\sim$ ’ denotes ‘is distributed approximately as’). If the test statistic is greater than the critical value, the null hypothesis is rejected, implying that the arrival series does not follow the corresponding distribution. Table 5.2

presents the chi-square test results. It can be seen that the test statistics for normal distribution are not always smaller than the critical values at the 0.01 significance level, such as growth in total arrivals and arrivals from Malaysia and Myanmar; and actual arrivals from South Korea. As for the other series, even though the normal distribution can pass the chi-square test, the Cauchy or Logistic distribution possesses a lower chi-square statistic than the normal. Using the smallest chi-square statistic as the criterion for distribution selection, the Cauchy distribution is selected for total arrivals, arrivals from Australia, Myanmar, Malaysia, New Zealand and South Korea; and the Logistic distribution for Bangladesh, Canada, Germany, Japan, Switzerland and the United States, 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). The latter can be tested by conducting diagnostic checking on the residuals for serial correlation using the Breusch–Godfrey LM test. The null hypothesis of the LM test is that the residuals are not correlated. Table 5.2 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 first-order Markov processes.

Table 5.2: Chi-Square Tests for Distributions and LM Test for Serial Correlation

Chi-Square Test Statistics					Serial Correlation Test	
Tourist Arrivals Series	Specified Distribution			Critical Value $\alpha=0.01$	LM(12) ^a	P-value
	Normal	Cauchy	Logistic			
<i>Arrivals: Growth</i>				21.670		
Total Arrivals	31.742	9.824	20.994		13.479	0.335
Australia	19.382	9.962	11.777		14.595	0.264
Bangladesh	13.954	24.185	7.619		16.859	0.155
Japan	18.507	11.557	6.434		18.712	0.095
Malaysia	27.183	14.202	15.779		16.176	0.183
Myanmar	52.309	7.549	35.489		16.184	0.183
New Zealand	17.317	4.373	7.766		11.985	0.447
Switzerland	13.887	22.967	9.549		19.498	0.077
<i>Arrivals: Actual</i>						
Canada	8.917	18.587	6.718		15.223	0.230
Germany	13.682	15.560	9.148		18.001	0.116
South Korea	23.646	4.183	14.527		17.090	0.146
USA	19.959	30.778	13.982		14.699	0.258

Note: ^a Include 12 lags for monthly data.

Analysis of Within Dependence Structure

To show the existence of Within Dependence for the arrivals series, we can simply look at the serial association coefficients. The three commonly used association coefficients are Pearson's coefficient (r), Spearman's rho coefficient (ρ), and Kendall's tau coefficient (τ). The Pearson's coefficient measures the strength of the linear relationship and is very powerful for a bivariate normal distribution, while the Spearman and

Kendall are nonparametric rank coefficients which measure either linear correlation or nonlinear monotonic associations (Nelsen 2007; Hauke and Kossowski 2011).

Table 5.3 presents the three association coefficients of tourism demand Y_t and its first lag Y_{t-1} for different series. It can be seen that all the association coefficients are significant for most series, which is the evidence of the existence of Within Dependence for the arrivals series. The two exceptions are total arrivals and arrivals from Australia. For these two series, the Pearson's coefficients are not significant, suggesting that there is no linear correlation between Y_t and Y_{t-1} . However, the Spearman's rho and Kendall's tau for these series are all significant, suggesting that their serial dependences are likely nonlinear associations. We can also conclude from Table 5.3 that the serial dependences for the arrivals' growths are negative, while those of the actual arrivals series are positive.

Table 5.3: Associations between Tourist Arrivals and Their First Lags

Tourist Arrivals Series	Pearson Correlation Coefficient	Kendall's Tau	Spearman's Rho
<i>Arrivals: Growth</i>			
Total	0.003	-0.201**	-0.287**
Australia	-0.114	-0.177**	-0.249**
Bangladesh	-0.406**	-0.266**	-0.365**
Japan	-0.186*	-0.181**	-0.251**
Malaysia	-0.339**	-0.292**	-0.405**
Myanmar	-0.366**	-0.193**	-0.284**
New Zealand	-0.309**	-0.255**	-0.367**
Switzerland	-0.255**	-0.204**	-0.292**
<i>Arrivals: Actual</i>			
Canada	0.853**	0.701**	0.880**
Germany	0.840**	0.648**	0.831**
South Korea	0.855**	0.660**	0.840**
USA	0.886**	0.669**	0.849**
* indicates results are statistically significant at 0.05 level; ** indicates results are statistically significant at 0.01 level.			

The association coefficients only reveal to what extent Y_t and Y_{t-1} are associated with each other. 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. The scatter plots in Figure 5.1 provide evidence that tourist arrivals exhibit various serial dependence structures in different situations. The series which correspond to arrivals' growth seem to show symmetric dependence structures with fat tails. In contrast, the actual arrivals series, especially from the United States, exhibit relatively strong right tail dependence, as their right tails seem heavier than the left.

Figure 5.1 reveals that there is little support for using only one type of serial dependence specification, namely the linear structure, for tourism demand modeling and forecasting. The copula method is thus more appealing as it flexibly accounts for different serial dependence structure.

Figure 5.1: Scatter Plots of Cumulative Densities for Tourist Arrivals to Singapore

Copula Model Specification

The scatter plots show that different serial dependence structures exist for different tourist arrivals series. To determine the appropriate type of structure for each series and the copula to use, a certain selection criterion is needed. Genest et al. (2009) summarized several goodness-of-fit testing for copula models, and the $S_n^{(C)}$ statistic recommended by them as the selection criterion for the copulas, is 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. I use 0.05 significant level in this study and the goodness-of-fit test results are reported in Table 5.4. 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 copula is not a good choice and should not be considered for any series except Myanmar. The Gumbel copula is another reasonable choice for the actual arrivals from Canada, Germany, South Korea and the United States, but not the other series.

Table 5.4: Goodness-of-fit Test for Copula Selection

Tourist Arrivals Series	P-value of S_n^C Statistics				
	Gaussian	Student-t	Frank	Clayton	Gumbel
<i>Arrivals: Growth</i>					
Total	0.390	0.427	0.504	0.000	0.000
Australia	0.367	0.437	0.402	0.008	0.003
Bangladesh	0.290	0.382	0.374	0.000	0.000
Japan	0.397	0.478	0.423	0.000	0.002
Malaysia	0.394	0.457	0.522	0.000	0.000
Myanmar	0.542	0.572	0.589	0.265	0.002
New Zealand	0.499	0.558	0.566	0.000	0.000
Switzerland	0.445	0.468	0.499	0.001	0.001
<i>Arrivals: Actual</i>					
Canada	0.319	0.383	0.640	0.000	0.400
Germany	0.514	0.553	0.746	0.013	0.539
South Korea	0.447	0.506	0.735	0.032	0.517
USA	0.355	0.438	0.641	0.001	0.444

To select the optimal copula for each series, I compare the log-likelihood (LL) values of the five models. The LL values of the copula models are reported in Table 5.5 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 the arrivals' growth series. The Student-t copula is also a better choice for Canada and South Korea, while the Gumbel copula is optimal for Germany and the United States. These selections are consistent with the patterns shown in Figure 5.1. The selected copulas will be used to generate demand forecasts, which will be compared with the forecasting results of the benchmark AR(1) model.

Table 5.5: Log-likelihood Values Comparison for Optimal Copula

Tourist Arrivals Series	Log-likelihood Value				
	Gaussian	Student-t	Frank	Clayton	Gumbel
<i>Arrivals: Growth</i>					
Total	7.436	19.670	8.692	-----	-----
Australia	6.450	19.870	6.652	-----	-----
Bangladesh	16.000	25.100	16.180	-----	-----
Japan	4.328	16.380	7.001	-----	-----
Malaysia	17.690	29.450	16.400	-----	-----
Myanmar	9.068	14.290	7.902	8.779	-----
New Zealand	15.630	25.380	14.420	-----	-----
Switzerland	7.459	9.986	8.808	-----	-----
<i>Arrivals: Actual</i>					
Canada	133.600	139.600	139.500	-----	127.300
Germany	121.600	122.100	119.200	-----	131.400
South Korea	125.100	139.800	109.500	-----	123.000
USA	134.500	137.900	126.800	-----	145.300

Table 5.6 reports the association parameters of the optimal copulas selected. The parameters for most series are significant at 0.001 level, albeit Switzerland at 0.05 significance level. This provides further information about the serial dependence of 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 growth is stronger in the tails than the association in the centre. It can also be inferred 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.

Table 5.6: Association Parameters of Optimal Copula

Tourist Arrivals Series	Optimal Copula	Association Parameter
<i>Arrivals: Growth</i>		
Total	Student-t	-0.341***
Australia	Student-t	-0.301***
Bangladesh	Student-t	-0.455***
Japan	Student-t	-0.300***
Malaysia	Student-t	-0.437***
Myanmar	Student-t	-0.304***
New Zealand	Student-t	-0.301***
Switzerland	Student-t	-0.335*
<i>Arrivals: Actual</i>		
Canada	Student-t	0.882***
Germany	Gumbel	2.772***
South Korea	Student-t	0.848***
USA	Gumbel	3.422***

*** indicates results are statistically significant at 0.001 level;

* indicates results are statistically significant at 0.05 level

For the actual arrivals series, the current tourist arrivals have a positive association with future arrivals. For Canada and South Korea, the dependence structure is described by the Student-t copula. Therefore, the positive association between current and future numbers of arrivals from Canada or South Korea is stronger in both the left and right tails than in the centre. However, a strong positive association can only be expected in the right tail for Germany and the United States, as their serial dependence structures are modelled by the Gumbel copula.

To examine the contribution of well specified serial dependence structures to tourism demand forecasting, I compare the forecasting results of copula models with those of AR(1) model. The AR(1) model is an ideal benchmark because: 1) the LM test indicates that the residuals of AR(1) model are not serially correlated for each arrivals series; 2) the AR(1) model is more parsimonious compared with other ARMA models for each series; 3) the AR(1) model is more comparable with the two-dimensional copula models ($k=1$) than the other univariate time series models.

Forecasting Results

Figures 5.2 and 5.3 provide a graphical illustration of the forecasting results and present the comparison between the actual total arrivals growth and the forecasts of AR(1) and copula models at Horizons $i=1$ and $i=24$. It can be seen that the point predictions of both the copula and AR(1) models are close to the values of actual total arrivals for both short and long term forecasts. However, the advantage of using copula for forecasting is not apparent in this figure and a detailed comparison of forecast errors will be conducted.

Figure 5.2 Actual Total Arrivals and Forecasts of AR(1) and Copula Models (Horizon $i=1$)

Figure 5.3: Actual Total Arrivals and Forecasts of AR(1) and Copula Models (Horizon $i=24$)

The common measures of forecast errors, namely the mean absolute percentage error (MAPE) and root mean square error (RMSE), are used to evaluate tourism forecast

performance. Tables 5.7 and 5.8 display the out-of-sample MAPE and RMSE results for the 12 arrivals series to Singapore, respectively.

According to Vergori (2012), MAPE values lower than 50% are acceptable and values lower than 20% correspond to good forecasts. Table 5.7 shows that all MAPEs are lower than 50% and most of them are under 20%, which indicates that both the copula method and AR(1) model perform well in forecasting tourism demand to Singapore. Moreover, the copula models perform marginally better than AR(1) model for arrivals' growth series with the exceptions of Japan and Myanmar. For actual arrivals series, the selected copula models generate smaller MAPEs than the AR(1) model at almost all forecasting horizons.

Similar results can be seen for most of the series in Table 5.8 which reports the RMSEs. In the arrivals' growth series, the selected copulas have outperformed the AR(1) model with the exception of Japan and Malaysia. In terms of the actual arrivals series, it is obvious that the forecasting performance of copulas is better than that of the AR(1) model.

Given the results of both MAPE and RMSE, it can be concluded that in general, the well-specified copula-based serial dependence structure contributes to greater forecast accuracy. The AR(1) model is a better forecasting approach than copulas only for tourism demand from Japan to Singapore. For Malaysia and Myanmar, both the copula method and AR(1) model are reasonable options. It is worth noting that the copula only performs marginally better than the AR(1) model in forecasting arrivals growths though it is a more advanced method. The possible reason could be that the nonlinear serial dependence structures of arrivals growths are close to the linear structure. Embrechts et al. (2001) pointed out that the linear dependence indicates the elliptical joint distribution of the

random variables. In this study, the dependence analyses show that the serial dependence structures of arrivals growths are elliptical (described by the Student-t copula in particular). It is possible that the serial dependence of arrivals growths can be approximated by the linear structure. Therefore, the benefits of using copula are not very obvious for the growths series.

Different sample splits are used as robustness checks for the forecasting results. Specifically, two validation periods (Period 1: from 2011M01 to 2013M6; Period 2: from 2011M07 to 2013M12) are considered and the forecasting results are shown in Tables D1 to D4 in Appendix D. In general, the robustness checks support the results of Tables 5.7 and 5.8, with the only exception of Myanmar.

Table 5.7: MAPEs of Copula and AR(1) Models over Five Time Horizons

Tourist Arrivals		Forecast Horizon					
Series	Model	<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24	
<i>Arrivals: Growth</i>	Total	Student-t	2.337	2.752	2.664	2.901	2.582
		AR(1)	2.639	2.797	2.719	2.940	2.596
	Australia	Student-t	3.509	3.839	4.490	4.598	5.168
		AR(1)	3.678	3.844	4.492	4.579	5.142
	Bangladesh	Student-t	4.970	5.528	5.176	4.654	5.354
		AR(1)	4.684	5.525	5.177	4.688	5.433
	Japan	Student-t	4.846	4.399	4.326	3.533	3.222
		AR(1)	4.692	4.405	4.328	3.490	3.193
	Malaysia	Student-t	5.281	6.391	6.538	7.909	7.372
		AR(1)	5.149	6.434	6.560	7.931	7.376
	Myanmar	Student-t	5.065	5.772	6.336	5.899	6.193
		AR(1)	4.849	5.769	6.328	5.894	6.194
	New Zealand	Student-t	6.457	6.425	4.863	5.229	5.084
		AR(1)	6.491	6.480	4.909	5.273	5.130
	Switzerland	Student-t	4.162	4.577	5.117	5.229	6.653
		AR(1)	4.197	4.583	5.108	5.241	6.655
<i>Arrivals: Actual</i>	Canada	Student-t	4.939	10.529	16.711	20.295	22.796
		AR(1)	5.191	12.392	18.016	21.023	23.158
	Germany	Gumbel	5.901	12.767	20.126	24.518	32.046
		AR(1)	6.463	16.071	24.004	27.420	31.115
	South Korea	Student-t	5.152	13.028	18.321	21.098	23.134
		AR(1)	5.335	14.594	20.365	23.325	25.450
	USA	Gumbel	3.267	9.624	16.929	21.568	24.483
		AR(1)	3.662	11.724	19.053	22.956	25.298

Note: MAPE is the mean absolute percentage error.

Table 5.8: RMSEs of Copula and AR(1) Models over Five Time Horizons

Tourist Arrivals		Forecast Horizon					
Series	Model	<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24	
Arrivals: Growth	Total	Student-t	38264.291	42588.455	42695.731	46917.574	44556.677
		AR(1)	40982.676	42940.162	43120.679	47127.447	44265.168
	Australia	Student-t	3984.277	4384.456	4844.245	4951.618	5363.873
		AR(1)	4227.993	4400.507	4865.859	4977.037	5360.563
	Bangladesh	Student-t	592.936	688.901	674.388	606.058	692.296
		AR(1)	581.611	691.144	676.307	608.393	694.102
	Japan	Student-t	3827.343	3991.022	4148.263	3425.585	3653.982
		AR(1)	3848.939	3965.760	4129.604	3403.477	3642.868
	Malaysia	Student-t	7485.903	9246.508	9617.467	10948.043	10893.580
		AR(1)	7432.003	9268.789	9622.085	10946.768	10873.124
	Myanmar	Student-t	524.229	593.447	627.941	578.590	597.630
		AR(1)	502.574	593.617	627.945	578.661	598.271
	New Zealand	Student-t	734.660	696.105	648.877	836.425	924.980
		AR(1)	734.743	696.482	647.836	836.134	929.018
	Switzerland	Student-t	397.233	460.850	505.429	533.870	632.844
		AR(1)	403.596	462.460	506.342	535.620	634.398
Arrivals: Actual	Canada	Student-t	513.938	881.995	1305.526	1567.469	1776.864
		AR(1)	523.902	1000.781	1403.737	1621.937	1804.492
	Germany	Gumbel	1749.518	3357.541	4846.775	5430.730	6923.418
		AR(1)	1799.479	3944.492	5515.095	5822.365	6706.744
	South Korea	Student-t	2624.643	5405.243	7443.901	8553.568	9545.515
	USA	AR(1)	2705.685	5975.181	8199.545	9386.634	10418.304
		Gumbel	1761.529	4249.515	7104.113	8873.309	10158.790
	AR(1)	1967.688	5043.800	7901.891	9429.093	10490.335	

Note: RMSE is the root mean square error.

High-Dimensional Copula Model for Within Dependence (Study 2)

Study 1 presents the simple two-dimensional case using seasonally adjusted tourist arrivals data. In Study 2, the sophisticated application of copula method will be discussed.

According to Ibragimov (2009), the copula method can be used to analyze high-dimensional univariate time series if the series are stationary high-order Markov processes. Study 2 uses the high-dimensional Gaussian copula to specify Within Dependence Structure of tourism demand. The forecasting results of the copula will be compared with AR(p) and SARIMA, which are linear models and they belong to ARMA family.

Methodology

High-dimensional Gaussian Copula

Although there are many families of copula models available, choosing one which allows for variations in the association of 3 or more variables is challenging. Parsa and Klugman (2011) argue that the Gaussian copula is appropriate. As mentioned in Chapter 3, the Gaussian copula model is a generalization of the conventional linear models. Specifically, the Gaussian copula function belongs to the elliptical family and can be combined with normal marginals to represent the linear dependence structure; while Gaussian copula function with other marginals represents certain nonlinear structure. For univariate time series, the high-dimensional Gaussian copula takes the form of:

$$C_{\text{Gaussian}}(u_t, u_{t-1}, \dots, u_{t-k}) = \Phi_G[\Phi^{-1}(u_t), \Phi^{-1}(u_{t-1}), \dots, \Phi^{-1}(u_{t-k}); \mathbf{R}_k], \quad (5.7)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution and Φ_G is the standard multivariate normal distribution with the correlation matrix $\mathbf{R}$ of $\mathbf{u} = (u_t, u_{t-1}, u_{t-2}, \dots, u_{t-k})$. Moreover, the correlation matrix can be partitioned as $\mathbf{R} = \begin{bmatrix} 1 & \mathbf{r}' \\ \mathbf{r} & \mathbf{R}_k \end{bmatrix}$ with $\mathbf{R}_k$ being the correlation matrix of $\mathbf{u}_k = (u_{t-1}, u_{t-2}, \dots, u_{t-k})$ and $\mathbf{r}$ being the $k \times 1$ vector of correlations between $\mathbf{u}_k$ and u_t .

A likelihood based method is used to estimate the correlation matrix $\mathbf{R}$. The Gaussian copula function $C_{\text{Gaussian}}(\cdot)$ has the probability density function $c(\cdot)$ given by (Clemen and Reilly 1999; Nelsen 2007):

$$c(y_t, y_{t-1}, \dots, y_{t-k}) = f_m(y_t) f_m(y_{t-1}) \dots f_m(y_{t-k}) \exp \left\{ -\frac{\mathbf{w}'(\mathbf{R}^{-1} - \mathbf{I})\mathbf{w}}{2} \right\} \times |\mathbf{R}|^{-0.5}, \quad (5.8)$$

where $f_m(y_{t-i})$ is the probability density corresponding to $F_m(y_{t-i})$ and $\mathbf{w}$ is a vector with i th element $w_{t-i} = \Phi^{-1}(u_{t-i})$ for $i = 0, 1, \dots, k$; $\mathbf{I}$ is the identity matrix. The log-likelihood function can then be written as:

$$\begin{aligned} L &= \sum_{t=1}^T \ln c(y_t, y_{t-1}, \dots, y_{t-k}) \\ &= \sum_{t=1}^T \left\{ -\frac{\mathbf{w}'(\mathbf{R}^{-1} - \mathbf{I})\mathbf{w}}{2} \right\} \times |\mathbf{R}|^{-0.5} + \sum_{t=1}^T \sum_{i=0}^k f_m(y_{t-i}). \end{aligned} \quad (5.9)$$

The correlation matrix $\mathbf{R}$ can be estimated by maximizing Equation (5.9). After that, the predicted value $\hat{y}_t$ can be generated by (Parsa and Klugman 2011):

$$\hat{y}_t = \int_a^b f_t(y_t) \exp \left\{ -0.5 \left[\frac{(\Phi^{-1}(u_t) - \mathbf{r}'\mathbf{R}_k^{-1}\mathbf{w}^*)^2}{1 - \mathbf{r}'\mathbf{R}_k^{-1}\mathbf{r}} - \left(\Phi^{-1}(u_t) \right)^2 \right] \right\} \times (1 - \mathbf{r}'\mathbf{R}_k^{-1}\mathbf{r})^{-\frac{1}{2}} dy_t, \quad (5.10)$$

where $\mathbf{w}^* = (w_{t-1}, w_{t-2}, \dots, w_{t-k})$.

Seasonal ARIMA (SARIMA) Model

The autoregressive moving average (ARMA) model popularized by George E. P. Box and Gwilym M. Jenkins (Box and Jenkins 1970) is commonly used to predict the future value of a stationary time series. Box and Jenkins (1970) further developed the ARMA model into the autoregressive integrated moving average (ARIMA) model to address non-stationarity in time series data.

When a time series displays a seasonal pattern, the seasonal ARIMA (SARIMA) process should be taken into account. As a more complex form of general ARIMA models, the SARIMA model not only contains AR term, MA term and non-seasonal integrated part, but also incorporates seasonal AR (SAR) term, seasonal MA (SMA) term and seasonal difference terms. The AR and MA terms account for the serial correlations at low seasonal lags, while the seasonal difference terms reduce the non-stationarity caused by the seasonal trend. The general SARIMA model is specified as $(p, d, q)(P, D, Q)_S$. The p (q) and P (Q) represent the orders of the non-seasonal and seasonal AR (MA) terms, respectively. The d and D are the degrees of non-seasonal and seasonal integrated (I) parts, respectively, and the S is the length of seasonal period. The SARIMA model is formulated as:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D y_t = a + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (5.11)$$

where y_t is the time series data at time t ; a is the constant term; B and B^S are the non-seasonal and seasonal backshift operators, respectively; $\phi_p(B)$ is the ordinary autoregressive components and $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$; $\theta_q(B)$ is the ordinary moving average components and $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$; $\Phi_P(B^S)$ is the seasonal autoregressive components and $\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{S^2} - \dots - \Phi_P B^{S^P}$; $\Theta_Q(B^S)$ is the seasonal moving average components and $\Theta_Q(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{S^2} + \dots + \Theta_Q B^{S^Q}$; and ε_t is the error term at time t with the normal distribution $N(0, \delta^2)$.

Data Description

Like Study 1, this study uses monthly tourist arrivals data from January 1995 to December 2013. Four sample series are examined as they are high-order Markov processes (their Markov property will be shown later). The four series comprise total tourist arrivals to Singapore, as well as arrivals from its major short, medium and long haul source markets, namely Indonesia, Australia and the United States, respectively.

Again, tourist arrivals from January 1995 to December 2010 are used as the sample estimation period, from which I conduct the data analysis, and specify the Gaussian copula and benchmark models. These models are used to generate 1-, 6-, 12-, 18- and 24-month-ahead forecasts, which can then be compared using the retained/available data in 2011 to 2013 for validation.

Preliminary Test

The Augmented Dickey-Fuller (ADF) tests are conducted to examine whether the monthly tourist arrivals series are stationary. Based on the ADF statistics reported in Table 5.9, all the untransformed series have unit roots but the seasonally differenced series are stationary. Therefore, the transformed series are used for data analysis in which the probability distributions are stable over time.

Table 5.9: Unit Root Tests for Tourist Arrivals to Singapore

Tourist Arrivals Series	ADF Statistic	
	Level	Seasonal Difference
Total	0.032	-4.294***
Arrivals by Source Market		
Indonesia	0.631	-5.027***
Australia	0.433	-5.466***
USA	-1.416	-5.480***

* indicates results are statistically significant based on ADF Test at 0.05 level;

** indicates results are statistically significant at 0.01 level;

*** indicates results are statistically significant at 0.001 level.

To substantiate the choice of Gaussian copula in the study, I also want to identify the distribution of the stationary tourist arrivals series under investigation. Similar to Study 1, I apply the chi-square goodness-of-fit test in this study to determine the distributions of tourist arrivals Y , namely: $Y \sim \text{Normal}$, $Y \sim \text{Logistic}$ or $Y \sim \text{Cauchy}$. According to Table 5.10, the test statistics for normal distribution are all higher than the critical value. The total tourist arrivals and arrivals from Indonesia and Australia seem to be associated with the Cauchy distribution, while the hypothesis that the US arrivals follow the Logistic distribution cannot be rejected. I depart from the conventional approach of assuming the normal distribution, as it is evident that the Logistic and Cauchy distributions are better fits for tourist arrivals from the US and the other markets, respectively. As defined by the Gaussian copula in Equation (5.7), I take into account the univariate marginals and joint distribution for the various observed arrivals series. After the parameter estimation stage, the forecasts for total arrivals and arrivals from each source market are generated using the conditional expectation of y_t in Equation (5.10).

Table 5.10: Chi-Square Test for Distribution of Tourism Demand Series

Chi-Square Test Statistics				Critical Value $\alpha=0.01$
Series	Specified Distribution			
	Normal	Cauchy	Logistic	
Total	32.713	8.405	23.812	20.090
Indonesia	33.232	4.117	21.261	
Australia	49.533	17.837	41.133	
USA	24.450	22.238	18.730	

If a series follows an AR(p) specification with independent residuals, the series is a p-order Markov process (Deco and Schürmann 2012). Again, I examine the Markov property of the tourist arrivals series involved by conducting the Breusch–Godfrey LM test on the residuals of AR(p) model for serial correlation. Table 5.11 presents the results of the LM test. There is no serial correlation in the residuals of the AR(13) models with the 1st, 12th and 13th lags (Y_{t-1} , Y_{t-12} and Y_{t-13}) for the total arrivals and arrivals from Australia and the United States, since the p-values of LM statistics are greater than 0.05. For arrivals from Indonesia, the AR(11) model with the 1st, 2nd and 11th lags (Y_{t-1} , Y_{t-2} and Y_{t-11}), has uncorrelated residuals. Hence, it can be concluded that the arrivals series from Indonesia is eleventh-order Markov process, while the other series are thirteenth-order Markov processes.

Table 5.11: Breusch–Godfrey LM Test for Serial Correlation

Series	AR(p) Model	Lags Included ^a	LM(12) ^b	P-value
Total	AR(13)	1 st , 12 th , 13 th	19.009	0.088
Indonesia	AR(11)	1 st , 2 nd , 11 th	10.451	0.576
Australia	AR(13)	1 st , 12 th , 13 th	20.401	0.060
USA	AR(13)	1 st , 12 th , 13 th	16.924	0.153

Note: ^a Lags included in the AR(p) model; ^b include 12 lags for monthly data.

Forecasting Results

As mentioned earlier, the LM test indicates that the arrivals series from Indonesia is eleventh-order Markov process in which Y_t is associated with Y_{t-1} , Y_{t-2} and Y_{t-11} , and that the other series are thirteenth-order Markov process in which Y_t is associated with Y_{t-1} , Y_{t-12} and Y_{t-13} . The Gaussian copula model is constructed accordingly with corresponding lags associated with Y_t for tourism demand forecasting. To illustrate the advantage of using the Gaussian copula for forecasting, I compare the forecasting results of Gaussian copula with AR(11) and AR(13) models.

Table 5.12 presents the optimal SARIMA models that describe the total tourist arrivals as well as arrivals from Australia, Indonesia and the United States to Singapore. Both the Akaike information criterion (AIC) and Schwarz Bayesian criterion (SBIC) are employed in the determination of the parameters. SARIMA models with the smallest AIC and SBC are thus selected.

Table 5.12: SARIMA Model Parameters for Arrivals Series

Series	SARIMA (p,d,q)×(P,D,Q) _s	Series	SARIMA (p,d,q)×(P,D,Q) _s
Total	(1,0,1)×(1,1,1) ₁₂	Indonesia	(1,0,1)×(1,1,1) ₁₂
Australia	(1,0,1)×(0,1,1) ₁₂	USA	(1,0,1)×(0,1,1) ₁₂

p and P are the number of AR and SAR terms, respectively;

q and Q are the number of MA and SMA terms, respectively;

d and D are the difference and seasonal difference orders, respectively;

s is the seasonal period

Figures 5.4 and 5.5 provide the comparison between the actual total arrivals and the forecasts of AR(13) and copula models at Horizons $i=1$ and $i=24$. It can be seen that the point predictions of both the copula and AR(13) models are close to the values of actual total arrivals for both short and long term forecasts. The advantage of using copula for forecasting is not apparent in Figure 5.4. However, most of the point predictions of the copula model are close to the actual total arrivals values than those of the AR(13) model in Figure 5.5.

Figure 5.4: Actual Total Arrivals and Forecasts of AR(13) and Copula Models (Horizon $i=1$)

Figure 5.5: Actual Total Arrivals and Forecasts of AR(13) and Copula Models (Horizon $i=24$)

Figure 5.6 reports the MAPE and RMSE results for the out-of-sample forecasts of total tourist arrivals. In general, the average MAPEs and RMSEs show that the Gaussian copula model has outperformed the rest in forecasting, while the performance of the SARIMA model is the worst. If we evaluate the performance of all the forecasting models at each horizon, the MAPE values of the Gaussian copula remain below 5% for the horizons $i = 6$ -, 12 -, 18 - and 24 -month ahead, while the other models have higher MAPEs. The SARIMA's MAPE even exceeds 10% for the horizons $i = 12$ -, 18 - and 24 -month ahead. For horizon $i=1$, the SARIMA model has performed marginally better than the other models. The RMSE results lead to a similar conclusion. Though the SARIMA model performs slightly better than the other two models for horizon $i = 1$, the Gaussian copula model yields the smallest RMSE for horizons $i = 6, 12, 18$ and 24 .

Figure 5.6: Histograms of MAPE & RMSE for Total Tourist Arrivals by Forecast Horizons

The out-of-sample MAPE and RMSE results for arrivals from Australia, Indonesia and the United States are also compared. Tables 5.13 and 5.14 shows that the Gaussian copula model yields smaller MAPE and RMSE than the AR(p) models for the three

markets at almost all the forecast horizons. These results are particularly useful; as the competing models have the same lags, they indicate that it is more reliable to use the Gaussian copula with a flexible specification of dependence structure for forecasting than one only with the linear association. When all the models are compared, it is evident that the MAPE and RMSE values of Gaussian copula are the smallest in most situations. Accordingly, the results show that the Gaussian copula model outperforms the benchmark models in forecasting tourist arrivals from the major short-, medium- and long-haul source markets. The SARIMA model provides better MAPE and RMSE results for short horizons such as 1-month ahead forecast for Australia and the United States or 6-month ahead for Indonesia.

Table 5.13: MAPEs of Copula and Benchmark Models over Five Time Horizons

Series	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Australia	Copula	5.879	6.471	6.799	6.213	9.040
	AR(13)	6.114	7.027	8.571	9.957	8.809
	SARIMA	5.768	7.141	9.162	10.780	9.406
Indonesia	Copula	6.892	8.109	7.438	7.069	6.257
	AR(11)	7.524	12.135	13.173	9.546	7.348
	SARIMA	7.048	6.999	8.238	8.743	9.869
USA	Copula	5.302	5.001	4.906	4.798	5.260
	AR(13)	5.070	5.114	6.289	5.670	5.484
	SARIMA	4.629	7.381	11.139	7.314	5.852

Table 5.14: RMSEs of Copula and Benchmark Models over Five Time Horizons

Series	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Australia	Copula	6640.354	7606.757	8182.038	8008.771	9824.577
	AR(13)	6380.778	8099.523	10020.740	11193.469	10695.223
	SARIMA	6213.168	8166.953	10870.849	12211.616	11695.472
Indonesia	Copula	21520.283	25028.812	21653.511	20951.849	18658.217
	AR(11)	22706.583	33479.569	36297.719	28647.199	23051.385
	SARIMA	21674.825	22944.780	24152.831	26805.001	28660.154
USA	Copula	2565.674	2655.550	2710.038	2418.246	2665.157
	AR(13)	2573.805	3267.406	3403.681	3037.664	3229.997
	SARIMA	2342.615	3795.846	5003.059	3673.905	3368.638

Different sample splits are used as robustness checks for the forecasting results. Specifically, two validation periods (Period 1: from 2011M01 to 2013M6; Period 2: from 2011M07 to 2013M12) are considered and the forecasting results are shown in Tables D5 to D8 in Appendix D. In general, the robustness checks are consistent with the results of Tables 5.13 and 5.14.

It is important to investigate Within Dependence in tourism demand forecasting. While the simple two-dimensional case is analyzed in Study 1, Study 2 applies the more sophisticated high-dimensional Gaussian copula for tourism forecasting. The two studies show that incorporating the analysis of Within dependence structures benefits tourism forecasts with greater forecasting accuracy. In next chapter, I will focus on the Between Dependence and demonstrate its contribution to tourism forecasting.

CHAPTER 6. BETWEEN DEPENDENCE AND MULTIPLE TOURISM FLOWS FORECASTING

In this chapter, my focus is on the Between Dependence structures of multiple tourism flows and to propose a copula-based tourism demand model as an alternative approach for forecasting with pairwise dependence structure. Du Preez and Witt (2003) argued that international travel demand by origin countries located in the same region with similar economic, political or social conditions, are more likely to be influenced by common determining factors affecting the region, such as real GDP growth, economic recessions and changing trend or consumer tastes in the region. If strong cross-correlation among the tourism flows exists, they postulated that the incorporation of interdependence between tourism flows can improve forecasting accuracy. I will combine the copula method with two traditional structural models, namely autoregressive distributed lag (ARDL) model and the equilibrium correction (or error correction) model (ECM), to analyze the Between Dependence and forecast multiple tourism flows. I will first introduce the copula-based approach for between dependence. After that, tourism flows from six origin countries located in three regions are analyzed. Specifically, I examine tourism flows from Canada, Japan, Malaysia, South Korea, Thailand and the United States to Singapore.

Copula-based Approach for Between Dependence of Tourism Flows

The general international tourism demand model typically estimated can be written as (variables are expressed in logarithms):

$$Y_{it} = g_i(\mathbf{X}_{it}), \quad i=1, 2, \dots, n \quad (6.1)$$

where Y_{it} is the logarithm of tourist arrivals from origin country i in time t ; g_i denotes the demand function of the i th flow; $\mathbf{X}_{it}$ is the vector of explanatory variables for origin country i at time t , $i \in (1, 2, \dots, n)$ and $t \in (1, \dots, T)$. The log-log specification is chosen for the tourism model so that the coefficients of the independent variables can be interpreted as demand elasticity with respect to the changes in the variables.

Equation (6.1) can be specified by different types of structural models, such as the ARDL model and ECM, which are commonly used in the tourism demand forecasting literature (Song and Li 2008). The ARDL model developed by Pesaran and Shin (1998) can address the cointegrating relationship among variables of $I(0)$ or $I(1)$ or a mixture of both (De Vita and Abbott 2002). If the variables pass the bounds test for the ARDL model, they are cointegrated and there is a stable long term relationship between tourism demand and its predictor variables. The ARDL model for Equation (6.1) is specified as:

$$Y_{it} = \alpha_{i0} + \sum_{j=1}^{p_1} \alpha_{ij} Y_{i,t-j} + \sum_{j=0}^{p_2} \beta_{iY,j} RI_{i,t-j} + \sum_{j=0}^{p_3} \beta_{iP,j} P_{i,t-j} + \sum_{j=0}^{p_4} \beta_{iOP,j} OP_{t-j} + \text{Dummies} + \mu_{it} \quad (6.2)$$

where RI_{it} is the real income of origin country i at time t ; P_{it} is the relative price; OP_t is the transportation cost; p_i ($i = 1, 2, 3, 4$) refers to the number of lags of the independent variables; and μ_{it} is the error term. Dummies are also included to capture the effects of one-off events, such as natural disasters or economic crisis, on tourism demand.

Cointegration and error correction models are closely related. An ECM relates the change in one variable to past equilibrium errors. The model provides a way of combining both the dynamics of the short run (changes) and long run (levels) adjustments processes simultaneously. After generating the error correction term, the ECM to be estimated is expressed as follows:

$$\Delta \ln Y_{it} = \pi_{i0} + \pi_{iY} \Delta \ln RI_{it} + \pi_{iP} \Delta \ln P_{it} + \pi_{iOP} \Delta \ln OP_{it} + \Delta \text{Dummies} + \pi_{ie} \hat{\varepsilon}_{i,t-1} + \tau_{it} \quad (6.3)$$

with the symbol Δ represents the first-difference operator and τ_{it} is the error term. $\hat{\varepsilon}_{i,t-1}$ is the error correction term which is given by:

$$\hat{\varepsilon}_{i,t-1} = Y_{i,t-1} - \alpha_{i0} - \beta_{iY,1} RI_{i,t-1} - \beta_{iP,1} P_{i,t-1} - \beta_{iOP,1} OP_{i,t-1} - \text{Dummies} \quad (6.4)$$

The copula-based approach adopted in this study combines the marginal distribution of each flow and their joint distribution. Following Sklar's Theorem, the joint distribution of tourist arrivals from n origin countries ($Y_{1t}, Y_{2t}, \dots, Y_{nt}$) can be uniquely represented by a copula multivariate distribution function:

$$F_t(y_{1t}, y_{2t}, \dots, y_{nt}) = C[F_{1t}(y_{1t}), F_{2t}(y_{2t}), \dots, F_{nt}(y_{nt}); \boldsymbol{\theta}] \quad (6.5)$$

where F_t refers to the joint distribution of $Y_{1t}, Y_{2t}, \dots, Y_{nt}$; F_{it} is the marginal distribution of Y_{it} ; y_{it} is the observation of Y_{it} ; and $C(\cdot; \boldsymbol{\theta})$ denotes the copula function with parameter vector $\boldsymbol{\theta}$ which captures information about the dependence among the Y_i 's.

Equation (6.5) can also be written as:

$$F_t(y_{1t}, y_{2t}, \dots, y_{nt}) = C(u_{1t}, u_{2t}, \dots, u_{nt}; \boldsymbol{\theta}) \quad (6.6)$$

where u_{it} is the cumulative density and $u_{it} = F_{it}(y_{it})$ for $i = 0, 1, \dots, n$.

A likelihood based estimation method is used to estimate Equations (6.1) and (6.6) simultaneously. Let f_{it} denotes the density of marginal distribution F_{it} . The log-likelihood function for the n flows is:

$$L = \sum_{t=1}^T (\ln f_{1t} + \ln f_{2t} + \dots + \ln f_{nt}) + \sum_{t=1}^T \ln c[F_{1t}(y_{1t}), F_{2t}(y_{2t}), \dots, F_{nt}(y_{nt}); \boldsymbol{\theta}] \quad (6.7)$$

where $c(\cdot)$ denotes the probability density function corresponding to the copula distribution function $C(\cdot)$ and is given by:

$$c(\cdot) = \frac{\partial^n C}{\partial F_{1t} \partial F_{2t} \dots \partial F_{nt}} \quad (6.8)$$

This copula-based approach is applicable for modeling dependence among two or more tourism flows. In this study, I will investigate pairwise dependence patterns ($n = 2$) between tourism flows.

Several copula functions are used to model the interdependence between tourism flows, namely Product copula, Frank copula and Clayton copula. In Equation (6.7), the dependence between two flows is captured by the association parameter θ . If two tourism flows are independent from each other, the structure of independence corresponds to the Product copula and the latter part on the right hand side of Equation (6.7) is equal to zero. Accordingly, the log-likelihood function for the Product copula-based model can be expressed as:

$$L = \sum_{t=1}^T (\ln f_{1t} + \ln f_{2t} + \dots + \ln f_{nt}). \quad (6.9)$$

Thus Equation (6.9) is nested within Equation (6.7). The former is the reduced model and Equation (6.7) is the full model. The estimation of Equation (6.7) will generate the coefficients of Equation (6.2) or Equation (6.3). With these coefficients, we can calculate the forecasts of tourism demand $\hat{Y}_{it}$.

Empirical Analysis

Data and Variables

As mentioned in Chapter 2, the Singapore Tourism Board (STB) has established its network of 20 regional offices located in 7 regions (STB 2015) to globally promote Singapore as a compelling leisure and business destination. In this study, I investigate tourism arrivals to Singapore from 6 source countries, including Canada (CA), United States (US), Japan (JP), South Korea (KO), Malaysia (MY) and Thailand (TH), over a 19-

year period (1995-2013). Specifically, I examine the tourism flows association or the dependence structure between two flows from 3 regions, namely CA&US (Pair 1) from the Americas' region, JP&KO (Pair 2) from the North Asia region, and MY&TH (Pair 3) from the Southeast Asia region. Other major source countries in the regions were also considered, such as Indonesia and the Philippines, but were not selected because they did not pass the cointegration test required by both the ARDL and ECM models.

Quarterly data are used to develop the optimal models and the estimation sample is 1995Q1 to 2008Q4. These models are used to generate out-of-sample forecast which can then be compared with the series from 2009Q1 to 2013Q4. In this way, I can evaluate how well the optimal dependence structure can contribute to tourism forecasting. The data used are quarterly tourist arrivals obtained from the Singapore Tourism Board (STB); real GDP per capita from Oxford Economics; CPI and exchange rate from the International Monetary Fund; and crude oil price (West Texas Intermediate (WTI) oil price) from the U.S. Energy Information Administration.

With reference to Equation (6.2) and (6.3), real income, Y_{it} , is measured by the logarithm of real gross domestic product (GDP) per capita. The logarithm of international crude oil price is the proxy for transportation cost OP_t and the logarithm of the real exchange rate is used to measure the relative price variable:

$$P_{it} = \ln \frac{(CPI_D/ER_D)}{(CPI_i/ER_i)} \quad (6.10)$$

where CPI_D and CPI_i are the consumer price index for the destination and origin i , respectively; ER_D and ER_i are the exchange rate of the local currency of the destination and origin i against the US dollar, respectively. The dummy variables include D1997 for the Asian financial crisis of 1997, D2003 for the SARS epidemic in 2003, and D2008 for

the global financial crisis of 2008. D1997 is only included in the tourism demand models for inbound tourists from Asia. Since quarterly data are used, the seasonal dummies, S1, S2 and S3, are also included as determinants.

Preliminary Test

Before estimating the copula models, I conduct unit root tests for the tourist arrivals, real GDP per capita, transportation costs and relative prices series. The unit root tests help us understand whether the variables involved are I(0) or I(1). The unit root tests using the augmented Dickey-Fuller (ADF) test procedure indicate that all the related variables (Y_{it} , RI_{it} , P_{it} , OP_t) are I(1) as each of the calculated ADF statistics exceeds the critical value of -2.89 at the 0.05 significance level. Taking first differences renders each series stationary (see Table 6.1).

Table 6.1: Unit Root Tests for Related Variables

Augmented Dickey-Fuller statistic (Level)						
Variable	CA	US	JP	KO	MY	TH
Y _{it}	-1.979	-1.648	-1.593	-2.484	-1.850	-1.555
RI _{it}	-2.385	-2.273	-1.257	-0.792	0.089	-0.510
P _{it}	-1.651	-0.671	-0.680	-1.528	-2.035	-0.304
OP _t	-1.147					
Augmented Dickey-Fuller statistic (1 st Difference)						
Variable	CA	US	JP	KO	MY	TH
Y _{it}	-5.142***	-13.077***	-13.060***	-10.747***	-11.667***	-13.601***
RI _{it}	-4.421***	-3.512*	-6.386***	-5.914***	-3.989**	-6.776***
P _{it}	-6.964***	-6.475***	-3.203*	-4.044**	-4.858***	-5.291***
OP _t	-8.034***					

*** statistical significance at 0.001 level; ** statistical significance at 0.01 level;

* statistical significance at 0.05 level

Unlike the studies in Chapter 5 whereby seasonal adjustment is made to the arrivals data, I use seasonal dummy here because the seasonality in the quarterly arrivals data is deterministic but not stochastic (Franses et al. 1995). The evidence is provided by the Hylleberg-Engle-Granger-Yoo (HEGY) tests for seasonal and nonseasonal unit roots in the tourist arrivals series. If the HEGY tests show that there is no seasonal unit root in the tourist arrivals, the seasonality in the series is deterministic (Abeyasinghe 1994). It can be seen in Table 6.2 that all the tourism series have nonseasonal (zero frequency) unit roots, which is consistent with the ADF test results in Table 6.1. However, we can reject the null hypothesis of seasonal unit root and argue that deterministic seasonality is present in the tourism series. The results of HEGY test support the use of seasonal dummies in the copula-ARDL and copula-ECM models.

Table 6.2: HEGY Test for Tourist Arrivals Series^a

	CA	US	JP	KO	MY	TH	Critical value (5%)
$t(\pi_1)$	-2.049	-2.228	-1.857	-2.569	-2.027	-1.355	-3.073
$t(\pi_2)$	-3.223 ^b	-3.436 ^b	-3.365 ^b	-4.039 ^b	-3.452 ^b	-3.466 ^b	-3.034
$F(\pi_3, \pi_4)$	14.598 ^b	13.075 ^b	12.876 ^b	22.230 ^b	11.778 ^b	14.628 ^b	6.598
^a π_1 = zero frequency; π_2 = biannual cycle; π_3 = annual cycle. A time series contains no seasonal unit roots if the t test of $\pi_2 = 0$ and the F test of $\pi_3 = \pi_4 = 0$ are both rejected. ^b Rejection of the unit root hypothesis at the 5% level of significance.							

Another necessary test for using the ARDL or ECM models is the cointegration test between the dependent and independent variables. Two types of cointegration test are implemented: the bounds test for the ARDL model and the Johansen test for ECM. Table 6.3 shows that the variables for CA, US and JP pass the bounds tests. Moreover, the table also reports the number of lags of the independent variables for the ARDL model of each origin country, which are determined by Schwarz criterion (SBIC). The Johansen tests in Table 6.4 show that the variables are cointegrated for KO, MY and TH. Therefore, the underlying variables meet the requirements of the ARDL model or ECM.

Table 6.3: Bounds Test for Cointegration and ARDL Specification

	F-statistic	ARDL (p_1, p_2, p_3, p_4) ^a	Critical Values of Bound Test			
			0.1		0.05	
			Lower	Upper	Lower	Upper
CA	7.291	(2,0,3,1)	2.72	3.77	3.23	4.35
US	4.583	(1,0,0,0)				
JP	5.786	(0,2,0,0)				
KO	1.683	-----				
MY	3.379	-----				
TH	2.615	-----				

^a p_i ($i = 1, 2, 3, 4$) refers to the number of lags of tourist arrivals, real GDP per capita, relative price and crude oil price, respectively.

Table 6.4: Johansen Cointegration Test

KO			MY		TH	
Hypothesized	Trace		Trace		Trace	
No. of CE(s)	Statistic	P	Statistic	P	Statistic	P
Unrestricted Cointegration Rank Test (Trace)						
None	53.210	0.014	42.320	0.150	63.334	0.001
At most 1	22.439	0.275	14.164	0.831	29.833	0.050
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)						
None	30.771	0.019	28.156	0.042	33.501	0.008
At most 1	17.963	0.131	10.278	0.718	21.313	0.047

I estimate CA and US (Pair 1) tourism demand using the ARDL model (Equation (6.2)); MY and TH (Pair 3) tourism demand with ECM (Equation (6.3)). To demonstrate

the flexibility of the copula-based approach, a combination of the ARDL model and ECM is used to estimate JP and KO (Pair 2) tourism demand.

Model Estimation

The scatter plots of the percentile ranks of residuals from the three country-pairs are shown in Figure 6.1. While the scatter plot in Panel (a) indicates a positive non-linear relationship between the tourism flows in Pair 1, a similar relationship is less apparent in Panels (b) and (c) for Pairs 2 and 3, respectively. Nonetheless, the corresponding Pearson correlation coefficients of the residuals for Pairs 1, 2 and 3 are 0.858, 0.486 and 0.575, respectively. These coefficients are significant at the 0.01 level and provide support for the positive association between the flows in each pair.

Figure 6.1: Scatter Plots of Percentile Ranks of Residuals

To formally determine the dependence structure for each pair, the model selection is based on the values of the log-likelihood (LL) function, AIC and SBIC for the Frank and Clayton copula models, with the Product copula as the benchmark. The Product copula, which exhibits independence relations between tourism flows, corresponds to the

common assumption of independence structure in prior tourism research. A larger (smaller) relative LL (AIC and SBIC) represents a better model fit. The goodness-of-fit statistics for the copula models are reported in Table 6.5. Undoubtedly, the Frank copula yields the highest LL value and the lowest values of AIC and SBIC for all the three pairs. The Frank model is thus the best option to accommodate the dependence structure of these tourism flows.

Table 6.5: Goodness-of-fit of Copula Models

		LL	AIC	SBIC
Pair 1 (CA&US)	Product Copula	195.833	-286.570	-188.856
	Frank Copula	214.042	-366.085	-268.371
	Clayton Copula	211.993	-361.987	-264.273
Pair 2 (JP&KO)	Product Copula	104.355	-156.711	-74.757
	Frank Copula	122.431	-190.862	-150.756
	Clayton Copula	116.602	-179.205	-94.099
Pair 3 (MY&TH)	Product Copula	162.957	-269.915	-181.657
	Frank Copula	174.475	-290.950	-199.540
	Clayton Copula	168.440	-278.879	-187.470

Table 6.6 shows that the positive and significant associations between tourism flows for the three country-pairs are supported by the estimated dependence parameter of the Frank copula model (given by θ). Since Equation (6.9) is nested within Equation (6.7), I also conduct a likelihood ratio test to examine the model fit for the Frank and Product copula-based models. According to Table 6.6, the test statistic for each pair is significant at the 0.001 level. This suggests that the full model (Frank copula-based) is preferable to the nested model (Product copula-based). The Frank copula-based model contributes

additional information about the association between flows which should be incorporated in tourism demand analysis for each country-pair.

Table 6.6: Estimation Results of Dependence Structure of Copula Models

	Model	Dependence Parameter θ	Likelihood Ratio Test Statistic ^a
Pair 1 (UK&US)	Frank	15.340***	36.418***
Pair 2 (JP&KO)	Frank	14.278***	36.152***
Pair 3 (MY&TH)	Frank	11.577***	23.036***

***statistical significance at 0.001 level

^a Frank copula model is the full model, while Product copula model is the nested.

Forecasting Results

Prior to generating tourism demand forecasts, various smoothing methods are used to predict the independent variables (income, relative price and transportation cost). They are simple and widely used techniques in tourism demand forecasting. The econometric software program EViews 9 incorporates several exponential smoothing methods, namely single, double, Holt-Winters-No Seasonal, Holt-Winters-Additive, Holt-Winters-Multiplicative and Error-Trend-Seasonality (ETS) exponential smoothing models. Among the various smoothing techniques, ETS framework encompasses the other exponential smoothing models and proposes the optimal combination of error, trend and seasonal components based on criterion such as AIC (Hyndman et al. 2002). Therefore, the ETS framework is applied to the logarithmic series of the independent variables in this study. Since the lagged variables of tourism demand (Y_{it}) are also included as the determinants of Y_{it} in the ARDL (ECM) model, I use the predicted values of Y_{it} generated by the

copula-ARDL (copula-ECM) approach as the observations for these lagged variables to calculate multiple-step forecasts.

The Frank and Product copula-based models are subsequently used to generate the out-of-sample joint forecast for the tourism flows in each pair which cover the period 2009-2013. I evaluate the performance of the copula models for 4-, 12-, and 20-quarter-ahead forecasts by analyzing the MAPE and RMSE. MAPE values lower than 50% are acceptable and values lower than 20% correspond to good forecasts (see Chapter 5). Thus, the MAPEs in Table 6.7 show that both the Frank and Product models generate good joint forecasts of tourism demand for each pair, while the Frank model performs even better than the Product model with smaller MAPEs. Moreover, the RMSE results for the Frank model are also predominantly smaller than the Product model, which is consistent with the MAPE results. Hence, the Frank model outperforms the Product model in forecasting joint tourist arrivals to Singapore from the countries in each pair.

Table 6.7: MAPE & RMSE of Different Forecasting Models

Pair	Model	4-Quarter-Ahead		12-Quarter-Ahead		20-Quarter-Ahead	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
1 (CA&US)	Product	0.709	0.099	1.162	0.131	1.678	0.180
	Frank	0.614	0.069	0.687	0.076	1.211	0.137
2 (JP&KO)	Product	7.624	0.784	7.038	0.721	7.855	0.806
	Frank	1.899	0.223	1.158	0.151	0.855	0.121
3 (MY&TH)	Product	0.953	0.130	3.206	0.385	3.821	0.441
	Frank	0.466	0.059	1.592	0.196	1.930	0.224

According to Figures 6.2, 6.3 and 6.4, most of the point predictions of the Frank model are closer to the actual tourist arrival values than those of the Product model.

Additionally, the Frank model predicts higher arrivals relative to the Product model in all three figures. The use of the Product model can underestimate tourism demand from the three country-pairs in most forecasting horizons.

Figure 6.2: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 1)

Figure 6.3: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 2)

Figure 6.4: Actual Tourist Arrivals and Forecasts of Copula Models (Pair 3)

This chapter examined the copula-ARDL and copula-ECM approaches with optimal dependence structure. They are found to possess better model fitness values and the copula-based approaches outperform the benchmark model which corresponds to independence structure (no association) of tourism flows from three regions. The Within and Between Dependences have been discussed in Chapter 5 and Chapter 6, respectively. Further discussion will be undertaken in next chapter to establish a comprehensive framework in which both Within and Between Dependences are taken into account.

CHAPTER 7. WITHIN-BETWEEN DEPENDENCE

The previous two chapters have analyzed the Within Dependence and Between Dependence separately for tourism demand forecasting. In this chapter, the statistical tool of copulas is used to integrate Within Dependence with the copula-based approach for Between Dependence to derive a copula-Within-Between Dependence (or copula-WB) framework. The purpose of this study is to investigate if this comprehensive framework can benefit tourism forecasting.

Copula-WB framework

To construct the copula-WB framework, I use copulas to specify the serial dependence structure of each tourism flow, as well as the dependence structures between flows from different origin countries. The proposed framework is designed as follows:

$$H_j(y_{jt}, y_{j,t-1}, \dots, y_{j,t-k}) = C_j(u_{jt}, u_{j,t-1}, \dots, u_{j,t-k}; \boldsymbol{\theta}) \quad (7.1)$$

$$H_j(y_{1t}, y_{2t}, \dots, y_{nt}) = C_j(u_{1t}, u_{2t}, \dots, u_{nt}; \boldsymbol{\delta}) \quad (7.2)$$

where $y_{j,t-i}$ is the observation of $Y_{j,t-i}$ and $Y_{j,t-i}$ denotes the tourism flow from origin country j ($j=1, 2, \dots, n$) at time $t-i$ ($i=0, 1, 2, \dots, k$); H_j refers to the conditional distribution of Y_{it} on $Y_{i,t-1}, \dots, Y_{i,t-k}$ and H_j refers to the joint distribution of $Y_{1t}, Y_{2t}, \dots, Y_{nt}$; $u_{j,t-i} = F_m(y_{j,t-i})$ is the cumulative density and $F_m(y_{i,t-j})$ denotes the marginal distribution of $Y_{j,t-i}$; and $C(\cdot; \boldsymbol{\theta})$ and $C(\cdot; \boldsymbol{\delta})$ denote the copula functions with parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, which capture information about the Within Dependence of $Y_{jt}, Y_{j,t-1}, \dots, Y_{j,t-k}$ and the Between Dependence of $Y_{1t}, Y_{2t}, \dots, Y_{nt}$, respectively.

In the simple two-dimensional case, the above framework can be written in the form:

$$H_j(y_{jt}, y_{j,t-1}) = C_j(u_{jt}, u_{j,t-1}; \theta_j), j=1, 2 \quad (7.3)$$

$$H_j(y_{1t}, y_{2t}) = C_j(u_{1t}, u_{2t}; \delta) \quad (7.4)$$

A likelihood based estimation method is used to estimate Equations (7.3) and (7.4) simultaneously. Since in univariate time series, it can only be expected that Y_t depends on $Y_{t-1}, \dots, Y_{t-k}$, but not the other way round, the copula modeling of the vector $(Y_t, Y_{t-1}, \dots, Y_{t-k})$ can be regarded as to model the conditional mean and volatility of the vector (Vaz de Melo Mendes and Aíube 2011). Therefore, the log-likelihood function for the bivariate case can be written as:

$$L = \sum_{j=1}^2 \sum_{t=1}^T \ln c_j[u_{jt}|u_{j,t-1}; \theta_j] + \sum_{t=1}^T \ln c_j[u_{1t}|u_{1,t-1}, u_{2t}|u_{2,t-1}; \delta] \quad (7.5)$$

where $c_j(\cdot)$ is the conditional probability density function of the copula function $C_j(\cdot)$ and it is given by:

$$c_j = \frac{\partial^2 C_j}{\partial u_{jt} \partial u_{j,t-1}} * \frac{1}{u_{j,t-1}} \quad (7.6)$$

and $c_j(\cdot)$ denotes the probability density function of the copula function $C_j(\cdot)$ and given by

$$c_j = \frac{\partial^2 C_j}{\partial u_{1t} \partial u_{2t}} \quad (7.7)$$

After estimating the dependence parameter θ , the binary search algorithm is applied to generate the forecasts of u_{jt} .

Empirical Analysis

Data Description

Like in Chapter 5, seasonally-adjusted monthly tourist arrivals data from January 1995 to December 2013 are used. The same country-pairs examined in Chapter 6 are considered in this study. However, I have to disregard the MY&TH country-pair because the arrivals series from Thailand is not first-order Markov process. The two country-pairs under investigation are CA&US (Pair 1) from the Americas' region and JP&KO (Pair 2) from the North Asia region. Using the copula-WB framework, I examine Within-Between Dependence structures of the two flows in each country-pair.

Similar to the studies in Chapter 5, the monthly data for each arrival series are split into the estimation period (from 1995M1 to 2010M12) and the validation period (from 2011M1 to 2013M12). The estimation sample is used to specify the copula and the benchmark models. Subsequently, the specified models are used to generate 1-, 6-, 12-, 18- and 24-month-ahead forecasts. The forecasts are compared using the expanding windows (or recursive) forecasting technique with the data in the validation sample.

Preliminary Test

The use of copulas for the Within-Between dependence necessitates the unit root test for stationarity, chi-square test for marginal distribution selection and LM test for Markov property.

The unit root tests in Chapter 5 have shown that tourist arrivals from Canada, South Korea and the United States are stationary in levels. The Japan series has a unit root and taking first difference renders this series stationary. Therefore, I analyze actual

arrivals (untransformed series) from Canada, South Korea and the United States, and arrivals' growth for Japan (in first differenced).

According to the chi-square test results in Table 5.2, the Cauchy distribution is selected for arrivals from South Korea, and the Logistic distribution for Canada, Japan and the United States, to generate the cumulative density u_t . Table 5.2 also shows that there is no serial correlation in the residuals of the AR(1) models for all the arrivals series, and thus these series are first-order Markov processes.

Copula Model Specification

The five copulas introduced in Chapter 4, namely the Gaussian, Student-t, Frank, Clayton and Gumbel copulas, are applied to construct the forecasting models under the copula-WB framework. However, the Clayton and Gumbel copulas require the association parameter $\theta \in (0, +\infty)$ and $\theta \in [1, +\infty)$, respectively. This means that these two copulas should not be used for negative associations. It can be seen from Table 5.3 that the Within Dependence of arrivals series from Japan is negative, as the association coefficients are all negative. It seems that the Clayton and Gumbel copulas are not appropriate to model the Within Dependence structure for this series. Hence, the five copulas are used to specify the Between Dependence structures for the two country-pairs and the Within Dependence structures for Canada, South Korea and the United States. The Gaussian, Student-t and Frank copulas are applied to the Within Dependence structure for Japan.

The Within-Between Dependence structure for each pair can be presented as W_i - W_j - B_{ij} , in which W_i and W_j are the Within Dependence structures of the Flow i and Flow j

and B_{ij} is the Between Dependence structure. As W_i , W_j and B_{ij} can be specified by different copula functions, a certain criterion is needed to select an optimal W_i - W_j - B_{ij} combination. Since I use the likelihood based method for model estimation, the log-likelihood (LL) values of copula-WB for each country-pair can be compared to determine the optimal model. A larger relative LL represents a better model fit. Table 7.1 reports the LL values for the optimal copula-WB for the two country-pairs. The LL values for all copula combinations are listed in Appendix E.

According to Table 7.1, the Gumbel-Gumbel-Clayton (or GB-GB-CT) combination generates the largest LL value and thus should be selected for copula-WB model construction of Pair 1. This GB-GB-CT combination means that the Within Dependence of CA and US are both specified by the Gumbel (GB) copula, and the Between Dependence of these flows is described by the Clayton (CT) copula. For Pair 2, the optimal choices should be the FK-GB-GB combination.

Table 7.1: Log-likelihood Values of Copula-WB Framework

Pair 1				Pair 2			
W_1	W_2	B_{12}	LL	W_3	W_4	B_{34}	LL
GB	GB	CT	4274.062	FK	Gb	Gb	2063.864
CT refers to the Clayton copula; FK refers to the Frank copula; GB refers to the Gumbel copula. W1 = Canada (CA); W2 = United States (US); W3 = Japan (JP); W4 = South Korea (KO).							

Forecasting Results

Since the series involved are all first-order Markov process in which Y_t is associated with Y_{t-1} , I compare the joint forecasts generated by the copula-WB framework with those generated by AR(1) model. The two methods include the same lag information. The only difference between them is that the former flexibly specifies the Within-Between Dependence structure, while the latter assumes linear Within Dependence structure with no Between Dependence structure. This comparison shows whether the well specified Within-Between Dependence structures contribute to tourism demand forecasting.

The ARIMA model is another benchmark which is linear and contains more information from high order AR and/or MA terms. The ARIMA model is given by:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = a + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t \quad (7.8)$$

where y_t is the time series data at time t ; a is the constant term; B is the backshift operators; and ε_t is the error term at time t with the normal distribution. The Akaike information criterion (AIC) is employed in the determination of the parameters, and optimal ARIMA models with the smallest AIC are thus selected. Table 7.2 presents the optimal ARIMA models that describe the arrivals from the 4 origin countries.

Table 7.2: ARIMA Model Parameters for Arrivals Series

Series	ARIMA (p,d,q)	Series	ARIMA (p,d,q)
CA	(1,0,1)	US	(2,0,0)
JP	(4,0,4)	KO	(2,0,0)

p is the number of AR terms; q is the number of MA terms;

d is the difference order.

Figures 7.1 and 7.2 provide the comparison between the actual total arrivals and the forecasts of AR(1) and copula-WB models at Horizons $i=1$ and $i=24$ for Pair 1. The advantage of using copula-WB for forecasting is not apparent in Figure 7.1. However, most of the point predictions of the copula-WB model are close to the actual total arrivals values than those of the AR(1) model in Figure 7.2. Similar patterns can be seen for Pair 2 in Figures 7.3 and 7.4.

Figure 7.1: Actual Arrivals of Pair 1 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=1$)

Figure 7.2: Actual Arrivals of Pair 1 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=24$)

Figure 7.3: Actual Arrivals of Pair 2 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=1$)

Figure 7.4: Actual Arrivals of Pair 2 and Joint Forecasts of AR(1) and Copula-WB Models (Horizon $i=24$)

I also evaluate tourism forecast performance by comparing the MAPEs and RMSEs of the copula-WB framework and benchmark models. Tables 7.3 and 7.4 display the out-of-sample MAPE and RMSE results for the 2 country-pairs, respectively. In general, the copula-WB framework generates MAPE values less than 20% for all horizons, indicating that this method generates good joint forecasts for the 2 country-pairs. Moreover, both the MAPE and RMSE results demonstrate that the copula-WB framework outperforms the AR(1) and even ARIMA models for almost all the forecast horizons.

Table 7.3: MAPEs of Copula and Benchmark Models over Five Time Horizons

Country-Pair	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Pair 1 (CA&US)	Copula-WB	3.399	4.667	5.644	8.088	11.029
	AR(1)	3.727	11.798	18.908	22.584	24.960
	ARIMA	3.467	9.999	17.130	21.337	24.297
Pair 2 (JP&KO)	Copula-WB	4.109	3.853	4.201	4.698	5.435
	AR(1)	4.089	6.253	7.965	8.885	9.295
	ARIMA	4.062	5.145	6.933	8.106	8.867

Table 7.4: RMSEs of Copula and Benchmark Models over Five Time Horizons

Country-Pair	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Pair 1 (CA&US)	Copula-WB	2026.793	2838.844	3605.916	4506.976	5930.059
	AR(1)	2272.568	5993.324	9274.276	11003.868	12289.767
	ARIMA	2174.829	5228.341	8454.338	10415.667	11979.522
Pair 2 (JP&KO)	Copula-WB	5357.500	5572.634	6200.284	5803.391	7152.598
	AR(1)	5376.216	7719.203	9672.806	10148.760	10990.557
	ARIMA	5296.197	6911.073	8744.485	9300.088	10503.370

Different sample splits are used as robustness checks for the forecasting results. Specifically, two validation periods (Period 1: from 2011M01 to 2013M6; Period 2: from 2011M07 to 2013M12) are considered and the forecasting results are shown in Tables D9 to D12 in Appendix D. In general, the robustness checks are consistent with the results of Tables 7.3 and 7.4.

This chapter proposes the comprehensive copula-WB framework which allows both the Within and Between Dependence structures to be specified by optimal copulas. Compared to the AR(1) and ARIMA models, the copula-WB framework generates better joint forecasts of tourism demand to Singapore for the two country-pairs. The contributions of the thesis, as well as some limitations of the empirical studies, will be discussed in the concluding chapter.

CHAPTER 8. CONCLUSION & RECOMMENDATIONS

The purpose of this thesis is to advance research in tourism demand modeling by initiating the application of the copula technique to address the dependence issues for tourism forecasting. This chapter will summarise the key findings of the empirical studies and comment on the contributions of the thesis. Some remarks about the limitations of this thesis and recommendations for future research are provided.

Overall Summary of the Findings

Four studies have been conducted to demonstrate the flexibility of the copula technique in analyzing different Within and Between dependence structures. In many past studies, the linear specification is employed to model the serial dependence (or Within Dependence) structure of tourist arrivals, as a proxy for tourism demand. Studies 1 and 2 in Chapter 5 support the use of the copula method to specify the Within Dependence of tourism demand, but not just the linear model. Study 1 employs the copula method to analyze the association between growth in tourist arrivals and actual arrivals to Singapore, and their first-lagged variables. The comparison of forecasting performance between copula and AR(1) models indicates that well specified Within dependence structure increases the accuracy of forecasts. This increase is especially obvious when we forecast the actual arrivals series. Study 2 investigates a more sophisticated situation for tourism demand series with seasonality and proposes the high-dimensional Gaussian copula model. The empirical results show that the Gaussian copula performs better than the AR(p) and even SARIMA models in forecasting tourist arrivals to Singapore. These results further

support the use of well specified Within dependence structures in tourism demand forecasting.

The third study in Chapter 6 examines the Between Dependence structure of tourism flows from different origin countries located in the same region. It is apparent that between dependence exists in the country-pairs of CA & US, JP & KO and MY & TH, and the dependence structures of the 3 country-pairs are exhibited by the Frank copula. Evidently, the copula-based Between Dependence approach generates more accurate joint forecasts, compared with the traditional models which assume independence between tourism flows.

The fourth study constructs a copula-WB framework in which the Within-Between Dependence structure can be specified by copula functions. Under this framework, the Gaussian, Student-t, Clayton, Frank and Gumbel copulas can all be used for the specification of the Within Dependence structure. Furthermore, these five copulas can also be applied to describe the Between Dependence structure, making the framework even more flexible than the approach proposed in the third study. The copula-WB framework generates better joint forecasts than the AR(p) and ARIMA models for tourism flows to Singapore from the 2 country-pairs of CA & US and JP & KO.

Contributions of the Research

The contribution of the thesis to the existing literature is to present the core ideas of copula-based models for tourism demand modeling and forecasting. When analysing individual tourism flow, the linear specification is traditionally used to model the Within dependence structures of tourism demand. This thesis shows that there is much to be

gained by investigating various Within Dependence structures using copulas. The analysis of Between dependence structure extends the tourism demand literature, in which the independence as well as the linear and nonlinear dependence structures among tourism flows can be appropriately addressed.

Besides methodological contributions, the thesis also has policy relevance. Analysis of Within Dependence structure provides some tourism insight to decision makers of the destination government, as well as various businesses supplying tourism products. The findings of Study 1 in Chapter 5 show that negative tail dependence structure exists in most 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.

When it comes to the actual arrivals, the positive dependence is also stronger in the tail/s than in the center. Specifically, for Canada, Germany, South Korea and the United States, a very strong positive dependence is present in the upper tail. This means that, if the current arrivals are at a high level, it is easy to maintain it in the next period for these four origin countries. Moreover, for Germany and the United States, we do not observe the lower tail dependence when the Gumbel copula is selected to specify the Within Dependence structures. This infers that the low level current arrivals are not expected to remain low in the future. It pays for STB to undertake market research and implement promotional strategies to attract tourists from Germany and the United States.

Analysis of Between Dependence structure in Chapter 6 provides insight on possible linkage between tourism flows. Knowledge of tourism flow dependence can help the governments and tourism businesses identify regional markets to implement joint tourism promotion. Destination promotion by region is less expensive compared to targeting the generating countries separately. Coordinated and/or combined marketing efforts based on informed decision-making can be appropriately implemented by tourism authorities and the travel industry to optimize returns on promotional expenditures. As mentioned in Chapter 2, STB which oversees all aspects of tourism in the country, has established 20 regional offices worldwide in Americas, Europe, Greater China, North Asia, Oceania, Southeast Asia, and South Asia, Middle East & Africa. The findings of Chapters 6 and 7 on the pairwise association between tourism flows provide insights to STB that joint marketing and promotion can be considered for some origin countries through its regional offices.

Calantone et al. (1987) argued that accurate tourism demand forecasting may reduce the opportunity cost, leading to efficient policies and decisions of governments or private sectors. Athiyaman and Robertson (1992) also concluded that the effectiveness of operational decisions, such as the allocation of government budgets in tourism, relies on accurate forecasts. Since demand forecasting is about the future of tourism, it is a necessary but challenging undertaking. In 2004, the Singapore Government established a S\$2 billion Tourism Development Fund for tourism development, sustainability and growth trajectory. The new five-year Tourism Development Fund has also been earmarked for tourism product, event and capability developments as disclosed at the 2016 Singapore Tourism Industry Conference. Accurate forecasts are necessary to aid

government make tourism spending decisions for these investments. The findings across the four studies show that the copula technique, a cutting edge methodology in empirical tourism research, can be used to benefit tourism policy development.

While the forecast of total tourism demand can inform the Singapore government on policy development in tourism investment, arrivals forecasts from different countries or regions are also useful for exploring potential markets. Singapore has experienced rapid tourism growth and the infrastructure development points to a positive tourism future for this country. Although tourism growth is expected to continue, the industry has encountered some volatility in tourist traffic recently. As some of the economies of Singapore's major tourist markets are slowing down, consumer discretionary spending on travel and tourism will be adversely affected. As mentioned in Chapter 2, STB has gradually expanded its marketing focus to potential short and long haul markets. Rather than relying on linear models, the copula technique can provide insightful dependence structure of tourism demand and more reliable forecasts for Singapore. The latter can inform STB in their search for new market opportunities and to make sensible decisions on tourism promotion and destination marketing.

Limitations of the Research & Future Studies

This thesis has shown the flexibility of using the copula technique for tackling different dependence issues. Nonetheless, this technique is still at the developing stage in tourism research due to modeling/programming difficulties, and its implication and extension in tourism research are still limited.

Many copula functions do not allow for variation in the association of more than two variables; thus using copulas for high dimensional cases is a difficult task. I can only apply the Gaussian copula for the higher-order Markov processes, although I am aware that some methods have been proposed recently to take into account high dimensional dependence structures, such as the vine (Kurowicka 2011), hierarchical (Okhrin et al. 2013; Brechmann 2014) or factor (Krupskii and Joe 2013; Oh and Patton 2015) copula modeling methods. The basic idea of these methods is to decompose or reduce the high dimensional dependence structures into lower dimensions, so that more copula functions can be incorporated to construct high dimensional copula models. Future research can employ these methods to include the information from more than one lags of tourism demand or combine more tourism flows for forecasting purposes.

Another limitation of copula is the lack of a clear-cut criterion for optimal copula model selection. The goodness-of-fit test usually recommends several reasonable choices, and the log-likelihood value is just a reference but not a specialized criterion. If a specific criterion can be developed for optimal copula selection, the dependence structure analysis will be more objective and convincing.

The use of copula technique suffers from some programming difficulties. While various copula functions can be used to specify dependence structures, there are different requirements for estimating the association parameters of different copulas. How to accommodate these requirements is a demanding task in programming. When we attempt to combine the various copula functions with the traditional methods to construct new forecasting models, the programming requirements become even more complicated. For example, the programming for the estimation of copula-ARDL and copula-ECM models

with the Frank and Clayton copulas can easily be achieved (which has been done in Chapter 6). However, designing programs for the copula-ARDL and copula-ECM with other copulas is still problematic. The Between Dependence structures of tourism demand will be better understood if the programming problem can be solved in future studies.

While recognising 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. Furthermore, there is no guarantee that the use of a sophisticated method like copula will generate a more accurate forecast. This is demonstrated in Study 1 of Chapter 5 of this thesis in which the forecast performance of the copula method is only marginally better than AR(1) model for the arrivals' growth series. This information, which perhaps supports the use of linear model, is established through the application of copula technique. 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. Moreover, it is obvious that the copula-based Between Dependence approach and the copula-WB framework generate better joint forecasts for different country-pairs than the linear benchmarks.

Tourism demand forecasting is complicated and challenging. The findings of this thesis can have limited contributions to policy development/formulation due to other constraints not included in the models proposed. Tourism behavior is complex and

tourism demand is sensitive to many factors and sporadic events such as natural and human-made disasters. Forecasting models cannot include all possible factors or predict the unexpected shocks. Besides conducting tourism forecasting, it also pays to draw on the insights of industry experts to establish sensible policy formulation and implementation.

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 (Witt and Song 2001; Li et al. 2005). The research undertaken in this thesis uses monthly and quarterly data, and focuses only on tourism demand for Singapore. Hence, the findings should not be generalised and it is necessary to conduct more studies across different countries in the future to validate the forecasting performance of the copula models.

REFERENCES

- Abeyasinghe, T. 1994. Deterministic seasonal models and spurious regressions. *Journal of Econometrics* 61 (2):259-272.
- Archer, B. 1995. Importance of tourism for the economy of Bermuda. *Annals of Tourism Research* 22 (4):918-930.
- Archer, B., and J. Fletcher. 1996. The economic impact of tourism in the Seychelles. *Annals of Tourism Research* 23 (1):32-47.
- Athanasopoulos, G., and A. De Silva. 2012. Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research* 51 (5):640-652.
- Athanasopoulos, G., and R. J. Hyndman. 2008. Modelling and forecasting Australian domestic tourism. *Tourism management* 29 (1):19-31.
- Athiyaman, A., and R. Robertson. 1992. Time series forecasting techniques: Short-term planning in tourism. *International Journal of Contemporary Hospitality Management* 4 (4):8-11.
- Ba-Fail, A. O., S. Y. Abed, and S. Jasimuddin. 2000. The determinants of domestic air travel demand in the Kingdom of Saudi Arabia. *Journal of Air Transportation World Wide* 5 (2):72-86.
- Balaguer, J., and M. Cantavella-Jorda. 2002. Tourism as a long-run economic growth factor: The Spanish case. *Applied Economics* 34 (7):877-884.
- Barrowclough, D. 2007. Foreign investment in tourism and small island developing states. *Tourism Economics* 13 (4):615-638.
- Beare, B. K. 2010. Copulas and temporal dependence. *Econometrica* 78 (1):395-410.
- Beneki, C., B. Eeckels, and C. Leon. 2012. Signal extraction and forecasting of the UK tourism income time series: a singular spectrum analysis approach. *Journal of Forecasting* 31 (5):391-400.
- Blake, A., R. Durbarry, T. Sinclair, and G. Sugiyarto. 2001. Modeling Tourism and Travel Using Tourism Satellite Accounts and Tourism Policy and Forecasting Models. In *Tourism and Travel Research Institute Discussion Paper*.
- BMI. 2015. Singapore Tourism Report Q2 2015. Singapore: BMI Research.
- Box, G. E., and G. M. Jenkins. 1970. *Time series analysis; forecasting and control*. San Francisco: Holden-Day.
- Brechmann, E. C. 2014. Hierarchical Kendall copulas: Properties and inference. *Canadian Journal of Statistics* 42 (1):78-108.
- Brooks, C. 2014. *Introductory econometrics for finance*. United Kingdom: Cambridge University Press.
- Brown, R. G. 1959. *Statistical forecasting for inventory control*. New York: McGraw-Hill Book Co., Inc.
- Caicedo-Torres, W., and F. Payares. 2016. A Machine Learning Model for Occupancy Rates and Demand Forecasting in the Hospitality Industry. Paper read at Ibero-American Conference on Artificial Intelligence.
- Calantone, R. J., C. A. Di Benedetto, and D. Bojanic. 1987. A comprehensive review of the tourism forecasting literature. *Journal of Travel Research* 26 (2):28-39.
- Chan, F., C. Lim, and M. McAleer. 2005. Modelling multivariate international tourism demand and volatility. *Tourism management* 26 (3):459-471.

- Chan, Y.-M. 1993. Forecasting tourism: A sine wave time series regression approach. *Journal of Travel Research* 32 (2):58-60.
- Chen, C.-F., M.-C. Lai, and C.-C. Yeh. 2012. Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems* 26:281-287.
- Chen, R., C.-Y. Liang, W.-C. Hong, and D.-X. Gu. 2015. Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing* 26:435-443.
- Chen, R. J., P. Bloomfield, and F. W. Cabbage. 2007. Comparing forecasting models in tourism. *Journal of Hospitality & Tourism Research* 32 (1):3-21.
- Chen, X., and Y. Fan. 2006. Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130 (2):307-335.
- Chen, X., W. B. Wu, and Y. Yi. 2009. Efficient estimation of copula-based semiparametric Markov models. In *Cowles Foundation Discussion Paper*.
- Cho, V. 2001. Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality & Tourism Research* 25 (4):399-420.
- Chong, T. 2010. The state and the new society: the role of the arts in Singapore nation-building. *Asian studies review* 34 (2):131-149.
- Chu, F.-L. 1998. Forecasting tourist arrivals: Nonlinear sine wave or ARIMA? *Journal of Travel Research* 36 (3):79-84.
- . 2004. Forecasting tourism demand: A cubic polynomial approach. *Tourism management* 25 (2):209-218.
- . 2008. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism management* 29 (1):79-88.
- . 2009. Forecasting tourism demand with ARMA-based methods. *Tourism management* 30 (5):740-751.
- . 2014. Using a logistic growth regression model to forecast the demand for tourism in Las Vegas. *Tourism Management Perspectives* 12:62-67.
- Claveria, O., E. Monte, and S. Torra. 2015. A new forecasting approach for the hospitality industry. *International Journal of Contemporary Hospitality Management* 27 (7):1520-1538.
- Crane, G. J., and J. van dar Hoek. 2008. Conditional expectation formulae for copulas. *Australian & New Zealand Journal of Statistics* 50 (1):53-67.
- Croes, R. R., and M. Vanegas Sr. 2005. An econometric study of tourist arrivals in Aruba and its implications. *Tourism management* 26 (6):879-890.
- Da Costa Dias, A. 2004. Copula inference for finance and insurance, Diss., Mathematische Wissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 15283, 2004.
- Daniels, M. J., W. C. Norman, and M. S. Henry. 2004. Estimating income effects of a sport tourism event. *Annals of Tourism Research* 31 (1):180-199.
- De Vita, G., and A. Abbott. 2002. Are saving and investment cointegrated? An ARDL bounds testing approach. *Economics Letters* 77 (2):293-299.
- Deco, G., and B. Schürmann. 2012. *Information dynamics: foundations and applications*. Germany: Springer Science & Business Media.

- Domma, F., S. Giordano, and P. F. Perri. 2009. Statistical modeling of temporal dependence in financial data via a copula function. *Communications in Statistics-Simulation and Computation* 38 (4):703-728.
- Douglas, C. 2001. *Forecasting Tourism Demand: Methods and Strategies*. Oxford: The Haworth Hospitality Press.
- Du Preez, J., and S. F. Witt. 2003. Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting* 19 (3):435-451.
- Dwyer, L., and P. Forsyth. 1993. Assessing the benefits and costs of inbound tourism. *Annals of Tourism Research* 20 (4):751-768.
- Dwyer, L., and P. Forsyth. 1996. Economic impacts of cruise tourism in Australia. *Journal of Tourism Studies* 7 (2):36.
- Dwyer, L., and P. Forsyth. 1998. Estimating the employment impacts of tourism to a nation. *Tourism Recreation Research* 23 (2):3-12.
- Dwyer, L., P. Forsyth, and R. Spurr. 2003. Inter-industry effects of tourism growth: implications for destination managers. *Tourism Economics* 9 (2):117-132.
- Dwyer, L., R. Mellor, N. Mistilis, and T. Mules. 2000. A framework for assessing “tangible” and “intangible” impacts of events and conventions. *Event management* 6 (3):175-189.
- Embrechts, P., F. Lindskog, and A. McNeil. 2001. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.
- Endo, K. 2006. Foreign direct investment in tourism—flows and volumes. *Tourism management* 27 (4):600-614.
- Ferreira, L. 2008. New tools for spread trading. *Futures* 37 (12):38-41.
- Fletcher, J. E. 1989. Input-output analysis and tourism impact studies. *Annals of Tourism Research* 16 (4):514-529.
- Frahm, G., M. Junker, and A. Szimayer. 2003. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters* 63 (3):275-286.
- Franses, P. H., S. Hylleberg, and H. S. Lee. 1995. Spurious deterministic seasonality. *Economics Letters* 48 (3):249-256.
- Frees, E. W., and E. A. Valdez. 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2 (1):1-25.
- . 2008. Hierarchical insurance claims modeling. *Journal of the American Statistical Association* 103 (484):1457-1469.
- Genest, C., B. Rémillard, and D. Beaudoin. 2009. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44 (2):199-213.
- Geurts, M. D., and I. Ibrahim. 1975. Comparing the Box-Jenkins approach with the exponentially smoothed forecasting model application to Hawaii tourists. *Journal of Marketing Research* 12 (2):182-188.
- Goh, C. 2012. Exploring impact of climate on tourism demand. *Annals of Tourism Research* 39 (4):1859-1883.
- Goh, C., and R. Law. 2002. Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism management* 23 (5):499-510.

- . 2011. The methodological progress of tourism demand forecasting: a review of related literature. *Journal of Travel & Tourism Marketing* 28 (3):296-317.
- Gounopoulos, D., D. Petmezas, and D. Santamaria. 2012. Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists' origin. *Annals of Tourism Research* 39 (2):641-666.
- Gray, H. P. 1970. *International travel--international trade*: Heath Lexington Books.
- Gunter, U., and I. Önder. 2015. Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data. *Tourism management* 46:123-135.
- Hanly, P., and G. Wade. 2007. Modelling tourism demand—an econometric analysis of North American tourist expenditure in Ireland, 1985–2004. *Tourism Economics* 13 (2):319-327.
- Hassani, H., A. Webster, E. S. Silva, and S. Heravi. 2015. Forecasting US Tourist arrivals using optimal Singular Spectrum Analysis. *Tourism management* 46:322-335.
- Hauke, J., and T. Kossowski. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 30 (2):87-93.
- Hayes, D. K., and A. A. Miller. 2011. *Revenue management for the hospitality industry*: Wiley River Street Hoboken, NJ.
- Henderson, J. C. 2015. Destination development and transformation: 50 years of tourism after independence in Singapore. *International Journal of Tourism Cities* 1 (4):269-281.
- Hernández-López, M., and J. J. Cáceres-Hernández. 2007. Forecasting tourists' characteristics by a genetic algorithm with a transition matrix. *Tourism management* 28 (1):290-297.
- Holt, C. 1957. Forecasting trends and seasonals by exponentially weighted averages. Carnegie Institute of Technology: Pittsburgh ONR memorandum.
- Hu, C. 2002. Advanced tourism demand forecasting: Artificial neural network and Box-Jenkins modeling, Purdue University, Indiana.
- Huang, H.-C. 2013. Artificial Intelligence Technology in Travel Agency Operating Revenue Forecasts. *Journal of Theoretical and Applied Information Technology* 54 (1):92-95.
- Huarng, K.-H., T. H.-K. Yu, and F. Sole Parellada. 2011. An innovative regime switching model to forecast Taiwan tourism demand. *The Service Industries Journal* 31 (10):1603-1612.
- Huse, M., T. Gustavsen, and S. Almedal. 1998. Tourism impact comparisons among Norwegian towns. *Annals of Tourism Research* 25 (3):721-738.
- Hylleberg, S. 1992. *Modelling seasonality*. Oxford: Oxford University Press.
- Hyndman, R. J., A. B. Koehler, R. D. Snyder, and S. Grose. 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18 (3):439-454.
- Ibragimov, R. 2009. Copula-based characterizations for higher order Markov processes. *Econometric Theory* 25 (03):819-846.
- Icoz, O., T. Var, and M. Kozak. 1998. Tourism demand in Turkey. *Annals of Tourism Research* 25 (1):236-240.

- Ivanovic, Z., S. Baresa, and S. Bogdan. 2011. Influence of foreign direct investment on tourism in Croatia. *UTMS Journal of Economics* 2 (1):21-28.
- Keum, K. 2011. International tourism and trade flows: a causality analysis using panel data. *Tourism Economics* 17 (5):949-962.
- Khan, H., C. F. Seng, and W. K. Cheong. 1990. Tourism multiplier effects on Singapore. *Annals of Tourism Research* 17 (3):408-418.
- Khoshnevis Yazdi, S., K. Homa Salehi, and M. Soheilzad. 2015. The relationship between tourism, foreign direct investment and economic growth: evidence from Iran. *Current Issues in Tourism*:1-12.
- Kim, J. H., and T. Ngo. 2001. Modelling and forecasting monthly airline passenger flows among three major Australian cities. *Tourism Economics* 7 (4):397-412.
- Kjersti, A. 2004. Modelling the dependence structure of financial assets: A survey of four copulas. Oslo: Norwegian Computing Center.
- Kon, S. C., and L. W. Turner. 2005. Neural network forecasting of tourism demand. *Tourism Economics* 11 (3):301-328.
- Krupskii, P., and H. Joe. 2013. Factor copula models for multivariate data. *Journal of Multivariate Analysis* 120:85-101.
- Kulendran, N., and K. Wilson. 2000. Is there a relationship between international trade and international travel? *Applied Economics* 32 (8):1001-1009.
- Kulendran, N., and S. F. Witt. 2003. Forecasting the demand for international business tourism. *Journal of Travel Research* 41 (3):265-271.
- Kurowicka, D. 2011. *Dependence modeling: Vine copula handbook*. Singapore: World Scientific.
- Law, R. 2004. Initially testing an improved extrapolative hotel room occupancy rate forecasting technique. *Journal of Travel & Tourism Marketing* 16 (2-3):71-77.
- Lee, A. O. 1990. Airline reservations forecasting: Probabilistic and statistical models of the booking process: Cambridge, Mass.: Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology,[1990].
- Lee, C.-K., and K.-S. Kwon. 1995. Importance of secondary impact of foreign tourism receipts on the South Korean Economy. *Journal of Travel Research* 34 (2):50-54.
- Li, G. 2004. Modelling and forecasting UK tourism demand in Western Europe: illustrations of TVP-LAIDS models' superiority over other econometric approaches, University of Surrey, United Kingdom.
- Li, G., H. Song, and S. F. Witt. 2005. Recent developments in econometric modeling and forecasting. *Journal of Travel Research* 44 (1):82-99.
- Li, G., K. K. Wong, H. Song, and S. F. Witt. 2006. Tourism demand forecasting: A time varying parameter error correction model. *Journal of Travel Research* 45 (2):175-185.
- Liang, Y.-H. 2014. Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering* 74:111-119.
- Liew, R. Q., and Y. Wu. 2013. Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds* 19 (1):12-30.
- Lim, C. 1997. The functional specification of international tourism demand models. *Mathematics and Computers in Simulation* 43 (3):535-543.

- Lim, C., and M. McAleer. 2000. A seasonal analysis of Asian tourist arrivals to Australia. *Applied Economics* 32 (4):499-509.
- . 2002. Time series forecasts of international travel demand for Australia. *Tourism management* 23 (4):389-396.
- Lin, V. S., P. Goodwin, and H. Song. 2014. Accuracy and bias of experts' adjusted forecasts. *Annals of Tourism Research* 48:156-174.
- Liu, J., and S. Sriboonchitta. 2013. Analysis of volatility and dependence between the tourist arrivals from China to Thailand and Singapore: A copula-based GARCH approach. In *Uncertainty Analysis in Econometrics with Applications*. New York: Springer, 283-294.
- Liu, J., S. Sriboonchitta, H. T. Nguyen, and V. Kreinovich. 2014. Studying volatility and dependency of chinese outbound tourism demand in Singapore, Malaysia, and Thailand: A vine copula approach. In *Modeling Dependence in Econometrics*: Springer, 259-274.
- Maheswaran, R., and R. Khosa. 2015. Wavelet Volterra Coupled Models for forecasting of nonlinear and non-stationary time series. *Neurocomputing* 149:1074-1084.
- Narayan, P. K. 2004. Economic impact of tourism on Fiji's economy: empirical evidence from the computable general equilibrium model. *Tourism Economics* 10 (4):419-433.
- Nelsen, R. B. 2007. *An introduction to copulas*. New York: Springer.
- Newbold, P., and T. Bos. 1994. Introductory business and economic forecasting. *International Journal of Forecasting* 16:451-476.
- Oh, D. H., and A. J. Patton. 2015. Modelling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics* (Published online: 17 Jul 2015).
- Okhrin, O., Y. Okhrin, and W. Schmid. 2013. On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics* 173 (2):189-204.
- Oosterhaven, J., and T. Fan. 2006. Impact of international tourism on the Chinese economy. *International Journal of Tourism Research* 8 (5):347-354.
- Ouerfelli, C. 2008. Co-integration analysis of quarterly European tourism demand in Tunisia. *Tourism management* 29 (1):127-137.
- Papatheodorou, A., and H. Song. 2005. International tourism forecasts: time-series analysis of world and regional data. *Tourism Economics* 11 (1):11-23.
- Parsa, R. A., and S. A. Klugman. 2011. Copula regression. *Variance Advancing and Science of Risk* 5:45-54.
- Pérez-Rodríguez, J. V., F. Ledesma-Rodríguez, and M. Santana-Gallego. 2015. Testing dependence between GDP and tourism's growth rates. *Tourism management* 48:268-282.
- Peric, J., and M. Niksic Radic. 2011. Sustainable foreign direct investment in tourism sector of developing countries. In *1st International Scientific Conference - Tourism in South East Europe*. Croatia.
- Pesaran, M. H., and Y. Shin. 1998. An autoregressive distributed-lag modelling approach to cointegration analysis. *Econometric Society Monographs* 31:371-413.
- Pfeifer, D., and J. Nešlehová. 2003. Modeling dependence in finance and insurance: the copula approach. *Blätter der DGVFM* 26 (2):177-191.

- Poirier, R. 2001. The political economy of tourism in Algeria. In *Mediterranean Tourism: Facets of Socioeconomic Development and Cultural Change*. London: Routledge, 211-225.
- Popescu, R.-I., and R.-A. Corboş. 2011. Strategic Analysis of Singapore's Brand and Rebrand Campaigns. *Review of International Comparative Management* 12 (2):136-153.
- Puarattanaarunkorn, O., and S. Sriboonchitta. 2014a. Copula based GARCH dependence model of Chinese and Korean tourist arrivals to Thailand: implications for risk management. In *Modeling Dependence in Econometrics*. New York: Springer, 343-365.
- . 2014b. Modeling dependency in tourist arrivals to Thailand from China, Korea, and Japan using vine copulas. In *Modeling Dependence in Econometrics*: Springer, 383-398.
- Reboredo, J. C. 2011. How do crude oil prices co-move?: A copula approach. *Energy Economics* 33 (5):948-955.
- Schwartz, Z., and S. Hiemstra. 1997. Improving the accuracy of hotel reservations forecasting: curves similarity approach. *Journal of Travel Research* 36 (1):3-14.
- Seetanah, B. 2011. Assessing the dynamic economic impact of tourism for island economies. *Annals of Tourism Research* 38 (1):291-308.
- Shan, J., and K. Wilson. 2001. Causality between trade and tourism: empirical evidence from China. *Applied Economics Letters* 8 (4):279-283.
- Sklar, A. 1959. *Fonctions de Répartition À N Dimensions Et Leurs Marges*. France: Université Paris 8.
- Song, H., and G. Li. 2008. Tourism demand modelling and forecasting—A review of recent research. *Tourism management* 29 (2):203-220.
- Song, H., and S. F. Witt. 2006. Forecasting international tourist flows to Macau. *Tourism management* 27 (2):214-224.
- Song, H., S. F. Witt, and T. C. Jensen. 2003. Tourism forecasting: Accuracy of alternative econometric models. *International Journal of Forecasting* 19 (1):123-141.
- Srisaeng, P., G. Baxter, S. Richardson, and G. Wild. 2015. A Forecasting Tool for Predicting Australia's Domestic Airline Passenger Demand Using a Genetic Algorithm. *Journal of Aerospace Technology and Management* 7 (4):476-489.
- STB. 2014. Singapore Tourism Board Annual Report 2013/14. Singapore.
- . 2015. Singapore Tourism Board Annual Report 2014/15. Singapore.
- . 2016. *Destination brand - YourSingapore*, 18/05/2016 2016a [cited 20/07 2016]. Available from <https://www.stb.gov.sg/about-stb/what-we-do/Pages/Marketing-Singapore.aspx>.
- . 2016. *Tourism Industry Conference 2016 Materials* 2016b [cited 18/08 2016]. Available from <https://www.stb.gov.sg/Pages/TIC2016.aspx>.
- Tang, J., S. Sriboonchitta, V. Ramos, and W.-K. Wong. 2014. Modelling dependence between tourism demand and exchange rate using the copula-based GARCH model. *Current Issues in Tourism* 19 (9): 876-894.
- Tang, J., S. Sriboondita, X. Yuan, and B. Wu. 2013. Volatility co-movement of China outbound tourism: dynamic copula based GARCH model. *International Journal of Innovative Management, Information & Production* 4 (1):51-64.

- Tisdell, C., and J. Wen. 1991. Foreign tourism as an element in PR China's economic development strategy. *Tourism management* 12 (1):55-67.
- Tosun, C., D. J. Timothy, and Y. Öztürk. 2003. Tourism growth, national development and regional inequality in Turkey. *Journal of Sustainable Tourism* 11 (2-3):133-161.
- Trivedi, P. K., and D. M. Zimmer. 2007. *Copula modeling: an introduction for practitioners*. Netherlands: Now Publishers Inc.
- Turner, R., and Z. Sears. 2013. Travel & Tourism as a Driver of Employment Growth. In *The Travel & Tourism Competitiveness Report 2013*: World Economic Forum.
- Turtureanu, A. G. 2010. Tourism Products Characteristics and Forms. *Acta Universitatis Danubius. (Economica)* 1 (1).
- UNWTO. 2016. *International tourism generates US\$ 1.4 trillion in export earnings*. World Tourism Organization 2014 [cited 18/08 2016]. Available from <http://media.unwto.org/press-release/2014-05-13/international-tourism-generates-us-14-trillion-export-earnings>.
- . 2016. *Over 1.1 billion tourists travelled abroad in 2014*. World Tourism Organization 2015a [cited 18/08 2016]. Available from <http://media.unwto.org/press-release/2015-01-27/over-11-billion-tourists-travelled-abroad-2014>.
- . 2015b. UNWTO Tourism Highlights: World Tourism Organization.
- Var, T., and K. Imam. 2001. Tourism in Egypt: History, policies, and the state. In *Mediterranean Tourism: Facets of Socioeconomic Development and Cultural Change*. London: Routledge, 181-196.
- Vaz de Melo Mendes, B., and C. Aúbe. 2011. Copula based models for serial dependence. *International Journal of Managerial Finance* 7 (1):68-82.
- Vergori, A. S. 2012. Forecasting tourism demand: the role of seasonality. *Tourism Economics* 18 (5):915-930.
- Wilmot, N. A., and C. R. McIntosh. 2014. Forecasting recreational visitation at US National Parks. *Tourism Analysis* 19 (2):129-137.
- Winters, P. R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* 6 (3):324-342.
- Witt, S., and H. Song. 2001. Forecasting future tourism flows. In *Tourism and hospitality in the 21st century*. Oxford: Elsevier Butterworth-Heinemann, 106-118.
- Witt, S. F., H. Song, and S. Wanhill. 2004. Forecasting tourism-generated employment: The case of Denmark. *Tourism Economics* 10 (2):167-176.
- Witt, S. F., and C. A. Witt. 1995. Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting* 11 (3):447-475.
- Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. USA: MIT press.
- WTTC. 2015a. Benchmark Report - Singapore. United Kingdom.
- . 2015b. Economic Impact 2015 Singapore. United Kingdom.
- Yu, G., Z. Schwartz, and B. R. Humphreys. 2007. Data patterns and the accuracy of annual tourism demand forecasts. *Tourism Analysis* 12 (1-1):15-26.
- Zdravevski, E., P. Lameski, R. Mingov, A. Kulakov, and D. Gjorgjevikj. 2015. Robust histogram-based feature engineering of time series data. Paper read at 2015

Federated Conference on Computer Science and Information Systems (FedCSIS),
at Lodz.

Zins, A. H., S. Dolnicar, K.-H. Huarng, T. Hui-Kuang Yu, L. Moutinho, and Y.-C. Wang.
2012. Forecasting tourism demand by fuzzy time series models. *International
Journal of Culture, Tourism and Hospitality Research* 6 (4):377-388.

APPENDICES

Appendix A. Total Tourist Arrivals and Arrivals from Different Source Countries (1995M01 to 2013M12)

Appendix B. Probability Density Functions of Five Copulas

Gaussian Copula

$$c_{\text{Gaussian}} = \frac{1}{\sqrt{1 - \theta^2}} \exp\left(\frac{2\theta\Phi^{-1}(u_t)\Phi^{-1}(u_{t-1}) - \theta^2(\Phi^{-1}(u_t)^2 + \Phi^{-1}(u_{t-1})^2)}{2 * (1 - \theta^2)}\right)$$

Student-t Copula

$$c_{\text{Student-t}} = \frac{1}{\sqrt{1 - \theta^2}} \frac{\Gamma\left(\frac{(v+2)}{2}\right) \Gamma(v/2)}{\Gamma^2((v+1)/2)} \frac{\left(1 + \frac{(T_v^{-1}(u_t))^2 + (T_v^{-1}(u_{t-1}))^2 - 2\theta(T_v^{-1}(u_t))(T_v^{-1}(u_{t-1}))}{v(1 - \theta^2)}\right)^{-(v+2)/2}}{\left(1 + \frac{(T_v^{-1}(u_t))^2}{v}\right)^{-(v+1)/2} \left(1 + \frac{(T_v^{-1}(u_{t-1}))^2}{v}\right)^{-(v+1)/2}}$$

Frank Copula

$$c_{\text{Frank}} = \frac{-\theta(e^{-\theta} - 1)e^{-\theta u_t}e^{-\theta u_{t-1}}}{[(e^{-\theta} - 1) + (e^{-\theta u_t} - 1)(e^{-\theta u_{t-1}} - 1)]^2}$$

Clayton Copula

$$c_{\text{Clayton}} = (\theta + 1)(u_t u_{t-1})^{-(\theta+1)}(u_t^{-\theta} + u_{t-1}^{-\theta} - 1)^{-(2\theta+1)/\theta}$$

Gumbel Copula

$$c_{\text{Gumbel}} = \frac{(\ln u_t * \ln u_{t-1})^{\theta-1}}{u_t u_{t-1}} \left[(-\ln u_t)^{\theta} + (-\ln u_{t-1})^{\theta} \right]^{(1-2\theta)/\theta} e^{-[(-\ln u_t)^{\theta} + (-\ln u_{t-1})^{\theta}]^{1/\theta}} \left[((-\ln u_t)^{\theta} + (-\ln u_{t-1})^{\theta})^{\frac{1}{\theta}} + \theta - 1 \right]$$

Appendix C. Binary Search Algorithm

Let x denotes the median of u_t , and the initial value of x is 0.5. The procedures of using binary search algorithm to calculate x are as follow:

```
x<-0.5
cp<-conditional probability  $x|u_{t-1}$ 
absolute error<-  $|cp - 0.5|$ 
up limit < -1
lower limit < -0
While absolute error > 0.001{
  If  $cp > 0.5$ 
    upper limit <-  $x$ 
  If  $cp < 0.5$ 
    lower limit <-  $x$ 
   $x <- (upper\ limit + lower\ limit) / 2$ 
  cp<- conditional probability  $x|u_2$ 
  absolute error <-  $|cp - 0.5|$ 
}
Return ( $x$ )
```

Appendix D. Robustness Check of Forecasting Results

Table D1: MAPEs of Study 1 (2011M01 to 2013M06)

Tourist Arrivals		Model	Forecast Horizon				
Series			<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Arrivals: Growth	Total	Student-t	2.195	2.683	2.546	2.835	2.184
		AR(1)	2.554	2.739	2.618	2.894	2.216
	Australia	Student-t	3.378	3.430	4.154	4.157	4.839
		AR(1)	3.479	3.436	4.160	4.134	4.798
	Bangladesh	Student-t	5.227	5.924	5.603	5.033	6.643
		AR(1)	4.883	5.907	5.577	5.047	6.739
	Japan	Student-t	5.359	4.960	5.044	4.213	4.218
		AR(1)	5.183	4.969	5.045	4.152	4.169
	Malaysia	Student-t	4.243	5.410	5.288	6.708	4.685
		AR(1)	4.283	5.461	5.320	6.752	4.709
	Myanmar	Student-t	4.995	5.622	6.315	5.667	6.018
		AR(1)	4.885	5.619	6.301	5.654	6.004
	New Zealand	Student-t	7.074	7.051	5.186	5.848	6.097
		AR(1)	7.089	7.073	5.192	5.854	6.088
	Switzerland	Student-t	3.856	3.933	4.439	4.289	6.126
		AR(1)	3.847	3.941	4.428	4.308	6.135
Arrivals: Actual	Canada	Student-t	5.314	10.802	16.382	20.043	23.067
		AR(1)	5.511	12.524	17.631	20.735	23.411
	Germany	Gumbel	6.618	13.293	21.478	27.751	32.611
		AR(1)	7.146	16.272	24.551	28.481	30.487
	South Korea	Student-t	5.230	12.421	17.392	19.773	21.235
		AR(1)	5.411	13.900	19.401	22.043	23.646
	USA	Gumbel	3.795	8.057	14.921	21.529	26.218
		AR(1)	3.972	12.631	19.765	23.776	26.567

Note: MAPE is the mean absolute percentage error.

Table D2: MAPEs of Study 1 (2011M07 to 2013M12)

Tourist Arrivals		Model	Forecast Horizon				
Series			<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Arrivals: Growth	Total	Student-t	2.574	2.662	2.899	2.580	2.613
		AR(1)	2.821	2.719	2.940	2.596	2.643
	Australia	Student-t	3.323	4.490	4.598	5.167	5.707
		AR(1)	3.524	4.492	4.579	5.142	5.665
	Bangladesh	Student-t	5.398	5.184	4.653	5.343	3.377
		AR(1)	5.045	5.184	4.688	5.433	3.458
	Japan	Student-t	4.243	4.326	3.526	3.211	2.166
		AR(1)	4.095	4.328	3.490	3.193	2.116
	Malaysia	Student-t	5.882	6.521	7.901	7.371	10.093
		AR(1)	5.733	6.562	7.931	7.376	10.043
	Myanmar	Student-t	5.400	6.331	5.903	6.193	5.976
		AR(1)	5.178	6.325	5.894	6.194	5.981
	New Zealand	Student-t	5.740	4.846	5.203	5.071	4.173
		AR(1)	5.776	4.908	5.273	5.130	4.250
	Switzerland	Student-t	4.469	5.116	5.230	6.648	6.324
		AR(1)	4.483	5.108	5.241	6.655	6.330
Arrivals: Actual	Canada	Student-t	4.520	10.543	17.461	21.302	22.358
		AR(1)	4.864	12.504	18.832	22.060	22.736
	Germany	Gumbel	6.437	11.995	16.493	21.679	31.545
		AR(1)	6.974	15.477	21.466	26.558	31.689
	South Korea	Student-t	5.130	12.742	18.486	21.765	24.709
	USA	AR(1)	5.302	14.333	20.515	23.937	26.928
		Gumbel	3.460	7.047	13.158	19.275	23.414
	AR(1)	3.711	11.860	19.209	23.273	24.304	

Table D3: RMSEs of Study 1 (2011M01 to 2013M06)

Tourist Arrivals		Forecast Horizon					
Series	Model	<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24	
Arrivals: Growth	Total	Student-t	34741.154	39693.835	38904.144	43797.064	35775.025
		AR(1)	38385.955	40376.241	39789.752	44545.023	35993.461
	Australia	Student-t	3970.571	4104.058	4654.484	4729.963	5329.867
		AR(1)	4156.841	4113.248	4671.207	4750.027	5294.636
	Bangladesh	Student-t	617.406	734.685	731.609	666.811	847.126
		AR(1)	603.391	736.878	733.217	668.666	848.770
	Japan	Student-t	4114.620	4360.176	4657.148	3969.299	4711.326
		AR(1)	4141.362	4331.505	4633.099	3937.933	4690.751
	Malaysia	Student-t	5905.389	7629.770	7648.453	9090.287	6968.700
		AR(1)	6071.977	7639.708	7648.196	9087.455	6904.283
	Myanmar	Student-t	507.490	580.610	622.477	545.271	554.183
		AR(1)	486.451	580.651	622.146	544.803	554.136
	New Zealand	Student-t	986.895	878.300	652.374	722.409	808.663
		AR(1)	988.128	876.405	649.729	720.319	805.917
	Switzerland	Student-t	343.923	378.827	421.774	430.012	560.396
		AR(1)	346.050	379.851	422.177	432.147	561.742
Arrivals: Actual	Canada	Student-t	549.271	903.547	1278.661	1536.397	1796.121
		AR(1)	555.016	1010.243	1371.533	1588.294	1822.403
	Germany	Gumbel	1895.827	3480.022	5203.091	5886.311	6988.672
		AR(1)	1938.252	4044.992	5715.874	5990.082	6561.567
	South Korea	Student-t	2632.673	4942.784	6860.449	7733.406	8518.222
	USA	AR(1)	2696.000	5461.094	7584.316	8559.025	9401.670
		Gumbel	1964.935	3745.076	6424.342	8893.846	11035.471
		AR(1)	2112.114	5369.873	8197.422	9789.961	11142.844

Note: RMSE is the root mean square error.

Table D4: RMSEs of Study 1 (2011M07 to 2013M12)

Tourist Arrivals		Forecast Horizon					
Series	Model	<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24	
Arrivals: Growth	Total	Student-t	40772.313	42690.368	46888.720	44563.873	49062.533
		AR(1)	43304.296	43120.679	47127.447	44265.168	48427.450
	Australia	Student-t	3840.762	4842.794	4955.182	5361.960	5482.357
		AR(1)	4091.953	4865.847	4977.037	5360.563	5477.770
	Bangladesh	Student-t	639.181	675.036	607.603	691.218	415.765
		AR(1)	626.010	677.071	608.420	694.102	418.447
	Japan	Student-t	3635.112	4148.927	3422.795	3648.672	1771.628
		AR(1)	3680.989	4129.607	3403.477	3642.868	1751.477
	Malaysia	Student-t	8085.162	9599.811	10944.627	10897.000	13432.634
		AR(1)	8017.159	9624.876	10946.705	10873.124	13416.418
	Myanmar	Student-t	557.580	627.464	578.836	597.602	605.731
		AR(1)	534.969	627.571	578.664	598.271	606.597
	New Zealand	Student-t	793.434	648.178	696.528	733.858	623.918
		AR(1)	793.937	647.812	696.482	734.743	626.185
	Switzerland	Student-t	424.341	504.928	534.029	632.551	655.829
		AR(1)	430.517	506.331	535.620	634.398	657.550
Arrivals: Actual	Canada	Student-t	437.421	895.593	1352.911	1664.486	1738.389
		AR(1)	454.248	1023.054	1455.215	1721.340	1767.277
	Germany	Gumbel	1886.423	3371.353	3651.471	5013.690	6831.911
		AR(1)	1932.812	3992.029	4701.444	5758.098	6803.765
	South Korea	Student-t	2740.418	5469.819	7601.145	9042.368	10275.053
	USA	AR(1)	2822.373	6050.012	8363.630	9867.340	11143.866
		Gumbel	1872.353	3556.087	5820.516	8227.843	9613.480
	AR(1)	2004.840	5226.950	7963.461	9678.356	9950.457	

Table D5: MAPEs of Study 2 (2011M01 to 2013M06)

Series	Model	Forecast Horizon				
		<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Total	Copula	3.453	3.526	3.631	3.268	1.706
	AR(13)	3.901	7.665	10.359	8.072	7.312
	SARIMA	3.455	9.066	14.870	12.679	12.481
Australia	Copula	5.687	6.104	6.466	7.479	11.749
	AR(13)	5.803	7.775	9.785	12.579	12.698
	SARIMA	5.511	7.670	10.604	13.779	13.881
Indonesia	Copula	7.321	8.607	7.685	7.861	6.999
	AR(11)	8.478	12.890	13.935	9.341	7.293
	SARIMA	6.975	7.292	8.748	9.411	10.380
USA	Copula	5.555	5.317	5.284	5.432	6.753
	AR(13)	5.418	5.691	7.374	6.636	7.247
	SARIMA	4.804	8.296	12.342	8.592	8.035

Table D6: MAPEs of Study 2 (2011M07 to 2013M12)

Series	Model	Forecast Horizon				
		<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Total	Copula	4.046	4.177	4.646	3.910	5.178
	AR(13)	3.971	6.187	7.826	7.210	5.863
	SARIMA	3.841	8.094	12.115	11.163	8.932
Australia	Copula	5.753	6.872	8.395	5.341	5.593
	AR(13)	5.741	7.495	10.131	8.718	4.164
	SARIMA	5.412	7.645	10.923	9.530	3.778
Indonesia	Copula	7.163	8.040	8.031	5.696	6.010
	AR(11)	7.664	10.952	10.981	9.218	8.162
	SARIMA	7.641	7.150	7.971	9.022	10.334
USA	Copula	5.061	4.871	4.478	5.117	3.283
	AR(13)	4.756	5.684	5.873	5.604	3.717
	SARIMA	4.504	7.974	10.652	6.843	3.677

Table D7: RMSEs of Study 2 (2011M01 to 2013M06)

Series	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Total	Copula	51054.866	53182.151	56463.892	52859.129	26309.112
	AR(13)	58375.609	104114.302	137221.547	113525.639	99307.281
	SARIMA	55217.207	122590.456	188073.806	166280.915	164474.446
Australia	Copula	6526.822	7398.453	8148.949	9135.464	12138.864
	AR(13)	6046.727	8747.730	11068.472	13155.308	13926.048
	SARIMA	6049.749	8730.962	12106.905	14410.714	15332.606
Indonesia	Copula	22179.513	24450.305	20871.898	22094.774	18897.808
	AR(11)	24378.758	34591.787	37501.886	27708.233	21063.236
	SARIMA	20576.300	21194.537	23479.943	27608.560	29861.750
USA	Copula	2677.924	2830.543	2951.378	2722.164	3321.214
	AR(13)	2726.900	3573.888	3817.873	3506.749	4162.709
	SARIMA	2469.550	4160.196	5486.188	4171.963	4374.231

Table D8: RMSEs of Study 2 (2011M07 to 2013M12)

Series	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Total	Copula	65285.368	71438.164	78725.943	68254.568	85303.397
	AR(13)	61362.179	89580.774	114657.335	107692.014	97321.835
	SARIMA	63033.677	115768.313	167117.490	159935.067	141005.299
Australia	Copula	6778.819	8196.977	9311.882	7168.952	5804.172
	AR(13)	6235.757	8591.522	11348.594	10592.348	4468.809
	SARIMA	5957.717	8710.201	12303.696	11780.444	4376.215
Indonesia	Copula	22941.546	24911.240	23557.654	18061.799	20186.113
	AR(11)	23620.015	29979.001	32287.359	28242.027	27265.076
	SARIMA	23104.374	23887.758	24906.689	27400.033	31346.316
USA	Copula	2504.865	2705.277	2465.340	2651.919	1505.158
	AR(13)	2447.756	3600.687	3161.739	3251.292	1714.912
	SARIMA	2290.242	4067.980	4771.543	3818.026	1728.532

Table D9: MAPEs of the 4th Study (2011M01 to 2013M06)

Country-Pair	Model	Forecast Horizon				
		<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Pair 1 (CA&US)	Copula-WB	3.719	5.100	6.536	9.947	12.937
	AR(1)	4.034	12.650	19.497	23.297	26.230
	ARIMA	3.763	11.020	17.907	22.323	25.681
Pair 2 (JP&KO)	Copula-WB	4.565	3.979	4.653	5.088	6.343
	AR(1)	4.502	6.258	7.828	8.654	8.770
	ARIMA	4.544	5.139	6.842	7.908	8.429

Table D10: MAPEs of the 4th (2011M07 to 2013M12)

Country-Pair	Model	Forecast Horizon				
		<i>i</i> =1	<i>i</i> =6	<i>i</i> =12	<i>i</i> =18	<i>i</i> =24
Pair 1 (CA&US)	Copula-WB	3.209	4.930	5.461	8.079	9.466
	AR(1)	3.674	11.973	19.108	23.083	23.869
	ARIMA	3.442	10.022	17.133	21.651	23.101
Pair 2 (JP&KO)	Copula-WB	3.713	3.623	3.729	4.173	4.385
	AR(1)	3.684	6.067	7.971	8.843	9.891
	ARIMA	3.422	4.921	6.648	8.007	9.296

Table D11: RMSEs of the 4th (2011M01 to 2013M06)

Country-Pair	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Pair 1 (CA&US)	Copula-WB	2169.441	3021.889	4028.749	5234.985	6910.205
	AR(1)	2428.496	6365.357	9577.063	11385.658	13066.263
	ARIMA	2335.599	5641.082	8833.457	10914.684	12805.784
Pair 2 (JP&KO)	Copula-WB	5779.831	5836.485	6803.255	6184.247	8537.603
	AR(1)	5790.523	7811.255	9744.107	9973.358	10920.317
	ARIMA	5716.144	7054.926	8911.246	9112.277	10531.894

Table D12: RMSEs of the 4th (2011M07 to 2013M12)

Country-Pair	Model	Forecast Horizon				
		$i=1$	$i=6$	$i=12$	$i=18$	$i=24$
Pair 1 (CA&US)	Copula-WB	1968.720	3050.862	3456.998	4657.551	4889.307
	AR(1)	2229.351	6219.194	9371.773	11393.852	11571.577
	ARIMA	2139.174	5402.058	8457.701	10726.371	11205.314
Pair 2 (JP&KO)	Copula-WB	5201.033	5490.163	4966.863	5606.553	5037.305
	AR(1)	5245.971	7809.551	9195.290	10450.107	10992.249
	ARIMA	4886.358	6909.403	7859.296	9508.325	10330.163

Appendix E. Log-likelihood Values of Copula-WB Framework (Full Version)

Pair 1				Pair 2			
W_1	W_2	B_{12}	LL	W_3	W_4	B_{34}	LL
GS	GS	GS	368.322	GS	GS	GS	166.649
GS	GS	ST	353.500	GS	GS	ST	142.448
GS	GS	FK	339.201	GS	GS	FK	162.712
GS	GS	CT	352.127	GS	GS	CT	164.742
GS	GS	GB	1239.896	GS	GS	GB	1139.112
GS	ST	GS	511.220	GS	ST	GS	354.982
GS	ST	ST	499.443	GS	ST	ST	339.374
GS	ST	FK	494.716	GS	ST	FK	350.654
GS	ST	CT	501.791	GS	ST	CT	353.676
GS	ST	GB	1440.810	GS	ST	GB	1360.273
GS	FK	GS	493.231	GS	FK	GS	326.792
GS	FK	ST	481.845	GS	FK	ST	310.986
GS	FK	FK	484.354	GS	FK	FK	324.845
GS	FK	CT	479.551	GS	FK	CT	323.917
GS	FK	GB	1444.622	GS	FK	GB	1344.512
GS	CT	GS	450.086	GS	CT	GS	348.215
GS	CT	ST	434.910	GS	CT	ST	338.727
GS	CT	FK	441.390	GS	CT	FK	344.561

GS	CT	CT	429.608	GS	CT	CT	342.891
GS	CT	GB	1396.926	GS	CT	GB	1383.703
GS	GB	GS	1011.144	GS	GB	GS	977.101
GS	GB	ST	1027.749	GS	GB	ST	730.087
GS	GB	FK	1450.973	GS	GB	FK	1332.127
GS	GB	CT	1487.560	GS	GB	CT	1304.437
GS	GB	GB	1721.140	GS	GB	GB	1624.266
ST	GS	GS	533.306	ST	GS	GS	349.023
ST	GS	ST	523.180	ST	GS	ST	328.517
ST	GS	FK	518.187	ST	GS	FK	348.341
ST	GS	CT	524.020	ST	GS	CT	345.397
ST	GS	GB	1418.800	ST	GS	GB	1253.197
ST	ST	GS	687.888	ST	ST	GS	536.945
ST	ST	ST	688.859	ST	ST	ST	521.556
ST	ST	FK	675.549	ST	ST	FK	535.945
ST	ST	CT	685.588	ST	ST	CT	533.902
ST	ST	GB	1609.729	ST	ST	GB	1555.461
ST	FK	GS	670.962	ST	FK	GS	510.000
ST	FK	ST	668.270	ST	FK	ST	497.634
ST	FK	FK	664.398	ST	FK	FK	510.045
ST	FK	CT	663.591	ST	FK	CT	506.264

ST	FK	GB	1627.312	ST	FK	GB	1534.706
ST	CT	GS	632.028	ST	CT	GS	531.816
ST	CT	ST	624.821	ST	CT	ST	523.893
ST	CT	FK	624.634	ST	CT	FK	530.911
ST	CT	CT	630.774	ST	CT	CT	529.227
ST	CT	GB	1620.664	ST	CT	GB	1571.349
ST	GB	GS	1263.018	ST	GB	GS	1161.165
ST	GB	ST	1099.918	ST	GB	ST	885.154
ST	GB	FK	1601.453	ST	GB	FK	1517.471
ST	GB	CT	1637.041	ST	GB	CT	1515.236
ST	GB	GB	1856.169	ST	GB	GB	1745.088
FK	GS	GS	530.122	FK	GS	GS	342.444
FK	GS	ST	520.231	FK	GS	ST	330.740
FK	GS	FK	519.993	FK	GS	FK	340.810
FK	GS	CT	515.999	FK	GS	CT	338.465
FK	GS	GB	1446.878	FK	GS	GB	1309.060
FK	ST	GS	687.594	FK	ST	GS	531.398
FK	ST	ST	686.240	FK	ST	ST	527.038
FK	ST	FK	677.634	FK	ST	FK	528.287
FK	ST	CT	679.999	FK	ST	CT	529.291
FK	ST	GB	1620.103	FK	ST	GB	1509.846

FK	FK	GS	684.928	FK	FK	GS	503.974
FK	FK	ST	684.354	FK	FK	ST	498.604
FK	FK	FK	675.759	FK	FK	FK	502.390
FK	FK	CT	675.783	FK	FK	CT	500.009
FK	FK	GB	1616.625	FK	FK	GB	1470.985
FK	CT	GS	626.092	FK	CT	GS	526.321
FK	CT	ST	625.631	FK	CT	ST	525.727
FK	CT	FK	626.359	FK	CT	FK	523.794
FK	CT	CT	622.851	FK	CT	CT	524.445
FK	CT	GB	1603.171	FK	CT	GB	1510.618
FK	GB	GS	1094.639	FK	GB	GS	1151.228
FK	GB	ST	1076.791	FK	GB	ST	882.766
FK	GB	FK	1629.842	FK	GB	FK	1508.578
FK	GB	CT	1630.854	FK	GB	CT	1506.701
FK	GB	GB	1604.190	FK	GB	GB	<u>2063.864</u>
CT	GS	GS	506.813	CT	GS	GS	-----
CT	GS	ST	496.021	CT	GS	ST	-----
CT	GS	FK	495.114	CT	GS	FK	-----
CT	GS	CT	504.208	CT	GS	CT	-----
CT	GS	GB	1468.942	CT	GS	GB	-----
CT	ST	GS	662.794	CT	ST	GS	-----

CT	ST	ST	657.689	CT	ST	ST	-----
CT	ST	FK	655.329	CT	ST	FK	-----
CT	ST	CT	655.076	CT	ST	CT	-----
CT	ST	GB	1606.693	CT	ST	GB	-----
CT	FK	GS	639.474	CT	FK	GS	-----
CT	FK	ST	646.978	CT	FK	ST	-----
CT	FK	FK	643.608	CT	FK	FK	-----
CT	FK	CT	634.264	CT	FK	CT	-----
CT	FK	GB	1643.665	CT	FK	GB	-----
CT	CT	GS	624.104	CT	CT	GS	-----
CT	CT	ST	629.701	CT	CT	ST	-----
CT	CT	FK	618.836	CT	CT	FK	-----
CT	CT	CT	625.358	CT	CT	CT	-----
CT	CT	GB	1630.668	CT	CT	GB	-----
CT	GB	GS	1151.154	CT	GB	GS	-----
CT	GB	ST	1012.626	CT	GB	ST	-----
CT	GB	FK	1602.242	CT	GB	FK	-----
CT	GB	CT	1609.862	CT	GB	CT	-----
CT	GB	GB	2037.386	CT	GB	GB	-----
GB	GS	GS	1010.233	GB	GS	GS	-----
GB	GS	ST	936.615	GB	GS	ST	-----

GB	GS	FK	1430.949	GB	GS	FK	-----
GB	GS	CT	1269.181	GB	GS	CT	-----
GB	GS	GB	1491.590	GB	GS	GB	-----
GB	ST	GS	1122.457	GB	ST	GS	-----
GB	ST	ST	1040.677	GB	ST	ST	-----
GB	ST	FK	1583.921	GB	ST	FK	-----
GB	ST	CT	1586.161	GB	ST	CT	-----
GB	ST	GB	1749.685	GB	ST	GB	-----
GB	FK	GS	1078.666	GB	FK	GS	-----
GB	FK	ST	1007.834	GB	FK	ST	-----
GB	FK	FK	1571.904	GB	FK	FK	-----
GB	FK	CT	1567.989	GB	FK	CT	-----
GB	FK	GB	1506.806	GB	FK	GB	-----
GB	CT	GS	1078.510	GB	CT	GS	-----
GB	CT	ST	929.301	GB	CT	ST	-----
GB	CT	FK	1526.825	GB	CT	FK	-----
GB	CT	CT	1532.033	GB	CT	CT	-----
GB	CT	GB	1698.435	GB	CT	GB	-----
GB	GB	GS	1787.942	GB	GB	GS	-----
GB	GB	ST	1787.942	GB	GB	ST	-----
GB	GB	FK	2544.393	GB	GB	FK	-----

GB	GB	CT	<u>4274.062</u>	GB	GB	CT	-----
GB	GB	GB	2371.780	GB	GB	GB	-----
<p>GS refers to the Gaussian copula; ST refers to the Student-t copula; CT refers to the Clayton copula; FK refers to the Frank copula; GB refers to the Gumbel copula.</p> <p>W1 = Canada (CA); W2 = United States (US); W3 = Japan (JP); W4 = South Korea (KO).</p>							