國 立 交 通 大 學

運輸與物流管理學系

博 士 論 文

No.019

模化航空旅客預先購票行為

Modeling Advance Purchase Behaviors of Air Passengers

指導教授： 邱裕鈞 教授

研 究 生： 劉佳欣

中 華 民 國 一 〇 五 年 十 二 月

模化航空旅客預先購票行為

MODELING ADVANCE PURCHASE BEHAVIORS OF AIR PASSENGERS

研 究 生：劉佳欣　　Student: Chia-Hsin Liu

指導教授：邱裕鈞　　Advisor: Prof. Yu-Chiun Chiou

國 立 交 通 大 學

運 輸 與 物 流 管 理 學 系

博 士 論 文

A Dissertation

Submitted to Department of Transportation and Logistics Management

College of Management

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in

Traffic and Transportation

December 2016

Taipei, Taiwan, Republic of China

中 華 民 國 一 〇 五 年 十 二 月

模化航空旅客預先購票行為

研究生：劉佳欣　　　　　　　　　　　指導教授：邱裕鈞 教授

國立交通大學 運輸與物流管理學系博士班

摘要

模化航空旅客預先購票型態 (advance purchase patterns) 是了解航空旅客行為之關鍵，並可做為航空公司定價、產品銷售、行銷企劃與航班調度等策略規劃之依據。航空旅客預先購票行為 (advance purchase behaviors) 是複雜且多變的，受到許多因素如機票動態訂價、航空公司競爭、旅客行程規劃、以及旅遊季節性影響。回顧以往文獻，針對旅客預先購票行為之研究，大多以發放調查問卷之方式進行探討。然問卷調查之研究方法有調查成本高、費時，以及回收樣本數有限之缺點。近年來，航空公司營收於線上售票之占比逐年上升，航空營收管理策略與旅客之間互動行為也越趨複雜。因此，與傳統問卷調查研究方法相比較，透過售票歷史資料進行航空旅客行為之研究與預測，不僅更為直接、成本低廉，且更具有代表性。據此，本研究收集 2011 年台北澳門航線歷史售票資料，針對預先購票曲線 (advance purchase curves)，以函數資料分析 (functional data analysis, FDA) 探討不同特性航班預先售票曲線之型態與特性，並據以預測特定航班於特定時間之銷售；再進一步以離散多項羅吉特模式 (discrete logit model) 以及連續羅吉特模式 (continuous logit model)，探討旅客出發時間與價格偏好對預先購票行為之影響。本研究之預測結果，可進一步與實時售票狀況比較，提供航空公司機票訂價與銷售策略參考。此外，研究中發現不同特性航班之預先購票曲線，具有相當大之差異。

關鍵字：航空需求、提前購票型態、航空旅客行為、函數型資料分析、函數型同步迴歸分析、離散多項羅吉特模式、連續羅吉特模式

MODELING ADVANCE PURCHASE BEHAVIORS OF AIR PASSENGERS

Student: Chia-Hsin Liu Advisor: Prof. Yu-Chiun Chiou

Department of Transportation and Logistics Management
National Chiao Tung University

ABSTRACT

Accurate advance purchase patterns can provide valuable insights into air passengers' behaviors and can be used to support airline decision-making activities with respect to seat allocation, pricing, marketing and flight scheduling. Advance purchase behaviors are complex and compounded with lots of factors, such as price dynamics, airline competition, trip flexibility, seasonality, etc. Previous studies conduct questionnaire surveys on air passengers so as to develop advance purchase behaviors. The samples collected by questionnaire surveys can represent real individual advance purchase behaviors, but the number of valid samples is usually rather limited and required of high survey cost. In contrast, with the growing revenue share of online purchasing, to develop and predict the collective advance purchase behaviors of flights directly based on transaction data is obviously much more timeliness, cost economic, and representative. The predicted advance purchase levels at a specific time of a specific flight based on historical transaction data can be viewed as a reference level in comparing with current sales data so as to dynamically advise pricing and promotion strategies prior to departure. Additionally, according to our analyses on the air ticket transaction data, it is found that advance purchase patterns differ remarkably across flights.

To explore flight advance purchase patterns, a functional concurrent regression model was firstly proposed. Several factors contributing to aggregate advance purchase patterns of various types of flights including flight schedule attributes (such as time of day, day of week, months of year and special vacations) and historical load factors were examined based on the shape of the advance purchase curve of each flight. The ticket transaction data which containing 1,044 flights and 134,820 transaction records of Taipei-Macau (TPEMFM) route in 2011 was used for model estimation. With better learning of advance purchase patterns and passenger behaviors for sales flights, airlines are able to develop and make appropriate adjustments for current strategy more efficiently and compete more effectively in today's marketplace.

Furthermore, the advance purchase behaviors of individual air passengers are considered. As airlines dynamically adjust prices and sales strategy based on the learning sales patterns, passengers can also decide to purchase at the ongoing price or choose to delay their purchase decisions. Therefore, choice models including discrete multinomial logit model and continuous logit model are proposed for the empirical analysis of advance purchase behaviors of air passengers. By modeling both price and departure time preferences of air passengers, the individual choice model developed in this research is expected to offer a rich behavioral interpretation of advance purchase behaviors and allow airlines to evaluate potential impacts of the implementing strategies. The models developed in this research have the potential to both improve existing applications in seat allocation and extend the scope of applications to other areas of airline planning such as pricing and revenue management.

KEYWORDS: Air transport demand, Advance purchase patterns, Advance purchase behaviors, Functional data analysis, Functional concurrent regression model, Discrete multinomial logit model, Continuous logit model.

謝誌

二零一六年是精彩的一年，接連完成好多難得的里程碑。數數日子，從職場回校園將近四十個月，繞著博愛路這棟百年古蹟從交研所碩士、當兵、就業到完成博士學位也超過十年，謹將這本博士論文獻給默默奉獻的家人們，一切成就都歸功於你們的支持。秉持著老爸老媽「就算賣房子借錢也要受教育」的初衷，在完成博士學位的此時，內心不禁慶幸南投老家依然在風雨中屹立著，感謝大哥與姐姐們能夠承擔家務讓我無後顧之慮。要感謝禍福與共的愛妻姿宜，總是細心包容，充當我第二個腦袋，讓我能任性往前，接下來的日子就靠我了。

感謝邱裕鈞教授無私指導與鼓勵，讓佳欣得已順利完成博士學位；總是不吝給予機會與磨練，從一年級甫入學就負責接待外賓進行英文表達訓練、協助日本教授至臺灣進行研究，到義大利 *Bergamo* 城堡中、新加坡大學及香港進行研討會報告，這些都是在企業中奮鬥數年也未必能得到的寶貴歷練。而老師認真生活及嚴謹治學的態度，不論在學識或是待人處事方面，皆對佳欣人生有極大影響及助益，也將是佳欣終身學習的標竿。感謝藍武王教授、黃承傳教授、黃台生教授、馮正民教授、陳穆臻教授、鍾易詩教授、吳昆峰教授以及碩士班指導老師許鉅秉教授，在佳欣碩士及博士求學期間，不僅在課堂傳道授業，也在生活上給予關心與照顧。在口試階段，承蒙口試委員顏進儒教授、溫傑華教授、呂錦隆教授的細心斧正，使論文更臻完備。感謝所辦柳美智小姐與何玉鳳小姐，在諸多方面的支持與鼓勵。謝謝博士班學長姐：昱凱、世昌、昭弘、承憲、永祥、群明、文斌、素如、彥斐、姿慧、士軒、傳強、熙仁及裕文等的照顧與指導；有幸能與志偉及孜穎、郁珍、國濟、詮勳與暫時休息的郭菁、奚若與鄒亞諸位 *4995* 博士研究室成員一起努力打拼晚上吃便當；交大土地公真的有拜有靈，祝福各位投稿順利。

特別感謝彥蘅學長，全力支持與鼓勵佳欣就讀博士班，開啟佳欣人生另一條康莊道路。由職場返回校園，感謝江衍宏董事長、陳仁陳顧問、郭漢禹總經理、振全及亞芝，在佳欣就職期間的照顧與支援，並在求學、職場及生活中不厭其煩的提供協助與建言；感謝創業夥伴謝宗震博士於論文方法論支援，以及培仁同志大方的技術分享與指導，未來也請各位多多指教。最後，勉勵自己能不斷向前。

"Every passing minute is another chance to turn it all around."

Rafe @ Dec. 9, 2016

Table of Contents

Lists of Figures

Lists of Tables

Chapter 1 Introduction

The chapter consists of four sections. Section 1.1 addresses the principal concept on analyzing advance purchase behavior of air passenger in this study. The research problems, objectives, and flowchart are introduced in Section 1.2, 1.3, and 1.4, respectively.

1.1 Background and Motivation

How air passengers choose departure times and purchase tickets in advance is essential for airlines to develop revenue management (RM) strategies. The advance purchase pattern provides valuable insights that can be altered by airlines to support decision making activities such as seat allocation, pricing, marketing, sales and flight scheduling. With better learning of advance purchase patterns and passenger behaviors for sales flights, airlines are able to know which and when do flights need to be promoted to increase sales. Based on the learning purchase pattern, airlines can develop and make appropriate adjustments for current strategy more efficiently and compete more effectively in today's marketplace.

Effective revenue management strategies have clearly helped airlines to increase profit and allocate resources more efficiently. For the non-storable and perishable goods such as flight tickets, airlines generally implement RM strategies to optimize selling strategies and make the most profit possible based on the remaining capacity, current market conditions and anticipated demand. RM demand model has been proposed based on a hypothesized inverse demand function using traditional statistics techniques, such as time series, averaging methods, or simple probability distributions (McGill and Ryzin, 1999). Those demand models mostly assume passenger demand to be independent among fare products that created based on different restrictions for passenger segmentation. Through setting the booking limits and fare restrictions to each designed fare classes, airlines are able to segment the demand by distinguishing passengers with different levels of willingness to pay and applying price discrimination to obtain the highest revenue possible.

Airlines generally employ strikingly complex pricing and sales strategies for the intense market competition, differentiated demand patterns, and achieving effective customer segmentation to achieve price discrimination (Bilotkach et al., 2010). For example, previous studies have shown that advance purchase discounts can be altered as an effective means to shift demand (Gallego et al., 2008; Dana, 1999, 1998; Gale and Holmes, 1993). By setting advance purchase discounts to sales flight tickets, airlines are able to induce price-sensitive leisure passengers to purchase earlier. The less price-sensitive but time-sensitive business passengers may therefore decide to purchase later. In addition, to prevent business passengers from purchasing tickets at discounted fares that designed for leisure travelers, the airlines typically built complex fare rules and restrictions to make the deals less unattractive to business travelers. Moreover, airlines also dynamically adjust prices based on learning demand to achieve higher load factor and gain extra revenue (Escobari, 2012; Deneckere and Peck, 2012). However, on the other hand, passengers can decide to make advance purchase at the ongoing price or to delay their purchase decisions. Those price and sales strategies may decrease the product value that passengers may be forced to make trade-offs between price, product attributes and advance purchase deadlines, and therefore, change their purchasing behaviors (Hotle et al., 2015; Escobari, 2014). Without knowing the real advance purchase behaviors of air passengers, the typical hypothesized demand functions may lead to erroneous estimated results.

Additionally, because of the rapid growth in low cost carriers (LCCs) market, many airlines have virtually removed typical fare restrictions which are typically used for market segmentation. Low-cost carriers not only acquire market by offering lower fares, but also prevailing one-way pricing models that disrupt the traditional pricing and revenue management strategies of full service carriers by relaxing fare rules such as advance purchase requirements and the Saturday night stay requirement of discounted fare products. Moreover, online sales have become one of major distribution channels for airlines and travel agencies. From the passenger perspective, online purchases allow passengers to compare different product offerings more easily. Therefore, it increases price transparency among purchased flight tickets and competition airlines. Passenger nowadays may perceive fare classes as different prices for a seat on an airplane and purchase based on price rather than product characteristics (Garrow, 2009). These market changes have made traditional fare products less clearly defined, and assumptions of traditional RM models such as independence across fare classes may no longer be valid (Barnhart and Smith, 2012).

Furthermore, in order to trace individuals' advance purchase decisions, recent researches have introduced choice models to RM for its ability to accommodate passenger preferences in RM strategies that can better explain how individuals making trade-offs (Garrow, 2009; Talluri and van Ryzin, 2004a, 2004b). The decisions of passengers can be modeled based on either stated preferences survey data (Proussaloglou and Koppelman, 1999; Wen and Lai, 2010) or revealed preferences data. Most of previous studies based on stated preferences usually needed to conduct a questionnaire survey on air passengers so as to develop advance purchase behaviors. The samples collected by questionnaire surveys can represent real individual advance purchase behaviors, but the number of valid samples is usually rather limited and required of high survey cost. In contrast, with the growing revenue share of online purchasing, to develop and predict the collective advance purchase behaviors of flights directly based on transaction data is obviously much more timeliness, cost economic, and representative.

Despite that demand models based on discrete choice models may be more appropriate in RM applications, for the revealed preferences settings, there is limited empirical research due to data acquisition problems. Both chosen and non-chosen alternatives are needed for revealed preference model implementations. Although the support of computer systems lowers down data collection costs, most of firms can still only record the results of passengers of successful purchase and information about non-chosen alternatives had been difficult to obtain, which made inferring the true demand with available data remains a quite expensive and challenge issue. In this study, we will then focus on how to use existing data sources to develop and estimate a model of airline advance purchase behavior that can better reflects passenger interests.

Given the background, a better understanding of passenger choice behavior is crucial to support the decision making process of airline and compete more effectively in today's marketplace. The approach is expected to offer a rich behavioral interpretation for air passenger advance purchase behaviors. It is also important that the model be capable of supporting a broad range of policy implications by utilizing the available disaggregated data. Grabbing these challenges and opportunities may provide the potential to explore advance purchase behavior from real-time transaction databases and thus motivate this research.

1.2 Research Problems

Accurate advance purchase behaviors can provide valuable insights into air passengers' behaviors and that can be used to support airline decision-making activities with respect to seat allocation, pricing, marketing and flight scheduling. With the growing popularity of online purchasing, to develop and predict the collective advance purchase behaviors of flights directly based on transaction data is much more timeliness, cost economic, and representative. Instead of using data based on costly large-scale questionnaire surveys, we will focus in this research on analyzing actual advance purchase behaviors as reflected in past transaction records. The transaction data used for empirical estimation in this study was based on International Air Transport Association (IATA) billing and settlement plan (BSP), which is widely used by financial department of airlines and easily acquired comparing with other data sources. The flight schedule data was also integrated to the analysis dataset. This will provide the basis to explore the trade-off between the major dimensions of airline passenger preferences such as schedule convenience and price.

In sum, this research attempts to explore advance purchase patterns and behaviors of air passengers based on the transaction data by examining the following problems in sequence:

(1) How is advance purchase behaviors represented?

(2) Do patterns of advance purchase behaviors exist?

(3) What are the relevant factors affecting advance purchase behaviors of air passengers?

(4) How these patterns support airline decisions?

1.3 Research Objectives

Based on the abovementioned background and motivations, the objectives of this research are:

(1) Propose an approach for identifying advance purchase behavior patterns and exploring their characteristics:

The approach proposed in this research was aimed to explore the advance purchase patterns and its role in airline's decision process. Given this objective, two types of methodologies were employed. The first was the functional data analysis techniques including curve smoothing and functional descriptive statistics. They were adopt to transform the discretely collected transaction data into functional curves/objects for analyze. The second one was functional regression model that used to examine the potential contributing factors including flight schedule attributes (such as time of day, day of week, months of year and special vacations) and historical load factors that may influencing the advance purchase patterns.

(2) Propose an approach for examining advance purchase behavior of air passengers.

The second objective was aimed to examine the individual advance purchase decision of passengers. After identifying flights that might have poor sales performance, airlines can adjust prices and sales strategy dynamically based on learning patterns. However, the behavior of passengers might be also forced to change. Two models were proposed to empirical analysis of the advance purchase behaviors of air passengers. The discrete choice model was firstly used to explore which choice set construct scheme based on advance purchase time that can estimate choice model well. Taking account for the nature of continuous purchase time choices in this study, the continuous logit model was then proposed.

1.4 Research Methods

The research methods of this study consist of two parts. The first part related to functional data techniques and a functional regression models, which are adopted to explore and predict aggregated advance purchase patterns of flights at different departure time before investigating advance purchase behaviors of air passengers. The FDA techniques are appropriate for the data with ideal units of observation are defined as functions on continuous domain, that allows for the accurate estimation of parameters for the use in the analysis movement patterns, data noise reduction and interpolation through curve smoothing, and applicability to handle data with irregular time sampling periods. It provides a natural way to think through modeling problems in a functional form, as would be the case if finite data were used to estimate an entire function, its derivatives, or the values of other functional, that traditional multivariate analysis approaches lack (Ramsay and Silverman, 2002, 2005, 2009).

Second, choice models including discrete choice and continuous choice model are then used to examine contributing factors to advance purchase behaviors. The choice models can be applied to the analysis of individual choice behavior when they are faced with multiple alternatives. To date, unlike previous studies, we defined advance purchase period and took it as the response variable (alternatives) to obtain the probability of purchasing in each time period. To reduce the number of alternatives and explore which choice set construct scheme that can estimate choice model well, the multinomial logit model (MNL) was used. The MNL is the most basic choice model and is widely used due to its simple estimation resulting from its strong assumptions. However, the independence of irrelevant alternatives (IIA) property of MNL model has restricted in many practical situations.

Although the more advance choice models such as nested logit model (NL) and generalized nested logit model (GNL) can alleviate the IIA problem. Some choices are continuous response variables such as advance purchase time, departure time, and location. Arbitrarily discretizing these continuous choices variables may lead to an erroneous result. The continuous logit model is used. The continuous choice model represents a generalization form of the MNL for continuous response variable settings and can be derived directly from the random utility theory. In this study, the advance purchase time choices of the individual passenger are considered as continuous. The continuous logit model is applied for its advantage of offering strong theoretical supports based on the random utility framework without discretizing the decision time horizon.

1.5 Research Flowchart

Given the objectives, the research flowchart was illustrated in Figure 1-1. The remainder of the dissertation is organized as follows. Chapter 1 introduces the research background, objectives, and the expected for this research. In Chapter 2, the literatures regarding the advance purchase behaviors, functional data analysis and choice models were reviewed. Chapter 3 presents methodologies for the advance purchase patterns and passenger choice model applicable to disaggregate data including function-on-function regression model, discrete multinomial logit model (MNL), and continuous choice model.

Chapter 4 presents the explanatory analysis for analysis dataset. Existing data sources including air ticket transaction data and flight schedule dataset will firstly combined to investigate the departure time preferences of passengers such as time of day, days of week and months of year. Note that, our dataset also contains transaction records from both direct purchasing passengers and from multiple distribution channels, which makes it hard to distinguish passenger behaviors from travel agents. Therefore, for simplicity, only the direct purchasing passengers were considered for passenger choice behavior modeling. Figure 1-2 further provides the data processing flow and model framework for this study. The explanatory analysis mainly focuses on passenger demand, advance purchase patterns, and purchasing price distribution. The estimating results from proposed there models were presented in Chapter 5. The related issues and applications were later discussed in Chapter 6. Finally, the conclusions and recommendations were drawn in Chapter 7.

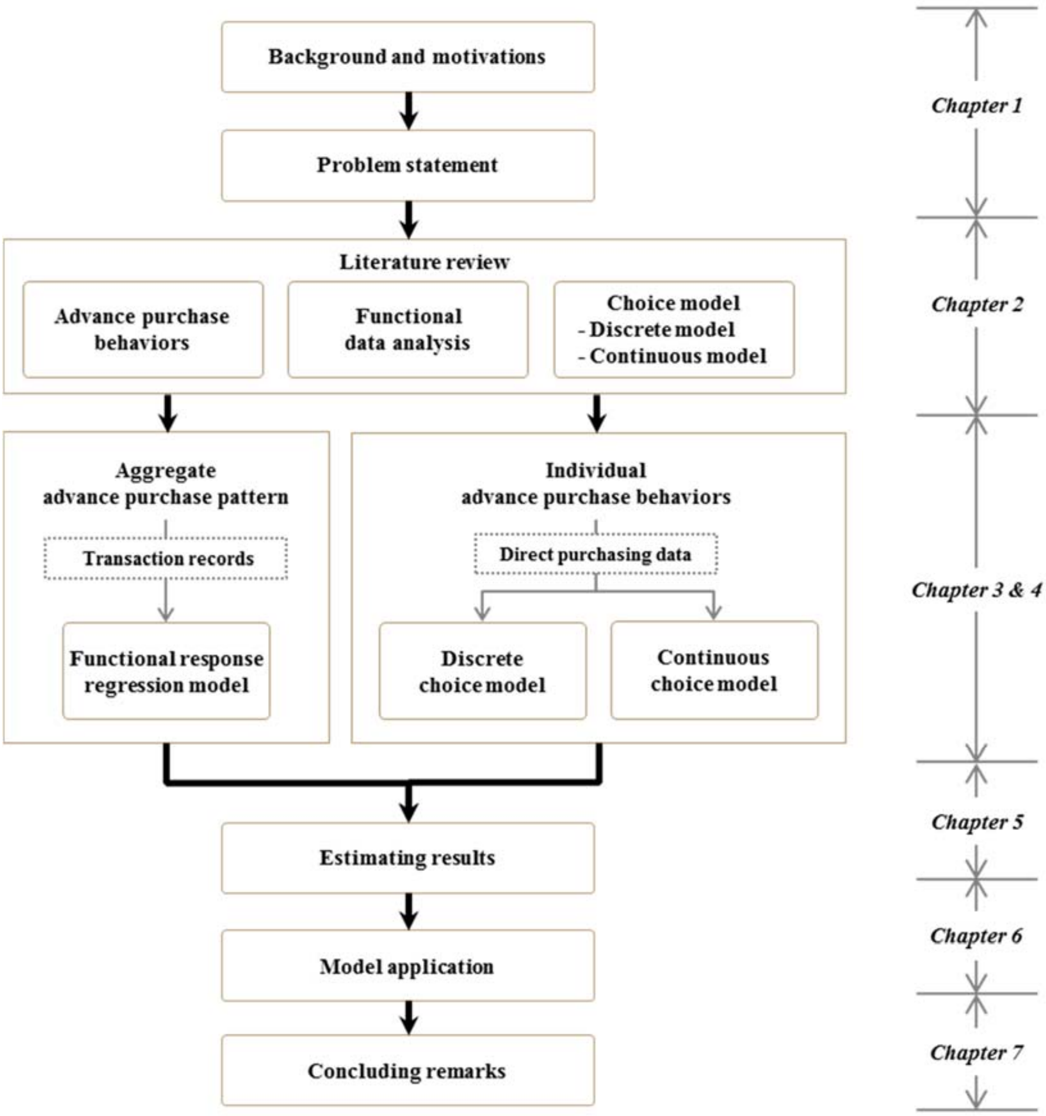


Figure 1-1: Research flowchart

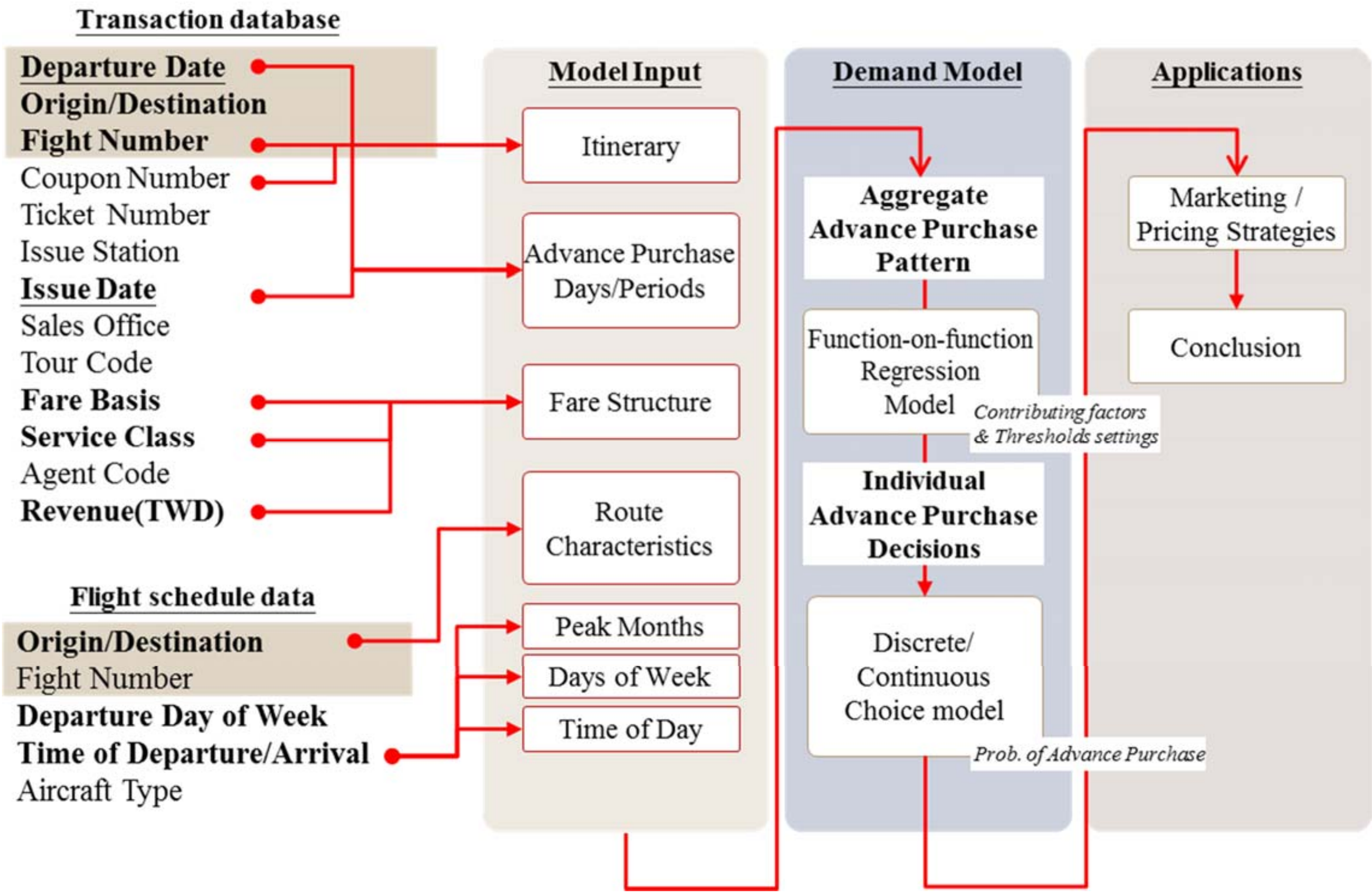


Figure 1-2: Modeling framework

Chapter 2 Literature Review

The aim of this chapter is to build a conceptual framework which explains the advance purchase patterns and its role in airline's decision process. First, a brief overview of advance purchase behavior analysis and applications on air transportation field are provided. Then, an extensive review on functional data analysis is presented, focusing on applications in pattern analysis. Finally, a comprehensive review in individual discrete and continuous choice model is presented, focusing on the setting and application related to this research.

2.1 Advance Purchase Behavior Analysis

The advance purchase behaviors of airline passengers have received increasing attention from researchers and marketing managers. In RM contexts, airlines employ different pricing strategies in response to intense market competition, differentiated demand patterns, and achieving effective customer segmentation.

Bilotkach et al. (2010) documented a set of stylized facts about price-setting dynamics across airlines who operating between New York City (NYC) and London (LON) area airports. A sample of daily fare quotes for non-stop travel from Expedia.com website were used. Their finding suggests that price-setting dynamics is indeed different across airlines on the research market. Airlines appear to employ strikingly different pricing strategies on this market. The estimated result also show that fares increase at an accelerated rate as the departure date approaches. In addition, the offered fares for last-minute travel were higher for the airlines with lower market share. Their conclusions may apply not only to the airline industry, but also to other markets for perishable products, such as hotels and car rentals markets that are characterized by fixed capacity and uncertain demand.

Escobari (2012) empirically investigated the dynamic pricing of inventories with uncertain demand as the departure date nears. In his study, a unique U.S. airlines panel data set was used. A dynamic pricing equation and a dynamic demand equation were jointly characterize with adjustment processes between prices and sales. The estimated results show that, at a fixed point prior to the departure date, the price increases as the inventory decreases. On the contrary, for a given inventory, the price decreases as there is less time to sell under, with breaks at 7 and 14 days to departure when price increases. Moreover, current decisions to price and purchase can be affected by prior realizations of fares and sales. Demand shocks have a positive and much larger effect on prices than the positive effect of anticipated sales. These findings are consistent with previous theoretical models of optimal pricing under uncertain demand and perishable inventories.

Deneckere and Peck (2012) formulated a dynamic model of perfectly competitive price posting under demand uncertainty. A dynamic trading model with uncertainty in demand and production in advance was developed. Information about aggregate demand is dispersed across different consumers, that resulting in information asymmetries among consumers, and between consumers and firms. Firms can dynamically adjust their prices to reflect information about aggregate demand extracted from observing the aggregate quantity sold, whereas consumers who have not yet purchased can decide whether to purchase or wait to try to purchase in later based on observed information from post history of sales and posted prices. The proposed model shown that high-valuation consumers purchase early and low-valuation consumers may delay their purchase decisions to exploit the option to refuse to purchase in the future if the sale price exceeds their valuation.

While airlines adjust prices dynamically based on learning demand from historical sale patterns, passengers can also decide to make advance purchase at the going price or to delay their purchase decision. Dynamic pricing strategies may decrease the value of the product and force passengers to make trade-offs between price, product attributes and deadlines, and therefore, change their advance purchase behaviors. In addition, airlines generally utilize advance purchase behaviors as a means of discriminating between passengers with high and low willingness-to-pay based on their time value and can improve the efficiency of seat allocation.

For example, by setting advance purchase discount, airlines can induce price-sensitive passengers to purchase tickets earlier and induce passengers who are less price-sensitive but more time-sensitive to make purchases later (Dana, 1999, 1998; Gale and Holmes, 1993). Dana (1998) examined a model in which consumers are heterogeneous in both their valuations and their demand uncertainty. The estimated results shown that consumers with more certain demands are willing to buy in advance because the presence of consumers with higher valuations and less certain demands could lead to an increase in prices. Advance-purchase sales were made to low-valuation customers, as results from traditional second-degree price discrimination model. Firms may use advance-purchase discounts or discriminatory pricing practices to affect the allocation of resources. Dana (1999) further demonstrates that when firms faced unforecastable demand fluctuations, equilibrium price dispersion could be used to efficiently shift demand and lower capacity costs even when the peak time is unknown, particularly when the costs to some consumers of changing departure day was high.

Escobari (2014) proposed a dynamic demand model with a panel dataset and analyzed how valuations change as the departure date nears. Their results support the claim that purchasing behavior changes as the departure date nears. They concluded that the lower valuations consumers become more price sensitive as their departure day approaches whereas high-valuation consumers tend to purchase earlier. Air passengers may sort themselves efficiently in equilibrium with low valuation travelers delay purchase decisions and even deciding not to buy if prices closer to departure are higher than their valuation.

Hotle and Garrow (2014) investigated how competitors' low-fare offerings influence the online search behavior of customers using unique website click stream data collected from the major carrier’s website. A truncated negative binomial model was proposed to predict the number of searches as a function of low-fare offerings for the same airport pair and competing airport pairs in the region. Clickstream data combined with detailed information about competitors' low-fare offerings for 10 directional markets was used to estimate the model. Their study found that the number of searches decreased as the difference between the carrier's lowest fare and competitors' lowest fare increased. The search intensity also increased as trip duration increased. In addition, trip characteristics had more impact on search behavior than the fare variables. The findings provide insight into the role of competitor pricing on multi-airport choice as it relates to customers' online search behavior.

Hotle et al. (2015) examined how passengers respond to advance purchase ticket deadlines and price uncertainties. They modeled the number of searches (and purchases) for specific search and departure dates. To correct price endogeneity problem, the instrumental variable was used. Results suggested that both search and purchase behaviors vary by search day of week, days from departure, lowest offered fares, variation in lowest offered fares across competitors, and market distance. Their works had showed that the number of searches increases just prior to an advance purchase deadline. The increase can be explained by passengers switching their desired departure dates to avoid higher fares after an advance purchase deadline.

2.2 Functional Data Analysis

To investigate and display the aggregated advance purchase patterns of air passengers, functional data analysis (FDA) techniques are used in this study. FDA is a collection of techniques in statistics for the analysis of curves or functions, which extends traditional statistical methods applications, such as functional ANOVA (fANOVA), functional principal component analysis (FPCA), canonical analysis, functional regression and functional clustering. The systematic overviews including Morris (2015), Ullah and Finch (2013), Ramsay and Silverman (2002, 2005, 2009) and Ferraty and Vieu (2006) had provided in-depth theoretical developments of functional data analysis, but also elaborated on the empirical applications in medicine, econometrics and biostatistics.

It has powerful ability of analyzing highly nonlinear and heterogeneous longitudinal data and is powerful in visualizing and capturing complex data patterns with a few simple measures (Dass and Shropshire, 2012). The techniques are appropriate for the data whose ideal units of observation are defined as functions on continuous domain. Under an FDA framework, each element is considered to be a function, the discretely collected observations are converted into functional curves/objects by specifying smooth basis functions before modeling and analysis. To date, FDA has better forecasting abilities than traditional models in dynamic environments (Dass, Jank, & Shmueli, 2011).

Wang, Jank and Shmueli (2008) applied functional data analysis to explain and predict the price of an ongoing online auction. In addition, the auction's price velocity and acceleration with other auction-related information were considered. To investigate the relationship between eBay's auction dynamics and other auction-related information, functional regression analysis was used. A novel set of Harry Potter and Microsoft Xbox data from eBay were applied to their proposed model. The estimated results had shown that the forecasting model based on functional data analysis outperformed traditional methods such as double exponential smoothing model, that do not take into account the dramatic change in auction dynamics.

Sood, James and Tellis (2008) proposed functional regression model to analysis and predict the market penetration of new products. Functional data analysis techniques including spline regression, functional principal components, functional clustering, and functional regression were applied. Several models including the Classic Bass model, Estimated Means, Last Observation Projection, a Meta-Bass model, and an Augmented Meta-Bass model were also proposed to compare its performance and predict eight aspects of market penetration. Data about market penetration from most of 21 products across 70 countries, for a total of 760 categories were used for model estimation. Results shown that functional regression model was superior to the abovementioned models. The product-specific effects were more important than country-specific effects when predicting penetration of an evolving new product.

Dass and Shropshire (2012) demonstrated FDA techniques on firm performance measures. Methodologies including functional principal component analysis, functional regression, and functional clustering were used to investigate measures of firm financial performance based on panel data set of the 1,000 largest U.S. firms by revenues from 1992 to 2008. The FDA techniques were adopted to explore how these measures vary across firms over time, common trends or factors across performance of all firms, the effect of various measures of firm size on these performance measures, and clusters of firms based on the dynamics of performance measures. Finally, the forecasting results obtained from FDA were compared with hierarchical linear modeling and the FDA-based forecasting model had better accuracy than the model based on HLM.

In terms of pattern analysis, Gastón et al. (2008) illustrated how FDA can be used in the simulation of time-varying arrival processes. The study focused particularly in the estimation of the cumulative mean function of a non-homogeneous Poisson Process (NHPP). The arrival processes are usually seen as discrete processes that can be described by using appropriate stochastic point processes. The dataset of observed arrival times of patients to the primary health center during 150 days was used. Functional principal component analysis and functional ANOVA methods were applied to estimate it from observed independent streams of arrival times. The results exhibited that FDA provides a useful framework for studying problems related with non-homogeneous Poisson process.

Gao and Niemeier (2008) investigated daily patterns for diurnal ozone and nitrogen oxides cycles, their interrelationships, and the multilevel spatio-temporal variations of ozone and nitrogen oxides measurements from Southern California. Functional data analysis techniques take account for the continuous nature of diurnal ozone/nitrogen oxides processes by converting discrete observed values into functional diurnal curves. Representative summer diurnal ozone profiles are constructed using functional clustering. Variability in hourly distribution of traffic activities and emissions is also discussed. The results provide valuable insights for identifying optimal transportation emissions control strategies.

Chiou (2012) presents a methodological framework for uncovering traffic flow patterns and prediction. Functional data techniques were applied for classification and prediction of traffic flow pattern, identify clusters with similar traffic flow patterns, facilitating accurate prediction of daily traffic flow. The methodology not only assist in predicting traffic flow trajectories, but also identify distinct patterns in daily traffic flow of typical temporal trends and variabilities. The empirical results shown that the proposed functional mixture prediction model can work reasonably well to predict traffic flow.

Chiou (2014) studied the missing values and outliers problems that frequently encountered in traffic monitoring data. The missing values were imputed by sampling the daily traffic flow rate trajectories from random functions using the conditional expectation approach to functional principal component analysis (FPCA). Based on the FPCA approach, the FPCA scores can be applied to the functional bagplot and functional highest density region boxplot, which makes outlier detection possible for incomplete functional data. The simulation study had shown that the FPCA approach performs better than two commonly discussed methods in the literature, the probabilistic principal component analysis and the Bayesian PCA. The proposed functional data methods for missing value imputation and outlier detection can be used in many applications with longitudinally recorded functional data.

Guardiola and Mallor (2014) analysis the daily traffic flow profiles based on the employment functional data techniques. 1-min traffic data from the I-94 Freeway in the Twin Cities, Minnesota (U.S.) metroplex ranging from 2004 to 2011 was used. To clustering recognized traffic patterns and also to identify outliers (bad performance in the recording of data or special circumstances that affected the traffic), functional principal component analysis model was proposed. In addition, multivariate control charts were adopted to monitor the daily flow traffic pattern over time and to be able to recognize major changes in the pattern's behavior. The functional analysis allows a maximum exploitation of the recorded historical data in daily traffic flow monitoring that would otherwise be difficult to detect via classical statistical methods.

Tastambekov et al. (2015) studied the short to mid-term aircraft trajectory prediction problem, which is crucial to conflict detection and resolution algorithms of Air Traffic Management (ATM) applications. To predict where an aircraft will be located over a 10–30 min time horizon, a local linear functional regression model was proposed. The validation of the proposed model had been strengthening with extensive simulations. In addition, a learning process had been used to adjust parameters. The approach considered data preprocessing, localizing and solved by using wavelet decomposition. One year of historical trajectories records between airports over France was used. The estimation shown efficient results with high robustness.

Jamaludin and Zulkifli (2016) studied spatial and temporal variabilities of rainfall patterns for 32 rainfall observation stations in the East Peninsula Malaysia using functional data analysis. Functional concepts such as functional descriptive statistics and functional analysis of variance were applied to describe the spatial and temporal rainfall fluctuations at the stations and at any time throughout the year. The discretely collected rainfall records of 32 stations for 32 years were used for estimation. The estimated results suggested that the rainfall profiles of studied regions were very dependent on their geographical and spatial locations of the regions, as well as the monsoon effect, which reflects the months of the year.

2.3 Revealed Preference Discrete Choice Models

To trace the individual advance purchase decision of passengers, recent researches have introduced choice models to revenue management for its ability to accommodate passenger preferences in RM strategies. The approach supports RM decisions by replacing typical demand forecasting models of probability and time-series models with models based on discrete choice theory. Though demand models based on choice models may be more appropriate in RM applications, empirical studies are limited due to the high cost of data acquisition.

In terms of passenger choice modeling, the decision process of passengers can be modeled with either stated preferences data or using the revealed preferences data. The stated preferences approach is estimated with dataset that collected through designed scenario surveys, whereas the revealed preferences approach is typically based on the real booking/transactions. One advantage of using revealed preferences data is that transaction data provide a direct record of the actual choices of air passengers and are easily collected by airlines. Using revealed preferences data also avoids the risk of response bias from the questionnaire surveys associated with the hypothetical nature of stated preference data (Carrier, 2008). In addition, with the growing online purchasing, to estimate and predict the advance purchase behaviors of flights directly based on transaction data is much more intuitive, cost economic, and representative.

However, for revealed preference model implementations, both chosen and non-chosen alternatives are needed to replicate the purchase scenario. Although the support of computer systems have reduced the cost of data collection, most firms only record data for passengers who had decided to purchase and information about non-chosen alternatives had been difficult to obtain. Therefore, inferring the true demand with available data remains a challenge issue.

Previous studies have used logit models of demand to analyze advance purchase behavior based on revealed preferences data for the airline industry (Escobari and Mellado, 2014; Vulcano et al., 2010; Carrier, 2008), hotel (Newman et al., 2014) and railway industry (Hetrakul and Cirillo, 2015, 2014, 2013). Within the airline industry, Talluri and Ryzin (2004b) analyzed a single-leg, multiple-fare-class RM problem under a general discrete choice model of demand. The choice model specifies the probability of purchasing each fare product as a function of the set of available fare products. The model is then incorporated into objective function of capacity allocation problem. To estimate choice models when no-purchase data are unobservable, an additional estimation procedure based on the expectation-maximization (EM) was also developed. Results of a simulation study were provided to compare choice-based method to a traditional single-leg method and shown a significant improvement in revenue.

Carrier (2008) modeled time-of-travel choice for airline travelers based on the latent class model with booking and seat availability data from Amadeus database. The choice set for each booking was reconstituted from data for booking, fare rules, and seat availability. To date, to represent time as a continuous variable, a trigonometric function was used. Estimation results of 2,000 bookings from three European short-haul markets shown that the latent class model and a continuous function of time led to a significant improvement in the fit of the model compared to previous models based on a deterministic segmentation of the demand and time-period dummies. This research had extended the scope of potential applications of passenger choice models to airline planning decisions such as pricing and revenue management.

Vulcano et al. (2010) proposed a choice-based RM model with readily available airline data such as data for flight schedules, revenue accounting, seat availability and screen scrape (sample information about alternatives and fares offered by competitors at different points in time during the booking horizon). To exploit passenger preferences, a single-segment MNL model was constructed. To account for unobservable data, a maximum likelihood estimation algorithm that uses a variation of the expectation-maximization method was developed. The selected market was New York City to Florida. The estimated results were then used in a simulation study to assess the revenue performance of the EMSR-b (expected marginal seat revenue, version b) capacity control policies. Their simulation result showed significant improvements (1%–5%) in average revenue in the tested markets.

Hetrakul and Cirillo (2013, 2014) applied multinomial logit, latent class, and mixed logit models to investigate heterogeneous characteristics of railway passenger behavior that differ by the length of haul based on internet booking data with limited individual variables. Their analysis quantifies the importance of fare, advanced booking, departure time of day, and day of week in purchase timing decision. They found the latent class model is found to be superior to mixed logit model in term of prediction capability. The empirical result shows that segmenting price sensitivity by booking period is more appropriate than by socioeconomic information. They further delivered RM optimization result that shows revenue improvement from 16.24% to 24.96% in multinomial logit models and from 13.82% to 21.39% in latent class models respectively.

Escobari and Mellado (2014) empirically studied advance purchase behaviors of air tickets in a dynamic setting. To date, it is the first study that estimates the itinerary choice (i.e., flight choice) in a revealed preference setting where information on choices and all the alternative flights is available. Their dataset included detailed data for contemporaneous prices and for characteristics of both chosen and non-chosen flights. The data contained all the 317 flights from the six carriers that served between New York City (Newark Liberty, John F. Kennedy, and La Guardia) and the main airport in Toronto (Toronto Pearson International) from December 19 to 24, 2008. The estimated results shown that in a 100-seat flight a 10% increase in prices throughout a 100-period selling season reduced quantity demanded by 7.7 seats. The quantity demanded was more responsive to prices for departures in the morning and evening when compared to departures in the afternoon.

2.4 Continuous Choice Models

Despite the many advantages of discrete choice models, choices response variables such as advance purchase time, departure time, and location are continuous and must be transformed to discrete alternatives for the model estimation. To directly discretize those continuous variables may suffer from some limitations. For example, the discretization for discrete choice models is usually based on different research purposes to divide the study period into a limited number of intervals. The interval settings usually do not cover the entire study period and to cause loss of temporal resolution. Additionally, two points close in time are likely to be perceived as similar alternatives by passengers but may be possibly misclassified into two discrete time periods. Although some studies have calculated correlations among alternatives, continuous treatment of time variable seems more intuitive and preferable (Bhat and Steed, 2002). Furthermore, different discrete interval settings would also lead to different and unstable estimation results (Chiou and Liu, 2016a).

The continuous logit model was first proposed and applied in location choice model (Ben-Akiva et al., 1985; McFadden, 1973). The method provides continuous setting grounded in random utility theory and that retains the key advantages of measuring the utilities. To model a continuous time-of-travel, Abou-Zeid et al. (2006) modeled tour-based time-of-day choice using multinomial logit models that consist of arrival time choice, departure time choice, and both. The model was estimated by using household activity survey data and transportation level-of-service information from the highway network. In addition, 35 time periods consisted of 33 half-hour intervals and two extreme periods of longer duration were used to model the choice process. Continuous utility functions including continuous arrival time functions, departure time functions, and duration functions as well as variables interacted with these functions were proposed. The test applications shown that the developed tour-based time-of-day modeling procedure during this research worked.

Popuri et al. (2008) developed time-of-day models framework for Tel Aviv tour-based model system of Israel's Ministry of Transport. A joint model of arrival and departure time choice with continuous trigonometric functions was presented. Instead of alternative-specific variables and constants settings, continuous trigonometric functions of arrival and departure time were proposed. The setting allowed for smooth and cyclic arrival and departure time profiles for each market segment. To predict door-to-door travel speeds and times for arrival or departure in each time slots, an additional regression methodology was used. The estimated results shown that the proposed modeling framework that using the commonly available household survey data and some basic level-of-service data could provide more estimation details and better suited for policy testing.

Lemp and Kockelman (2010) empirically investigated departure time choices with continuous logit model using Bayesian estimation techniques. The home-based work tour departure time had been modeled in a continuous fashion. Additionally, ordinary least squares regression models were used to estimate travel times and their variance across times of day for the auto and transit modes. These network variables were used to inform estimation of the continuous logit model of departure time. The model was estimated on work tour data from the 2000 San Francisco Bay Area Travel Survey (BATS). The estimated results were reasonable and meaningful for multiple applications. Their work can be extended to a two-dimensional choice construct so that the departure and return times can be modeled simultaneously.

Lemp et al. (2010) further extended the method to a continuous cross-nested logit model (CCNL). The model resulted from generalizing the discrete cross-nested logit model (CNL) for continuous response settings and formulated to come from the generalized extreme value (GEV) class of models. Bayesian estimation techniques and San Francisco Bay Area data were used for parameter estimation. The CCNL model also conforms to random utility theory, which offers a strong theoretical basis by offering the economic welfare implications for evaluation of policy alternatives. The empirical results showed that the model predictions were very similar to those obtained by the continuous logit model, the CCNL performs better in terms of out-of-sample prediction. In addition, the proposed CCNL model allows a more flexible choice behavior to emerge.

Ben-Akiva and Abou-Zeid (2013) addressed methodological issues that arise when modelling time-of-travel preferences. Three approaches including unequal period lengths, schedule delay and the 24-hour cycle were reviewed. These methodologies were then applied in a tour-based travel demand model and estimated with the 2000 Bay Area travel survey dataset that included survey data of 36,680 individuals from 15,064 households. The estimated models tested with various scenarios such as highway and transit improvements and congestion pricing. The estimated results showed that the time-of-travel distributions were reasonable. The peak spreading was observed when congestion levels increased. Additionally, the time-of-travel distributions predicted by the model for a baseline scenario compared favorably with the observed patterns.

However, none of abovementioned models have ever been used to analyze the airline industry or advance purchase behavior. Empirical studies of how departure time preferences of air passengers affect advance purchase behaviors are also limited. The market conditions today have more complicated the purchase decisions process for air passengers. Without a clear understanding of the advance purchase behavior, the potential benefits of RM may be limited.

2.5 Review Summary

Based on the literature review, this section summarizes major studies related to this research and identifies promising research directions for this dissertation. Table 2-1 lists major studies focusing on pricing dynamics of airlines in revenue management applications and advance purchase behavior changes of air passengers. As airlines learning sales patterns from existing stock changes, airlines can dynamically adjust their sales and marketing strategies. Moreover, by setting advance purchase discounts, airlines can further shift demand effectively. On the other hand, air passengers can choose to purchase at ongoing price based on their preferences, or wait for the acceptable price, or decide to drop the deal. Based on the behavior dynamics, this research aims to empirically analyze factors influencing advance purchase behaviors of air passengers and provide empirical evidences to support existing decision theory. Table 2-2 further summarizes major researches applying functional data analysis where the framework is relevant to this research. Function data analysis techniques including parameter smoothing and functional regression model were used to explore flight advance purchase patterns based on the shape of the advance purchase curves of each flight. Table 2-3 represents major studies of discrete choice models with respect advance purchase behaviors, whereas Table 2-4 lists the continuous choices models related to time-of-travel choice and purchase timing decisions. Compared with previous studies that introduced time of purchasing as explanatory variables into choice model, the advance purchase days are seen as the response variable (alternatives) in this dissertation. To explore the extent to which choice set construct scheme based on the time of purchase that can estimate choice model well, discrete choice models were used. A continuous logit model was further proposed to account for the continuous nature of advance purchase time. These tables also provide comparison for the studies proposed in this dissertation to the existing studies reviewed.

Table 2-1: Summary of studies related to advance purchase behaviors

Authors	Topic	Behavior changes
Bilotkach et al. (2010)	Price-setting dynamics	· Price-setting dynamics is different across airlines and fares increases at an accelerated rate as the departure date approaches.
Escobari (2012)	Dynamic pricing with uncertain demand	· Current decisions to price and purchase can be affected by prior realizations of fares and sales.
Deneckere and Peck (2012)	Dynamic model of perfectly competitive price posting under demand uncertainty	· High-valuation consumers purchase early and low-valuation consumers may delay their purchase decisions or refuse to purchase.
Dana (1998)	Advance-purchase discounts and price discrimination in competitive markets.	· Firms may use advance-purchase discounts or discriminatory pricing practices to affect the allocation of resources.
Dana (1999)	Using yield management to shift demand when the peak time is unknown	· Equilibrium price dispersion could be used to efficiently shift demand and lower capacity costs, particularly when the costs of changing departure day was high for customers.
Escobari (2014)	How valuations change as the departure date near?	· Lower valuations consumers become more price sensitive as their departure day approaches whereas high-valuation consumers tend to purchase earlier.
Hotle and Garrow (2014)	How competitors' low-fare offerings influence the online search behavior?	· The number of searches decreased as the difference between the carrier's lowest fare and competitors' lowest fare increased.
Hotle et al. (2015)	How passengers respond to advance purchase deadlines and price uncertainties?	· The number of searches increases just prior to an advance purchase deadline.

Table 2-2: Summary of researches applying functional data analysis

Authors	Application	Functional data analysis techniques
Wang, Jank and Shmueli (2008)	Predict the price of an ongoing online auction	· Functional regression for a functional response variable
Sood, James and Tellis (2008)	Predict the market penetration of new products.	· Functional principal components analysis · Functional clustering · Augmented Functional Regression model
Dass and Shropshire (2012)	Firm performance measures	· Functional principal component analysis · Functional clustering · Functional regression for a functional response variable
Gastón et al. (2008)	Estimation of the cumulative mean function of a non-homogeneous Poisson arrival Process (NHPP)	· Functional principal component analysis · Functional ANOVA
Gao and Niemeier (2008)	Daily patterns for diurnal ozone and nitrogen oxides cycles,	· Functional principal components analysis · Functional clustering
Jeng-Min Chiou (2012)	Traffic flow patterns and prediction	· Functional mixture prediction · Functional clustering and discrimination
Jeng-Min Chiou (2014)	Missing values and outliers problems in traffic monitoring data	· Functional principal components analysis
Guardiola and Mallor (2014)	Daily traffic flow monitoring	· Functional principal components analysis
Tastambekov et al. (2015)	Short to mid-term aircraft trajectory prediction problem	· Local linear functional regression model
Jamaludin and Zulkifli (2016)	Spatial and temporal variabilities of rainfall patterns	· Functional descriptive statistics · Functional ANOVA

Table 2-3: Summary of revealed preference discrete choice models

Authors	Application	Model	Data	Choice-set
Talluri and Ryzin (2004b)	single-leg, multiple-fare-class RM problem	Multinomial logit choice model (MNL)	transaction data	Fare products
Carrier (2008)	time-of-travel choice for airline travelers	latent class model	booking and seat availability data from Amadeus database	Fare products
Vulcano et al. (2010)	choice-based RM model	Multinomial logit choice model (MNL)	flight schedules, revenue accounting, seat availability and screen scrape data	Fare products
Hetrakul and Cirillo (2013, 2014)	purchase timing decision of railway passenger	Multinomial logit, latent class, and mixed logit models	internet booking data with limited individual variables	31 alternatives, from 30 days before departure to departure day.
Escobari and Mellado (2014)	advance purchase behaviors of air tickets in a dynamic setting	Multinomial logic choice model (MNL)	transaction data from online travel agency Expedia.com	All of available fare products across airlines

Table 2-4: Summary of continuous choice models

Authors	Application	Model	Data	Choice-set
Abou-Zeid et al. (2006)	tour-based time-of-day choice	Multinomial logic choice model (MNL) with continuous utility function.	Household activity survey data and transportation level-of-service information from the highway network.	35 time periods (33 half-hour intervals and two extreme periods of longer duration)
Popuri et al. (2008)	time-of-day models	Multinomial logic choice model (MNL) with continuous utility function.	Israel National Travel Habits Survey (NTHS) conducted in 1996	666 alternative corresponding to the arrival and departure time combinations [36 × (36 + 1)/2]
Lemp and Kockelman (2010)	departure time choices	Continuous logit model using Bayesian estimation techniques	Work-tour data from the 2000 San Francisco Bay Area Travel Survey (BATS)	Continuous departure time
Lemp et al. (2010)	departure time choices	Continuous cross-nested logit model (CCNL)	Work-tour data from the 2000 San Francisco Bay Area Travel Survey (BATS)	Continuous departure time
Ben-Akiva and Abou-Zeid (2013)	Methodological issues that arise when modelling time-of-travel preferences.	Continuous logit model	2000 San Francisco Bay Area Travel Survey (BATS)	Continuous departure time

Chapter 3 Models

After reviewing related references, this chapter aims to present a model for identifying advance purchase patterns and exploring their characteristics, and to propose an approach for examining the advance purchase behaviors of air passengers. The remainder of this section is organized as follows. First, a functional concurrent regression model is presented in Section 3.1 for the advance purchase patterns of specific type of flights. To explore the extent to which choice set construct scheme based on advance purchase time that can estimate choice model well, the discrete choice model is presented in Section 3.2.1. Meanwhile, the continuous choice model is described in later Section 3.2.2.

3.1 Aggregate Pattern Model

Although the daily transaction data used in this study is discretely collected, the true nature of the data is continuous. Instead of considering the observed transactions as discrete values, the data is treated as a finite curve over a time period and further examined with a functional linear model. The approach provides a mean to investigate the effects of potential contributing factors on the shape of abovementioned advance purchase curves.

For functional linear models, there are three scenarios for response y_i and explanatory variables x_i. Functional linear models can be functional in one or both of the response variable y with argument t is functional and one or more of the independent variables x is functional. Table 3-1 summaries three types of functional linear models and corresponding relationship between response and explanatory variables. In this study, a function-on-function regression model is used to investigate advance purchase patterns of air passengers and predict the future pattern. Function-on-function regression model relates to a smoothed functional response variable, $y(t)$, the known independent variables, one or multiple functional explanatory variable $z(t)$ and multiple explanatory indicators x, by a linear combination of parameter functions. The simpler case of function-on-function model is also called *concurrent model*, where the value of the response variable $y(t)$ is predicted only by the values of one or more functional response variable at the same time t (Ramsay and Silverman, 2005).

Table 3-1: Function linear models

Response variable	Explanatory variables *Scalar*	*Function*
Scalar	linear models	functional predictor regression (scalar-on-function)
Function	functional response regression (function-on-scalar)	function-on-function regression

In this study, a concurrent model is proposed to predict the advance purchase patterns after seven days (1 week after). Assume that an observation interval t of 0~60 days (about 9 weeks), the time-varying cumulative daily transaction pattern for the flight j is given as Eq. (1):

$$y_j(t) = \beta_0(t) + \sum_{i=1}^{I} x_{ij}(t)\beta_i(t) + \varepsilon_i(t) \tag{1}$$

where $y_j(t)$ is a functional response of predicted load factors after 1 week, whereas $x_{ij}(t)$ consist of a functional explanatory variable of historical load factor (7 days before current day) and multiple categorical indicators related to flight characteristics such as time of day (MORNING and AFTERNOON), day of week (FRIDAY), month of year (PEAK.SEASON) and VACATION indicators. $\beta_j(t)$ are coefficients to be estimated. The first stage of FDA is to represent the discrete observe proportion into a functional form by a suitable basis function/system. Therefore, the variables are required to be specified with selected basis system to define the function in advance for modeling. The basis system is a linear combination of basis functions $\phi_k(t)$ as presented in Eq. (2).

$$x_i(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{2}$$

The basis functions $\phi_k(t)$ are assumed mathematically independent of each other and so that any function can be approximated arbitrarily well by a weighted sum of a large number K of basis functions (Ramsay and Silverman, 2005), whereas c_k refers to the estimated basis coefficient. The basis functions are required to have features as close as possible to the data, so that an accurate representation of the function can be obtained with only a few basis terms (Clarkson et al., 2005). For example, the Fourier basis system is the common choice for periodic data, whereas the B-spline generally serves well for non-periodic functions. The wavelet basis is suitable for sparse dataset, which is particularly good in presenting jumps, spikes or peaks in estimating data. The most commonly used basis system of FDA are Fourier basis and B-spline basis, while other basis are available.

A linear combination of K number B-spline basis is used to represent the curves for open-ended transaction data and a more flexible fitting in this study. The B-spline basis function is a piecewise polynomial function of order p, with the interior breakpoints (knots) at $t_1, t_2, \ldots, t_b$. The number of B-spline basis functions K can be determined by the relation of order of the polynomials plus number of interior breakpoints. Note that the observation time interval t of interest of 0 to 60 days is broken into 10 equally spaced sub-intervals and within each interval a polynomial of order 4 is employed. Based on this, the total number of B-spline can be settled with K=13. Here, the order of the B-spline is settled by the default value of our program as 4, meaning that piecewise cubic polynomials will be used. Although it is possible to use different order of polynomials in each sub-interval, in our study we will keep the order constant. To estimate the basis coefficient, the regression smoothing method is used. The polynomial smoothing spline may result in a potentially better fit but usually tends to have a poorer recovery of the underlying trend (Ramsay et al. 2009). Thus, to avoid potentially overfitting problems, the roughness penalty is incorporated into the least square criterion for a finer control over the amount of smoothing. A measure of a function's roughness (i.e. total curvature) is defined by the integrated squared second derivative and an additional smoothing parameter λ is also specified to control the degree of curvature penalty. The penalized sum of squares (PENSSE) is given as follows:

$$PENSSE = SSE + \int [D^m x_{ij}(t)]^2 dt \tag{3}$$

As shown in Eq. (3), m represents the mth derivation of the function $x_{ij}(t)$. In this study, the second derivative (m = 2) is used. When the roughness is included in the fitting process, the goal is to find a function that minimizes the penalized residual sum of squares error (PENSSE). In addition, the generalized cross-validation (GCV) criterion (Craven & Wahba 1979) is used to control the roughness of the estimate to prevent curve over-fitting and determine the best value for smoothing parameter λ. The criterion is defined by:

$$GCV(\lambda) = \left(\frac{n}{n - df(\lambda)}\right)\left(\frac{SSE}{n - df(\lambda)}\right) \tag{4}$$

The criterion shows a twice-discounted mean square error measure, where the degrees of freedom are controlled by λ. The left factor discounts this estimate by multiplying by $n/(n-df(\lambda))$. The right factor is the unbiased estimate of error variance familiar in regression analysis, and represents discounting by subtracting $df(\lambda)$ from n. The proposed FDA model is then estimated by R "fda" package and selected in terms of the GCV index. The estimated results are later described in Chapter 5. In order to evaluate the prediction accuracy of the proposed model, the mean absolute percentage error (MAPE) based on Lewis (1982) is employed and described as Eq. (5).

$$\mathrm{MAPE}(t) = \frac{1}{n}\sum_{j=1}^{J}\left|\frac{y_j(t) - \hat{y}_j(t)}{y_j(t)}\right| \tag{5}$$

where $\hat{y}_j(t)$ is the forecasted output at time t for the flight j and n represents total number of transaction records. The smaller the MAPE value, the less forecast error and thus the more accurate is the forecasted result. Based on MAPE and applying Lewis's scale, provides some framework as shown in Table 3-2 to judge the model.

Table 3-2: A scale of judgment of forecast accuracy (Lewis, 1982)

MAPE value	Judgment of Forecast Accuracy
Less than 10%	highly accurate
11 to 20%	good forecast
21 to 50%	reasonable forecast
51% or more	inaccurate forecast

3.2 Individual Choice Model

To further investigate the individual advance purchase behaviors of air passengers, choice models are used. However, our numerical dataset not only contains transaction records from both direct purchasing passengers, but also from diverse distribution channels, including travel agencies, direct Internet sales and airline counters. It is hard to distinguish individual passenger behaviors from other purchasing channels. Therefore, for individual advance purchase behavior modeling, this study considers only the subset of transactions that had been made through the direct purchasing channel (website and airline counters).

3.2.1 Discrete Choice Model

To study the purchase timing decision of advance purchase behaviors, the advance purchase days was defined in this study as days between tickets issued date and departure date. Noted that previous studies had introduced time of booking/purchasing as explanatory variables into choice model with developed different segmentation schemes (Carrier, 2008; Garrow, 2012; Hetrakul & Cirillo, 2013, 2014). Compared with those approaches, here we take the advance purchase days as the response variable (alternatives) that the goal of our proposed model is to obtain the probability of advance purchasing in each time segment. In order to reduce the number of alternatives and facilitate model development, the advance purchase horizon is further divided into five time periods according to three segmentation methods: the first method is to divide the horizon into five equal time periods (each period is of 12 days). The second method is to divide the horizon into five time periods with equal number of purchases. Finally, the third method is to divide the horizon according to subjective judgment from experience managers of the study airline.

This study models the advance purchase behaviors in static settings and from airline perspective; hence we assume that all five purchasing period alternatives are available to passengers at the same time under perfect information. Additionally, since our data contains transactions data of only one carrier, it is not able to account for the choices of other flights or carriers. The settings here only consider the choice of advance purchase period within the same flight. However, for transaction data, only the time period of successful transaction was recorded and none of information regarding unsuccessful transactions in other time periods was available. A data-intensive method is used to impute the values of generic variables of other time periods. The purchase prices in the other periods of flight *j* were imputed by advance purchase days, service classes and purchase months. Passengers are assumed to be myopic that they purchase at the price whenever their valuation exceeds it. Finally, factors including price, and travel time preferences (time of day, day of week, and months of year) are then further examined by using multinomial logit (MNL) model.

As presented in Eq. (6), the utility of passenger *n* who purchased at the alternative advance purchase period *i* for flight *j* is given by,

$$U_{nij} = \alpha_{nij} + \beta \log(PRICE)_{nij} + \gamma \log(PRICE)_{nij} \times x_{nj} + \delta x_{nj} + \varepsilon_{nij} \tag{6}$$

Here, the systematic component part of utility is modeled a linear function of observed characteristics, $V_{nij} = \alpha_{nij} + \beta \log(PRICE)_{nij} + \gamma \log(PRICE)_{nij} \times x_{nj} + \delta x_{nj}$, whereas the unobserved random component is expressed as ε_{nij}. β, γ and δ are the coefficients to be estimated. In Eq. (5), α_{nij} is the alternative specific constant for the alternative i, $i \in Cn \in \{1, \dots, 5\}$ which captures the average effect on utility of all variables that are not included in the model. The $\log(PRICE)_{nij}$ and its interaction terms of travel time preferences are settled as *generic variables*, that the marginal effect of the variable is assumed to have same impact on the utility of each alternative. The interaction term specification is helpful to account for the relationships between purchase price and flight preferences. Notably, if the carrier learns more about the demand as departure day approaches and dynamically adjust price strategies, the correlation between $\log(PRICE)_{nij}$ and the unobserved $x_{nj} + \varepsilon_{nij}$ may cause potential price endogeneity problem. Escobari (2012, 2014) controlled for the potential endogeneity with internal instruments and flight fixed effect. Since the price dynamic is not the current issue of this study, we assume ε_{nij} is uncorrelated with $\log(PRICE)_{nij}$. Finally, the probability P_{nij} of passenger n choosing advance purchase period i can be derived as Eq. (7).

$$P_{nij} = \frac{e^{V_{nij}}}{\sum_{k \in C_n} e^{V_{nik}}} \tag{7}$$

The selected explanatory variables are purchase price (in logarithmic form) and flight schedule preferences x_j. Flight specific attributes such as morning flight (MORNING), flights on Friday (FRIDAY), flights in peak months including July and August (PEAK.SEASON) according to flight schedule database are treated as alternative specific variables to capture the time of day, days of week and month of year preferences of air passengers. The setting allows us to observe the marginal effect of flight preference changes across advance purchase periods. Furthermore, to identify passengers who often travel around consecutive holidays (more than three days) and special vacation such as Chinese New Year and spring vacation, are also marked as vacation (VACATION) tourists.

3.2.2 Continuous Choice Model

The discrete choice models provide a methodology for tracing individual decision making processes and for profitably exploiting their preferences for product attributes. However, some choices are continuous response variables such as advance purchase time, departure time, and location. Arbitrarily discretizing these continuous choices variables may lead to an erroneous result. This section further present continuous choice model, which takes advance purchase time alternatives as continuous. The continuous logit model is advantageous at treating advance purchase time in a continuous fashion and offers strong theoretical supports based on the random utility framework without discretizing the decision time horizon.

The continuous choice model represents a generalization form of the MNL for continuous response variable settings and can be derived directly from the random utility theory. Here, assume a passenger n purchasing a ticket for flight j at the advance purchase time t, where the continuous advance purchase time period of interest z is bounded by b_1 and b_2. For the discrete choice model, the g^{th} discrete alternative of z can be presented as t_g (g = 1, 2, ..., G). Suppose the number of discrete advance purchase period alternatives K is defined as G=1+[(b_2 - b_1)/s], where s denotes the distance of each alternative. If the limit of distance is close to 0, such that $s \rightarrow 0$, the continuous choice probability function can be written as Eq. (8).

$$P_{nj}(t)=\frac{\exp\left(V_{nj}(t)\right)}{\int_{b1}^{b2}\exp\left(V_{nj}(z)\right)dt} \tag{8}$$

The utility of passenger n who purchased at the continuous advance purchase time t for flight j can be expressed as the sum of a deterministic part of the utility $V_{nj}(t)$, and an unobserved random component $\varepsilon_{nj}(t)$ as discrete choice models:

$$U_{nj}(t)=V_{nj}(t)+\varepsilon_{nj}(t) \tag{9}$$

Here, since the utility is presented as continuous form in advance purchase time, the explanatory variables that do not vary over time alternatives (e.g., departure time preference of an individual) are specified by sinusoidal functions of advance purchase time interacted with variables. This study models the advance purchase behaviors in static settings and from airline perspective. All purchase time alternatives are assumed to be available to passengers at the same time under perfect information. Passengers are assumed to chosen flights already and only have to make their purchase decision at particular time t based on their preferences. The deterministic part of the utility is given as Eq. (10) and the sinusoidal function used in our model is presented as Eq. (11).

$$V_{nj}(t)=\alpha PRICE_{nj}(t)+\beta X_{nj}\gamma(t) \tag{10}$$

$$\gamma(t)=\left[\sin\left(\frac{2\pi t}{T}\right),\sin\left(\frac{4\pi t}{T}\right),...,\sin\left(\frac{2L\pi t}{T}\right),\cos\left(\frac{2\pi t}{T}\right),\cos\left(\frac{4\pi t}{T}\right),...,\cos\left(\frac{2L\pi t}{T}\right)\right]' \tag{11}$$

As presented in Eq. (10), the time-varying variables (i.e. purchase price) are presented as $PRICE_{nj}(t)$, whereas X_{nj} represents the time-invariant flight preferences of the flight j for passenger n. Flight specific attributes such as morning flight (MORNING and AFTERNOON), flights on Friday (FRIDAY), flights in peak months including July and August (PEAK.SEASON) according to flight schedule database and vacation (VACATION) are incorporated to capture departure time preferences of air passengers. The setting is helpful to observe the marginal effect of flight preference changes across advance purchase time. To further consider the price uncertainties that passengers may likely face while making purchase decisions, the coefficient of variation (CV) of observed historical prices across flights within the purchasing month is also incorporated, which reflects potential price fluctuation caused by airline promotions. If price uncertainties were higher, passengers may wait for a better price and delay their purchase. α and β are the coefficients to be estimated.

In the Eq. (11), $\gamma(t)$ is a collection of sine and cosine functions of advance purchase period of T (b_1 to b_2). The model specification allows the utility function to take on a variety of shapes and reflect the passenger preference variations over time. Additionally, some variables may be interacted with fewer than $2L$ sinusoidal functions, similar settings can be seen in Abou-Zeid et al. (2006), Popuri et al. (2008), and Lemp et al. (2010a, 2010b). Notably, the correlation between $PRICE_{nj}(t)$ and the unobserved $X_{nj}+\varepsilon_{nj}(t)$ may cause potential price endogeneity problem if the airline learns more about the demand as departure day approaches and dynamically adjusts price. Some researches had controlled the potential endogeneity with internal instruments and flight fixed effect (Escobari, 2012, 2014a). This study treats $PRICE_{nj}(t)$ as endogenous and assumes $\varepsilon_{nj}(t)$ is uncorrelated with price.

Unfortunately, the study dataset only contains transactions data of one airline; therefore, it is not able to account for choice behaviors among airlines. Additionally, since the transaction dataset only records successful transactions, to obtain time-varying prices for continuous logit model estimation, an additional ordinary least squares regression model is constructed. While some airlines have applied dynamic pricing strategy to change fares from time to time, others are typically accounting for designed fare classes and restrictions and remaining capacity of competitors in the market. Several factors influencing the price settings of the study airline were estimated by a regression model, as shown below:

$$Price_j(t) = \varphi x_j + \left(\sum_{k=1}^{2} \omega \left(\gamma'(t) \right)^h \right) * SEATS_j + \varepsilon_j \tag{12}$$

where x_j represents the explanatory variables for the flight j, such as the maximum stay limitation (YEE1M), fare classes, and the number of remaining seats ($SEATS_j$). To further described how remaining capacity varied over time t, a number of sinusoidal functions $\gamma'(t)$ were also introduced to the regression model. The setting here were aimed to capture the effect of the number of remaining seats on airline pricing settings. Passengers are assumed to be myopic that they purchase at the price whenever their valuation exceeds it. Finally, φ and ω are the coefficients to be estimated.

Chapter 4 Data

Prior to estimates of the abovementioned models, this chapter presents a numerical analysis of available datasets used for model estimations in this research. The analysis focuses on elements of historical air ticket transaction data for proposed aggregate advance purchase pattern model and individual advance purchase choice model proposed in the previous chapter. The analysis focuses on exploring the distribution of passenger demand, advance purchase patterns, and purchasing price.

4.1 Numerical Dataset

To empirically investigate key factors contributing to advance purchase patterns, the ticket transaction data of Taipei-Macau (TPEMFM) route in 2011 are used. Taipei-Macau (TPEMFM) route was selected because of its high flight frequencies. The flight length of the selected route is approximately 840 km, and the flight time is about 2 hours. Notably, the Taipei-Macau route had the largest number of annual passengers in the studied airline. The transaction dataset was based on International Air Transport Association (IATA) billing and settlement plan (BSP), aka "Bank Settlement Plan", which is an electronic billing system. The system is built to facilitate the payment and data flow between multiple travel agencies and airlines. Instead of every agent having an individual relationship with each airline, all of the information is consolidated through the system. The dataset is widely used by financial department of airlines and easily acquired comparing with other data sources. It provides records of all transactions between airlines and diverse distribution channels, including travel agencies, direct Internet sales and airline counters. Table 4-1 presents a sample record of airline revenue accounting data.

The revenue accounting data has one record per ticket issued. Each ticket has unique ticket number and different flight coupons for the itinerary. Here, a coupon represents a single flight leg in the itinerary, which is the sequence number of the coupon in the ticket. For example, coupon number of 1 identifies a ticket for the first flight of the whole itinerary, whereas coupon number of 2 represents a ticket for the second leg. The data fields related to this study were departure date, itinerary origin and destination, fight number, coupon number, ticket number, issue station, issue date (purchase date), sales office, tour code, fare basis, service class, agent and price. The seat availability data is then derived by ordering every issued ticket by issued date, as well as the itinerary information (round trip or one-way trip) can be also obtained through mapping ticket number and coupon number. To date, to study the time of advance purchase behaviors of air passengers, the advance purchase days was defined as days between the flight ticket issued (purchased) date and departure date.

Table 4-1: Airline revenue accounting data sample record

Item	Information
Departure Date	2011/12/1
Origin/Destination	TPEMFM
Fight Number	351
Coupon Number	1
Ticket Number	2440792666
Issue Station	TPE
Issue Date	2011/11/11
Sales Office	22473
Tour Code	403XIN2I162554
Fare Basis	*YEE3M/IN90
Service Class	K
Agent Code	34305585
Price(TWD)	2,500

*YEE3M/IN: Economy excursion fare, valid 90 days for Infant

Table 4-2 further presents the detail descriptions of service class and fare-basis code, which represented designed fare products and rules for numerous distribution channels and passenger value segments. While those fields may be useful for segment the demand and identify differences of air travel passenger choice behavior over purchasing, it is difficult to apply for the study due to the fact of complexity fare rules. Only a subset of data can be found in most of previous literatures (Carrier, 2008). In addition, the BSP dataset also contains transaction records from both direct purchasing passengers and from multiple distribution channels, which makes it hard to distinguish passenger behaviors from travel agents. Therefore, we only considered outbound flights and regular economy class tickets to reduce the complexity of problem.

Table 4-2: Descriptions of service class and fare-basis

Service Class Code	Identifies
C, J, D	Business class
Y	Economy class
G	Group Passengers
Fare-basis Code	**Identifies**
EE	Excursion fare
OW	One Way Journey
RT	Round Trip
2	Fare for 2 persons
VP	Value Package
AD	Agent Discount Fare
BB	Budget fare
BD	Budget fare Discount
CB	Cabin Baggage
CG	Tour Conductor Fare
CH	Accompanied Child fare
DG	Discounted Government Fares
ID	Industry Discount
IN	Infant fare (Not occupying a seat)

4.2 Passenger demand

With the purposes to complete the purchasing information, the flight schedule data was also integrated to the analysis dataset. Figure 4-1 shows the total passengers arranged by months and Table 4-3 lists the detail information of descriptive statistics for analysis flights. The dataset contains 1,044 flights and 134,820 transaction records. The most popular months flying to Macau were June, July and August, whereas March and October had the fewest passengers. In terms of average load factor (total passengers/ seat capacity, here we referred as 185 seats of Airbus A320 family), July and August had the highest value, followed by June. The summary statistics shows that July and August were the most popular travel months.

The BSP dataset provides records of all transactions between airlines and diverse distribution channels, including travel agencies, direct Internet sales and airline counters. Noted that the transaction records of group passengers were also categorized in the direct purchase category of the original dataset. In our study, we only considered the individual passenger who purchase through the airline website and counters. Overall speaking, the percentage of direct purchase passenger takes 18.37% of the total sales, which was relatively higher compared with other operating routes according to our sponsor airline. It is believed that long-haul routes should have lower percentage of direct passengers. December and June were the top 2 months that had the most direct purchase passengers of 2,935 (22.84%) and 2,680 (26.76%) respectively.

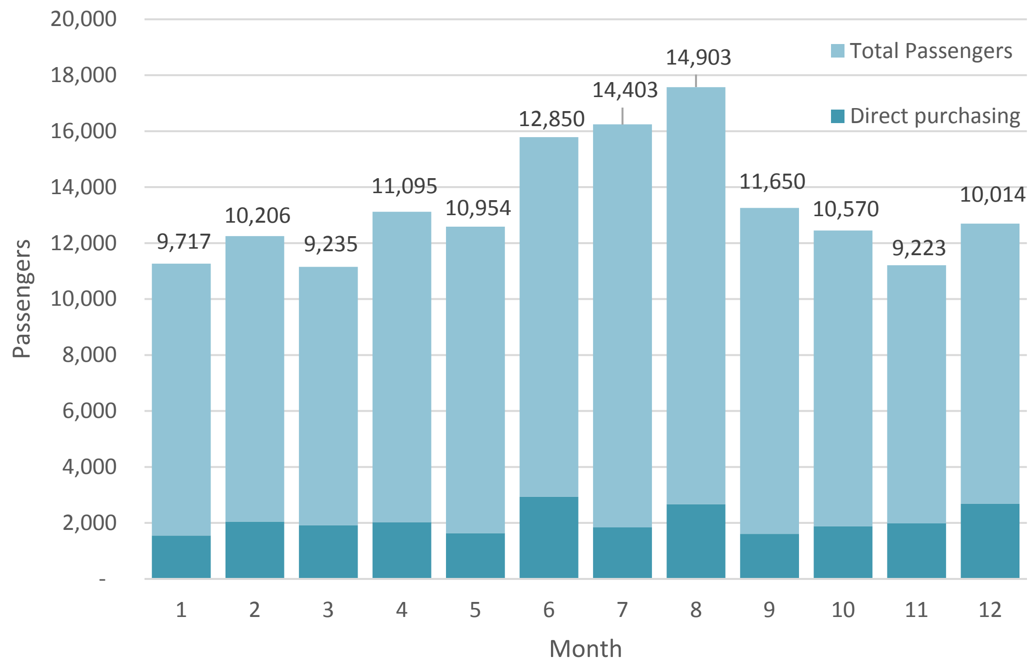


Figure 4-1: Total passengers of TPEMFM in 2011 by months

The study airline offered three daily flights that departed in the morning, afternoon and evening (Departure at 08:10, 13:30 and 18:20; arrival at 09:45, 15:10, and 20:10, respectively). The two datasets were combined to investigate the departure time preferences of passengers such as time of day, days of week and months of year. Data anomalies including outliers or incomplete records are also removed, which results in the final subset of 1,044 flights and 134,820 transaction records. Table 4-4 further lists the selected flights and correspondent load factor (LF) information for this study. The afternoon flights had the most passengers of 49,064 and highest mean load factor of

0.747. The morning flights had medium values, whereas the evening flights had lowest mean load factors of 0.622. The table suggests that passengers may prefer afternoon flights the most, followed by morning flights.

Table 4-3: Passenger statistics by months

Months	Flights	Total Passengers	Average Load Factor	Direct purchasing
1	89	9,717	59.02%	1,550 (15.95%)
2	81	10,206	68.11%	2,044 (20.03%)
3	88	9,235	56.73%	1,916 (20.75%)
4	86	11,095	69.74%	2,025 (18.25%)
5	87	10,954	68.06%	1,632 (14.90%)
6	89	12,850	78.04%	2,935 (22.84%)
7	87	14,403	89.49%	1,842 (12.79%)
8	93	14,903	86.62%	2,670 (17.92%)
9	86	11,650	73.22%	1,608 (13.80%)
10	90	10,570	63.48%	1,883 (17.81%)
11	81	9,223	61.55%	1,983 (21.50%)
12	87	10,014	62.22%	2,680 (26.76%)
Total	1,044	134,820	69.69%	24,768 (18.37%)

Table 4-4: Load factor statistics by flight schedule

Flights	Flights	Passengers	mean.LF	max.LF	min.LF	med.LF
Morning	342	45,835	0.724	1.000	0.097	0.784
Afternoon	355	49,064	0.747	0.995	0.173	0.795
Evening	347	39,921	0.622	0.995	0.049	0.605
Total	1,044	134,820	0.698	0.995	0.049	0.751

4.3 Advance purchase behaviors

In order to investigate advance purchase behaviors of air passenger, the advance purchase days was defined as days between the flight ticket purchased date and departure date. Figure 4-2 exhibits the overall advance purchase pattern of direct sale channels prior to departure for the Taipei-Macau route. The left axis of figure 4-2 shows the cumulative percentages of advance purchase passengers, whereas right axis presents the number of passengers in percentages by advance purchase days. Since 97.05% of passengers purchased their tickets within 60 days, for individual choice models, an observation interval of 60 days (9 weeks) is studied. Table 4-5 further lists the detailed cumulative percentage of direct purchase passengers within 7 advance purchase days. Data indicates that 2 days prior to departure day had the most advance purchase passengers, followed by the 1 day before departure day. About 2% of passengers purchased at the departure day, and approximately 50% of passengers purchased tickets 1 week before departure. It is belief that advance purchase patterns are different depending on the route characteristics, flight characteristics and price dynamics.

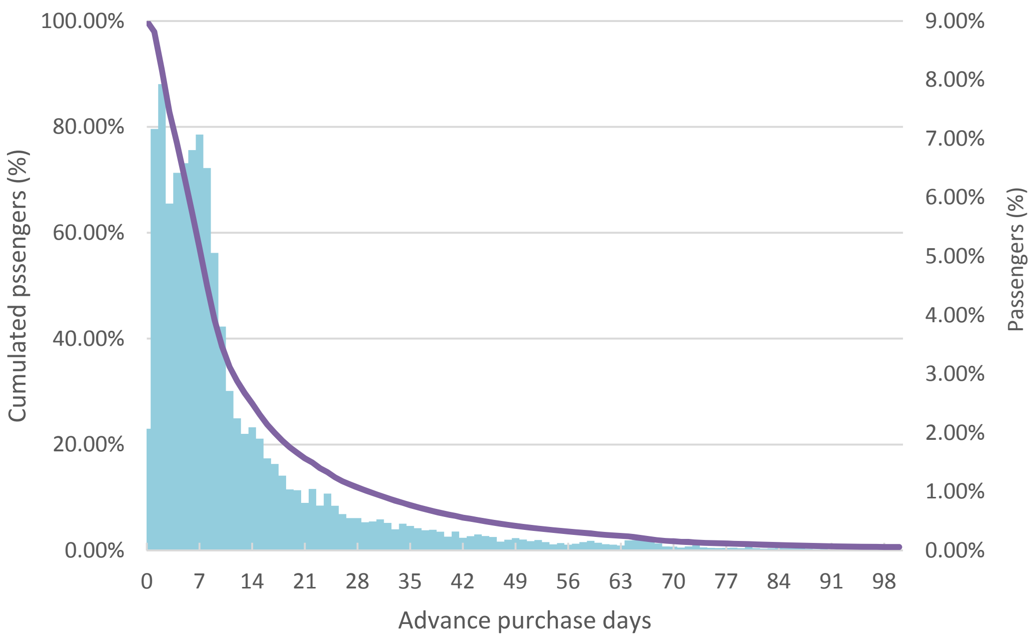


Figure 4-2: Aggregate advance purchase pattern prior to departure

Table 4-5: Cumulative percentage of advance purchase passengers

Advance purchase days	Passengers	Percentage	Cumulative Percentage
0	2,797	2.07%	1.34%
1	9,686	7.17%	8.56%
2	10,716	7.93%	16.54%
3	7,970	5.90%	22.48%
4	8,678	6.42%	28.95%
5	8,897	6.58%	35.58%
6	9,201	6.81%	42.44%
7	9,557	7.07%	49.56%

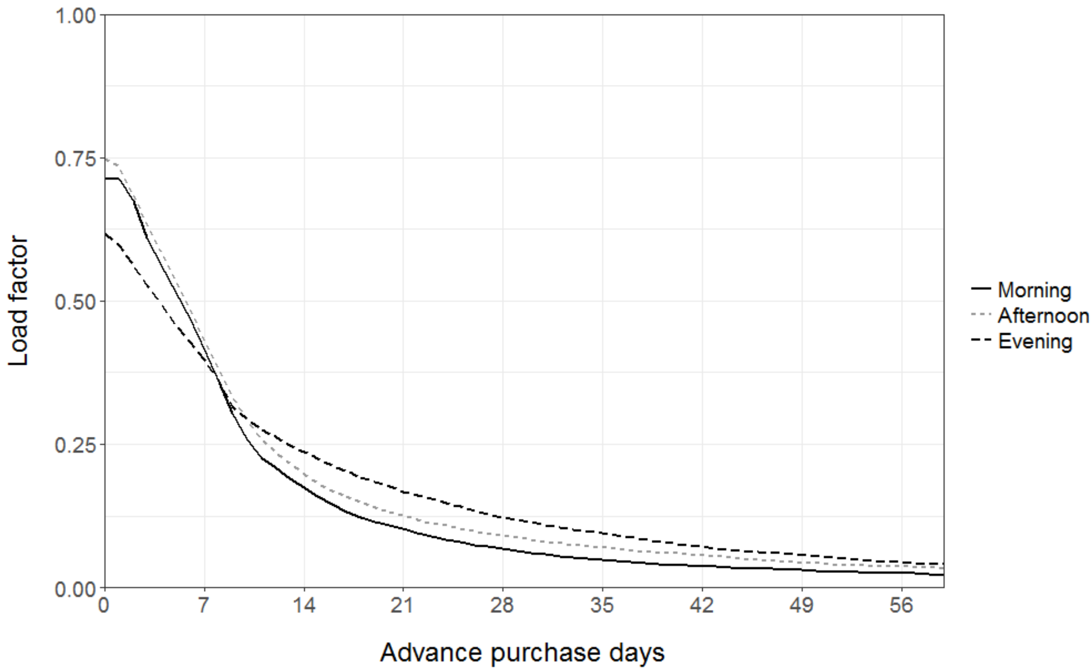


Figure 4-3: Aggregate mean advance purchase pattern by time of day

By combining the flight schedule and BSP transaction data, we can investigate the departure time preferences of passengers such as time of day, days of week and months of year. Passengers are assumed to make advance purchase decisions for a particular flight which represent their travel preferences. Figure 4-3 illustrates the different advance purchase patterns between morning, afternoon and evening flights. As expected, the afternoon flights have the highest final load factor as presented in Table 4-4. The advance purchases for morning and afternoon are identical and both increasing sharply before approximately 2 weeks prior to departure. The evening flights have relatively stable advance purchase pattern and lower final load factor compare with morning and afternoon flights. Figure 4-3 further illustrates the different advance purchase patterns by month of year. Interestingly, February afternoon flights had much higher value compared with other months, suggesting passengers who preferred afternoon flights in February may purchase much earlier. The evening flights in June, July, August and December had shown identical patterns as February afternoon flights, suggesting in these months, passengers may exhibit same behaviors. Based on that, it is belief that different departure time preferences of air passengers may present different advance purchase patterns.

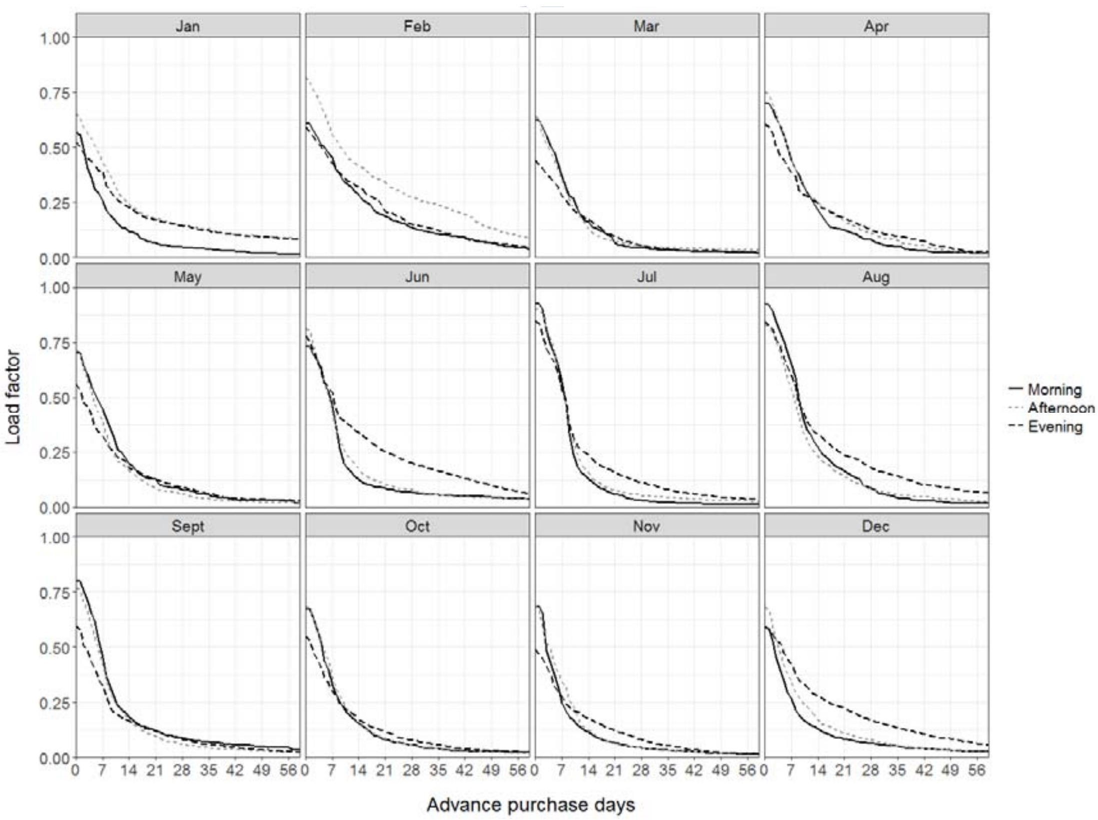


Figure 4-4: Aggregate advance purchase pattern by month of year

4.4 Price distributions

While service class and fare basis are typically used for designing fare products, it is difficult to be applied in the study because of the complexity of various fare rules. Therefore, for simplicity, this study considers only the subset of economic class of fare basis YEE1M and YEE3M round-trip tickets purchased through the direct purchasing channel (website and airline counters). Table 4-6 lists service classes and its corresponding average prices, and Figure 4-5 shows the price distribution across the advance purchase horizon. Notably, as departure day approaches, the observed price range tends to widen, and the number of expensive airline tickets sold increases. However, since passengers nowadays may less aware of designed fare classes instead of purchase price, passengers are assumed to make advance purchase timing decisions based on ticket price and their departure time preferences when purchasing the flight tickets.

Table 4-6: Average prices with respect to service class

Service classes	Avg.price	SD.price
W	4.702	0.160
L	4.568	0.203
B	4.519	0.053
T	3.889	0.114
Q	3.796	0.017
X	3.426	0.120
Z	3.292	0.038

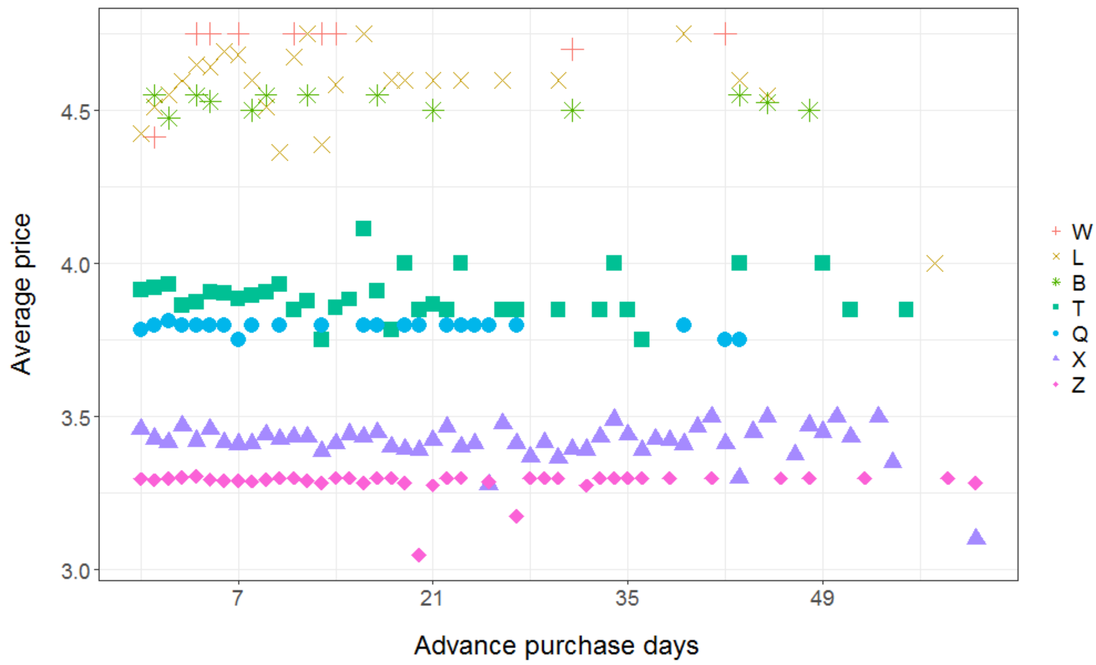


Figure 4-5: Price distribution of selected classes

Chapter 5 Results

Based on the data analysis in the previous chapter, the objective of this chapter is to demonstrate the estimation results of proposed models. The methodologies presented in Chapter 3, including the functional concurrent regression, multinomial logit, and continuous logit models are applied in this chapter to explore the advance purchase behaviors of air passengers.

5.1 Aggregate Pattern Model

The proposed functional concurrent regression model for advance purchase pattern analysis was estimated with R package "fda". An observation interval of 60 days (about 9 weeks) is studied. Data anomalies including outliers or incomplete records are also removed, which results in the final subset of 1,044 flights and 134,820 transaction records. 80% of data (836 flights) were randomly selected as training set and 20% of data (209 flights) were used for validation. By combining BSP and flight schedule dataset, the flight-specific attributes obtained from flight schedule database were used to capture the time of day (morning, afternoon and evening flight), days of week (flight on Friday), and month of year (flights in peak months) preferences of air passengers. Additionally, consecutive holidays and special vacations are also marked for model estimation. To date, the advance purchase days was defined as days between the flight ticket issued (purchased) date and departure date. Table 5-1 presents the descriptive statistics of selected variables for the aggregate advance purchase pattern model.

Table 5-1: Descriptive statistics for functionl concurrent model

Variables	Description	%
HISTORICAL LF	Functional objects, load factor at 7 days before current day	
MORNING	Dummy, 1 if morning flight; 0 if others.	32.76%
AFTERNOON	Dummy, 1 if afternoon flight; 0 if others.	34.00%
FRIDAY	Dummy, 1 if flight on Friday; 0 if others.	14.66%
PEAK SEASON	Dummy, 1 if travel in July and August; 0 if others.	17.24%
VACATION	Dummy, 1 if consecutive holidays; 0 if others.	10.54%

Prior to model estimation, functional data analysis techniques are adopted to transform discretely collected transaction dataset into the function form for model estimation. The discrete observed response variable of daily transactions are further presented as time-varying cumulative load factors. The observation time interval t of interest of 0 to 60 days is further broken into 10 equally spaced sub-intervals of 9 breakpoints and within each interval a polynomial of order 4. The daily transaction data are therefore smoothed by the total number of K=13 (9 breakpoint with default order 4 polynomials) B-spline basis curves in advance for functional regression modeling. The B-Spline functions are the most common choice of approximation system for non-periodic functional data, whereas Fourier series basis systems are popular for periodic data and functions (Ramsay and Silverman, 2005).

Here, Figure 5-1 exhibits the 13 B-spline curves used in this study. The x-axis represents the observation interval and the y-axis represents the B-spline function values. The nine interior knots are also exhibited as vertical dashed lines in the figure. The B-spline basis function has a property of the sum of the function value of time t is equal to one, which is known as the partition of unity property. For example, the value of the first and last basis functions are exactly one at the boundaries of t equals 0 and 60, whereas all the other basis functions will go to zero at these end points. The property assures the invariance of the shape of the B-spline curve under translation and rotation, which is useful when fitting the curve. Based on the B-spline basis system, the best value for smoothing parameter λ is then determined by generalized cross-validation (GCV) criterion. Figure 5-2 further shows how GCV values varies as a function of $\log_{10}(\lambda)$ for the daily transaction data. The dashed line represents the minimum GCV value for advance purchase patterns, where the value of $\log_{10}(\lambda)$ is -0.25 and the minimum GCV value is 0.8724. Therefore, the best value for smoothing parameter λ is determined with value of 0.5623.

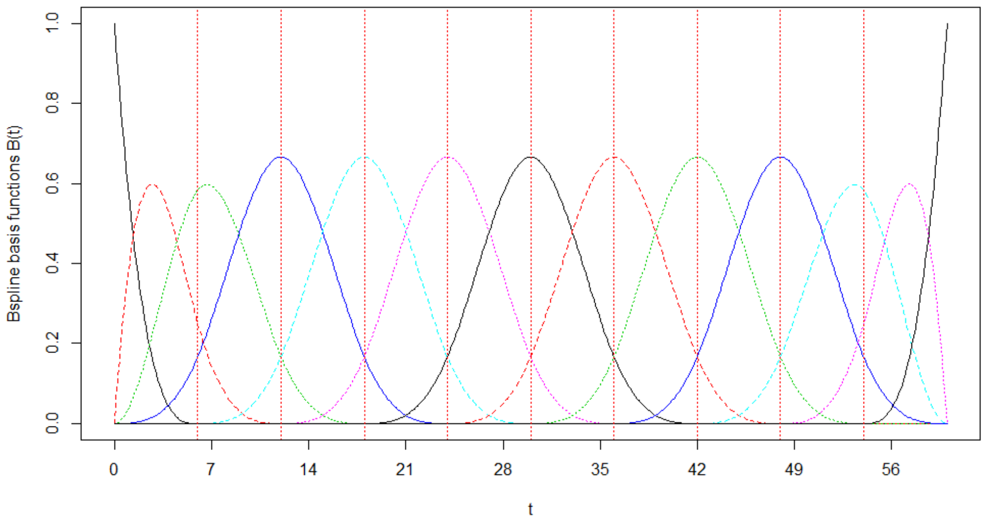


Figure 5-1: B-spline system used in this study

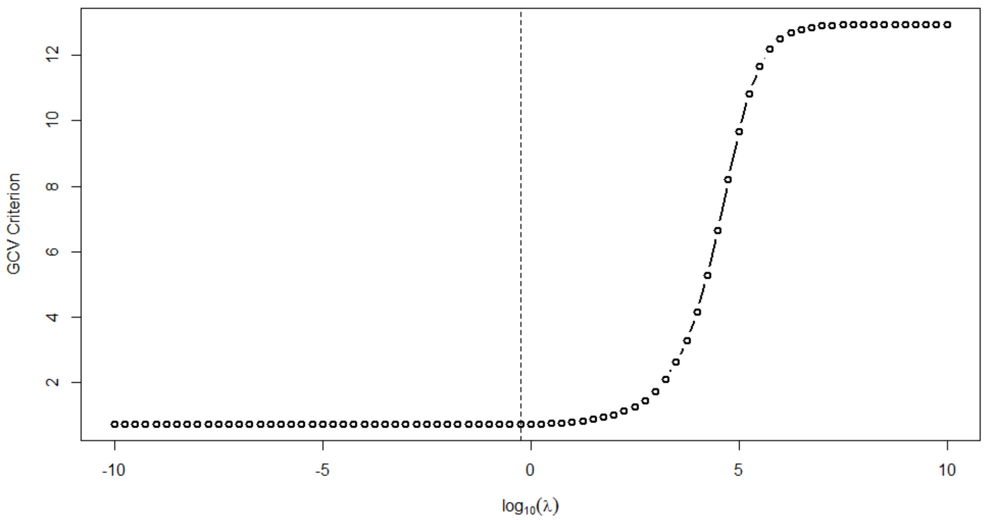
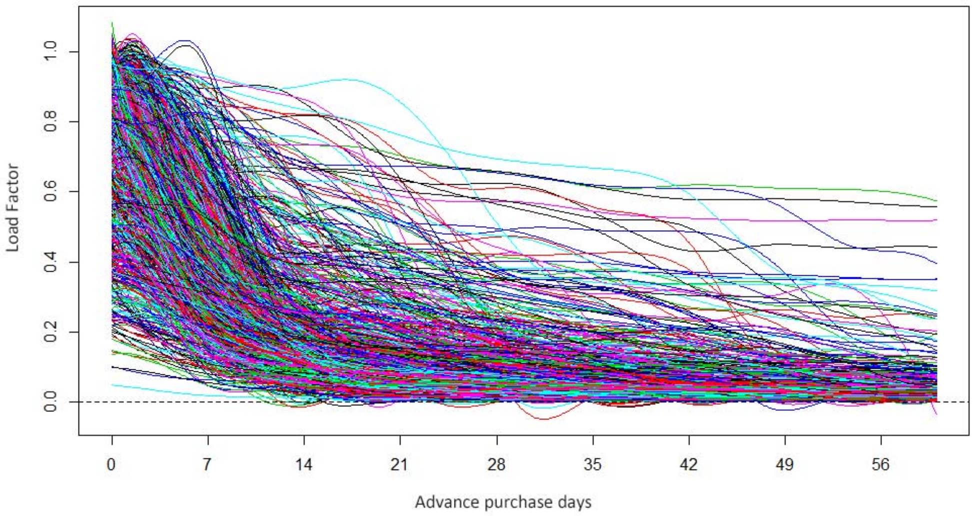
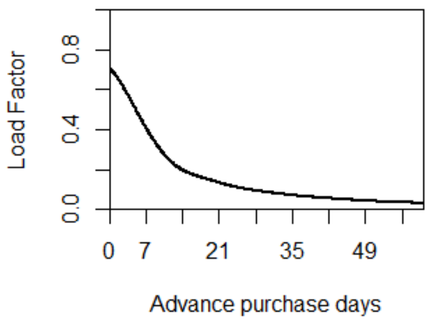


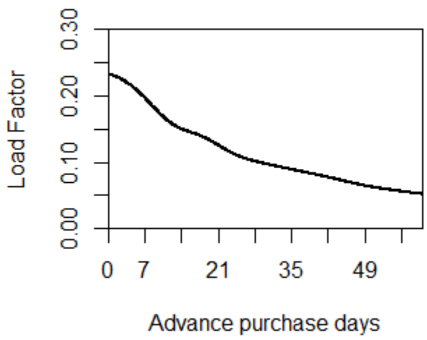
Figure 5-2: The value of GCV criterion



(a) Smoothed advance purchase curves



(b) Mean



(c) Standard Deviation

Figure 5-3: Advance purchase smoothing curves

Figure 5-3 (a) displays the smoothed advance purchase curves over the advance purchase period, whereas Figure 5-3 (b) and (c) shows the summary statistics for the functional data in terms of their mean and standard deviation for flights. It shows that the final load factor is about 0.7059. The mean advance purchase curve has gradually increased as the departure day approaches, and has drastically increase around 2 weeks. The pattern suggests passengers may make their purchase decisions late. On the other hand, the plot of the standard deviation function shows a stable increase and reaches the highest value of 0.2313, implying that the load factors are relatively unstable toward the departure day.

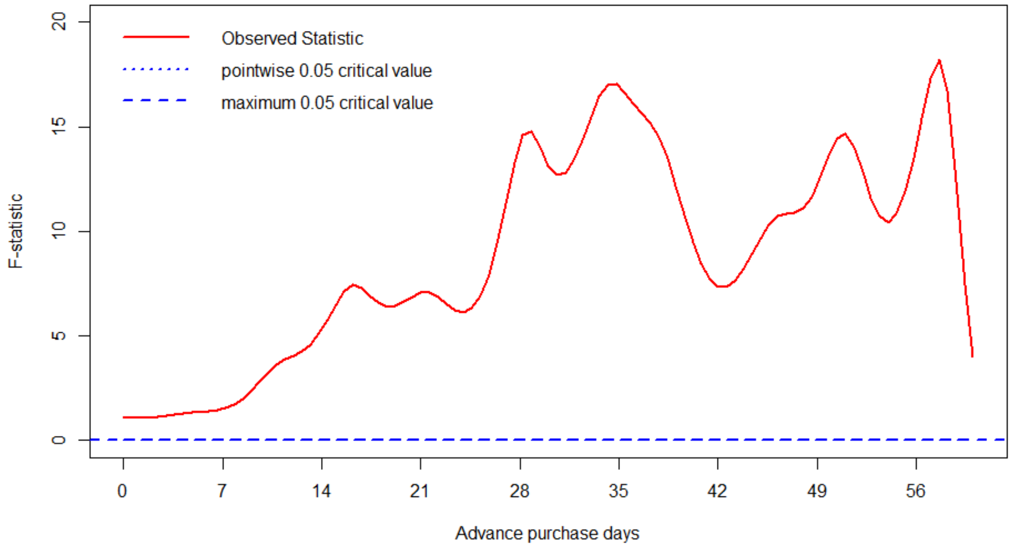


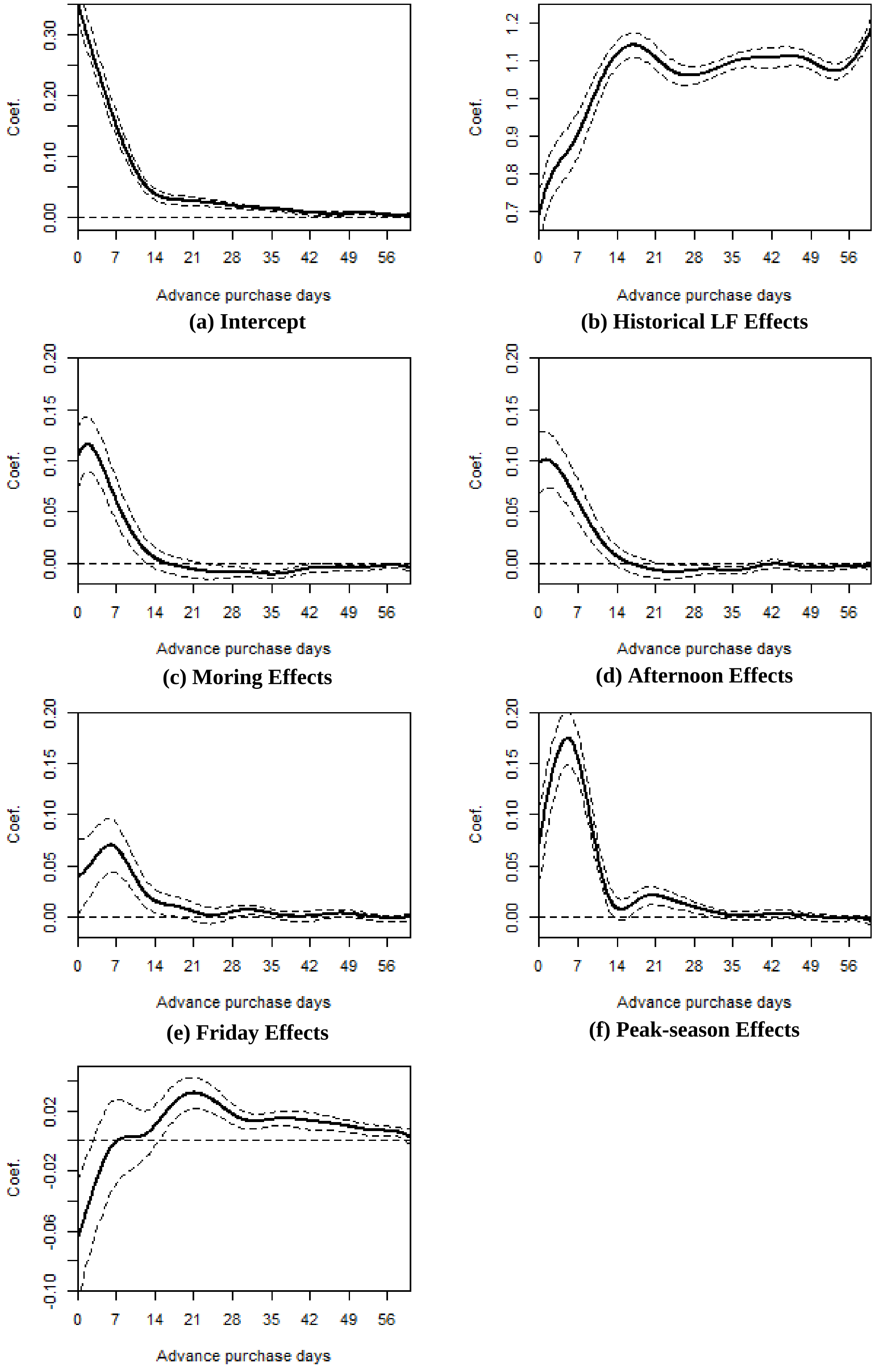
Figure 5-4: Permutation F-test

Figure 5-4 shows the results of permutation F-test for a predictive relationship between future load factors and investigating factors. Here, the dashed line represents the permutation 0.05 critical value for the maximum of F statistic and the dotted line the permutation critical value for the point-wise statistic. Result shows the max observed F value is 16.6491, indicating that there is a significant difference with the functional response and investigating covariates. The smoothed curves are then examined by the functional concurrent regression model with pointwise approach, that is, to apply least square for a fixed t=t*, and repeat the process for the entire observation interval to obtain the time-varying estimate coefficient, $\hat{\beta}(t)$ (Ramsay and Silverman, 2005). To date, because of the response variable is a continuous curve, so is the estimated $\hat{\beta}(t)$, which makes interpreting the results different from classical regression model.

The estimated regression coefficients for intercepts and flight specific attribute effects on advance purchase patterns are shown in Figure 5-5. The dashed lines represent pointwise 95% confidence intervals for these effects, where the confidence bounds are calculated by adding ±2 standard errors at each point of the parameter curve, which indicating the significance of each effects. The first panel 5-5(a) shows the time-varying coefficients for the intercept. The coefficient is positive throughout the entire observation intervals, and increasing sharply around 2 week prior to departure. The pattern corresponds to the mean load factor for all advance purchase patterns over the observation intervals of 0~60 days as Figure 5-3(b). Figures 5-5(b) ~ (g) exhibit contributing effects of factors on the mean load factor respectively.

Figure 5-5(b) illustrates the time-varying effects of load factors at 7 days before current day on predicting the current load factor. Throughout the entire advance purchase observation interval, the coefficient is positive, indicating a positive relationship between the historical load factors and predicting load factor at any time during the advance purchase. The relationship has sharp decreases around 2.5 weeks before departure. One possible explanation is that, for our study short-haul route, approximately 80% passengers made their purchases within 3 weeks prior to departure. That leads to very sparse final load factors as Figure 5-3(c) presented, thereby causing the sharply decreasing slope and weaken relationship between historical and predicting load factor.

The coefficients for morning and afternoon flights are further presented as Figure 5-5(c) and 5-5(d), where both figures exhibit a similar pattern. The coefficient of morning flight shows more negative effects than noon flights, but both turn to positive effects around 2 weeks before departure. The pattern indicates that passengers who preferred morning and noon flights may tend to purchase around 2 weeks prior to departure. The afternoon flights have relatively stable effects than that of morning flights, suggesting that afternoon flights may have more stable demand. However, morning flights have shaper incensement and higher final coefficient toward departure date, indicating that morning flights are more popular. Figure 5-5(e), 5-5(f) and 5-5(g) illustrate the coefficient plots for Friday, peak-season and vacation flights, respectively. The estimated coefficients of three flight specific attributes all have increasing positive effects on load factor throughout the entire observation interval, suggesting that those three types of flights have gradually increasing demand as the departure day arrives.



(a) Intercept

(b) Historical LF Effects

(c) Moring Effects

(d) Afternoon Effects

(e) Friday Effects

(f) Peak-season Effects

(g) Vacation Effects

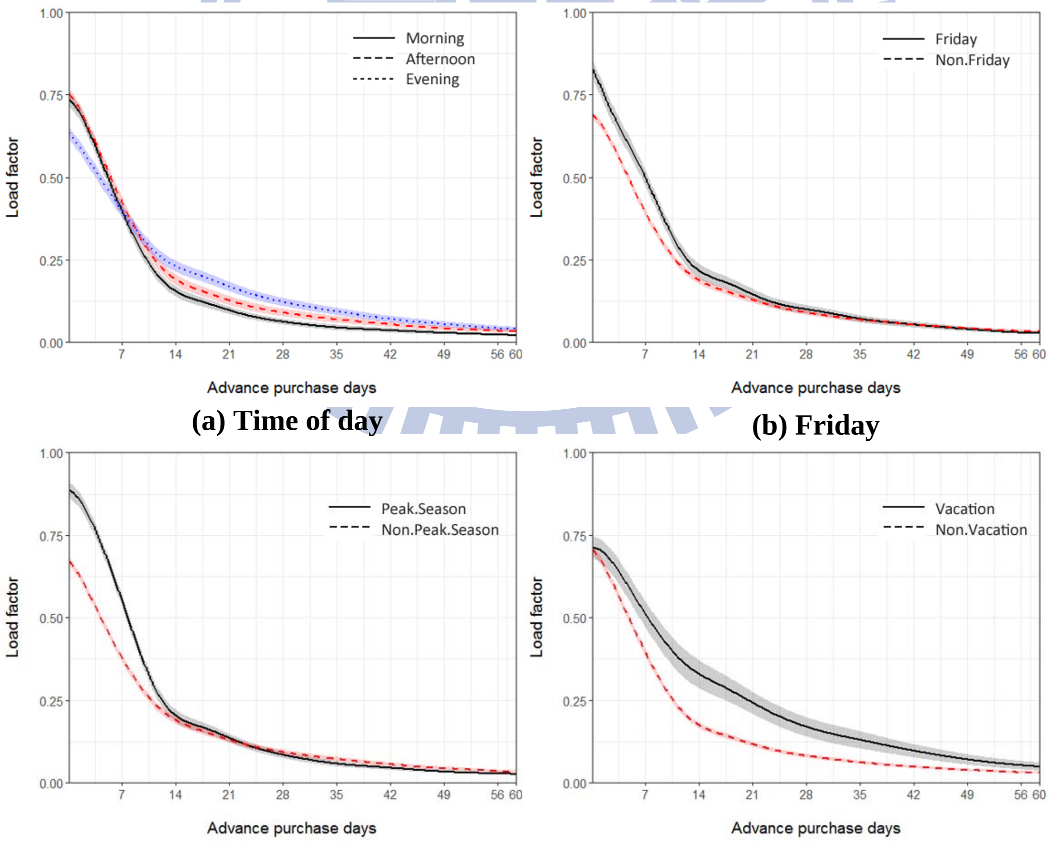
Figure 5-5: Coefficient plots $\beta_j(t)$ for flight specific attributes.

Interestingly, the coefficient for vacations effect has all increasing positive effects on the average load factor, but a sudden drop pattern around 3 weeks before departure day. The estimated results suggest that passengers tend to purchase early for vacation trips and may hardly purchase tickets on departure day. When the tickets were getting fewer and fewer transaction would be made, thus the vacation flights have sudden decreasing trend and minor effects. The same pattern can be observed for Friday and peak-season flights around 1 week prior to departure. Further insights can be gained from Figure 5-6, which depict the predicted advance purchase curves under four time-invariant variables, including time of day (MORNING and AFTERNOON), day of week (FRIDAY), month of year (PEAK.SEASON) and VACATION indicators.

Figure 5-6(a) shows the advance purchase curves for morning, afternoon and evening flights. In terms of load factors, morning and afternoon flights have higher final predicted load factor, suggesting those two types of flights are more popular. The predicted curve of morning flights (the solid curve) is slightly smaller than afternoon flights, but with identical same shapes. Around 2 weeks prior to departure, the predicted values for those two types of flights have a significant increment, and become larger than evening flights approximately 1 week before departure. Surprisingly, the predicted curve of evening flights (the dotted curve) are larger than other two types of flights before 2 weeks prior to departure, implying that passengers may purchase earlier. One possible explanation is that airlines may pay more attention on popular fights and tend to hold the seats, which may result in phenomenon of late purchases and delay increases of load factor.

Figure 5-6(b) further depicts advance purchase curves for the Friday and non-Friday flights, whereas Figure 5-6(c) shows how the load factor changes in the peak-season (flights in July and August). Figure 5-6(d) demonstrates the predicted curves for flights in consecutive holidays and special vacations. As expected, Friday and peak-season fights (the solid curves) exhibit the same pattern as popular flights, which have larger final predicted load factors and sharp upward trend. Although the advance purchase pattern of non-Friday flights (the dashed curves) has slightly difference compared with Friday flights, but shown diverge at around 2 weeks before departure.

The non-Friday flights (the dashed curves) have slightly different shape compared with Friday flights, but diverge at 2 weeks before departure which result in the lower final load factor of non-Friday flights. The non-peak season flights present the same advance purchase pattern as evening flights, and also lower load factor. For those two types of not popular flights, airline could create promotion plans or discounts around 2 weeks to enhance final load factor. Notably, vacation flights shows the similar early purchase patterns, but the curve much steeper as departure day approaches. The modeling results suggest that passengers who travel on consecutive holidays tend to make their purchases earlier (planned travelers), and there are also some passengers prefer to purchase tickets close to the departure date (spontaneous travelers). One explanation is that the study route is short-haul and therefore has more spontaneous travelers than long-haul air routes. Hence, for the vacation short-haul flights, airlines can appropriately raise the price to gain extra revenue. We believe that the advance purchase behaviors for long-haul air routes should be different.



(a) Time of day (b) Friday

(c) Peak-season (d) Vacation

Figure 5-6: Aggregate advance purchase pattern prior to departure

Finally, Figure 5-7 exhibits the time-varying MAPE value for the forecast and the eventual outcomes. The calculated overall MAPE is 27.85%, whereas the maximum and minimum value is 36.55% and 20.80%, respectively. Based on Lewis (1982) forecasting accuracy scale, the MAPE values of the proposed model are ranging from 21 to 50%, indicating that the model has reasonable forecast performance for future load factor.

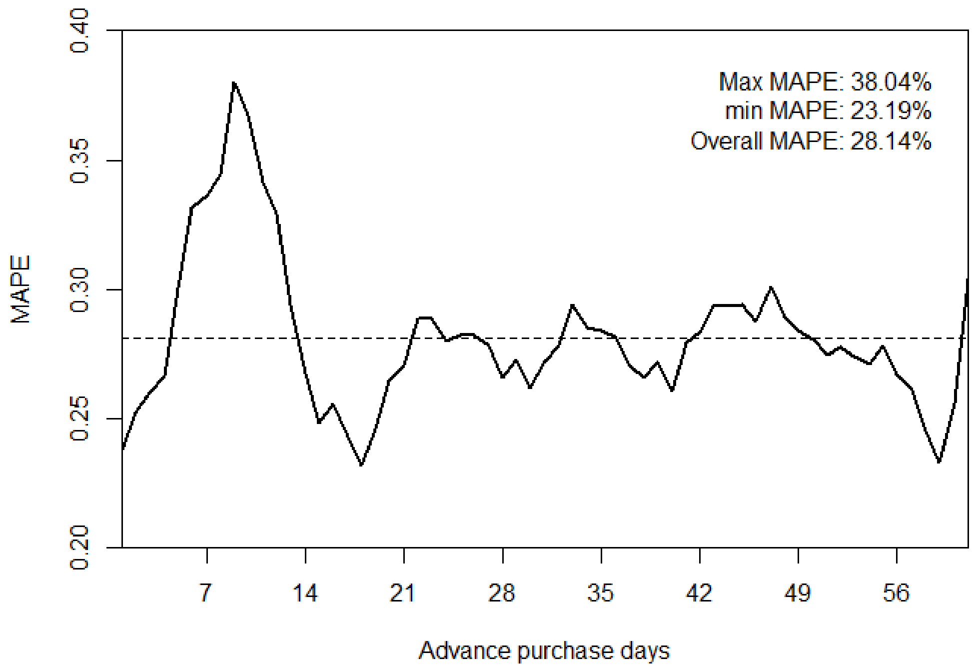


Figure 5-7: Mean absolute percentage error (MAPE)

5.2 Individual Choice model

5.2.1 Discrete Choice Model

To study the individual advance purchase decisions of air passengers, a subset of economic round-trip tickets that purchased through the direct purchasing channel (website and airline counters) was used. Compared with previous studies that introduced time of booking/purchasing as explanatory variables into choice model, the advance purchase days are seen as the response variable (alternatives) in our study. Moreover, to reduce the number of alternatives and facilitate model development, the advance purchase horizon is divided into five time periods according to three segmentation methods, including equal time periods (each period is of 12 days), time periods with equal number of purchases and time periods according to subjective judgments from the study airlines.

For the subjective judgment method that suggested by experts from the study airline, as departure day approaches, the airline will generally begin to check the seat reservations and decide to have discounts and promotions to raise sale volume or not. The airline will announce promotion information to travel agencies around 1 to 2 months from the departure day. Two weeks prior to departure, they will start to ask travel agencies to pay for group passengers, or return the remaining seats, so the remaining seats of the flight will change dramatically. In the last week prior to departure, promotions such as "last minute sale" and internet advertisements will be performed to attract individual passengers. Finally, we expect passengers purchasing tickets at the departure day have different choice behaviors. Passengers are assumed to make advance purchase decisions for a particular flight which represent their travel preferences. Table 3-3 outlines the defined advance purchase time periods and number of passengers. Figure 5-8 demonstrates the relationships between defined periods and advance purchase days for the selected dataset.

Table 5-2: Defined advance purchase periods

Period	Equal time period		Equal purchase number		Subjective judgment	
	Number	Days before Dep.	Number	Days before Dep.	Number	Days before Dep.
P1	2,048 (71%)	0~12	744 (26%)	0~2	164 (6%)	0
P2	509 (18%)	13~24	676 (23%)	3~5	1,396 (48%)	1~7
P3	204 (7%)	25~36	405 (14%)	6~9	632 (22%)	8~14
P4	105 (4%)	37~48	559 (19%)	10~17	497 (17%)	15~31
P5	33 (1%)	>49	515 (18%)	>17	210 (7%)	>31

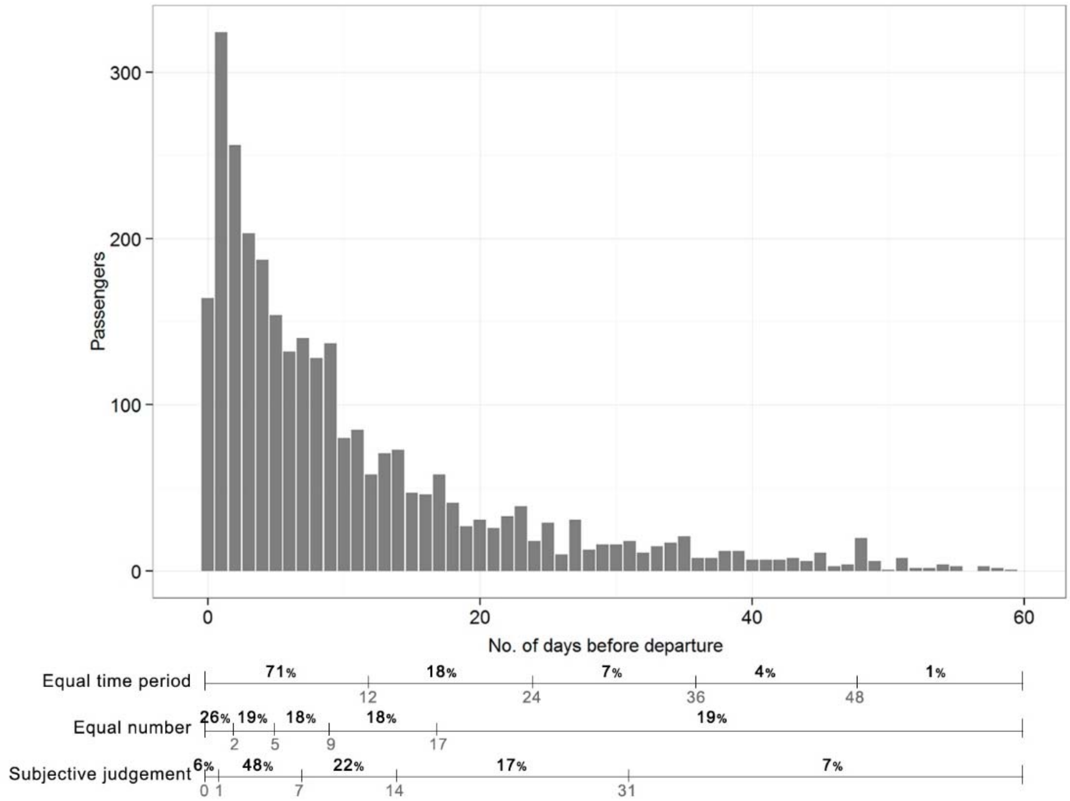


Figure 5-8: Time periods of advance purchase of three segmentation methods

To obtain the probability of purchasing in each time segment, multinomial logit (MNL) model is used. Several factors contributing to advance purchase behaviors are further examined. The selected generic variable is the average purchase price in logarithmic form (PRICE), whereas flight schedule preferences including morning flight (MORNING), flight on Friday (FRIDAY), flights in peak months including July and August (PEAK.SEASON), and travel on vacation (VACATION) are treated as alternative specific variables to capture the time of day, days of week and month of year preferences of air passengers across advance purchase periods. Additionally, to account for the relationships between purchase price and flight preferences, the interaction term specification is used. All the alternative specific variables are expected to decrease as departure day approaches. For simplicity, this study considers only the subset of economic class (service classes of W, L, B, T, Q, X) of fare basis YEE3M tickets. Data anomalies including outliers or incomplete records were also removed, which results in the final of 2,899 transaction records for model estimations. The descriptive statistics of explanatory variables for discrete choice model are reported as Table 5-3.

Table 5-3: Descriptive statistics for discrete choice model

Variables	Description	Mean/%	Sd	Med.	Max	Min.
PRICE	Purchase price in thousands New Taiwan Dollars	3.546	0.346	3.450	4.750	3.050
ADVDAYS	Advance purchase days before departure	10.548	11.374	7	60	0
MORNING	Dummy, 1 if morning flight; 0 if others.	31.46%				
FRIDAY	Dummy, 1 if Friday; 0 if others.	21.52%				
PEAK SEASON	Dummy, 1 if July and August; 0 if others.	16.97%				
VACATION	Dummy, 1 if consecutive holidays; 0 if others.	11.69%				

Table 5-4 presents the estimation results of the three MNL models by using LIMDEP NLogit software. Results demonstrate that all abovementioned variables are significantly tested with the expected sign. For model comparisons, several performance indices, including log-likelihood statistic at optimal, adjusted rho-square and Akaikes Information Criterion (AIC) are selected. The optimal log-likelihood values of three models are -2,524.61, -4,485.71 and -3,738.01, respectively; whereas the adjusted rho-square are 0.455, 0.034 and 0.195. In term of two indices, the model based on the equal time period performs best. Additionally, the equal time period model also has the smallest AIC/N value of 2.010, suggesting the proposed temporal segmentation based on equally of 12 days can better explain the advance purchase behaviors.

Table 5-4: Estimated results of three models

Variables / Experiments	Equal time period		Equal number		Subjective judgement	
	Coeff.	t-value	Coeff.	t-value	Coeff.	t-value
ASC2	-1.855	-24.230 ***	-0.137	-2.280 **	2.023	21.470 ***
ASC3	-2.769	-24.360 ***	-0.857	-9.910 ***	0.918	8.520 ***
ASC4	-3.559	-23.320 ***	-0.729	-9.320 ***	0.422	3.780 ***
ASC5	-4.301	-21.350 ***	-0.938	-10.910 ***	-0.295	-2.220 **
log(PRICE)	-1.042	-2.160 **			-1.025	-2.660 ***
log(PRICE)*MORNING	-2.358	-3.210 ***	-2.436	-4.480 ***	-1.860	-2.980 ***
log(PRICE)*PEAK.SEASON	2.713	3.160 ***	2.306	3.940 ***	2.740	3.930 ***
log(PRICE)*VACATION	4.853	5.680 ***	2.670	4.140 ***	3.890	5.360 ***
MORNING (P2)	0.498	4.700 ***			0.928	3.840 ***
MORNING (P3)	0.526	3.410 ***	0.330	2.680 ***	1.354	5.440 ***
MORNING (P4)	0.691	3.360 ***	0.557	5.170 ***	1.429	5.660 ***
MORNING (P5)			0.615	5.520 ***	1.551	5.680 ***
FRIDAY (P2)	0.528	4.570 ***				
FRIDAY (P3)	0.558	3.360 ***	0.413	3.050 ***	0.228	1.970 **
FRIDAY (P4)			0.360	2.910 ***	0.666	5.680 ***
FRIDAY (P5)			0.503	4.040 ***		
PEAK.SEASON (P2)	0.784	6.650 ***	0.561	3.410 ***	0.443	2.100 **
PEAK.SEASON (P3)			0.793	4.440 ***	1.076	4.930 ***
PEAK.SEASON (P4)			1.149	7.220 ***	1.196	5.360 ***
PEAK.SEASON (P5)			0.852	4.850 ***		
VACATION (P2)			-0.295	-1.760 *	-0.811	-5.490 ***
VACATION (P3)	0.924	5.050 ***	-0.499	-2.260 **	-0.690	-3.770 ***
VACATION (P4)	1.599	7.370 ***				
VACATION (P5)	1.048	2.540 **	0.897	6.260 ***	0.671	3.560 ***
Goodness of fit measures						
No. of observation		2,899		2,899		2,899
No. of parameters		17		20		20
Log-likelihood at zero		-4,665.76		-4,665.76		-4,665.76
Log-likelihood at constant		-2,634.70		-4,603.24		-3,881.59
Log-likelihood at optimal		-2,524.61		-4,485.71		-3,738.01
ρ^2		0.459		0.039		0.199
Adj-ρ^2		0.455		0.034		0.195
AIC/N		2.010		3.563		2.972

Note: ***, **, and * represent reaching 1%, 5% and 10% significance level, respectively.

For the generic variables, the log(PRICE) coefficients in the equal time period model has significantly negative marginal effect of -1.042 on advance purchase as expected, suggesting the higher purchase price, the lower utility of passengers and thus the probability of the airline being chosen decreases. The interaction term between log(PRICE) and morning flights has a negative coefficient of -2.358 which makes that the total of log(PRICE) marginal effect becomes -3.4. The result indicates that purchase price has larger negative effect for the morning flight. In contrast, the interaction terms of PEAK.SEASON and VACATION flights have a positive effect of 2.713 and 4.853, implies that the passengers' disutility of price effect is lower when passengers travel during peak season or vacation, which results in the price effect turns to be positive. One possible explanation for this could be that we used the transaction records that provides only purchased alternative for modeling. Only the passengers who accepted the higher purchasing price were being recorded in our dataset, passengers who chose to purchase later nor not purchase were unable to capture.

For the alternative specific dummy variables included in the model are aimed to capture the flight schedule preferences across advance purchase periods. All alternative specific variables have the significantly positive effects on utility when compared with base alternative (advance purchase period 1: within 12 days prior to departure). Both MORNING and FRIDAY variables show the expected decreasing pattern. The utility decreases as the advance purchase period approaches departure day, implying that passengers who prefer morning and Friday flights tend to purchase ticket earlier. However, the coefficients of VACATION variables present the positive but irregular effect. One possible reason for this might be that airlines are believed to hold the seats of lower fare class and release them as late as possible before vacation times.

Tables 5-5 further summarizes aggregate direct price elasticities. The elasticities of ordinal flights are determined according to the estimated parameter of log(PRICE) excluding interaction effects. As expected, the longer the advance purchase days are, the higher the direct elasticity. The values of most elasticities are larger than one, indicating that the passengers are sensitive to price changes. The elasticity decreases closer to the date of departure, implying that passengers becomes less sensitive to price as closer to the date of departure. This phenomenon also reflects that once passengers have decided the purchase flight. They may have to make the purchase as departure day approaches, no matter how price changes. For ordinal flights, Period 5 has the highest elasticity of -1.297, suggesting 1% of log price increases will result in 1.297% decrease of choice probability in Period 5 (>49 days prior to departure), whereas Period 1 (0~12 days prior to departure) has the lowest direct elasticity of -0.371 which is less than one, suggesting the price inelasticity of Period 1 passengers. The elasticities of morning flights are 3 times higher than those of ordinal flights, indicating that passengers preferring morning flights are more price sensitive.

Interestingly, popular flights in peak season and vacation have positive elasticities, suggesting price increases will also result in increase of choice probability. This is because the positive estimated total log price effects of 1.671 and 3.811. The result suggests that passengers may need to spend more for purchasing peak season and vacation flights. However, our data only reflects the behavior of passengers who accepted the higher purchase price. Passengers may choose alternative flights from other carriers or choose not to purchase. The elasticity values of morning and vacation flights are larger than other flights, suggesting those two types of flights are more sensitive to price changes, leaving a large room for RM strategies.

Table 5-5: Direct price elasticities

Advance purchase periods	Ordinal flights	Morning flights	Peak season flights	Vacation flights
Period 1	-0.371	-1.209	0.594	1.355
Period 2	-1.040	-3.391	1.666	3.801
Period 3	-1.196	-3.899	1.915	4.370
Period 4	-1.231	-4.014	1.972	4.499
Period 5	-1.297	-4.231	2.078	4.742

Note: Ordinal flights are defined as the flights are not in the morning, peak season and vacation.

5.2.2 Continuous Choice Model

To explore what kind of segmentation methods can better reconstitute choice sets for individual advance purchase behavior modeling, previous section had provided the estimated results of three segmentation methods. However, some choices/response variables are continuous in nature such as advance purchase time, departure time, and location. Arbitrarily discretizing these continuous choices variables may lead to an erroneous result. The discretization for discrete choice models is based on different research purposes to divide the study period into a limited number of intervals, which may not able to cover the entire time period and to cause loss of temporal resolution. In our study, previous section had shown that different discrete interval settings would also lead to different and unstable estimation results. Although previous studies had prevented this situation by considering correlations among alternatives, continuous treatment of time variable seems more intuitive and preferable.

In order to avoid the subjective segmentation of advance purchase horizon, a continuous logit model was constructed. The estimated results are presented in this section. For simplicity, this study considers only the subset of economic class (service classes of W, L, T, X) of fare basis YEE1M and YEE3M round-trip tickets purchased through the direct purchasing channel. After excluding anomalies such as outliers or incomplete records, the final subset contained 2,534 transaction records. The flight-specific attributes obtained from flight schedule database were used to capture the time of day (morning, afternoon and evening flight), days of week (flight on Friday), and month of year (flights in peak months) preferences of air passengers. Consecutive holidays and special vacations are also marked for model estimation. The coefficient of variation (CV) of observed historical prices across flights within the purchasing month is also incorporated, which reflects potential price fluctuation caused by airline promotions. Table 5-6 presents the descriptive data for selected variables for the proposed continuous logit model.

Table 5-6: Descriptive statistics for continuous choice model

Variables	Description	Mean/%	Std
PRICE	Purchasing price in thousands (NTD)	3.53	0.33
ADV	Advance purchase days	7.92	11.36
SEATS	Number of remaining seats at the time of purchase	114.50	47.58
CLASS	Dummy for service classes as shown in Table 2.		
YEE1M	Dummy, 1 if YEE1M; 0 if YEE3M.	95.70%	
CV	Dummy, 1 if CV of observed prices across flights within the purchase month > 0.1; 0 if others.	23.68%	
MORNING	Dummy, 1 if morning flight; 0 if others.	32.91%	
AFTERNOON	Dummy, 1 if afternoon flight; 0 if others.	48.34%	
FRIDAY	Dummy, 1 if flight on Friday; 0 if others.	21.39%	
PEAK.SEASON	Dummy, 1 if travel in July and August; 0 if others.	26.60%	
VACATION	Dummy, 1 if consecutive holidays; 0 if others.	10.97%	

In addition, to obtain time-varying purchase prices for continuous logit model estimation, an additional ordinary least squares regression model is firstly constructed. Table 5-7 reports the estimation results for purchase price at various advance purchase time. The R-squared and adjusted R-squared values of 0.875 and 0.874, respectively, indicate a good model fit. For the explanatory variables, advance purchase days (ADV) has significantly negative effect of -0.002, suggesting that price decreases as the number of advance purchase days increases. As expected, maximum stay limitation within 30 days (YEE1M) and lower service classes (comparing with highest class W) have significant negative effects, suggesting the lower service class the lower purchase price. The result reflects the airline revenue management strategy that utilized fare levels to distinguish passengers with different price elasticities. Although some of the SEATS variables that interacted with sinusoidal functions were not tested for statistical significance, the overall coefficients present a reasonable negative effect of -0.056 to -0.255. This suggests as the number of seats available in the market increases, the ticket price decreases. Figure 5-9 further exhibits the fitted price corresponding with each service class is overlaid with the observed price. The time-varying purchase prices were imputed by the estimated regression model, which are then incorporated with flight specific attributes for passenger advance purchase time choice modeling.

Table 5-7: Estimated purchase price model

Variable	Coeff.	t-value	
(Intercept)	4.7746	143.8120	***
ADV	-0.0022	-3.0870	**
YEE1M	-0.0775	-4.9240	***
CLASS_L	-0.1662	-5.1120	***
CLASS_B	-0.2284	-5.8170	***
CLASS_T	-0.8067	-26.2130	***
CLASS_Q	-0.8931	-25.4620	***
CLASS_X	-1.2646	-41.7270	***
CLASS_Z	-1.3955	-45.2150	***
Seats	-0.0088	-1.8030	.
Seats*sin(2πt/60)	0.0029	2.1740	*
Seats*cos(2πt/60)	0.0003	1.1750	
Seats*sin(4πt/60)	0.0017	1.3490	
Seats*cos(4πt/60)	0.0017	2.0520	*
Seats*sin(6πt/60)	0.0008	2.4160	*
Seats*cos(6πt/60)	0.0001	0.6490	
Seats*sin(8πt/60)	-0.0018	-2.2130	*
Seats*cos(8πt/60)	0.0030	1.7540	.
Seats*sin(2πt/60)2	-0.0011	-2.2220	*
Seats*cos(2πt/60)2	-0.0001	-1.1690	
Seats*sin(4πt/60)2	0.0001	0.6020	
Seats*cos(4πt/60)2	-0.0003	-2.1880	*
Seats*sin(6πt/60)2	-0.0002	-2.1650	*
Seats*cos(6πt/60)2	0.0000	-0.3470	
Seats*sin(8πt/60)2	0.0003	2.4290	*
Seats*cos(8πt/60)2	-0.0088	-1.8030	.
Observations		2,534	
R-squared		0.8754	
Adj. R-squared		0.8742	

Note: *** represents 1% level; ** represents 5% level; * represents 10% level

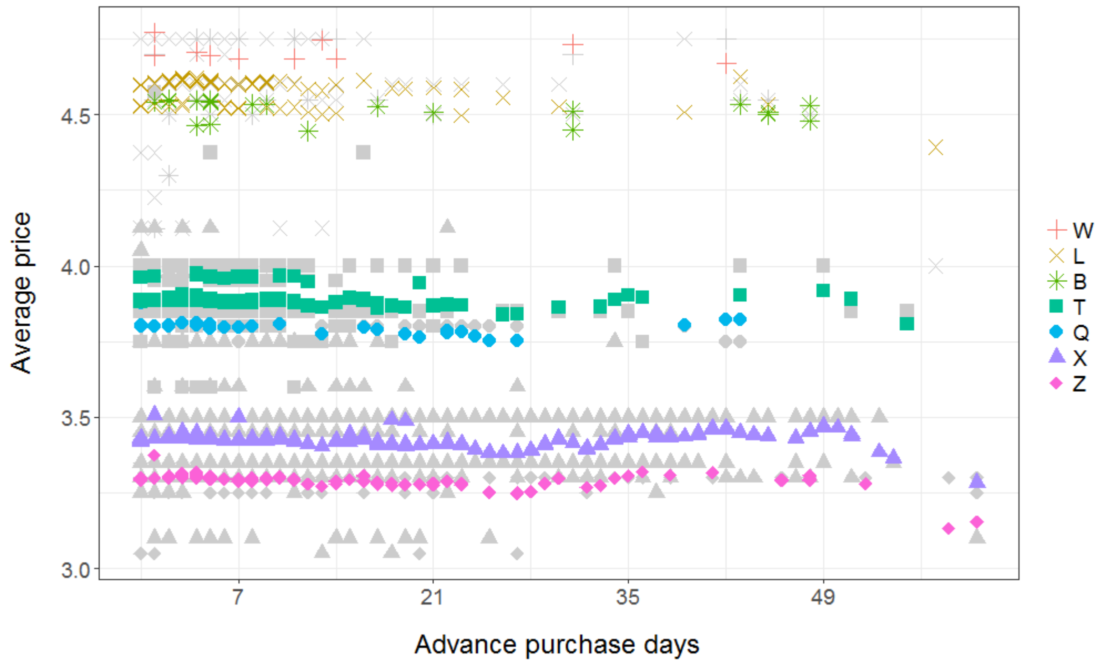


Figure 5-9: The observed and fitted purchase curve of purchase price

For the choice model estimation, GAUSS software (Aptech Systems, 1995) was used to estimate the continuous logit model with the maximum likelihood method. Table 5-8 shows the estimation results for continuous choice model whereas Figure 5-10 depicts the actual and estimated arrival curves for advance purchase passengers. Figure 5-10(a) shows that the estimated arrival curve is close to the actual one, suggesting a good model fit. From the estimation results, most of the selected explanatory variables are significantly tested with an expected sign. The time-variant variable, PRICE, has a negative effect of -0.292 on advance purchase time decision, which suggests that, as price increases, the passenger utility decreases. Thus, the probability of making purchase at time t decreases. That is, passengers could possibly delay their purchase due to the higher purchase price, and vice versa. These modeling results correspond with previous RM research showing that airlines can use price strategies to shift demand.

For the interacting effects of price with other time-invariant variables, including CV, time of day (MORNING and AFTERNOON), day of week (FRIDAY), month of year (PEAKSEASON) and VACATION indicators, are not intuitive to be interpreted. Accordingly, Figure 5-10(b)-(f) depict the predicted arrival curves of advance purchase

Table 5-8: Estimated continuous logit model

Variable	Coeff.	t-value	
PRICE	-0.292	-16.612	***
Departure time function			
sin(2πt/60)	-0.544	-2.464	**
cos(2πt/60)	0.613	4.576	***
sin(4πt/60)	-0.318	-2.492	**
cos(4πt/60)	0.341	2.602	***
sin(6πt/60)	-0.188	-3.334	***
cos(6πt/60)	-0.043	-0.658	
CV			
sin(2πt/60)	-0.296	-2.349	**
cos(2πt/60)	0.063	0.588	
sin(4πt/60)	-0.132	-1.305	
cos(4πt/60)	-0.243	-2.723	***
Morning flight indicator interaction			
sin(2πt/60)	-0.007	-0.031	
cos(2πt/60)	-0.588	-4.521	***
sin(4πt/60)	0.336	2.640	***
cos(4πt/60)	-0.442	-3.452	***
Afternoon flight indicator interaction			
sin(2πt/60)	0.264	1.390	
cos(2πt/60)	-0.388	-2.888	***
sin(4πt/60)	0.291	2.050	**
cos(4πt/60)	-0.405	-3.102	***
sin(6πt/60)	0.362	4.289	***
cos(6πt/60)	-0.068	-0.795	
Friday indicator interaction			
sin(2πt/60)	0.395	3.380	***
cos(2πt/60)	-0.316	-3.984	***
Peak season indicator interaction			
sin(2πt/60)	0.338	3.158	***
cos(2πt/60)	-0.231	-3.074	***
Vacation indicator interaction			
sin(2πt/60)	-0.953	-7.540	***
cos(2πt/60)	-0.324	-2.948	***
sin(4πt/60)	-0.365	-3.060	***
cos(4πt/60)	-0.123	-1.220	

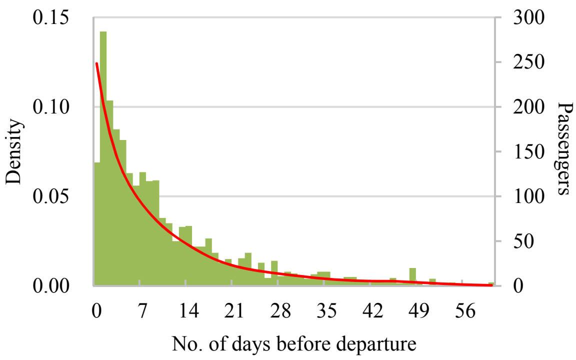
Note: *** represents 1% level; ** represents 5% level; * represents 10% level passengers under four time-invariant variables, where Figure 5-10(b) presents the predicted arrival curves for different price uncertainty within the purchase month. The density curves are identical before 21 days before departure day, suggesting the price uncertainty has minor effect on advance purchase time. The predicted values of high-CV are lower than low-CV within 10 days prior to departure, implying that passenger tend to delay purchases when price fluctuation is high. A possible explanation is that, as the departure day nears, only the expensive fare products are available for passengers. Passengers may wait for cheaper flight tickets and thus results in delay purchasing.

Figure 5-10(c) further shows the advance purchase time decision for passengers choosing morning, afternoon or evening flights. In terms of density before 2 weeks prior to departure, the predicted values for morning flights (solid curve) are higher than those of other two flights. These results imply that passengers who prefer morning flights tend to purchase tickets earlier. In contrast, passengers preferring evening flights have a steeper advance purchase curve, especially within 10 days before departure, indicating the passengers of evening flights tend to purchase tickets much later. Airlines may yield a low-high pricing mechanism for evening flights, and high-low pricing strategy for passengers who prefer morning flights.

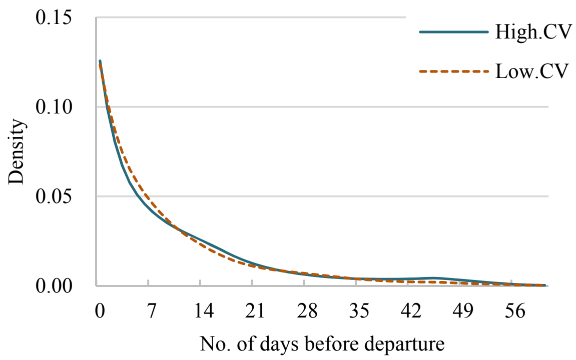
Meanwhile, Figure 5-10(d) depicts advance purchase time choice differences for the Friday flight. Figure 5-10(e) shows how the advance purchase time decision changes in the peak-season (flights in July and August). Finally, Figure 5-10(f) demonstrates the predicted arrival curve for traveling in consecutive holidays more than three days. As expected, passengers tend to purchase earlier for the Friday and peak-season fights (the solid curves). In contrast, the non-Friday and non-peak season (the dashed curves) have relatively lower density value but increase sharply within 21 days before departure, which suggests that passengers tend to purchase later. Airline could create promotion plans for weekdays and non-peak flights around 3 weeks to induce passengers to purchase earlier. Those results also correspond to the phenomenon of that leisure passengers usually plan their trips in advance and purchase flight tickets earlier whereas business travelers tend to purchase tickets close to the departure date.

The predicted density curve of vacation flights in Figure 5-10(f) shows the similar early arrival patterns as those of Friday and peak-season flights, but the curve surprisingly becomes much steeper within approximately seven days before departure. The modeling results suggest that passengers who travel on consecutive holidays tend to make their purchases earlier (planned travelers) and there are also some passengers prefer to purchase tickets close to the departure date (spontaneous travelers). One possible explanation is that the study route is short-haul and has more spontaneous travelers than long-haul air routes. Hence, for the vacation short-haul flights, airlines can deploy different seat management strategies or appropriately raise the price for individual passengers as departure day approaches to gain extra revenue. We believe that the advance purchase behaviors for long-haul air routes should be different.

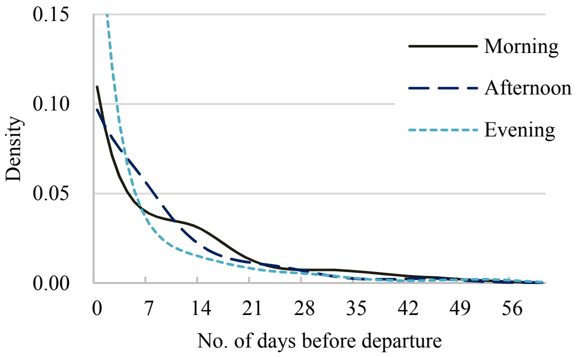
In sum, compared with the estimated results of aggregate advance purchase pattern model (as presented in Figure 5-6), the individual passengers are tend to purchase popular flights (Morning, Afternoon, Friday, Peak season and Vacation) earlier as expected. On the other hand, the value of predicted density is lower as departure day approaches, suggesting that the probability of purchasing popular flights is gradually decreasing. Interestingly, the predicted final load factor of vacation flights is not as high as expected (Figure 5-6d) which might be resulted from the revenue management operations. According to our study airlines, they usually reserve seats of vacation flights for group travelers rather than for individual passengers, and released seats approximately 1 weeks prior to departure day. Based on the strong increment of predicted density of vacation flight as shown in Figure 5-10(f), the seat-release timing could be earlier for the vacation flight individual passengers.



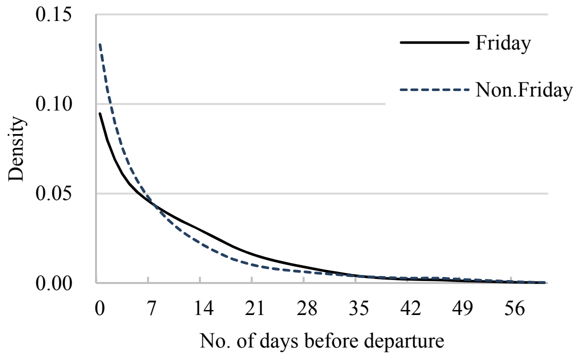
(a) Real and predicted arrival curves



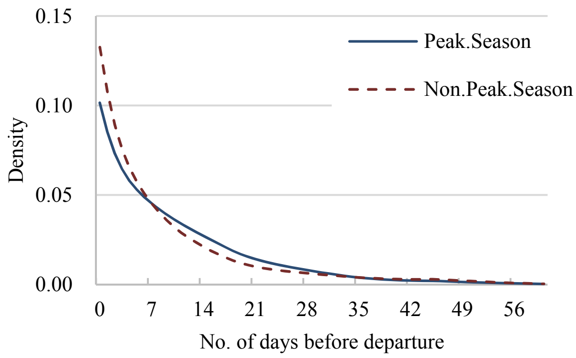
(b) Price CV effect



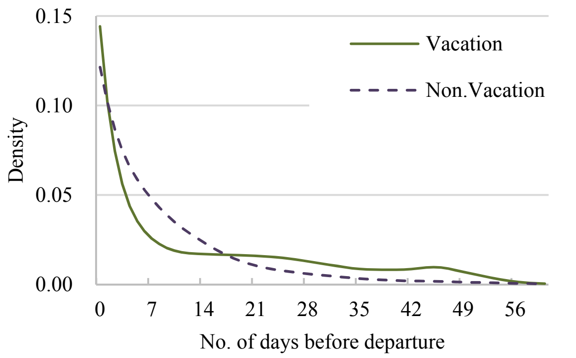
(c) Time-of-day effect



(d) Friday effect



(e) Peak season effect



(f) Vacation effect

Figure 5-10: The predicted arrival curves of advance purchase passengers

Chapter 6 Applications

In previous chapters, we have developed and estimated models for the advance purchase patterns and air passenger choice, methodologies including functional concurrent regression model, discrete multinomial logit model (MNL), and continuous choice were used. In this chapter, we discuss how the parameter estimates of both continuous models can be applied to monitor and forecast future advance purchase pattern for the selected flight and to predict passengers' purchase behaviors under different pricing strategies. In particular, we will show how these models can be used to support a wide range of short to medium term airline planning decisions. The rest of this chapter is organized as follows. The introduction of model application and flowchart is described in Section 6.1. An example scenario is introduced in Section 6.2.

6.1 Application Flowchart

Accurate advance purchase behaviors can provide valuable insights that can be used to support airline decision-making activities with respect to seat allocation, pricing, marketing and flight scheduling. The purposes of this study are aimed to identify the usual patterns of flights with different attributes, and to find a proper representation of advance purchase patterns for analyzing air passengers' advance purchase process. If we can identify and monitor the advance purchase pattern for the specific type of flight during the sales horizon, airlines are able to develop or make appropriate adjustments in time for a more effective sales strategy. For example, knowing accurate advance purchase patterns are able to allow airlines to know which and when do flights need to be promoted to increase sales. More importantly, to assist planners to evaluate potential impacts of the implementing strategy for increasing sales and revenue. This section will discuss these potential implementations. Figure 6-1 provides the flowchart for application at both the aggregate and individual levels of the proposed models.

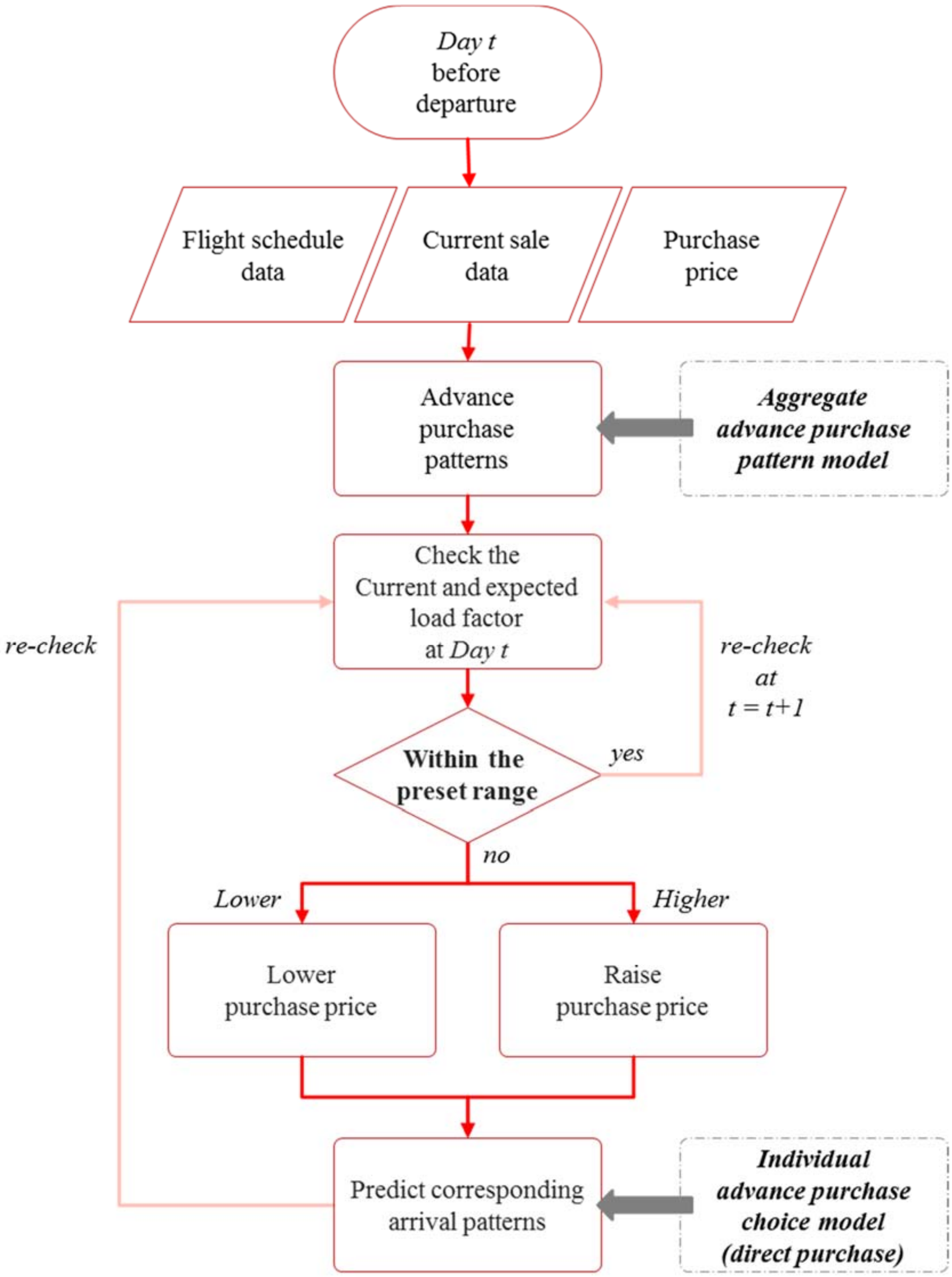


Figure 6-1: Model application flowchart

As illustrated in Figure 6-1, in order to identify the various advance purchase patterns of flights with different characteristics, datasets including flight schedule, current sales and purchase prices is collected at advance purchase day t. To investigate and predict the aggregate advance purchase patterns of flights, a functional concurrent model was firstly proposed. The aggregate pattern model provides a mechanism for us to analyze the relationship of factors that contribute to the heterogeneous patterns of distinct types of flights and forecast the future advance purchase pattern based on historical load factors. The output predicted load factors are then further monitored and compared with a preset goal, or with flights of similar attributes. Different pricing strategies should be used for different flights according to maximize sales volume and revenue. If the predicted values and patterns are within the expected range, indicating that the result meets the performance expectation and therefore we can keep observing and re-checking the sales status at the next period. If the predicted pattern is lower than prior expectation, an additional marketing promotion or discounts can be developed and implemented to the direct purchase channel to increase sales. On the contrary, if the predicted pattern is lower than prior expectation, airline could further consider raise the purchase price to gain an extra revenue.

Furthermore, for a clear understanding of how purchase price changes may influence the advance purchase behaviors, the individual advance purchase choice model was developed. The proposed continuous logit model has shown that advance purchase behaviors are significantly affected by price, price uncertainty, time of day (morning, afternoon and evening flight), days of week (flight on Friday), months of year (peak or off-peak seasons), and consecutive holiday. The predictive densities of the individual choice model not only allow us to observe and explain passengers' advance purchase decisions based on the purchase price and departure time preferences, but also can be used to evaluate pricing policy influences for airlines. Finally, after predicting corresponding arrival patterns of individual passengers, the updated load factors could be able to compare with a preset goal. The design process should be repeat and to refine or improve the solution if necessary.

6.2 Application Scenarios

Follow the flowchart presented in Figure 6-1, an example scenario was developed in this section. The basic idea of proposed framework is to provide a mechanism for airlines to identify the sales/advance purchase patterns of specific type of flights during any time of the sales horizon. In doing so, historical load factors and flight characters were used. The proposed model is able to monitor and forecast the future advance purchase pattern of study flights. If the preset goal were failed to reach preset targets, different pricing and marketing strategies should be altered to increase the advance purchase probability of customers. A continuous logit model was used here to analysis the advance purchase behaviors of air passengers. According to the predicted densities, the choice model allows airlines to evaluate pricing policy influences and support the development of more effective strategies.

The application scenario present in this session will be used to provide more detailed descriptions of how the proposed framework for advance purchase patterns and passengers' behaviors by providing suitable examples. Here, we assume there is a non-vacation, non-peak season, morning flight at 28 days prior to departure. Based on aggregate pattern model presented in Chapter 3.1, the historical load factors of 7 days (29~35 days) before current day were collected to forecast the advance purchase pattern for next 7 days (22~28 days). The historical load factors for the scenario testing were randomly selected from the validation dataset. The model input variables for the aggregate advance purchase pattern model are summarized as Table 6-1.

Table 6-1: Descriptive statistics for the design scenario

Variables	Description
Historical LF	Historical load factor of 29~35 days before current day
Morning	Morning, 1
Afternoon	Non-afternoon, 0
Friday	Friday, 0
Peak Season	Non-peak season, 0
Vacation	Non-vacation, 0

Figure 6-2 depicts the predicted aggregate advance purchase pattern for simulation scenario. Based on the aggregate pattern model, the "*" points marked on Figure 6-2 represent the seven historical load factors of 29~35 days before current day (day 28) that used for the prediction. The black triangles "▲" are the aggregate pattern model forecasting results for 22~28 advance purchase days. In addition, to predict future pattern before 21 days prior to departure, an additional rolling-window prediction technique based on the aggregate pattern model prediction results is used. The results of rolling prediction are marked with "•" symbol.

Moreover, in order to monitor the sales performance, the confidence bounds are used on the propose flow. The confidence bounds can be seen as the threshold values for the adjustment of current ongoing strategies. Of course, the preset monitoring thresholds are based on different requirements of airlines. The example provided here is a reference to describe how our propose models works. The solid line presents the mean average advance purchase pattern based on the predicted functional curves of the morning flight at non-Friday, non-vacation and non-peak season. The dashed lines serve as pointwise 95% confidence intervals by adding ±2 standard errors of functional values at each observation points. The mean average pattern curve presented here had eliminated the effect of estimating load factors to present the general condition for the specific type of flights.

Both historical and predicted load factors presented in Figure 6-2 are slightly below the average advance purchase pattern and the preset threshold. The rolling prediction results of observing flights also shows that the future advance purchase pattern will be lower than the average performance. Additionally, the simulation result indicates that the final forecasting load factor would be 0.6733, and that is lower than the average load factor of 0.6963, if the current sales or marketing strategy for the flight remains unchanged. Based on the simulation results of aggregate pattern model, airlines can develop more efficient sales strategies or adjust purchase prices in time to increase sales volume.

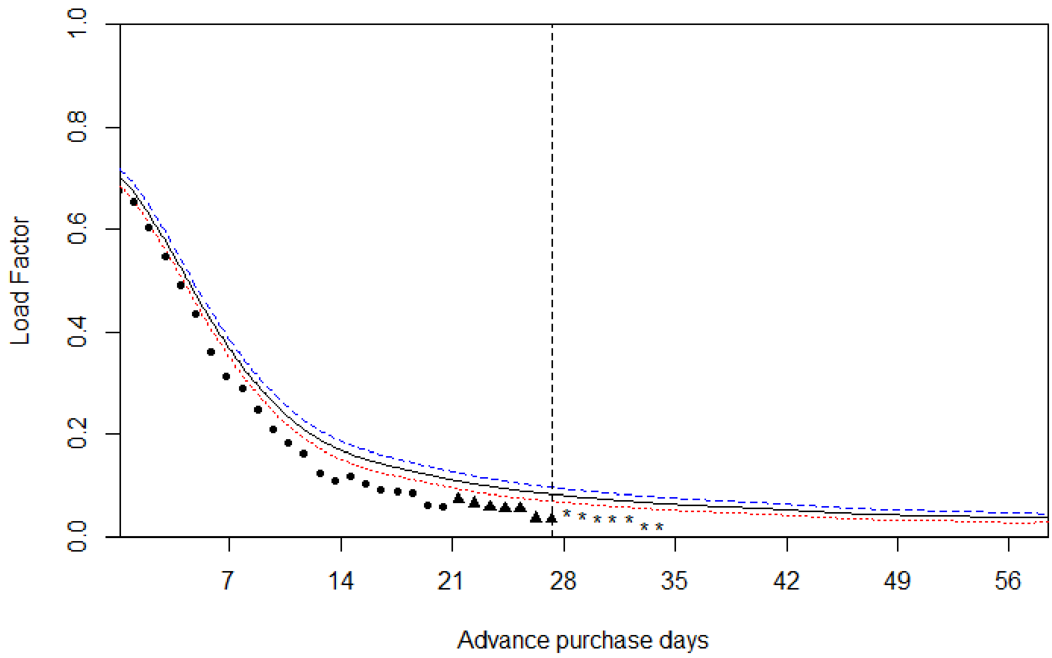


Figure 6-2: Aggregate advance purchase pattern for simulation scenario

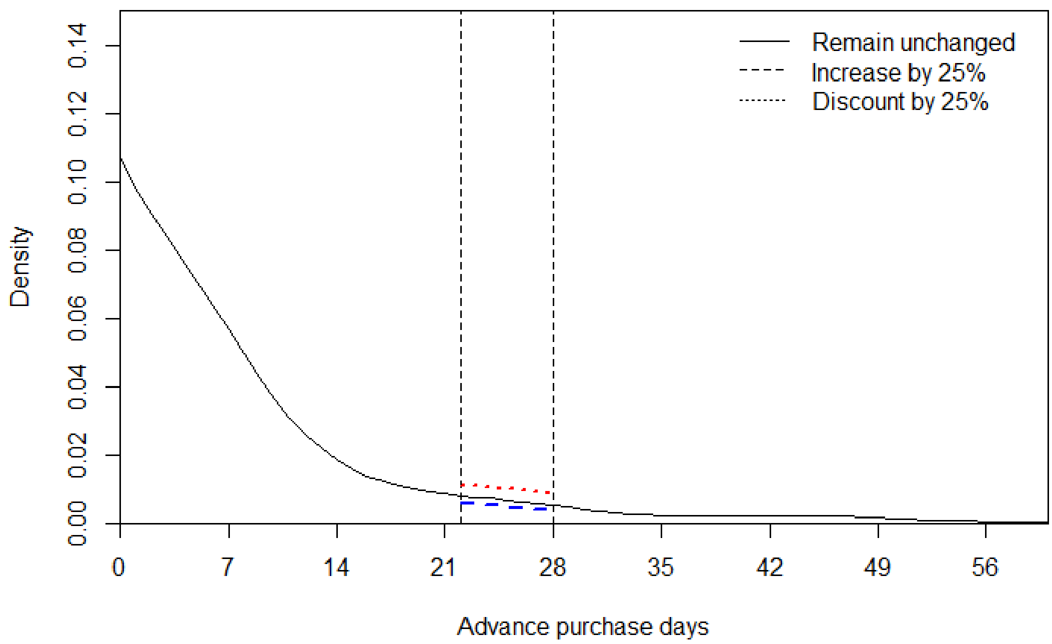


Figure 6-3: Advance purchase time distributions of pricing scenarios

However, passengers can also decide to make advance purchase at the ongoing price or choose to delay their purchase decision. Those price changes may force passengers to make trade-offs between price and flight attributes of desired departure time, and therefore, change their purchasing behaviors. In this research, a continuous choice model had proposed to empirically investigated advance purchase behaviors of air passengers considering both price and departure time preferences. Follow the scenario, a simple price-setting example with two different strategies is provided.

The first price adjustment policy assumes a purchase price increase of 25%, and the second one assumes a purchase price decrease of 25%. Figure 6-3 further presents the estimated densities for the individual air passenger. The dashed lines represents the simulate result of the first price adjustment policy of increasing purchase price by 25%. The probability of an individual passenger making purchases during from 22 to 28 days prior to departure will decrease from 4.639% to 3.041%. On the other hand, the dotted line shows if the purchase price had 25% discount, the probability of advance purchase will increase from 4.639% to 6.669%, suggesting about a 2% improvement. The predictive densities in the above two price adjustment policy show the expected effects on air passenger advance purchase time decisions, and can be used to evaluate pricing policy influences by airlines. Figure 6-4 further shows the predicted aggregate advance purchase pattern for the original prediction and 25% price discount scenarios. After refining the sales or marketing strategy, airlines should re-check load factor at the next advance purchase day *t-1* and repeat the process to refine or improve the solution.

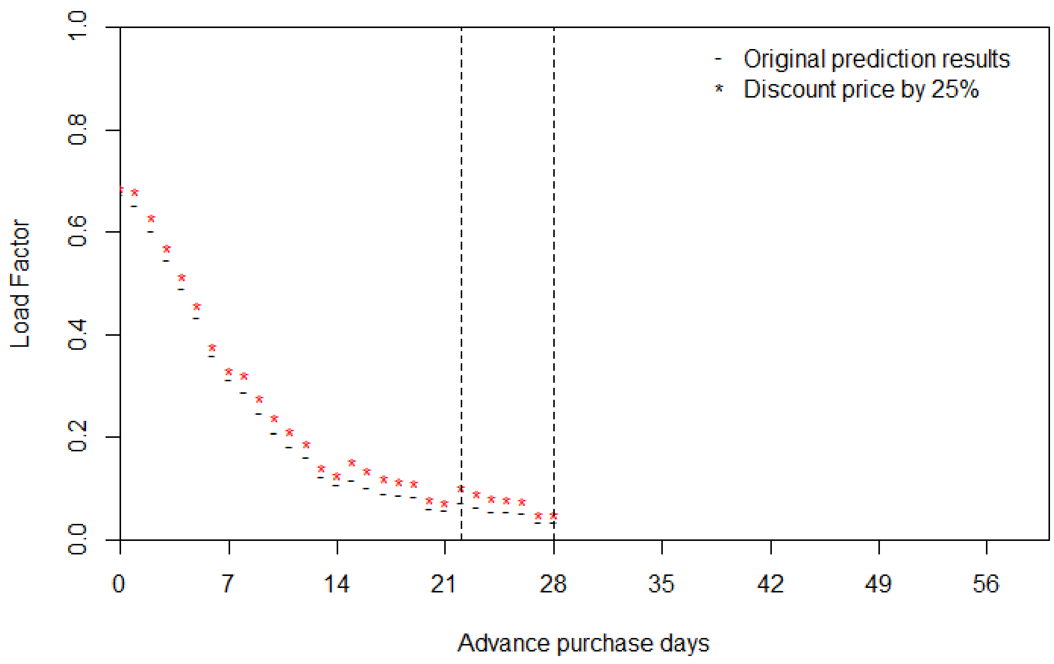


Figure 6-4: Predicted aggregate advance purchase patterns

Chapter 7 Concluding Remarks

As mentioned earlier, a better understanding of passenger choice behavior is crucial to support the decision making process of airline. The objectives of this research were to propose an approach for identifying advance purchase patterns and exploring their characteristics, and to propose an approach for examining the advance purchase behaviors of air passengers. This chapter concludes this dissertation by summarizing the important findings of this research. The summary of the work performed in this research was described in Section 7.1. Recommendations for further research were drawn in Section 7.2.

7.1 Summary

7.1.1 Contributions to practice

The primary contributions related to practices with regarding to advance purchase patterns of air passengers are summarized in the following points:

1. The proposed aggregate advance purchase pattern model is aimed to identify and forecast the advance purchase patterns of various types of flights during any time of the sales horizon. The predicted advance purchase levels at a specific time of a specific flight based on historical transaction data can be viewed as a reference level in comparing with current sales data so as to dynamically advise pricing and promotion strategies prior to departure. Additionally, according to our analyses on the air ticket transaction data, it is found that advance purchase patterns differ remarkably across flights. To explore flight advance purchase patterns, ***functional data analysis*** (FDA) techniques are used. Several factors contributing to aggregate advance purchase patterns of various types of flights including flight schedule attributes (such as time of day, day of week, months of year and special vacations) and historical load factors are examined.

2. A ***functional concurrent regression model*** was used to investigate the effects

of abovementioned variables on the shape of the advance purchase pattern curve. The empirical results has showed how abovementioned variables affected advance purchase patterns. In terms of time-varying coefficient, the coefficient of morning flight shows more negative effects than noon flights, but both turn to positive effects around 2 weeks before departure, indicating passengers who preferred morning and noon flights tend to purchase later. The estimated coefficients of Friday, peak-season and vacation flights have increasing positive effects on load factor throughout the observation interval, suggesting those flights have gradually increasing effect as the departure day approaches. The coefficients for special vacation has a decreasing negative effect toward the departure day, suggesting passengers tend to purchase early for vacation trips and may hardly purchase tickets on the departure day.

3. With better learning of advance purchase patterns for sales flights, airlines are able to develop and make appropriate adjustments for current strategy more efficiently and compete more effectively in today's marketplace. For example, based on the estimated result of the proposed aggregate pattern model, the popular flights such as morning, afternoon, Friday and peak-season flights have identical sharp upward advance purchase patterns around 2 weeks prior to departure. Airlines may offer advance purchase discount to induce passengers to purchase earlier and further to reduce the rationing risk. On the other hand, around 2 weeks before departure, airlines can raise the purchase price for late-purchasing passengers to gain extra revenue.

4. Furthermore, the advance purchase behaviors of individual air passengers are considered. As airlines adjust prices and sales strategy dynamically based on learning patterns, passengers can also decide to purchase at the going price or choose to delay their purchase decisions. The ***discrete choice model*** was firstly used to explore which choice set construct scheme based on advance purchase time that can estimate choice model well. To facilitate model development, the advance purchase horizon is divided into five time periods according to three segmentation methods, including equal time periods, time

periods with equal number of purchases and time periods according to subjective judgments. Explanatory variables including price, flight schedule (time of day, day of week, and months of year) and fare class preferences are examined. In terms of adjusted rho square, AIC value and log-likelihood statistics, the equal time segmentation performs best. Based on the estimated coefficients, the log price has negative effect on advance purchase, suggesting the higher price, the lower utility of passengers. The interaction terms for peak season and vacation have positive effects, implying that the passengers' disutility of price is lower when passengers travel during vacation time. As for dummy variables, the morning and Friday flights show the expected decreasing pattern as departure day approaches, indicating that passengers prefer morning and Friday flights generally purchase ticket earlier in advance. The irregular pattern of vacation variable reflects the behavior of the airline strategies by holding and releasing seats for vacation flights. Additionally, based on the direct and cross-elasticity analysis, the extent of advance purchase behavior with respect to price strategy is also revealed.

5. However, some choices are continuous response variables, to arbitrarily discretize these continuous choices variables may lead to an erroneous result. Different discrete interval settings would also lead to different and unstable estimation results. The ***continuous logit model*** has advantage of treating advance purchase time in a continuous setting and offers theoretical supports based on the random utility. Therefore, a continuous logit model was further proposed in this research for empirical analysis of the advance purchase behaviors of air passengers. Additionally, an additional ordinary least squares regression model was also used to obtain time-varying purchase prices for continuous logit model estimation. The study identified several contributing factors in flight ticket prices set by the studied airline, including the advance purchase days, maximum stay, fare class, percentage of group passengers, and the number of remaining seats at the time of purchase, whereas the purchase price, time of day (morning, afternoon and evening flight), days of week (flight on Friday), months of year (flights in peak season) and

consecutive holiday preferences of air passengers are then examined by the continuous logit model.

6. The estimated results show that as the departure day nears, passengers may delay their purchases if prices widely vary. The individual passenger who preferring morning or afternoon flights tends to purchase ticket earlier, while passengers of evening flights appear to purchase flight tickets within 10 days before departure. Passengers also make advance purchase earlier for the Friday and peak season. For the effect of vacations on short-haul flights, some passengers who travel on consecutive holidays prefer to make their purchases earlier (planned travelers), but some in contrast are likely to purchase tickets close to the departure date (spontaneous travelers). By modeling both price and departure time preferences of air passengers, the individual choice model developed in this research is expected to offer a rich behavioral interpretation of advance purchase behaviors and allow airlines to evaluate potential impacts of the implementing strategies.

7. Finally, a ***scenario application*** example is provided to show how the proposed models can be applied to monitor and forecast future advance purchase pattern for the selected flight, and to evaluate passengers' purchase behaviors under different pricing and promotion strategies. The basic idea of proposed framework is to provide a mechanism for airlines to identify the sales/advance purchase patterns of specific type of flights during any time of the sales horizon. The aggregate pattern model is able to monitor and forecast the future advance purchase pattern of study flights. If the preset goal were failed to reach preset targets, different pricing and marketing strategies should be altered to increase the advance purchase probability of customers. The individual choice model is used for analyzing the advance purchase behaviors of air passengers. According to the predicted densities of choice model, airlines are allowed to evaluate pricing policy influences and support the development of more effective strategies. The models developed in this research have the potential to both improve existing applications in seat

allocation and extend the scope of applications to other areas of airline planning such as pricing and revenue management. The proposed framework may not only apply to the airline industry, but also to other online sales markets for perishable products, such as train, hotel rooms, car rentals, and entertainment and sporting events.

7.1.2 Contributions to methodologies

The primary contributions related to methodologies in this study are summarized in the following points:

1. This study is a new attempt to apply functional data analysis for advance purchase pattern analysis. This study proposed a functional concurrent regression model which could effectively examine the characteristics that contribute to the heterogeneous advance purchase patterns of distinct types of flights and air passengers. The estimation results indicated that significant differences exist between all variables and advance purchase patterns differ remarkably across flights. These characteristics also provide valuable insights into air passengers' behaviors and can be used to support airline decision-making activities.

2. To empirically investigate of the advance purchase behaviors of air passengers, the individual advance purchase choice models that accounting for departure time preferences heterogeneity are developed. Instead of segmenting passengers by trip purpose or by socio-economic attributes of air passengers, which is not available in the transaction data, advance purchase time horizons are used to classify passenger into groups. With this approach, passenger segments can be identified, and the differences in their behavioral preferences among groups can be captured.

3. Moreover, advance purchase behaviors were modeled using an easily acquired and continuously growing transaction dataset to prevent the cost of a large-scale questionnaire survey. With the growing revenue share of online

purchasing, to estimate and predict the advance purchase behaviors on individual passenger directly based on transaction data is believed much intuitive, cost economic, and representative.

4. In particular, to facilitate model development, the advance purchase choice of individual passenger is modeled from both discrete and continuous approaches. The discrete approach had divided the advance purchase horizon into five time periods according to three segmentation methods. On the contrary, the continuous logit model is advantageous at treating advance purchase time in a continuous fashion and offers strong theoretical supports based on the random utility theory. The study shown that different discrete interval settings would lead to different and unstable estimation errors, whereas a continuous logit model can provide reasonable results and allow more intuitive interpretations of advance purchase behaviors.

7.2 Recommendations

This analysis clarified the advance purchase behaviors of air passengers. The following recommendations reflect the need for an improved understanding of complex advance purchase behaviors.

1. The study is based on a short-haul air route (Taipei-Macau), the advance purchase behaviors of long-haul air routes are believed to be remarkably different. The comparisons deserve for further study.

2. For the aggregate advance purchase model, the advance purchase curves are diverse in shape and the standard deviation of load factors increase sharply as the departure day approaches. Airlines might adjust price and sales strategy based on changing load factors of different flights in practice. Future studies can apply functional cluster analysis that clustering smoothed curves based on the different advance purchase patterns. It will be desirable to categorize the flights in advance of estimating the functional regression model for better estimated and prediction results.

3. For the individual advance purchase model, the online ticket transactions data investigated in this study only account for low percentage of whole advance purchase records, but with a growing share of online sales in airline revenues, the proposed model can be used to examine a larger dataset for advance purchase behaviors.

4. Since the dataset does not contain the socio-economic variables and trip characteristics of air passengers, similar models that consider more valuable explanatory variables should be estimated based on a questionnaire survey on air passengers so as to draw more policy implications for RM strategies.

5. The discrete logit model proposed in this study is aimed to explore which choice set construct scheme that can estimate choice model well. The multinomial logit model (MNL) was applied due to its simple estimation and strong assumptions based on the random utility framework. Future researches can apply more advanced model such as nested logit model (NL) or generalized extreme value (GEV) models that can alleviate the IIA problem of the standard logit model.

6. The average price regression methodology presented in the continuous logit model is only a substitute until fully dynamic models are developed. The regression model could be improved by incorporating dynamic pricing models for a better reflection of airline competition (Bilotkach et al., 2010) and the price change in the advance purchase time variation (Escobari, 2012; Deneckere and Peck, 2012).

7. The dataset used in this study did not enable analysis of choices among alternative flights and carriers nor “not fly” alternative. Future studies could consider other data sources such as web searches log data that provide information of all options is available for better model estimation. To date, Escobari and Mellado (2014b) have empirically studied advance purchase behaviors of air tickets in a dynamic setting with revealed preference data where the information of all options is available.

References

Abou-Zeid, M., Rossi, T. and Gardner, B. 2006. Modeling time-of-day choice in context of tour- and activity-based models. Transportation Research Record 1981, 42–49.

Aptech Systems, 1995. GAUSS Applications: Constrained Maximum Likelihood. Aptech Systems, Inc., Maple Valley, WA.

Barnhart, C. and Smith, B. 2012. Quantitative problem solving methods in the airline industry - Driven demand and operations management models. Springer Science & Business Media, NY.

Ben-Akiva, M., Abou-Zeid, M., 2013. Methodological issues in modelling time-of-travel preferences. Transportmetrica A: Transport Science 9(9), 846–859.

Ben-Akiva, M., Litinas, N. and Tsunokawa, K. 1985. Continuous spatial choice: The continuous logit model and distributions of trips and urban densities. Transportation Research Part A: General 19 (2), 119–154.

Ben-Akiva, M., Watanatada, T., 1981. Application of a continuous spatial choice logit model. In Structural Analysis of Discrete Choice Data with Econometric Applications (Manski, C.F., McFadden, D. eds.), MIT Press, Cambridge, MA, 320-343.

Bhat, C.R., Steed, J.L. 2002. A continuous-time model of departure time choice for urban shopping trips. Transportation Research Part B: Methodological 36 (3), 207–224.

Bilotkach, V., Gorodnichenko, Y., and Talavera, O., 2010. Are airlines' price-setting strategies different? Journal of Air Transport Management 16(1), 1–6.

Carrier, E., 2008. Modeling the choice of an airline itinerary and fare product using booking and seat availability data. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology. Available at: http://agifors.org/award/submissions2008/Carrier_paper.pdf.

Chiou, J. M., 2012. Dynamical functional prediction and classification, with application to traffic flow prediction. The Annals of Applied Statistics 6, 1588–1614.

Chiou, J. M., Zhang, Y. C., Chen, W. H. and Chang, C.-W., 2014. A functional data approach to missing value imputation and outlier detection for traffic flow data.

Transportmetrica B: Transport Dynamics. 2(2), 106–129.

Chiou, Y.C., Liu, C.H., 2016a. Advance purchase behaviors of air tickets. Journal of Air Transport Management 57, 62–69.

Chiou, Y.C., Liu, C.H., 2016b. Advance purchase behaviors of air passengers: A continuous logit model. Transportation Research Part E: Logistics and Transportation Review 93, 474–484.

Clarkson, D. B., Fraley, C., Gu, C. C. and Ramsey, J. O., 2005, S+Functional data analysis user's manual for Windows. New York: Springer.

Dana, J. D., 1998. Advance-purchase discounts and price discrimination in competitive markets. Journal of Political Economy 106(2), 395.

Dana, J. D., 1999. Using yield management to shift demand when the peak time is unknown. Rand Journal of Economics 30(3), 456–474.

Dass M., Jank W. and Shmueli G., 2011. Maximizing bidder surplus in simultaneous online art auctions via dynamic forecasting. International Journal of Forecasting, 27(4), 1259–1270.

Dass, M., Shropshire, C., 2012. Introducing functional data analysis to managerial science. Organization Research Methods 15, 693–721.

Deneckere, R., Peck, J., 2012. Dynamic competition with random demand and costless search: A theory of price posting. Econometrica 80(3), 1185–1247.

Escobari, D., 2012. Dynamic pricing, advance sales and aggregate demand learning in airlines. The Journal of Industrial Economics 60(4), 697–724.

Escobari, D., 2014. Estimating dynamic demand for airlines. Economics Letters 124(1), 26–29.

Escobari, D., Mellado C., 2014, The choice of airport, airline, and departure date and time: Estimating the demand for flights, in James Peoples (ed.) The Economics of International Airline Transport (Advances in Airline Economics, Volume 4) Emerald Group Publishing Limited, 177 – 198

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis. New York: Springer.

Gale, I.L., Holmes, T.J. 1993. Advance-purchase discounts and monopoly allocation of capacity. The American Economic Review 83 (1), 135–146.

Gallego, G., Kou, S. G. and Phillips, R., 2008. Revenue management of callable products. Management Science 54(3), 550–564.

Gao, H.O., Niemeier, D. A., 2008. Using functional data analysis of diurnal ozone and NOx cycles to inform transportation emissions control. Transportation Research Part D: Transport and Environment 13, 221–238.

Garrow, L. 2009. Online travel data: A goldmine of new opportunities. Journal of Revenue and Pricing Management 8 (2-3), 247–254.

Gastón, M., León, T. and Mallor, F., 2008. Functional data analysis for non-homogeneous poisson processes, in: Proceedings of the 40th Conference on Winter Simulation, WSC '08. Winter Simulation Conference, 337–343.

Guardiola, I.G., Leon, T. and Mallor, F., 2014. A functional approach to monitor and recognize patterns of daily traffic profiles.Transportation Research Part B: Methodological 65, 119–136.

Hetrakul, P., Cirillo, C. 2013. Accommodating taste heterogeneity in railway passenger choice models based on internet booking data. Journal of Choice Modelling 6, 1–16.

Hetrakul, P., Cirillo, C. 2014. A latent class choice based model system for railway optimal pricing and seat allocation. Transportation Research Part E: Logistics and Transportation Review 61, 68–83.

Hetrakul, P., Cirillo, C. 2015. Customer heterogeneity in revenue management for railway services. Journal of Revenue and Pricing Management 14 (1), 28–49.

Hotle, S.L., Castillo, M., Garrow, L.A. and Higgins, M.J. 2015. The impact of advance purchase deadlines on airline consumers' search and purchase behaviors. Transportation Research Part A: Policy and Practice 82, 1–16.

Lemp, J.D., Kockelman, K.M. and Damien, P. 2010. The continuous cross-nested logit model: Formulation and application for departure time choice. Transportation Research Part B: Methodological 44 (5), 646–661.

Lemp, J.D., Kockelman, K.M., 2010. Empirical investigation of continuous logit for departure time choice with bayesian methods. Transportation Research Record 2165, 59–68.

Lewis, C. D., 1982. International and Business Forecasting Methods. London:

Butterworths.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior, in P. Zarembka (Ed.), Frontiers of Econometrics, Academic Press, 105–142

McGill, J. I., van Ryzin, G., 1999. Revenue management: Research overview and prospects. Transportation Science 33(2), 233–256.

Morris, J.S., 2015. Functional Regression. Annual Review of Statistics and Its Application 2, 321–359.

Newman, J.P., Ferguson, M.E., Garrow, L.A. and Jacobs, T.L. 2014. Estimation of choice-based models using sales data from a single firm. Manufacturing & Service Operations Management 16 (2), 184–197.

Popuri, Y., Ben-Akiva, M. and Proussaloglou, K. 2008. Time-of-day modeling in a tour-based context: tel aviv experience. Transportation Research Record 2076, 88–96.

Proussaloglou, K., Koppelman, F.S., 1999. The choice of air carrier, flight, and fare class. Journal of Air Transport Management 5, 193–201.

Ramsay, J., Hooker, G. and Graves, S., 2009. Functional Data Analysis with R and MATLAB. , New York: Springer.

Ramsay, J., Silverman, B., 2002. Applied functional data analysis, New York: Springer.

Ramsay, J., Silverman, B., 2005. Functional data analysis, New York: Springer.

Reddy, S.K. and Dass, M., 2006. Modeling on-line art auction dynamics using functional data analysis. Statistical Science 21, 179–193.

Sood, A., James, G.M. and Tellis, G.J., 2008. Functional Regression: A New Model for Predicting Market Penetration of New Products. Marketing Science 28, 36–51.

Suhaila, J., Yusop, Z., 2016. Spatial and temporal variabilities of rainfall data using functional data analysis. Theoretical and Applied Climatology, 1-14.

Talluri, K., van Ryzin, G. 2004a. Revenue management under a general discrete choice model of consumer behavior. Management Science 50 (1), 15–33.

Talluri, K., van Ryzin, G. 2004b. The Theory and Practice of Revenue Management, International Series in Operations Research and Management Science, Vol. 68, Springer Science & Business Media, NY.

Tastambekov, K., Puechmorel, S., Delahaye, D. and Rabut, C., 2015. Aircraft trajectory forecasting using local functional regression in Sobolev space. Transportation

Research Part C: Emerging Technologies 39, 1–22.

Ullah, S., Finch, C.F., 2013. Applications of functional data analysis: A systematic review. BMC Medical Research Methodology 13 (1), 43-55.

Vulcano, G., van Ryzin, G. and Chaar, W. 2010. Choice-based revenue management: An empirical study of estimation and optimization. Manufacturing & Service Operations Management 12 (3), 371–392.

Wang, S., Jank, W. and Shmueli, G., 2008. Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis. Journal of Business & Economic Statistics 26, 144–160.

Wen, C. H., Lai, S. C., 2010. Latent class models of international air carrier choice. Transportation Research Part E, 46(2), 211-221.

Curriculum Vitae

劉佳欣

Rafe C. H. Liu

snexuz@gmail.com | 0912-120811

***

⌘ Education

Ph.D. Candidate
Department of Transportation and Logistics Management
National Chiao Tung University
Taipei, Taiwan, Republic of China (2012 - Est. final defense: 2016/11)
Advisor: Prof. Yu-Chiun Chiou
Dissertation: Modeling Advance Purchase Behaviors of Air Passenger

MS
Major: Institute of Traffic and Transportation
Minor: Institute of Information Management
National Chiao Tung University
Taipei, Taiwan, Republic of China (2005 - 2007)
Advisor: Prof. Jiuh-Biing Sheu
Thesis: Consuming Emergency Relief Demand Forecasting

BS
Department of Transportation and Communication Management
Program: Telecommunication Program
National Cheng Kung University
Tainan, Taiwan, Republic of China (2001 - 2005)

⌘ Publication

Journal Papers

Chiou, Y.C., Liu, C.H., 2016. Advance purchase behaviors of air passengers: A

continuous logit model. Transportation Research Part E: Logistics and Transportation Review 93, 474–484. (Impact Factor: 2.279)

Chiou, Y.C., Liu, C.H., 2016. Advance purchase behaviors of air tickets. Journal of Air Transport Management. 57, 62–69. (Impact Factor: 1.084)

Chiou, Y.C., Liu, C.H., and Hsieh C.W., 2015. Key factors contributing to operating cost of short-haul domestic air routes. International Journal of Aviation Management 2, 3-4.

Conference Papers

Chiou, Y.C., Liu, C.H., 2016. Exploring flight advance purchase patterns by functional data analysis. Poster presentation at the 21st international conference of Hong Kong Society for Transportation Studies (HKSTS), December 10-12, Hong Kong.

倪如霖、邱裕鈞、劉佳欣，2016，「公共自行車租借量之影響因素分析-地理加權迴歸和函數資料分析方法之應用」，105 年中華民國運輸年會暨國際學術研討會，國立東華大學，12 月 8 日

Chiou, Y.C., Liu, C.H., 2015. Advance purchase behaviors of air tickets. Presented at the 19th Air Transport Research Society (ATRS) World Conference, July 2-5, Singapore.

Chiou, Y.C., Liu, C.H. and Hsieh, C.W., 2013, Key factors contributing to the total operating cost of Taiwan domestic air routes. Presented at the 17th Air Transport Research Society (ATRS) World Conference, June 27-30, Bergamo, Italy.

Reports

邱裕鈞，周榮昌，鍾易詩，謝志偉，陳孜穎，劉佳欣，白舜豪，林佳樺，張開國，葉祖宏，陳凱斌，田養民，賴靜慧，2013，「道路交通事故之能源消耗與碳排放量推估研究」，交通部交通運輸研究所 (編號：MOTC-IOT-102-SDB001)。

⌘ Experience

Individual Consultant, 2015 - Present

Focusing on data team organization and data analysis with R.

- ✧ **Enterprise consultant: Gogolook (2015, 6 months)**
  - Junior data analysts & data team training program
  - Dashboard renewal, Data project kick-off
- ✧ **Teaching experiences:**
  - Data Thinking with R - NTUST (2016, 24 hours, 3 classes)
  - Doing Research with R - NCTU (2016, 3 hours)

R Lecturer and Consultant at DSP Company, 2014 - Present

DSP provides data scientist training programs, consultant and data portal service for commercial and nonprofit organizations.

- ✧ **Front-end designer & engineer:**
  - Wordpress customizing, Companion page design (CSS, LESS)
  - Data portal theming, Multi-language plugins (Python, Pylons, Jinja2)
- ✧ **Sales and Marketing manager:**
  - Develop and implementing marketing strategies / campaigns.
  - Identify and coordinate with potential strategic partners for business developments.
  - Cofounder of D4SG (Data for Social Goods) community.
- ✧ **Teaching experiences:**
  - Enterprise: Pixnet, CHT (2015)
  - Data Thinking with R - NCCU (2015, 2016), PU (2015)
  - Web Scraping with R - Institute of Sociology Academia Sinica (2015)

Senior Product and Marketing Manager at YESTRIP.com travel agency of China Airlines Group, November 2013 - May 2015

- ✧ **Product manager:**

1013.com.tw, the first ***<u>flight-hotel dynamic packaging service</u>*** in Taiwan.

- Responsible for developing, modifying and maintaining a brain-new travel site, in charging for website launching, marketing and sales planning.
- Managing the entire product line life cycle from website development to tactical marketing activities.
- Operation team recruitment, training and leading.
- Cover operating cost (salary and advertisements) <u>after launching 3 months.</u>
- The project was retrieved by China Airlines as "Dynasty lite" and "Backpacker holiday" (for travel agents) services.

✧ **Marketing manager:**

- Working closely with the all departments to develop company-wide sales, marketing and business development strategies.
- Support graphic design and campaign site development works.
- Organize and support travel show expo (over 30 expos).

***

<u>Marketing Specialist</u> at TransAsia Airways, 2010 – Aug 2012

- Responsible for maintaining, marketing and online sales of TransAsia.com.
- Develop and implementing online marketing, advertising and sales plans based on monthly transaction analysis.
- Social media and customer relationship management.
- Surpass historical online sales record (2012/3)

<u>Office Manager</u> at Fu-Hsiung Restaurant Consultant Corp, 2008 - 2010

- Develop and implementing operation, marketing and sales plans.

- Organizing and developing standard operation procedure (SOP) documents.
- Point-of-Sales system managements, job including developing sales and marketing plans for franchises based on POS data.
- Support graphic design and campaign site development works.

<u>Project Manager</u> at 3S Pocketnet Tech Inc, 2007 - 2008

- Responsible for developing a LBS-based food and dining recommendation service site, in charging for website and mobile APP planning launching, marketing and sales planning.