An integrated Markov decision process and nested logit consumer response model of air ticket pricing

Citation for published version (APA):

DOI:
10.1080/23249935.2017.1306727

Document status and date:
Published: 03/07/2017

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 23. Aug. 2019
An integrated Markov decision process and nested logit consumer response model of air ticket pricing

Jing Lu, Tao Feng, Harry Timmermans & Zhongzhen Yang


To link to this article: http://dx.doi.org/10.1080/23249935.2017.1306727

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Accepted author version posted online: 20 Mar 2017.
Published online: 17 Apr 2017.

Submit your article to this journal

Article views: 182

View related articles

View Crossmark data
An integrated Markov decision process and nested logit consumer response model of air ticket pricing

Jing Lu\textsuperscript{a}, Tao Feng\textsuperscript{b}, Harry Timmermans\textsuperscript{b} and Zhongzhen Yang\textsuperscript{c}

\textsuperscript{a}College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, People’s Republic of China; \textsuperscript{b}Department of the Built Environment, Eindhoven University of Technology, Eindhoven, Netherlands; \textsuperscript{c}College of transportation management, Dalian maritime university, Dalian, People’s Republic of China

ABSTRACT

The paper attempts to propose an optimal air ticket pricing model during the booking horizon by taking into account passengers’ purchasing behavior of air tickets. A Markov decision process incorporating a nested logit consumer response model is established to modeling the dynamic pricing process. The proposed model is estimated and applied using the data collected in a multi-airport region with competition in China. Results indicate that, by considering air ticket purchasing behavior, air ticket price can be set dynamically and optically in response to the changes in exogenous factors which are not controlled by airlines.

ARTICLE HISTORY

Received 29 July 2015
Accepted 11 March 2017

KEYWORDS

Pricing strategy; air ticket purchasing behavior; Markov decision process; nested logit model; multi-airport area

1. Introduction

‘Yield-management’ in air ticket pricing is one of the most complicated pricing strategies (Gillen and Hazledine 2012). Basically, airlines adjust their ticket prices over time prior to the flight departure in an attempt to gain the maximum profit (Gallego and Hu 2014). A variety of factors, influencing ticket pricing, have been reported in the literature, including type of air route, flight time and the number of unsold seats (Zhang et al. 2014). Any pricing strategy model, which is based on these factors, has the advantage that all information is available in the booking system (Lin 2006). However, such models are established on the assumption that passengers only consider the factors which are controlled by airlines when purchasing their tickets.

In reality, however, previous research has shown that passengers’ air travel behavior is mainly influenced by air ticket price, flight time, access service, trip purpose and social demography (Pels, Nijkamp, and Rietveld 2003; Başar and Bhat 2004; Hess and Polak 2005; Lian and Ronnevik 2011; Wang et al. 2014). Passengers’ ticket purchasing behavior in reality, however, has been ignored for long in existing research about air ticket pricing (Zhang and Cooper 2005). As a result, inappropriate pricing strategies are commonly applied in case of failing to predict consumers’ response to the change of air ticket prices, leading to profit loss (Etzioni et al. 2003).
Therefore, this paper attempts to propose an improved air ticket pricing model which differs from earlier work in two important aspects: first, it takes into account passengers’ purchasing behavior in response to air ticket price; second, it explicitly considers the whole process of ticket purchase in the context of air travel. The modeling framework builds on the Markov decision process incorporating a value iteration algorithm. The process of price updating incorporates a dynamic mechanism with an emphasis of reflecting the real behavior of air ticket purchasing. Moreover, in order to further understand the purchasing behavior, a nested logit model is estimated using the data collected in a stated choice experiment. The proposed model is then applied in a multi-airport region to optimize the pricing strategy.

The remainder of this paper is as follows: the second section reviews the literatures on dynamic pricing and air travel choice behavior. The third section presents the dynamic pricing model, which incorporates the air ticket purchasing behavior. The fourth section presents the design of the stated choice experiment and the estimation of the air ticket purchasing model. Finally, the paper is summarized and concluded with the main findings and future works.

2. Literature review

Airlines, as an example of service industries, have some unique features that influence their business strategies and marketing: (i) seats cannot be sold once the flight takes off, (ii) the number of seats is fixed in advance and adding extra seats is impossible or expensive. It means that the marginal cost of increasing the inventory is extremely high (McAfee and Te Velde 2006). Moreover, Donovan (2005) pointed out that the traditional pricing in which airlines determine the price of each fare class in advance is less appropriate nowadays because of web sales. The reason is that passengers can easily compare the prices of all available flights and do their shopping quickly through Internet (Malighetti, Paleari, and Redondi 2015). Due to these features, most airlines set the air ticket price dynamically subject to a set of general pricing constraints, aiming at maximizing the profits of a flight (Sengupta and Wiggins 2014).

Dynamic pricing, which is the most important component of yield management techniques, can set highly flexible prices on different time points by taking into account current supply-and-demand conditions (Gallego and Van Ryzin 1997). The quintessence of yield management is to examine consumer behavior to achieve the maximum amount of profit from a particular good or service within some fixed time horizon. The modeling of consumer behavior determines a dynamic price setting such as to entice consumer to buy the service at a particular moment. The challenge here is on the dynamics between timing, pricing, and purchasing in order to achieve the maximum profit.

This approach has been reported to be effective in optimizing the price for industries that provide products with strict expiration-time limits and fixed inventory, such as hospitality, travel and retail (Gosavi, Bandla, and Das 2002). As an example, Davis (1994) argued that American airlines earned an extra 500 million dollars per year by using the dynamic pricing technique.

In the context of airlines’ dynamic pricing, Botimer (1996) studied the daily price issue during the booking horizon of one flight. He found that ticket price increases when the number of open seat is low but demand is high, thus avoiding profit waste resulting from
under-estimating the price that passengers are willing to pay. Reversely, when the current demand is low but supply is high, airlines intend to cut down the price to investigate whether tickets are over-priced. One of the problems here is that the demand for a flight is hard to predict.

As a potential solution of the demand forecast for passengers’ purchasing flow, dynamic pricing offers such an opportunity that it allows airlines to adjust their prices based on the actual supply–demand condition, with a purpose of maximizing profit for each seat. For instance, Feng and Gallego (2000) developed a dynamic pricing model based on the Markov decision process (MDP). By using MDP, airlines can decide whether to raise or reduce the price at the beginning of each period. The results of MDP can optimize the process of air ticket pricing and the pricing policy to maximize profit.

Here, the booking horizon in MDP structure of a flight is normally divided into several discrete time periods. At the beginning of each period (‘decision point’), unsold seats are counted to describe the situation, which is defined as ‘state’; relevant to a particular state, the ticket price will be set for this specific time period, which is defined as ‘action’. However, the effect of a certain price value on passengers’ purchasing behavior is unknown before pricing. For example, passengers will buy tickets in case of a low price but actually can accept a higher price. In case of a high price, passengers may switch to other air options or transportation modes. Therefore, airlines cannot predict the number of tickets and the relevant profit that can be earned at the end of a period (also called ‘reward’ in MDP). In MDP, this phenomenon can be described using the transition probability where the chance of transition from one state to another of a certain action is under controlled and the reward is not simultaneously predictable (Wu and Huang 2010). Due to the fact that the transition probability is uncontrolled in reality, airlines, who pursue the maximum profit across the whole booking horizon, need to find the best action to a certain state on every decision point, taking into account all probable future rewards.

Kulkarni et al. (2011) proposed an air ticket pricing model using a semi-Markov decision process (SMDP) instead of MDP, because the length of each decision epoch is set random in the SMDP. However, according to the analysis conducted by Zhang and Cooper (2005), the revenue loss by updating the price with certain pre-specified interval is very small in comparison to real-time pricing policies. Therefore, airlines are able to fix the interval between consecutive time points of price updating. Thus, these studies have shown evidence that the MDP is a suitable approach to model airlines’ pricing process.

Within the framework of MDP, the methods that have been developed to describe the air ticket purchases mainly refer to the ‘airline-determines’ method and the ‘passenger-determines’ method (McAfee and Te Velde 2006). The ‘airline-determines’ method assumes that passengers’ arrival time follows a Poisson distribution, and they should state their preferred price to the airlines before purchase. Then, airlines have to accept or reject passengers based on the passenger-stated price (Van Slyke and Young 2000). In contrast, in the ‘passenger-determines’ method, airlines set an air ticket price for different time periods during the booking horizon, and passengers arriving at different time points decide whether to purchase or not (Gallego and Van Ryzin 1997). Because the second method is more realistic than the first, the ‘passenger-determine’ method is adopted in this paper.

The ‘passenger-determines’ method has been used to address the airlines’ pricing strategy by taking into account air ticket purchasing behavior. Song, Yang, and Wang (2014) optimized the pricing policy considering the effects of price and flight time on passengers’
purchasing behavior. Results show that passengers are not only price-sensitive, but also responsive to the combined effect of price and flight time. Boyd and Kallesen (2004) took two types of passengers into account, service-sensitive and price-sensitive passengers, and proposed an air ticket pricing model to simulate the purchasing behavior of different passengers.

Because the decision on airports and flights is the consequence of air ticket choice, the purchasing behavior actually also reflects the air travel choice decision, which is influenced by many distinct factors, such as flight time, air ticket price, airport access time, type of flight, class of membership in frequent-flier programs (Skinner 1976; Hess 2008; Tam, Lam, and Lo 2010; Chiou and Chen 2012).

In addition to the factors related to the ‘air’ alternative itself, in reality, passengers may also take into account substitutable travel options, such as high-speed railway, normal railway and cars. (Yang and Zhang 2012). Because high-speed railway has been a strong competitor for airlines (Adler, Pels, and Nash 2010), and the price is one of the most important factors that influences the mode choice decision of passengers (Fu, Zhang, and Lei 2012), the option of high-speed railway will be included in this study.

Moreover, people may give up the travel option when none of the travel modes satisfies their needs (Zhang and Yan 2006). An extra choice option defined as ‘non-travel’ needs to be considered (Bergantino and Capozza 2015). Therefore, we propose an integrated modeling framework in order to formulate such a process by incorporating the air ticket purchasing behavior into the ticket pricing model.

The air ticket purchasing behavior is complex because it involves trip generation, travel mode choice and joint choice of airlines and airports (Tam, Lam, and Lo 2008). It is normally treated as a hierarchical decision process, resulting into different modeling approaches. Bhat (1995) estimated the share of existing intercity travelers who are diverted to a new or upgraded service using a nested logit model. Hess et al. (2013) used the cross-nested logit structure to model ‘airport-airline-access mode’ joint choice behavior. In this study, we use a nested logit model to allow for substitutable relationships between different travel modes and between air travel plans.

3. Model formulation and choice of solution algorithm

In this section, we propose a dynamic pricing model developed for multi-flight air routes incorporating air ticket purchasing behavior. The model is composed of two sub-models, a dynamic pricing model and a ticket purchase model. The dynamic pricing model is constructed on the basis of MDP, and the ticket purchase model is developed based on the nested logit structure with the purpose of describing passengers’ responses to different pricing strategies.

3.1. Dynamic pricing model based on MDP

A Markov decision process is a discrete time stochastic control process. It contains 4 tuples and can be formulated as $(S, A, P(s, a, s'), R(s, a, s'))$, in which $S$ is set of states, $A$ is set of actions, $P(s, a, s')$ is probability of moving from $s$ to $s'$ if taking action $a$, and $R(s, a, s')$ is the corresponding immediate reward.
Referring to the structure of MDP, the dynamic pricing model is established. The variables of the model are explained as follows. \( E \) is the set of flights between an origin and a destination, and \( e \) is the identifier of each flight, \( e \in E \). The capacity of flight \( e \) is \( F_e \); the length of the booking horizon of every flight is \( T \); \( t \) is identifier of time, which is counted backward. The start time point of the booking horizon is \( T \) and the end time point is 0. \( T \) is divided into \( M \) period and \( m \) is identifier of each period, \( m \in \{1, \ldots, M\} \); \( t_m \) is the start time point of period \( m \) and \( t_{m-1} \) is the corresponding end time point. For the whole booking horizon, \( t_0 = T \) and \( t_{M+1} = 0 \).

For each period \( m \), \( y_m \) is the vector of air ticket prices on the air route: \( \vec{y}_m = [y_{m1}, \ldots, y_{m2}, \ldots, y_{mE}] \), where \( y_{me} \) is the price of flight \( e \) in time period \( m \), and \( y_{me} \in \{y_0, y_1, \ldots, y_2, \ldots, y_z\} \). \( y \) is the set of all available air ticket prices of the air route, and \( y_z(z = 1 \) to \( Z) \) is the \( z \)th kind of air ticket price which is generated from the real data; \( y_0 \) is defined as a null price, it will be used when the flight is closed or the capacity is depleted (Zhang and Cooper 2005).

\( O \) is the set of passengers who have the desire to travel from the origin to the destination on a specific day, and \( o \) is the identifier of each passenger, \( o \in O \). The arrival time of passenger \( o \) during the booking horizon is \( t_o \). \( \lambda_{mo} \) is a 0–1 variable, if passenger \( o \)'s arrival time satisfies the condition \( t_m \leq t_o < t_{m-1} \), then \( \lambda_{mo} = 1 \), otherwise \( \lambda_{mo} = 0 \).

Suppose that passenger \( o \) arrives in period \( m \), passenger \( o \) will decide to buy or not buy the air ticket based on the price vector, \( \vec{y}_m \). The probability to buy the air ticket of flight \( e \) for passenger \( o \) with price \( y_{me} \) is \( p_o(y_{me}) \), and the probability to not buy an air ticket for passenger \( o \) is \( p'_{mo} \). \( p_o(y_{me}) \) and \( p'_{mo} \) can be calculated using the model of air ticket purchasing behavior. The value of the two variables should satisfy the following condition: \( \sum_{e=1}^{E} p_o(y_{me}) + p'_{mo} = 1 \).

When formulating the air ticket pricing model based on MDP, \( s_e \) is the state of flight \( e \). It represents the volume of unsold seats on flight \( e \) and \( s_e \in \{0, 1, \ldots, F_e\} \). \( A(s_e) \) is the set of actions which can be taken in \( s_e \), it is composed of the available air ticket prices in \( s_e \) and \( A(s_e) = \{y \in Y|s_e > 0; y_{me} = y_0|s_e = 0\} \) and we set that the available price is \( y_0 \) when the aircraft is full. We define \( C_e \) as the fixed cost of flight \( e \), and \( c_e \) as the service cost per transported passenger.

The volume of passengers arriving in period \( m \) is \( N_m = \sum_{o=1}^{O} \lambda_{mo} \). By sorting the travelers in set \( O \) based on arrival time, let \( n_b \) be the first \( b \) travelers. Then, we know that the state \( s_e \) at the start period \( m \) may transfer to any new state in the following set \((s_e - 0, (s_e - 1), \ldots, (s_e - n_b), \ldots, (s_e - N_m))\). The probability of the transition from \( s_e \) to \((s_e - n_b)\) is \( p_m(n_b|y_{me}) \), which is a stochastic value determined by the passenger volume arrived in period \( m \) and the corresponding purchasing probability \( p_o(y_{me}) \). More specifically, the function of \( p_m(n_b|y_{me}) \) can be expressed as below.

\[
p_m(n_b|y_{me}) = p_1(y_{me}) \cdots p_b(y_{me}) \cdot p'_{m,b+1} \cdots p'_{m,N_m}.
\]

Let \( \pi \) denote the policy of pricing and \( \gamma \) denote the discounted rate. We propose Equation (2) to formulate the long-run expected profit of flight \( e \) during the booking horizon with policy \( \pi \). In the equation, \( w_{\pi m}(s_e) \) is the expected profit by adopting policy \( \pi \) for \( m \) periods with the origin state \( s_e \).

\[
w_{\pi m}(s_e) = \sum_{n_b=0}^{N_m} \left[ p_m(n_b|y_{me}) \cdot (y_{me} - c_e) \cdot n_b + \gamma \cdot w_{\pi m-1}(s_e - n_b) \right] - C_e/M.
\]
In the equation, $y_{me} \cdot n_b$ is the immediate revenue if there are $n_b$ passengers purchasing the ticket of flight $e$ with price $y_{me}$ in the $m$th period. In this situation, the state of the $(m - 1)$th period will be $(s_e - n_b)$, and $[(y_{me} - c_e) \cdot n_b - Ce/M]$ is the corresponding immediate profit. $w_{\pi,m-1}(s_e - n_b)$ is the expected profit by adopting policy $\pi$ for $m - 1$ periods with the origin state $(s_e - n_b)$. It should be noted that the fixed cost $C_e$ is divided to $M$ parts on average and each part is assigned to one period during the horizon. Thus, the cumulative fixed cost becomes larger and larger when $m$ approaches to $M$ period by period.

In this paper, we want to optimize the policy $\pi$ with the aim of maximizing the profit, $w^\pi_{\pi,m}(s_e)$. The optimal function can be formulated in Equation (3). The optimized $\pi$ will provide the best action (price) in the front of state $s_e$.

$$w_{\pi,m}(s_e) = \max_{y_{me} \in A(s_e)} \sum_{n_b=0}^{N_m} \{p_m(n_b|y_{me}) \cdot [(y_{me} - c_e) \cdot n_b + \gamma \cdot w_{\pi,m-1}(s_e - n_b)] - Ce/M\}$$

(3)

Here, we define the optimal pricing policy as $\pi^*_m(s_e)$. The corresponding value function is $w^{\ast}_{\pi,m}(s_e)$. Then, $\pi^*_m(s_e)$ can be calculated as

$$\pi^*_m(s_e) = \arg \max_{y_{me} \in A(s_e)} \sum_{n_b=0}^{N_m} \{p_m(n_b|y_{me}) \cdot [(y_{me} - c_e) \cdot n_b + \gamma \cdot w^{\ast}_{\pi,m-1}(s_e - n_b)] - Ce/M\}. \quad (4)$$

### 3.2. Model of air ticket purchasing behavior

In general, the process to buy an air ticket consists of three stages: trip making decision, travel mode choice and air travel choice. The decision tree in Figure 1 illustrates the process of air ticket purchase. It distinguishes between three levels: at the top level $f$, passengers decide whether to travel or not; at the second level $h$, passengers choose their travel mode. Similar to the air travel which is normally for long-distance trips, high-speed railway and normal-speed railway compromise the other choice options; at the third level $q$, conditional on the ‘air’ travel mode being chosen, passengers choose a ticket from a set of options which stand for an air travel plan specified by attributes such as flight time and access mode.

![Figure 1. Decision tree demonstrating air ticket purchase.](image)
According to the decision tree, a substitutable relationship exists between alternatives in branch 'Air' and 'Travel'. That is, error variances in the utility functions of these alternatives are not independently and identically distributed. Therefore, to capture this purchasing process, we develop a discrete choice model, taking the substitutable relationship between the alternatives into account. In this study, we use the nested logit model (NL) to describe the nested structure and handle the correlation between alternatives.

In the model, $i$, $j$, $k$ are the identifiers of the alternatives at level $f$, $h$, $q$, respectively. The probability of $f_i$ being chosen $p(f_i)$ is formulated as Equation (5). In the equation, $V_i$ is the observable portion of the utility of $f_i$ ($U_i$), which is formulated as Equation (6). In the utility function, $\epsilon$ is the error variance; $x_{il}$ is the $l$th attribute describing alternative $f_i$; and $\beta_{il}$ is the corresponding coefficient.

$$p(f_i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)}, \quad (5)$$

$$U_i = V_i + \epsilon_i = \beta_0 + \sum_{l \in L} \beta_{il} \cdot x_{il} + \epsilon_i. \quad (6)$$

At level $f$, there exists an aggregate alternative 'Travel'. $V_i$ of 'Travel' is given in Equation (7), and consists of two parts: (i) the observable utility $V_{i}^*$ of 'Travel' branch; (ii) the expected maximum utility $\Gamma_f$ of the alternatives grouped within the branch (de Dios Ortuzar 1983). $\theta_f$ is the inverse of $\mu_f$ which is the scale parameter of branch 'Travel', and $0 < \theta_f < 1$.

$$V_i = V_{i}^* + \theta_f \cdot \Gamma_f, \quad (7)$$

where $\Gamma_f$ is formulated as Equation (8), and $V_j$ is the observable utility of alternative $h_j$ in branch 'Travel'. Similar to Equation (6), Equation (9) is the utility function of $h_j$, $\epsilon_j$ is the error variance, and $x_{jl}$ is the $l$th attribute of alternative $h_j$; $\beta_{jl}$ is the corresponding coefficient.

$$\Gamma_f = \ln \sum_{j=1}^{J} \exp(V_j/\theta_f), \quad (8)$$

$$U_j = V_j + \epsilon_j = \sum_{l \in L} \beta_{jl} \cdot x_{jl} + \epsilon_j. \quad (9)$$

At level $h$, $p(h_j)$ is the probability that $h_j$ is selected on condition that 'travel' branch has been chosen on the top level. The observable utility of the aggregate alternative at level $h$ is given by Equations (11) and (12). $\theta_h$ is the inverse of scale parameter $\mu_h$ of branch 'Air', and $0 < \theta_h < 1$.

$$p(h_j) = P(h_j|f_i) \cdot p(f_i) = \frac{\exp(V_j/\theta_f)}{\sum_{j=1}^{J} \exp(V_j/\theta_f)} \cdot \frac{\exp(V_i)}{\sum_{i=1}^{I} \exp(V_i)}, \quad (10)$$

$$V_j = V_{j}^* + \theta_h \cdot \Gamma_h, \quad (11)$$

$$\Gamma_h = \ln \sum_{k=1}^{K} \exp(V_k/\theta_h). \quad (12)$$
At level \( q \), \( p(q_k) \) is the probability of \( q_k \) being selected on the condition that ‘travel’ and ‘Air’ branches have been chosen, \( p(q_k) \) is formulated as Equation (13). In the equation, \( V_k \) is the observable utility of the \( k \)th alternative in the branch. The utility function is as Equation (14), in which \( \epsilon_k \) is the error variance, and \( x_{kl} \) is the \( l \)th attribute describing alternative \( q_k \); \( \beta_{kl} \) is the corresponding coefficient. Thus, we can calculate the probability of buying the ticket \( P_n(y_{me}) \) in the MDP structure based on \( p(q_k) \).

\[
P(q_k) = P(q_k|h_j)P(h_j|f_i) \cdot P(f_i)
\]

\[
\sum_{k=1}^{K} \exp(V_k/\theta_h) \cdot \sum_{j=1}^{J} \exp(V_j/\theta_f) \cdot \sum_{i=1}^{I} \exp(V_i),
\]

(13)

\[
U_k = V_k + \epsilon_k = \sum_{l \in L} \beta_{kl} \cdot x_{kl} + \epsilon_k.
\]

(14)

3.3. Algorithm

The dynamic programming and reinforcement learning are the main methods of solving MDP model. The reinforcement learning is suitable for solving MDP when the transition probability is unknown. However, because we can observe the transition probability based on the air ticket purchasing model in this study, we use dynamic programming to solve the model.

Value iteration is a popular and commonly used algorithm in dynamic programming. The key of this algorithm is to find the optimal result \( w^*_m(s_n) \) of the corresponding value function, which is formulated as Equation (2). In the algorithm, the value function can be optimized through iteration.

Before the iteration starts, the transition matrix in which the transition probability \( p(s_e, y_{me}, s_e-n_b) \) from state \( s_e \) to \( (s_e-n_b) \) by taking action \( y_{me} \) is calculated based on \( p_m(n_b|y_{me}) \), and the reward matrix in which the immediate reward \( R(s_e, y_{me}, s_e-n_b) \) equals to \( [(y_{me} - c_e)n_b - C_e/M] \).

Let \( V_g(s_e) \) be the refined result of the value function at the \( g \)th stage during iteration; and \( Q_g(s_e, y_{me}) \) be the refined result of the \( Q \)-function at the \( g \)th stage. Here, \( g \) is counted backward. \( V_g(s_e) \) and \( Q_g(s_e, y_{me}) \) are formulated as Equations (15) and (16). From the equation, we can see that \( Q_g(s_e, y_{me}) \) calculates the reward in the situation that action \( y_{me} \) is taken in state \( s_e \), and \( V_g(s_e) \) is the maximum one of them.

\[
Q_g(s_e, y_{me}) = \sum_{n_b=0}^{N_m} p(s_e, y_{me}, s_e-n_b)
\]

\[
\cdot [R(s_e, y_{me}, s_e-n_b) + \gamma V_{g-1}(s_e-n_b)], \quad g \geq 0,
\]

(15)

\[
V_g(s_e) = \max_{y_{me}} Q_g(s_e, y_{me}), \quad g > 0.
\]

(16)

In Equation (15), \( V_{g-1}(s_e-n_b) \) is the future reward when taking action \( y_{me} \) to state \( s_e \) in the \( g \)th stage. However, as the iteration begins from the ‘bottom’ and then goes backward by refining vectors \( Q [s_e, y_{me}] \) and \( V[s_e] \), \( V_g(s_e) \) is updated based on \( V_{g-1}(s_e-n_b) \). During the iteration, \( V_0(s_e) \) is assumed to be 0. Then, let \( g = g + 1 \) and start the ‘loop’ until the
convergence condition is reached which means the optimal action of a certain state is finally found. \( \delta \) is a threshold value, and \( \delta > 0 \). The process of iteration is as follows:

<table>
<thead>
<tr>
<th>Step of value iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set ( w_0(s_e) = 0, g = 0 )</td>
</tr>
<tr>
<td>loop: ( g = g + 1 ); loop: for all ( s_e ); loop: for all ( y_{me} )</td>
</tr>
<tr>
<td>( Q_g(s_e, y_{me}) = \sum_{n_b=0}^{N_b} p(s_e, y_{me}, s_e - n_b) \cdot \left[ R(s_e, y_{me}, s_e - n_b) + \gamma V_{g-1}(s_e - n_b) \right] )</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>( V_g(s_e) = \max_{y_{me}} Q_g(s_e, y_{me}) )</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>until (</td>
</tr>
<tr>
<td>end loop</td>
</tr>
</tbody>
</table>

4. Survey design, data collection and model estimation

In order to estimate the parameters of the air ticket purchasing model, a survey for air ticket purchasing behavior was conducted to collect the data.

4.1. Survey design

The survey was designed into two parts. The first part collects data on the general information of respondents, including their socio-demographical information, preferred air ticket purchasing time and experiences in commonly used travel options. The second part of the survey is a stated choice experiment combining the travel or non-travel choice, travel mode choice and the joint choice of airports and airlines. To construct the experimental design, the factors influencing the choice of travel modes, airports and flights and the levels of each variable were elicited based on existing literatures. Table 1 gives an overview of the selected variables and their levels.

The values for each level were determined as close as to the reality. Because in general flights are for relatively longer trips, we assumed that the context of the travel mode choice relates to long-distance trips. For example, the minimum level of travel time for a normal train is nine hours. In addition, the values were set to incorporate the variations in the ticket price, type of seat, distance for different travel modes and airlines.

As of the majority of passengers are in the economic or second class for air-and/or high-speed trains, in the experiment, the fare class was not explicitly considered as a variable. Instead, to incorporate the different effects of fare class and discounting rate, we tried to vary the ticket price from very low (200 RMB) to very high (2000 RMB), which approximates the reality for deeply discounted price and one for business class, respectively.

These attribute levels were systematically varied according to an orthogonal fraction of the full factorial design. A total of 256 profiles using fourteen variables, including one two-level variable and thirteen four-level variables, were generated and used for the experiments.

Respondents were invited to finish the choice experiment in two steps. First, respondents were asked to select one from the six travel alternatives, including three air travel plans, high-speed train, normal train and the non-travel option. At this stage, only ticket
Table 1. Selected attributes and corresponding levels.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribute variables</strong></td>
<td></td>
</tr>
<tr>
<td>Air ticket price, unit: RMB (PRICE)</td>
<td>200, 800, 1400 and 2000</td>
</tr>
<tr>
<td>Train ticket price, unit: RMB (PRICE)</td>
<td>100, 300, 500 and 700</td>
</tr>
<tr>
<td>High-speed rail ticket price, unit: RMB (PRICE)</td>
<td>200, 600, 1000 and 1400</td>
</tr>
<tr>
<td>Air travel time, unit: hour (TTIME)</td>
<td>1.0, 1.5, 2.0 and 2.5</td>
</tr>
<tr>
<td>Train travel time, unit: hour (TTIME)</td>
<td>9, 18, 27 and 36</td>
</tr>
<tr>
<td>High-speed rail travel time, unit: hour (TTIME)</td>
<td>3, 6, 9 and 12</td>
</tr>
<tr>
<td>Access time to airport, unit: hour (ATIME)</td>
<td>0.5, 1, 1.5 and 2</td>
</tr>
<tr>
<td>Probable flight delay time, unit: minutes (DTIME)</td>
<td>No delay, 30, 60 and 90</td>
</tr>
<tr>
<td>Departure time of flight (FTIME)</td>
<td>early morning (before 7:00), morning (7:00–12:00), afternoon (12:00–21:00), late night (21:00–)</td>
</tr>
<tr>
<td>Cost of changing air ticket (percentage of the ticket price) (CFEE)</td>
<td>0, 30%, 50% and 100%</td>
</tr>
<tr>
<td>Type of aircraft (CATP)</td>
<td>Boeing or Airbus, Other branch airplane (Jet900)</td>
</tr>
<tr>
<td><strong>Context variables</strong></td>
<td></td>
</tr>
<tr>
<td>Travel purpose (PUR)</td>
<td>Visit family/friends, work, tourism and other</td>
</tr>
<tr>
<td>Class of frequent flyer membership (CMFF)</td>
<td>Non-member, regular membership, silver membership, Golden or higher membership</td>
</tr>
</tbody>
</table>

price and in-vehicle travel time are considered. If one of the three air travel plans was chosen, respondents were guided to an air travel choice experiment where the same three air plans were presented, with additional attributes related to the details of the flight and airport. In total, eight profiles for mode choice were randomly presented to each respondent. The individuals who choose air travel received additional information related to the air plans.

In order to avoid an insufficient number of observations for the choice of air plans, at the second stage, the choice experiment of air travel plans was conducted explicitly. The air travel plan is a combination of attributes related to flight and airport. For each profile, three air plans were presented. Again, each respondent received eight profiles which were randomly selected from the 256 profiles.

### 4.2. Data collection and descriptive analysis

The questionnaires were distributed to the respondents in three cities, Dalian, Yingkou and Dandong. The latter two cities have a connection with the airports, Shenyang airport and/or Dalian airport. The locations of the three cities are shown in Figure 2. All cities have a connection of normal railways. Dalian has an airport in the city boarder, and the city is located on a high-speed railway line. Yingkou city is also on the high-speed line, but does not have an airport. Similarly, Dandong city has no airport and no direct connection of the high-speed railways. Therefore, passengers in the two cities for air travel have to use Dalian or Shenyang airport.

Dalian and Shenyang airports are the two main airports in northeast China. The surrounding area of the airports is a representative region featured with multiple airports, resulted from a decentralized airport governance in the 1990s. Besides airborne travel, other travel modes (bus, train, high-speed train) are well reachable, making this area a suitable case to test our proposed model.

The survey was conducted through Internet in two ways: (i) to distribute the questionnaires directly in airports and railway stations, and (ii) to publish the Link address of our
Figure 2. Location of cities for survey.

online survey system through the social media tool, WeChat. The latter is efficient because of the low cost and the potential wideness of spread. In total, 350 questionnaires were collected, including 100 from airports, 110 from railway stations and 140 from social media. In the end, 300 valid questionnaires were collected, containing 2400 sets of choice data. The response rate is 85.7%.

The participants who joined the survey through interviews and social media are randomly selected. A preliminary descriptive analysis was conducted. It shows the percentage of women and men in the sample are almost evenly distributed, with 45.9% and 54.1% for women and men, respectively. To get the experience of respondents on different available travel options, a question regarding their available transport modes was asked. Result shows that the percentage of respondents who have experienced all travel modes, two travel modes and only one travel mode is 87.6%, 96.4% and 98.7%, respectively. This means most respondents are familiar with the travel alternatives in the choice experiment, making the survey results more realistic.

4.3. Model estimation

In spite that an orthogonal experimental design is used, the orthogonality between variables may not remain in the data. One may argue in reality that the correlation between
Table 2. Correlation between air ticket price and cost of changing ticket.

<table>
<thead>
<tr>
<th></th>
<th>Air ticket price (standard price)</th>
<th>Cost of changing ticket (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air ticket price</td>
<td>Pearson correlation (p-value)</td>
<td>N</td>
</tr>
<tr>
<td>(unit: standard price)</td>
<td>1</td>
<td>3325</td>
</tr>
<tr>
<td>Cost of changing ticket</td>
<td>0.015 (0.559)</td>
<td>1</td>
</tr>
</tbody>
</table>

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table 3. Results of the nested logit air ticket purchasing model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>INC_1</td>
<td>0.0371</td>
<td>DTIME_1</td>
<td>0.07268</td>
</tr>
<tr>
<td></td>
<td>(0.04**)</td>
<td></td>
<td>(0.82)</td>
</tr>
<tr>
<td>INC_2</td>
<td>0.0387</td>
<td>DTIME_2</td>
<td>−0.05564</td>
</tr>
<tr>
<td></td>
<td>(0.04**)</td>
<td></td>
<td>(−0.64)</td>
</tr>
<tr>
<td>INC_3</td>
<td>0.1651</td>
<td>DTIME_3</td>
<td>−0.12495</td>
</tr>
<tr>
<td></td>
<td>(0.04**)</td>
<td></td>
<td>(−1.51)</td>
</tr>
<tr>
<td>PUR_1</td>
<td>0.0074</td>
<td>FTIME_1</td>
<td>−0.12986</td>
</tr>
<tr>
<td></td>
<td>(0.04*)</td>
<td></td>
<td>(−1.51***)</td>
</tr>
<tr>
<td>PUR_2</td>
<td>0.0604</td>
<td>FTIME_2</td>
<td>0.13064</td>
</tr>
<tr>
<td></td>
<td>(0.04*)</td>
<td></td>
<td>(1.52***)</td>
</tr>
<tr>
<td>PUR_3</td>
<td>0.0077</td>
<td>FTIME_3</td>
<td>0.06886</td>
</tr>
<tr>
<td></td>
<td>(0.04*)</td>
<td></td>
<td>(0.78***)</td>
</tr>
<tr>
<td>PRICE</td>
<td>−0.00265</td>
<td>ATIME</td>
<td>−0.01237</td>
</tr>
<tr>
<td></td>
<td>(−26.35***)</td>
<td></td>
<td>(−0.23**)</td>
</tr>
<tr>
<td>TTIME</td>
<td>−0.20524</td>
<td>CATP_1</td>
<td>0.03341</td>
</tr>
<tr>
<td></td>
<td>(−19.08***)</td>
<td></td>
<td>(0.63)</td>
</tr>
<tr>
<td>CFEE_1</td>
<td>0.17259</td>
<td>CMFF_1</td>
<td>0.00168</td>
</tr>
<tr>
<td></td>
<td>(1.74*)</td>
<td></td>
<td>(0.02*)</td>
</tr>
<tr>
<td>CFEE_2</td>
<td>−0.05344</td>
<td>CMFF_2</td>
<td>0.05116</td>
</tr>
<tr>
<td></td>
<td>(−0.63*)</td>
<td></td>
<td>(0.61*)</td>
</tr>
<tr>
<td>CFEE_3</td>
<td>−0.12329</td>
<td>CMFF_3</td>
<td>0.06058</td>
</tr>
<tr>
<td></td>
<td>(−2.03***)</td>
<td></td>
<td>(0.12*)</td>
</tr>
</tbody>
</table>

Log-likelihood function: −1943.28
Pseudo $R^2$ ($\rho^2$): 0.45

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

variables has an effect on the model estimation. In order to avoid effects of this specification, we take the correlation between air ticket price and changing fee as an example. The result of correlation analysis (Table 2) shows that there is no significant correlation between the two variables.

The air ticket purchasing model is estimated based on the effect-coded data (except for ticket price, travel time and access time) using the maximum likelihood estimation method. For each variable, the last level is treated as the base level. Results are shown in Table 3. The goodness of fit is satisfactory, with the pseudo $R^2$ equal to 0.45. The inclusive value parameter is 0.45 for Travel nest, 0.32 for non-travel nest, 0.014 for Air nest, 0.033 for High-speed railway nest and 0.024 for Normal railway nest. All the inclusive value parameters fall within 0 and 1, which supports the nested logit structure of the air ticket purchasing model.
According to the estimation results, the utility functions of different travel options are shown in the following equation:

\[ U(Air) = 0.037 \cdot INC_1 + 0.039 \cdot INC_2 + 0.165 \cdot INC_3 + 0.007 \cdot PUR_1 + 0.060 \cdot PUR_2 + 0.007 \cdot PUR_3 - 0.003 \cdot PRICE - 0.205 \cdot TTIME + 0.173 \cdot CFEE_1 - 0.053 \cdot CFEE_2 - 1.123 \cdot CFEE_3 + 0.073 \cdot DTIME_1 - 0.056 \cdot DTIME_2 - 0.125 \cdot DTIME_3_1 - 0.130 \cdot FTIME_1 + 0.131 \cdot FTIME_2 + 0.069 \cdot FTIME_3 - 0.012 \cdot ATIME + 0.033 \cdot ACTP_1 + 0.002 \cdot CMFF_1 + 0.051 \cdot CMFF_2 + 0.061 \cdot CMFF_3 + \epsilon_{air}, \]

(17)

\[ U(Normal \ train) = 0.037 \cdot INC_1 + 0.039 \cdot INC_2 + 0.165 \cdot INC_3 + 0.007 \cdot PUR_1 + 0.060 \cdot PUR_2 + 0.007 \cdot PUR_3 - 0.003 \cdot PRICE - 0.205 \cdot TTIME + \epsilon_{normal} \]

(18)

\[ U(High\ - \ speed \ train) = 0.037 \cdot INC_1 + 0.039 \cdot INC_2 + 0.165 \cdot INC_3 + 0.007 \cdot PUR_1 + 0.060 \cdot PUR_2 + 0.007 \cdot PUR_3 - 0.003 \cdot PRICE - 0.205 \cdot TTIME + \epsilon_{high\-speed} \]

(19)

\[ U(Non\ - \ travel) = 0.037 \cdot INC_1 + 0.039 \cdot INC_2 + 0.165 \cdot INC_3 + 0.007 \cdot PUR_1 + 0.060 \cdot PUR_2 + 0.007 \cdot PUR_3 + \epsilon_{non\-travel}. \]

(20)

The positive signs of the income coefficients (INC_1, INC_2, INC_3) prove the fact that passengers with a higher income plan more travel and accept higher travel costs. INC_3 has the highest coefficient, which means the utility is higher for passengers with an income higher than 100,000 RMB per year. As for travel purpose, PUR_2 has the largest positive coefficient, indicating that the travel purpose ‘work’ leads to a highest increase of utility.

The negative signs of PRICE and TTIME indicate that the utility will decline with increasing ticket price and travel time. The coefficient of ATIME is statistically significant and its sign shows that an increase in access time leads to a decreasing utility. The table further shows that utility is higher if the flight departs on time, and then decreased with increasing flight delays, and the amplitude of decrease is related to the length of the delay time. Similarly, the utility will rise up if the changing fee of air ticket is free (CFEE_1), but will decrease if the changing fee is higher than 0 (CFEE_2, CFEE_3).

In the group of flight time, FTIME_1 leads to negative effects on the utility, which shows the phenomenon that passengers do not prefer the flights departing before 7:00 in morning. The sign of the coefficient for CATP_1 proves that passengers prefer to choose large aircrafts. Moreover, the class of membership (CMFF) leads to a positive effect, indicating that the higher the class, the higher the utility.

In order to explore the effectiveness of the proposed model, an elasticity analysis was conducted. Here, consumers’ income level, travel purpose and frequent-flier membership are set according to the statistical results in Travel Survey of Liaoning Province in 2014 and
Table 4. Data of the alternative travel modes.

<table>
<thead>
<tr>
<th>Modes</th>
<th>PRICE (yuan)</th>
<th>TTIME (hour)</th>
<th>ATIME (hour)</th>
<th>DTIME (hour)</th>
<th>FTIME</th>
<th>CFEE (yuan)</th>
<th>CATP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air (economy class)</td>
<td>600</td>
<td>1.5</td>
<td>1.5</td>
<td>1</td>
<td>10:00</td>
<td>300</td>
<td>Boeing</td>
</tr>
<tr>
<td>High-speed train (economy class)</td>
<td>400</td>
<td>4</td>
<td>1.5</td>
<td>0</td>
<td>9:00</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td>Normal train (economy class)</td>
<td>200</td>
<td>10</td>
<td>1.5</td>
<td>0</td>
<td>21:00</td>
<td>0</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 5. Demand response to air ticket price changes.

<table>
<thead>
<tr>
<th></th>
<th>Air demand</th>
<th>High-speed train</th>
<th>Normal train</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% increment in air ticket price</td>
<td>– 2.62%</td>
<td>+ 2.27%</td>
<td>+ 0.08%</td>
</tr>
</tbody>
</table>

the data provided by airlines, the information of alternative travel modes is set as indicated in Table 4. Table 5 demonstrates the effects of changing air ticket price on its own demand and the demand for other modes.

When increasing the air ticket price by 1%, its own demand reduces by 2.62%, but there is improvement in choosing high-speed train and normal train by 2.27% and 0.08%, respectively. It can be found that most travelers will shift to high-speed trains if the air ticket price increases, but will seldom switch to the normal train.

Comparing to the literatures focusing on travel mode choice in Asia, the ‘reduction of air-choice probability with 1% air ticket price increment’ is smaller than those reported in previous studies. For example, Park and Ha (2006) reported a 2.47% gain in Korea, while Fu, Oum, and Yan (2014) found 1.30% in Japan’s intercity mode choice between high-speed train and air. It may be largely attributable to income differences in different regions in Asia.

In the stated choice experiment, air ticket purchase behavior was designed as a sequential decision process. The first stage is the mode choice incorporating the main attributes only (travel time and ticket price); the second stage is the decision of air travel plan based on additional air travel factors. In order to examine the possible influences of factors at the second stage on the choice made at the first stage, we compare the aggregated modal split results calculated based on the attributes at different stages. By considering the additional factors in the second stage, the air demand increases 0.05%, and the demand of high-speed train and normal train decrease 0.02% and 0.01%, respectively. This coincides with our assumption that the additional attributes have little impact on mode choices.

5. Pricing for monopoly air routes

In order to optimize the pricing strategy of flights on one air route, our proposed pricing model is applied in the multi-airport area, which is served by Dalian and Shenyang airport in China. The multi-airport area is composed of seven cities (Dalian, Shenyang, Yingkou, Panjin, Anshan, Benxi and Dandong) with approximately 23.32 million people.

In this section, we will optimize the pricing strategy of the flights on the monopoly air route ‘Dalian – D1 city’. (D1 means a city using Flight 1). The ‘monopoly’ means the air route is the only alternative if passengers in the multi-airport area want to go to D1 city by air. There are three flights on this air route; the flights’ attributes such as flight time, aircraft type and capacity are shown in Table 6.
Table 6. Flights’ attributes on air route of ‘Dalian–M city’.

<table>
<thead>
<tr>
<th>Flight index</th>
<th>Flight 1</th>
<th>Flight 2</th>
<th>Flight 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air route</td>
<td>‘Dalian – D1 city’</td>
<td>‘Dalian – D1 city’</td>
<td>‘Dalian – D1 city’</td>
</tr>
<tr>
<td>Flight time</td>
<td>10:00–11:00, 21:00–22:00</td>
<td>13:00–14:00</td>
<td>13:00–14:00</td>
</tr>
<tr>
<td>Aircraft type</td>
<td>Airbus A320</td>
<td>Airbus A320</td>
<td>Airbus A320</td>
</tr>
<tr>
<td>Capacity</td>
<td>140 (economy class)</td>
<td>140 (economy class)</td>
<td>140 (economy class)</td>
</tr>
<tr>
<td>Travel time</td>
<td>2.5 hours (direct)</td>
<td>2.5 hours (direct)</td>
<td>2.5 hours (direct)</td>
</tr>
</tbody>
</table>


Figure 3. Statistical price variation of flights on ‘Dalian – D1 city’. Source: Booking data provided by Dalian airport.

The price variation during the booking horizon (60 days) for flights departed on 20 April 2014 is shown in Figure 3. In the figure, the horizontal axis stands for the days prior to departure and the vertical axis stands for the standard price of the flight, 1 means full price and 0 means free. It can be seen that the highest prices and the price variations of the three flights are different. More specifically, because the departure time of Flight 1 and Flight 3 is better than that of Flight 2, the price of the two formers are higher than that of the latter one. It indicates that the flight departure time has been taken into consideration in airlines’ price setting.

As required by the algorithm in the ticket pricing model, it is necessary to specify the total traveler volume who have the desire of traveling from the multi-airport area to D1 city. Therefore, we calculated this value using multiple data sources, including the Travel Survey of Liaoning Province in 2014, the statistical data of transport passenger volume by air, high-speed railway and normal railway per month in 2014, and the economic data in the multi-airport area.

Traveler’s income, travel purpose and class of frequent flyer membership are assumed the same as the data provided by the Travel Survey of Liaoning Province in 2014. To be specific, 56.4% of them travel for work, 13.7% travel for tourism, 13.7% travel for visiting family and friends and 11.3% travel for other purposes. Besides above variables, other variables in the air ticket purchasing model are set according to the real data of airport, flight and other travel modes. Thus, the transition probability in MDP structure can be obtained based on the purchasing probability calculated by the nested logit model.
Based on the ticket purchasing time and travel purpose, passengers are assigned to different time points prior to departure during the booking horizon, which constructs a booking curve. As required by MDP, the booking horizon is divided into 60 time periods \((m = 60)\); the set of available price \(Y = \{y_0, 0.05, 0.1, 0.15, \ldots, 0.95, 1\}\), and \(y_0\) is a null value; the set of unsold seats is \(\{0, 1, \ldots, 140\}\); the fixed cost of flight \(e (C_e)\) and service cost per person \(c_e\) is set according to the real data.

We solved the pricing model using the value iteration algorithm. It results in the optimized pricing policy of each flight from Dalian to \(D_1\) city. Finally, we calculated the price variation during the booking horizon. Because the pattern of the price variation of Flight 1 is similar to that of Flight 3, we take Flight 1 and Flight 2 as the example in next discussions.

For Flight 1, the price variation calculated by the optimized pricing policy is shown in Figure 4. In the figure, the square points stand for the real data and the round points stand for the optimized results. It can be seen that during the first 20 days of the booking horizon, the optimized price is lower than the real one. During the middle 20 days, the optimized price is almost equal to the real one, but during the last 20 days, the optimized price is higher than the real data.

The corresponding cumulative ratio of sold seats on different time points during the booking horizon is shown in Figure 5. In the figure, the round points stand for the real data, and the square points stand for the optimized results. We can find that the optimized load factor of Flight 1 increases by 5% than that of real data, and the cumulative ratio of sold seats on every time point is higher than the real ones.

In Figure 5, the two curves are ‘S-shaped’, which means the increasing speed of the cumulative ratio varies by time. For the purpose of analysis, the whole booking horizon is divided into 3 stages, 20 days for each stage. The first stage is from the 60th day to the 40th day prior to departure, passengers who buy the ticket at this stage account for 18.4% of the total transported passengers of Flight 1. Based on the assumed travel purpose in the data, most passengers who arrived in this stage are traveling for private affairs, they are probably sensitive to air ticket price. Therefore, cutting down ticket price in this time period may largely increase the probability of choosing Flight 1.

The second stage is from the 40th day to the 20th day; passengers who buy tickets at this stage account for 21.2% of the total passengers. At this stage, the difference between the
optimized cumulative ratio and the real data is almost equal to the value of difference in previous stage, which means the pricing policy at this stage keeps the advantages resulted from the pricing policy of stage 1.

The third stage is from the 20th day to the departure day; passengers who buy tickets at this stage account for 60.4% of the total passengers. Because most passengers at this stage are traveling for work, indicating that commuters can accept relatively high ticket price. In another words, passengers at this stage are not as price-sensitive as the ones at previous stages. Therefore, increasing price at this stage will not lead to a same reduction in the number of purchase as that at previous stages. As a result, the optimized pricing policy leads to the profit increase by 3.5% for Flight 1.

For Flight 2, the optimized price variation during its booking horizon is shown in Figure 6. It can be seen that the optimized price is lower than the real data during the first 30 days, but higher than the real one after the 30th day prior to departure.

The cumulative ratio of sold seats on every time point is shown in Figure 7. By using the optimized pricing policy, the load factor of Flight 2 increases from 83.3% to 92.4%.
The optimized price variation can be divided into two stages according to the increasing speed of the cumulative ratio. Stage 1 is from the 60th day to the 30th day, passengers choosing Flight 2 at this stage account for 63.2% of total transported passengers. In this stage, 82.3% of these passengers are traveling for leisure purpose. Thus, although the ratio of sold seats increases very quickly, the optimized pricing policy do not speed up the price increase by considering passengers’ price-sensitivity.

Stage 2 is from the 30th day to the departure day. Passengers who buy a ticket in this stage account for 36.8% of the total number of passengers. Because most passengers in this stage travel for work, the optimized pricing policy raises the price up to the second stage. After optimization, the profit of Flight 2 increases by 8.9%.

6. Pricing for competitive air route

In last section, the model is applied as an illustration to optimize the pricing strategy of a monopolistic air route owned by Dalian airport. The results indicate that airlines do not consider all factors influencing the air ticket purchasing behavior in price setting. Hence, the pricing policy used in reality is different from the one optimized by our proposed model. In order to simulate airlines’ real pricing policy, we adjust the attributes included in the utility function to simulate the purchase behavior perceived by airlines in the ticket pricing model. To simulate passengers’ responses to different pricing policy, the purchase model presented in Section 3.2 is also applied.

After comparing the utility functions specified by different combinations of attributes, the appropriate air ticket purchasing model is developed for calculating the transition matrix of the new pricing model. The estimated coefficients of the new air ticket purchase model are shown in Table 7.

In the utility function, ticket price, travel time as well as flight time are incorporated. Compared to Table 3, the value of Pseudo $R^2$ decreases, which means the new purchasing model cannot explain the survey data as well as the previous air ticket purchasing model in Section 3.2. It can be seen from the coefficients that the negative effect of air ticket price and travel time is attenuated.
Table 7. Current model of air ticket purchase.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRICE</td>
<td>-0.00223 (−21.49***)</td>
</tr>
<tr>
<td>TTIME</td>
<td>-0.26717 (−6.63***)</td>
</tr>
<tr>
<td>FTIME_1</td>
<td>-0.07566 (−0.95***)</td>
</tr>
<tr>
<td>FTIME_2</td>
<td>0.10430 (1.31***)</td>
</tr>
<tr>
<td>FTIME_3</td>
<td>0.06336 (0.78***)</td>
</tr>
</tbody>
</table>

Log-likelihood function: $-1483.41$
Pseudo $R^2 (\rho^2)$: 0.23

Figure 8. Price variations optimized by different models. Source: Realdata A is provided by Dalian airport.

We optimize the pricing strategy of Flight 4 (departing between 14:00 and 15:00) on the air route from Dalian to $D_2$ city using the new pricing model and the model established in Section 4, respectively. The optimized price variations and the cumulative ratio of sold seats during the booking horizon are shown in Figures 8 and 9, respectively. In the figure, the rhombus points stand for the real data, the square points stand for the prices optimized by the current model and the round points are the prices calculated by the original pricing model established in Section 3. It can be seen that the prices optimized by the current model fit the real data well. Therefore, the new pricing model can be used for describing airlines’ pricing behavior in reality.

From the above analysis, we found that airlines do not consider the influential factors, such as access time, flight delay and change fee, which affect the ticket purchasing behavior. Therefore, the current pricing policy may be not optimal if these exogenous factors change. In order to study the importance of considering the exogenous factors, the two models presented above are used to optimize the pricing policy of the flights on ‘Dalian – $D_2$ city’ air route with an assumption that the accessibility to Dalian airport is improved.

As Shenyang airport also has an air route to $D_2$ city, ‘Dalian – $D_2$ city’ air route is not the only choice in the multi-airport area. Table 8 describes the details of the flights on ‘Dalian – $D_2$ city’ and ‘Shenyang – $D_2$ city’.

We consider the following context: Dalian airport improves its access service, the travel time from the cities within the multi-airport area to Dalian airport is reduced by 30%,
Figure 9. Cumulative ratio of sold seats to capacity calculated by different models. Source: Realdata A is provided by Dalian airport.

Table 8. Detailed information of flights.

<table>
<thead>
<tr>
<th>Air route</th>
<th>'Dalian – D_2 city'</th>
<th>'Shenyang – D_2 city'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight time</td>
<td>9:00–10:00, 14:00–15:00, 21:00–22:00</td>
<td>10:00–11:00, 13:00–14:00</td>
</tr>
<tr>
<td>Aircraft type</td>
<td>Airbus A320</td>
<td>Airbus A320</td>
</tr>
<tr>
<td>Capacity</td>
<td>140 (economy class)</td>
<td>140 (economy class)</td>
</tr>
<tr>
<td>Travel time</td>
<td>1.5 hours (direct)</td>
<td>1.5 hours (direct)</td>
</tr>
</tbody>
</table>


Figure 10. Price variation calculated by different pricing models.

The volume of passengers choosing 'Dalian - D_2 city' air route increases because of the improved accessibility. We use the two models mentioned above to optimize the pricing policy of the flights departing from Dalian to D_2 city. Here we take Flight 4 as the example to analyze the results.

The price variation optimized by the two models is shown in Figure 10. In the figure, Markovdata 1 describes the price variation optimized by the current model and Markovdata 2 is the results optimized by the original model. It can be seen that during the first 40 days in the booking horizon, the price of Markovdata 1 is higher than that of Markovdata 2. However, the price of Markovdata 1 is lower than that of Markovdata 2 in the last 20 days.
The cumulative ratio of sold seats optimized by the two models are shown in Figure 11. Markovdata 1 is the cumulative ratio optimized by the current model and Markovdata 2 is the result optimized by the origin model.

During the first 40 days, the magnitude of price increase in Markovdata 1 is higher than that in Markovdata 2, but the quickly increased price leads to the situation that 23.9% of passengers who choose Flight 4 in Markovdata 2 are lost in Markovdata 1. Among the lost passengers, 12.4% transfer to other flights which depart from Dalian airport, 24.2% transfer to the flights owned by Shenyang airport, 14.5% choose high-speed rail, 26.7% choose the normal railway and 22.2% give up their travel. This is understandable because that the current model does not consider the improved utility caused by the shortened access time, which means the influence of the price on air ticket purchasing behavior is underestimated. Moreover, passengers arrive at this stage are price-sensitive, indicating that high price will lead to the use of other flights, other modes or giving up the travel. As a result, the profit of ‘Dalian – D2 city’ flight raises 4.3% by using our proposed pricing model.

Although the price optimized by the current model decreases in the last 20 days, the previous lost demand cannot be offset by the increased passengers at this stage. In addition, the reduced price will cut down the profit of Flight 4. One can see that the pricing policy optimized by the current model may lead to improper actions if the exogenous factors change. However, because the access time is incorporated in our proposed model, the effects of the improved access service will not be discounted by the quickly increased price. Therefore, the price policy keeps the increased passenger volume attracted by the improved access service for the airport and simultaneously improves the loading factor of the flight, which will finally lead to a higher profit for airlines.

7. Conclusion

The problem addressed in this study is the development and application of an improved air ticket pricing model. This has been seen important for airlines to get the maximum profit and market shares. In the present study, we proposed a Markov decision process to model airlines’ pricing mechanism. To represent the effects of various factors on the choices of air ticket purchases for passengers, the air ticket purchase behavior is modeled and incorporated using a nested logit model.
The model was estimated and applied using the data collected in a multi-airport region in China. As a result, the pricing strategy of a monopolistic air route departing from the multi-airport area is optimized. Results show that passengers with different travel purposes have different price-sensitivity and ticket-purchase time period, indicating that the price in each period should be set according to passengers’ characteristics. Furthermore, the optimal pricing strategy of one flight should not be on maximizing the immediate reward for each period, but the total profit during the whole booking horizon.

In the case study, the pricing strategy was simulated in the context that passengers can switch from one airport to another because of the improved airport access service. The results are compared specifically with that of the current model, which does not consider the exogenous influencing factors. The comparison results indicate that, when using the current model, the increased utility caused by the improved access service can be discounted by the quickly increased air ticket price. However, based on our proposal model, the air ticket price is not necessary to increase immediately because the effects of improved airport accessibility on passengers’ ticket purchase behavior is incorporated in the price setting. This indicates that, by knowing the choice behavior of air ticket purchase and the changes of influencing factors which are not controlled by airlines, policies regarding pricing and services can be made in a smart way to maximize profit and market share of airlines.

In spite of the value of incorporating the ticket purchasing model into the dynamic ticket pricing model, one should notice that further studies are needed. The present study is limited regarding the data used, e.g. number of airlines, rules on the use of the data, number of days of available data. Future studies with larger datasets are necessary to further examine the applicability of the proposed model in real cases. For example, if one has the data of historical pattern of seats sales, it would help simulate the real price settings. In addition, due to the competition between different multimodal travel options, it seems interesting to further explore the optimal pricing method considering the combined options, such as air-and-high-speed train. We will address these issues in our future works.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**


