Hindawi
Complexity
Volume 2019, Article ID 1568941, 24 pages
https://doi.org/10.1155/2019/1568941

Research Article

Estimation of Time-Varying Passenger Demand for High Speed Rail System

Tangjian Wei,[1,2] Feng Shi ,[1] and Guangming Xu [1]

[1]School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, China
[2]School of Transportation and Logistics, East China JiaoTong University, Nanchang, Jiangxi 330013, China

Correspondence should be addressed to Guangming Xu; xuguangming@csu.edu.cn

Received 15 May 2018; Revised 29 December 2018; Accepted 5 February 2019; Published 7 March 2019

Academic Editor: Dan Selişteanu

Passenger demand plays an important role in railway operation and organization, and this paper aims to estimate passenger time-varying demand by simulating the ticket-booking process for High Speed Rail (HSR) system. The ticket-booking process of each OD pair can be partition into discrete booking phases by the times when the tickets of any itinerary had sold out. The ticket booking volume of each itinerary is reversely assigned to its corresponding expected departure intervals to obtain the time-varying demand in each booking phase using the rooftop model, and the total time-varying demand are estimated by summing the time-varying demand distributions in all booking phases. Only with the data about the itinerary flow, the precedence relationship is introduced to constrain the ticket sold-out order of all itineraries for each OD pair. Based on the precedence relationships of itineraries, two typical situations are proposed, in which the Single Booking Phase Reverse Assignment (SBPRA) algorithm and the Multiple Booking Phases Reverse Assignment (MBPRA) algorithm are proposed to estimate the time-varying demand respectively. Case analysis on OD pair Beijing-Shanghai are presented, and the validity analysis demonstrates that the error rates of SBPRA algorithm and MBPRA algorithm are 8.64% and 6.37%, respectively.

1. Introduction

Passenger demand plays an important role in railway operation and organization. The conventional methods of line planning and scheduling generally aim to meet total passenger demand volumes of each OD pair [1–4]. However, High Speed Rail (HSR) is characterized as train operations with rapid speed and high frequency, and the transportation capacity of HSR is much larger than ordinary speed railway. For example, the distance from Beijing to Shanghai is about 1300 kilometers, and the service frequency between them in ordinary speed railway (speed≤160km/h) is 10 times a day in 2006. With the construction and operation of HSR (250km/h≤speed≤350km/h), the service frequency from Beijing to Shanghai is 42 times a day in 2018 (https://www.12306.cn). The HSR system can not only meet total passenger demand volumes with large transportation capacity, but also meet expected departure/arrival times of passengers with the high frequency of train operations. For a given OD pair, the demand rates with different expected departure/arrival times may differentiate within a day, which can be defined as the time-varying demand. With the improvement of HSR system, more and more studies focus on meeting the time-varying demand in the line planning and scheduling [5–8]. As for the above studies, the time-varying demand was adopted as input data, how to obtain that data has become an important common issue. However, little research has focused on this problem. This paper aims to fill in the research gap of the time-varying demand estimation of HSR. This paper focuses on the time-vary demand over the expected departure time. However, one may estimate the demand against the expected arrival time. If train travel time is constant (without delays and uncertainties), departure-based demand can be converted into arrival-based demand. However, if delays and uncertainties are to be considered, these two might not be simply converted to each other. We leave this for future research.

At present, there are different organization modes adopted by HSR among different countries, which cause distinctive estimation problems in terms of the time-varying demand. In some countries, such as China, passengers must book in advance, and sit according to their ticket number. Passenger flow in the train therefore is equal to the ticketing volume of this train. We are able to get all passenger flows of a OD pair from the Railway Ticketing System (RTS), and the transport volume of this OD pair is the sum of all passenger flows. In China, the ticket fare is fixed throughout the pre-sale period and will not be discounted due to multiple purchase of a single passenger or group purchase. There are also some countries, such as Japan, where HSR tickets have two types: free/non-reserved seats and reserved seats. Passengers who hold a free/non-reserved seat ticket can get on any train during the valid time. Passenger flow in each train therefore cannot be calculated by the ticketing data. In these areas, passengers who book tickets in advance, purchase round-trip tickets or group tickets may enjoy discounts. Hence, ticket discounts will affect the choice of HSR passengers. In this paper, we focus on solving estimation problem of HSR time-varying demand in situations like the case in China: all passengers must book in advance and sit according to their seat numbers marked on the tickets, i.e., the flow of each itinerary can be obtained from its corresponding ticketing volume; and the ticket fare for each itinerary during the whole pre-sale time window is fixed (booking time independent fare).

For HSR system, the time-varying demand of each OD pair has two features: total demand volume in one day and its corresponding time-varying distribution in the operation period of this day. At present, China has large transportation capacity and high frequency of HSR system. HSR passengers do not need to shift to other transport modes (except for some important holidays) due to insufficient capacity. Therefore, in our estimation problem, we assume that HSR system has enough capacity over the service time window to serve all passengers for each OD pair, and the flow shifting between transport modes is not considered. Then, the total demand volume of each OD can be obtained from Railway Ticketing System (RTS). In China, HSR passengers must book tickets in advance, and then, the real departure time of each passenger and the ticketing volume of each itinerary can be obtained from RTS. However, the real departure time of a passenger may deviate from his or her expected departure time. For instance, if there was no departing train at the expected departure time, or the ticket of that train had sold out, this passenger would have to adjust his or her departure time to another train. Hence, all passenger flows at their real departure time from the RTS cannot be regarded as the time-varying distribution directly. In this paper, we managed to tackle the issue that given the total demand volume of the OD pair and the ticketing volume of each itinerary, how to reverse the discrete ticketing volume to continuous time-varying distribution.

At present, there are little studies of estimation of HSR passenger time-varying demand. In the past few decades, previous studies on the time-varying demand estimation mainly focus on airline demand forecast, traffic dynamic OD estimation and public transit OD estimation.

In the aviation industry, accurate forecasts of passenger demand are the heart of a successful revenue management system [9]. The objective of revenue management or yield management is "selling the right seats to the right customers at the right prices" [10]. The forecasts are usually based on historical booking data. Thus, one of the main objectives of airline booking data analytics is to estimate unconstrained demand for each fare class using censored historical booking data [11]. The booking data are called censored because after a booking limit is reached, further booking attempts are rejected and not recorded by the system [9, 12]. Weatherford and Pölt [9], McGill [13], Mukhopadhyay et al., [14] and Ratliff et al., [15] have developed various remedial approaches for estimating unconstrained demand. Additionally, low-cost airlines irregularly launch ticket promotions, where fares may differ by day of the week and departure dates. The timing for purchasing air ticket is thus closely associated with fares. Passengers often do not buy airline tickets immediately when they determine their itinerary, and may choose to wait for fare promotions before marking reservations [16]. Therefore, Wen and Chen [16] and Chiou and Liu [17, 18] study the advance purchase behavior of air passengers using booking data. The result from Wen and Chen [16] indicated that lower fares increase the number of bookings and heterogeneous preferences in booking timing are present. Some travelers tend to book flights earlier than the other groups: these are the price-sensitive customers. The result from Chiou and Liu [18] indicates that advance purchase timing is associated with airfare, uncertainty of airfare, time of day, days of the week, months of the year and consecutive holidays. Diego [19] uses an original dataset with posted prices and sales to estimate the dynamic demand for airlines. They find that consumers become more price sensitive as time to departure nears which is consistent with having lower valuations and the number of active consumers increases closer to departure. However, HSR time-varying demand estimation problem is different from airline demand forecast. The ticket fare in airline changes dynamically during the pre-sale period, and the change of the ticket fare is associated with demand and advance purchase timing of passengers. The main purpose of demand forecasting in airline demand forecast is to obtain the demand corresponding to different fares during the pre-sale period in the segmentation market. However, for the estimation problem of time-varying demand in HSR system, the ticket fare is fixed throughout the pre-sale period and the effect of ticket fare changes on demand does not need to be considered. Passengers usually purchase their tickets as early as possible when they determine their itinerary, in order to purchase the tickets as close as possible to their expected departure time. Under the circumstance that HSR has enough capacity over the service time window to serve all passengers for each OD pair, we want to estimate the time-varying demand distribution in the operation period of each OD pair.

For traffic dynamic OD estimation problems, they mainly use some observation information, including link volumes, traffic counts and various forms of exogenous information, either in the forms of a priori knowledge or structural assumptions, to solve the estimation problems. A common approach is using autoregressive process to describe the dynamic process for the evolution of demand [20, 21]. Along this line, instead of an autoregressive process, Zhou and Mahmassani [22] developed a polynomial trend filter to capture the possible structural deviation in real-time demand. To improve unknown/equations ratio, Marzano and Papola [23] and Cascetta et al. [24] proposed a "quasi-dynamic" framework estimator. Djukic et al. [25] used principal component analysis to reduce the dimensionality of the estimation problem. In addition, for the dynamic demand estimation problem, not only within-day dynamic demand estimation, day-to-day dynamics has received much attention as well. For instance, Zhou and Mahmassani [22] modelled explicitly a day-to-day evolution process using a Kalman filter. Hazelton [26] used statistical estimation theory to estimate day-to-day OD matrices. Shao et al. [27] estimated the mean and covariance of peak hour OD demands from day-to-day traffic counts. However, HSR time-varying demand estimation problem is different from the above problem. Firstly, HSR trains operate according to timetable, and then the impact of timetable should be taken into consideration. Secondly, HSR trains operate during time-of-day periods; therefore, it is only necessary to analyze the within-day time-varying demand.

In Transit network, Wang et al. [28] and Chan et al. [29] used the boarding counts at every station from the Automatic Fare Collection system to generate the estimation problems, and some researchers use the boarding and alighting data from the Automatic Passenger Count systems and base on some assumptions and principles to estimate transit station-to-station OD matrices [30–36]. Although the transit network operates according to timetable, there are still some differences between transit network OD estimation and HSR time-varying demand estimation. In transit network, passengers don't need to book in advance, they purchase tickets when they arrive at the station, so the arriving or boarding time can be regarded as their expected departure time. However, in HSR system, specifically in China, passengers must book tickets in advance, only those passengers who hold tickets are allowed to get on. All passengers are scrambling for tickets, and occupying the train capacity, which are affected by several factors including timetable, travel cost and train capacity etc. The real departure time of passengers cannot be regarded as their expected departure time. In general, we need a proper method to resolve the HSR time-varying demand estimation problem.

The highlight of this paper is presented below. We utilize 'rooftop' model to figure out the relationship between itineraries and expected departure intervals, and then reversely assign the ticketing volume of itinerary to its corresponding expected departure intervals to obtain time-varying demand. By simulating the ticket-booking process of HSR, the precedence relationship is introduced to constrain the ticket sold-out order of all itineraries for each OD pair. Based on the precedence relationships of itineraries, we propose two typical situations of all preferable itineraries' tickets sold out order, i.e., for any itinerary, its tickets would be sold out in its first booking phase, and its tickets would be sold last from its first booking phase to last booking phase respectively. According to these two typical situations, two algorithms are proposed to estimate the time-varying demand respectively. Case analyses on OD pair Beijing-Shanghai are presented and the validity analyses of those two methods are further examined.

The rest of the paper is organized as follows. We propose the assumptions and state the details of HSR time-varying demand estimation problem in Section 2. Section 3 develops the Single Booking Phase Reverse Assignment (SBPRA) algorithm and the corresponding case analysis is presented. The Multiple Booking Phases Reverse Assignment (MBPRA) algorithm is proposed and the corresponding case analysis is given in Section 4. In Section 5, validity analysis is presented. Finally, Section 6 concludes the paper.

2. Problem Statement and Overview of Proposed Approach

In this section, we first summarize the major assumptions for the estimation problem of time-varying demand. Then, we describe the estimation problem of time-varying demand. After that, the rooftop model and simulated ticket-booking process will be introduced, respectively. At last, based on the booking phases, the reverse assignment method will be introduced.

2.1. Assumptions. The following assumptions are made for the demand estimation problem. Assumption (*A4*) reflects the current practice of the HSR system operations in China. However, different fares can be readily incorporated in our modeling framework.

(*A1*) HSR system has enough capacity over the service time window to serve all passengers for each OD pair, and the flow shifting between transport modes is not considered.

(*A2*) All passengers have the same value of time (homogeneous passengers).

(*A3*) Each passenger chooses the itinerary to minimize his or her travel cost (rational passengers).

(*A4*) The ticket fare for each itinerary during the whole pre-sale time window is fixed (booking time independent fare).

2.2. Problem of Time-Varying Demand Estimation. Before moving further, the major notations are shown in Appendix A. The time-varying demand estimation problem can be descripted as follows: Given each itinerary flow between each OD pair (r, s), we need to estimate the time-varying demand $Q_{rs}(x)$, $x \in [T^0_{rs}, T^1_{rs}]$, where x is the expected departure time for passengers, and $[T^0_{rs}, T^1_{rs}]$ is the operation period of OD pair (r, s).

For OD pair (r, s), let p^k_{rs} denote an itinerary, which means a travel scheme adopted by passengers, including trains and transfer stations from station r to station s. Denote P_{rs} as the itinerary set of OD pair (r, s). For any itinerary $p^k_{rs} \in P_{rs}$, its cost is defined as c^k_{rs}, which includes the in-train time costs, transfer time costs and ticket fees. The flow of p^k_{rs} is expressed as q^k_{rs}, which can be obtained from the RTS. With the large capacity and high frequency trains in the HSR network, the total demand volume could be obtained by summing all itinerary flows for OD pair (r, s). Therefore, the problem of time-varying demand estimation is how to reversely assign all itinerary flows to $[T^0_{rs}, T^1_{rs}]$ to obtain the time-varying distribution.

Next, we will describe the simulation of the ticket-booking process, and then reversely assign each itinerary flow to its corresponding expected departure time interval to estimate the time-varying demand.

2.3. Rooftop Model and Ticket-Booking Process of HSR Passengers. As HSR passengers must book tickets in advance, which is distinct from the conventional traffic assignment models of public transit, thus ticket-booking process need to be analyzed to model the passenger assignment.

In HSR system, passengers of each expected departure time book tickets of their preferable itineraries in the set of available itineraries. As the ticket-booking process goes on, some itineraries' tickets will be sold out. Then, some passengers have to book the tickets of their preferable itineraries in the set of remaining available itineraries. Hence, the ticket-booking process means that the above process is repeated until all passengers have booked their tickets.

From the above ticket-booking process, it is known that the set of available itineraries would be updated after the time division points when the tickets of any preferable itinerary had sold out. These time division points partition the pre-sale period into several pre-sale time intervals. In each pre-sale time interval, which also can be regarded as a booking phase, passengers choose their preferable itineraries in the current set of available itineraries. Thus, the continuous ticket-booking process can be partitioned into several discrete booking phases following the above method. In each booking phase, passengers' choice behaviors can be described as a rooftop model which will be introduced afterwards.

The rooftop model [37, 38] can be described based on the Assumption (*A2*) and (*A3*). If there was no itinerary at the time of x, he/she would adjust his/her departure time to another itinerary. Let t_{rs}^k denotes the departure time of itinerary $p_{rs}^k \in P_{rs}$ at station r. Define θ as the unit time fee for passengers who adjust expected departure time, and the travel cost of this passenger choosing itinerary p_{rs}^k is $\theta|x - t_{rs}^k| + c_{rs}^k$. Therefore, based on the Assumption (*A2*) and (*A3*), the preferable itinerary chosen by passengers with the expected departure time x can be expressed as:

$$p_{rs}(x) = \arg\min\left\{\theta\left|x - t_{rs}^k\right| + c_{rs}^k \mid p_{rs}^k \in P'_{rs}\right\}$$
$$P'_{rs} \subset P_{rs},\ (r,s) \in RS,\ x \in \left[T_{rs}^0, T_{rs}^1\right] \tag{1}$$

Besides, P'_{rs} is the current set of available itineraries of OD pair (r, s).

A simple example of preferable itineraries which are calculated by rooftop model is shown in Figure 1. For a given OD pair (r, s), there are 6 itineraries $p_{rs}^1, p_{rs}^2, \cdots, p_{rs}^6$, departure times are $t_{rs}^1, t_{rs}^2, \cdots, t_{rs}^6$ respectively, and the cost of each itinerary is $c_{rs}^1, c_{rs}^2, \cdots, c_{rs}^6$ respectively, which are shown by the height of the black vertical solid lines in Figure 1. For passengers who want to depart during $[T_{rs}^0, T_{rs}^1]$, the extra cost of adjusting their expected departure times for each itinerary is illustrated by red dotted line in Figure 1, and the slopes of those lines are $-\theta$ and θ. Based on Assumption (*A2*) and (*A3*), the set of preferable itineraries which calculated by Eq. (1) is $\overline{P}_{rs} = \{p_{rs}^1, p_{rs}^4, p_{rs}^6\}$. Passengers of all expected departure times would only book tickets with their current minimum travel costs in the preferable itinerary set $\overline{P}_{rs}$. For instance, passenger who wants to depart at t_{rs}^2 would book tickets of p_{rs}^1 rather than p_{rs}^2 due to the reason that the current travel cost of p_{rs}^1 is lower than p_{rs}^2 for him/her.

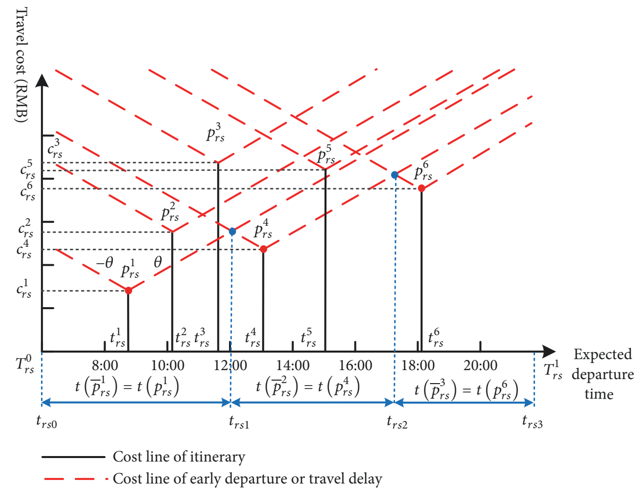

Figure 1: Preferable itineraries and expected departure intervals.

In addition, for OD pair (r, s), define $\overline{P}_{rs} \subset P'_{rs}$ as preferable itinerary set which includes all itineraries calculated by Eq. (1) for any $x \in [T_{rs}^0, T_{rs}^1]$. We sort every preferable itinerary in $\overline{P}_{rs}$ according to the departure time and still express it as $\overline{P}_{rs}$, i.e., $\overline{P}_{rs} = \{\overline{p}_{rs}^1, \overline{p}_{rs}^2, \cdots, \overline{p}_{rs}^i, \cdots, \overline{p}_{rs}^I\}$, with $\overline{t}_{rs}^{i-1} \le \overline{t}_{rs}^i, 1 < i \le I$.

In the above analysis, we can see that these preferable itineraries in $\overline{P}_{rs}$ divide the total expected departure time period $[T_{rs}^0, T_{rs}^1]$ into several intervals. For any preferable itinerary $\overline{p}_{rs}^i \in \overline{P}_{rs}$, we denote $t(\overline{p}_{rs}^i) = [t_{r,s,i-1}, t_{r,s,i}) \subset [T_{rs}^0, T_{rs}^1]$ as the expected departure interval of $\overline{p}_{rs}^i \in \overline{P}_{rs}$, which means that passengers whose expected departure times are within $[t_{r,s,i-1}, t_{r,s,i})$ would only choose the preferable itinerary $\overline{p}_{rs}^i$, and their actual departure times are $\overline{t}_{rs}^i$. Hence, $\overline{P}_{rs}$ divides $[T_{rs}^0, T_{rs}^1]$ into I expected departure intervals as follows.

$$\left[t_{r,s,0}, t_{r,s,1}\right), \left[t_{r,s,1}, t_{r,s,2}\right), \cdots, \left[t_{r,s,i-1}, t_{r,s,i}\right), \cdots,$$
$$\left[t_{r,s,I-1}, t_{r,s,I}\right] \tag{2}$$

For Eq. (2), t_{rsi} $(1 \le i \le I - 1)$ is the division point between the expected departure interval $t(\overline{p}_{rs}^i)$ and $t(\overline{p}_{rs}^{i+1})$. As shown in Figure 2, for the intersection point of line $y = \overline{c}_{rs}^i + \theta(x - \overline{t}_{rs}^i)$ and line $y = \overline{c}_{rs}^{i+1} - \theta(x - \overline{t}_{rs}^{i+1})$, its abscissa value is t_{rsi}, and its ordinate value satisfies the following equation:

$$\overline{c}_{rs}^i + \theta\left(t_{rsi} - \overline{t}_{rs}^i\right) = \overline{c}_{rs}^{i+1} - \theta\left(t_{rsi} - \overline{t}_{rs}^{i+1}\right) \tag{3}$$

Then, t_{rsi} can be calculated by the following equation:

$$t_{rsi} = \frac{1}{2\theta}\left(\theta\overline{t}_{rs}^{i+1} + \theta\overline{t}_{rs}^i + \overline{c}_{rs}^{i+1} - \overline{c}_{rs}^i\right),$$
$$i = 1, 2, \cdots, I - 1,\ (r,s) \in RS \tag{4}$$

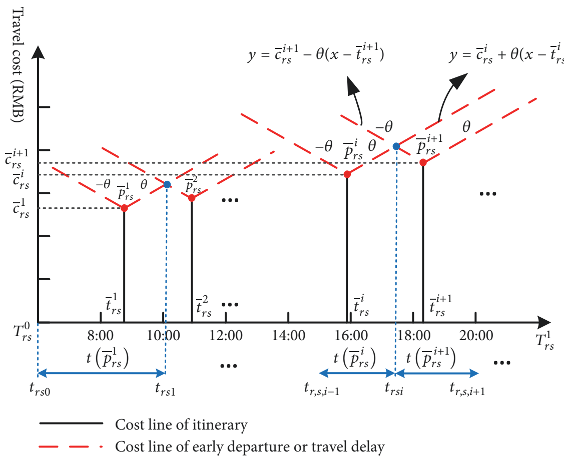

Figure 2: Division point between $t(\overline{p}_{rs}^{i})$ of $t(\overline{p}_{rs}^{i+1})$.

Besides, let $t_{rs0} = T_{rs}^{0}, t_{rsI} = T_{rs}^{1}$.

The ticket-booking process also can be partitioned into several booking phases. In the example of Figure 1, at the beginning of pre-sale period, denoted as the booking phase I, passengers can book tickets in the preferable itineraries set $\overline{P}_{rs} = \{\overline{p}_{rs}^{1}, \overline{p}_{rs}^{2}, \overline{p}_{rs}^{3}\} = \{p_{rs}^{1}, p_{rs}^{4}, p_{rs}^{6}\}$ which is calculated by Eq. (1). According to Eq. (2) and (4), $[T_{rs}^{0}, T_{rs}^{1}]$ would be divided into 3 expected departure intervals by $\overline{P}_{rs}$, $t(\overline{p}_{rs}^{1}) = t(p_{rs}^{1}) = [t_{rs0}, t_{rs1})$, $t(\overline{p}_{rs}^{2}) = t(p_{rs}^{4}) = [t_{rs1}, t_{rs2})$ and $t(\overline{p}_{rs}^{3}) = t(p_{rs}^{6}) = [t_{rs2}, t_{rs3}]$ respectively. Passengers who want to depart at $x \in t(\overline{p}_{rs}^{i}), i = 1, 2, 3$ would book tickets of preferable itinerary $\overline{p}_{rs}^{i}$. As the booking process proceeds, when any preferable itinerary's tickets had sold out, such as p_{rs}^{1}, this itinerary would be unavailable for passengers. The set of preferable itineraries would be updated to $\overline{P}_{rs} = \{\overline{p}_{rs}^{1}, \overline{p}_{rs}^{2}, \overline{p}_{rs}^{3}\} = \{p_{rs}^{2}, p_{rs}^{4}, p_{rs}^{6}\}$ by Eq. (1). Then the booking process moves on to booking phase II. In the booking phase II, $[T_{rs}^{0}, T_{rs}^{1}]$ would be partitioned by the new preferable itinerary set $\overline{P}_{rs} = \{\overline{p}_{rs}^{1}, \overline{p}_{rs}^{2}, \overline{p}_{rs}^{3}\} = \{p_{rs}^{2}, p_{rs}^{4}, p_{rs}^{6}\}$. The ticket-booking process is similar to the above process, and repeat it until all passengers have booked their tickets.

2.4. Reverse Assignment Based on the Booking Phases. As we analyzed in Section 2.3, ticket-booking process can be partitioned into several booking phases, and in each booking phase, passengers book tickets in the set of preferable itineraries. Hence, according to the booking phases, we adopt reverse assignment method to estimate the time-varying demand of HSR network.

The ticket booking volume of each preferable itinerary is reversely assigned to its corresponding expected departure interval in each booking phase, and the time-varying demand distribution of each booking phase can be calculated. The total time-varying demand can be obtained by summing all the time-varying demand distributions of all booking phases.

In this paper, without the data about the time division points when the tickets of each preferable itinerary had sold out, we only have the data of each itinerary flow for each OD pair to estimate the time-varying demand. Therefore, the key point of this problem is how to partition ticket-booking process into discrete booking phases, i.e., how to get the ticket sold-out order of all preferable itineraries, and how to determine the ticketing volume of each preferable itinerary in each booking phase.

The sold-out order of all preferable itineraries' tickets is various, and the ticketing volume of each preferable itinerary may be different in each booking phase. However, some itineraries' ticket sold-out order can be determined by their costs with Assumption $(A3)$.

For itinerary $p_{rs}^{k} \in P_{rs}$, if itinerary $p_{rs}^{h} \in P_{rs}, h \neq k$ satisfy the following Eq. (5) for any $x \in [T_{rs}^{0}, T_{rs}^{1}]$, then the tickets of p_{rs}^{h} wouldn't be sold until the tickets of p_{rs}^{k} had sold out.

$$\begin{aligned} &\theta \left| x - t_{rs}^{k} \right| + c_{rs}^{k} < \theta \left| x - t_{rs}^{h} \right| + c_{rs}^{h}, \\ &\qquad x \in \left[T_{rs}^{0}, T_{rs}^{1} \right], \; p_{rs}^{k}, p_{rs}^{h} \in P_{rs}, \; (r, s) \in RS \end{aligned} \tag{5}$$

Denote this precedence relationship as $p_{rs}^{k} \prec p_{rs}^{h}$, i.e., p_{rs}^{k} takes precedence over p_{rs}^{h}. From the precedence relationship constraint, the ticket sold-out order of them is that the tickets of p_{rs}^{k} had sold out earlier than that of p_{rs}^{h}. With the calculating formulation of travel cost of each itinerary, Eq. (5) can be equal to the following:

$$\theta \left| t_{rs}^{h} - t_{rs}^{k} \right| + c_{rs}^{k} < c_{rs}^{h}, \quad p_{rs}^{k}, p_{rs}^{h} \in P_{rs}, \; (r, s) \in RS \tag{6}$$

Thus, Eq. (6) can be used to easily check the precedence relationship between any two itineraries.

The precedence relationship has transitive property, i.e., if $p_{rs}^{k} \prec p_{rs}^{h}$ and $p_{rs}^{h} \prec p_{rs}^{g}$, then $p_{rs}^{k} \prec p_{rs}^{h} \prec p_{rs}^{g}$. With the precedence relationship, $p_{rs}^{k_1} \prec p_{rs}^{k_2} \cdots \prec p_{rs}^{k_n}$ can be obtained and regarded as a precedence relationship chain if there is no itinerary p_{rs}^{k} satisfying $p_{rs}^{k_n} \prec p_{rs}^{k}$, $p_{rs}^{k} \prec p_{rs}^{k_1}$ or $p_{rs}^{k_j} \prec p_{rs}^{k} \prec p_{rs}^{k_{j+1}}$ $(1 \leq j < n)$. Besides, if an itinerary has no precedence relationship with any other itineraries, this single itinerary also can be regarded as a precedence relationship chain. Hence, there are many precedence relationship chains in P_{rs}. The ticket sold-out order of all itineraries is constrained by the precedence relationship. The itinerary number of a precedence relationship chain can be regarded as its length. Denote the longest precedence relationship chain in P_{rs} as L_{rs}, and the itinerary set and itinerary number of L_{rs} are denoted as P_{rs}^{L} and M respectively. For further estimating the time-varying demand with the precedence relationship, we propose the following assumptions.

$(A5)$ For each expected departure time, the ticket-booking process is continuous and lasts the entire pre-sale time period, and for all expected departure times, the ticket-booking processes are synchronized during the pre-sale time period.

$(A6)$ Passengers' booking tickets of itineraries in the longest precedence relationship chain L_{rs} would last the entire pre-sale time period. The ticket-booking process would be partitioned into only M booking phases by the ticket sold-out time points of M itineraries in L_{rs}.

Input The effective operation period $[T_0, T_1]$ and the itinerary set P_{rs} of OD pair (r, s); the cost c_{rs}^k, and the departure time t_{rs}^k of itinerary $p_{rs}^k \in P_{rs}$; the unit time fee θ for adjusting expected departure time for passengers
Output M; $\check{m}_{rs}^k$, for $p_{rs}^k \in P_{rs}$
Begin
For any $p_{rs}^k, p_{rs}^h \in P_{rs}$
calculate the precedence relationship between p_{rs}^k and p_{rs}^h using Eq. (6);
do $M = 0, \check{P}_{rs}(1) = \emptyset$;
For any $p_{rs}^k \in P_{rs}$
do $d^k = |\{p_{rs}^h \in P_{rs} \mid p_{rs}^h \prec p_{rs}^k\}|$;
if $d^k = 0$
do $\check{m}_{rs}^k = 1, \check{P}_{rs}(1) = \check{P}_{rs}(1) \bigcup \{p_{rs}^k\}$;
do $M = 1, \check{P}_{rs}(2) = \emptyset$;
while $|P_{rs}| > \sum_{m=1}^{M} |\check{P}_{rs}(m)|$, do
Begin 1
for any $p_{rs}^h \prec p_{rs}^k, p_{rs}^k \in P_{rs} \setminus \bigcup_{m=1}^{M} \check{P}_{rs}(m), p_{rs}^h \in \check{P}_{rs}(M)$
if $d^k = 1$
do $\check{m}_{rs}^k = M + 1, \check{P}_{rs}(M + 1) = \check{P}_{rs}(M + 1) \cup \{p_{rs}^k\}$;
otherwise
do $d^k = d^k - 1$;
$M = M + 1, \check{P}_{rs}(M + 1) = \emptyset$;
Return 1
End

Algorithm 1: The First Booking Phase Partition algorithm.

For the Assumptions (*A5*) and (*A6*), it should be noted that the tickets of each itinerary which is not in P_{rs}^L would also have sold out in one of the above M ticket sold-out time points. For instance, in Figure 1, $p_{rs}^1 \prec p_{rs}^2 \prec p_{rs}^3$ is L_{rs}. Based on Assumption (*A3*), (*A5*) and (*A6*), the ticket-booking process would only be partitioned into 3 booking phases by the ticket sold-out time points of p_{rs}^1, p_{rs}^2 and p_{rs}^3. For precedence relationship chain p_{rs}^6, the ticket sold-out time points of p_{rs}^6 would be at the end of booking phase I or II or III.

Hence, the tickets of the $m^{th}, m = 1, 2, \cdots, M$ itinerary in L_{rs} would be sold only in booking phase m. For other itineraries not in P_{rs}^L, the sale of their tickets may last for more than one booking phase. For instance, in Figure 1, for itinerary p_{rs}^6, its tickets may be sold for more than one booking phase, and its ticket sold-out time point may be the end of any booking phase.

For any $p_{rs}^k \in P_{rs}$, denote the first and the last booking phase when its tickets can be sold as $\check{m}_{rs}^k$ and $\widehat{m}_{rs}^k$ respectively. In the following content, $\check{m}_{rs}^k$ and $\widehat{m}_{rs}^k$ is described as the first and the last booking phase of p_{rs}^k briefly. Obviously, if $p_{rs}^k \in P_{rs}^L$, its tickets would be sold out in one booking phase, i.e. $\check{m}_{rs}^k = \widehat{m}_{rs}^k$. If $p_{rs}^k \notin P_{rs}^L$, the sale of its tickets may last from its first booking phase $\check{m}_{rs}^k$ to booking phase $m = \check{m}_{rs}^k, \check{m}_{rs}^{k+1}, \cdots, \widehat{m}_{rs}^k$. For any $p_{rs}^k \in P_{rs}$, we designed the Algorithms 1 and 2, which is shown in Appendix B and C, to calculate $\check{m}_{rs}^k$ and $\widehat{m}_{rs}^k$ respectively.

It should be noted that $\check{P}_{rs}(m) = \{p_{rs}^k \mid p_{rs}^k \in P_{rs}, \check{m}_{rs}^k = m\}$ for $m = 1, 2, \cdots, M$ can be calculated by the Algorithm 1. We denote the first booking phase scheme of P_{rs} as $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$, which will be used to calculate the value of $\widehat{m}_{rs}^k$ for $p_{rs}^k \in P_{rs}$ in the Algorithm 2.

In Algorithm 2, $\widehat{P}_{rs}(m) = \{p_{rs}^k \mid p_{rs}^k \in P_{rs}, \widehat{m}_{rs}^k = m\}$ can be calculated for $m = 1, 2, \cdots, M$. We denote the last booking phase scheme of P_{rs} as $\widehat{P}_{rs}(1), \widehat{P}_{rs}(2), \cdots, \widehat{P}_{rs}(M)$.

The example of the calculations of the first and the last booking phase for each itinerary in Figure 1 can be described as follows. According to the first and the last booking phase partition algorithm, we can obtain the following first and the last booking phase scheme.

$$\begin{aligned}
\check{P}_{rs}(1) &= \left\{p_{rs}^1, p_{rs}^4, p_{rs}^6\right\}, \\
\check{P}_{rs}(2) &= \left\{p_{rs}^2, p_{rs}^5\right\}, \\
\check{P}_{rs}(3) &= \left\{p_{rs}^3\right\} \\
\widehat{P}_{rs}(1) &= \left\{p_{rs}^1, p_{rs}^4\right\}, \\
\widehat{P}_{rs}(2) &= \left\{p_{rs}^2\right\}, \\
\widehat{P}_{rs}(3) &= \left\{p_{rs}^3, p_{rs}^5, p_{rs}^6\right\}
\end{aligned} \tag{7}$$

The ticket sold-out order of itineraries in each precedence relationship chain is constrained by the precedence relationship. For instance, in Figure 1, for precedence relationship chain $p_{rs}^4 \prec p_{rs}^5$, due to the reason that $\check{m}_{rs}^4 = \widehat{m}_{rs}^4 = 1$, the tickets of p_{rs}^4 would only be sold in booking phase I. Due to the reason that $\check{m}_{rs}^5 = 2, \widehat{m}_{rs}^5 = 3$, the tickets of p_{rs}^5 can be sold in booking phase II or in booking phase II and III. In conclusion, due to the reason that the ticket

Input The effective operation period $[T_0, T_1]$ and the itinerary set P_{rs} of OD pair (r, s); the cost c_{rs}^k, and the departure time t_{rs}^k of itinerary $p_{rs}^k \in P_{rs}$; the unit time fee θ for adjusting expected departure time for passengers; the first booking phase scheme $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$;
Output $\widehat{m}_{rs}^k$ for $p_{rs}^k \in P_{rs}$
Begin
For any $p_{rs}^k, p_{rs}^h \in P_{rs}$
calculate the precedence relationship between p_{rs}^k and p_{rs}^h using Eq. (6);
do $\widehat{P}_{rs}(1) = \emptyset$;
For any $p_{rs}^k \in \check{P}_{rs}(1)$, do $\widehat{m}_{rs}^k = 1, \widehat{P}_{rs}(1) = \widehat{P}_{rs}(1) \cup \{p_{rs}^k\}$;
For $m = 2, 3, \cdots, M$, do
Begin 1
do $\widehat{P}_{rs}(m) = \emptyset$;
For any $p_{rs}^k \in \check{P}_{rs}(m)$
do $\widehat{m}_{rs}^k = m, \widehat{P}_{rs}(m) = \widehat{P}_{rs}(m) \cup \{p_{rs}^k\}$;
For any $p_{rs}^h \in \widehat{P}_{rs}(m-1)$,
if there is no $p_{rs}^k \in \check{P}_{rs}(m)$ satisfying $p_{rs}^h \prec p_{rs}^k$
do $\widehat{m}_{rs}^h = m, \widehat{P}_{rs}(m) = \widehat{P}_{rs}(m) \cup \{p_{rs}^h\}$,
$\widehat{P}_{rs}(m-1) = \widehat{P}_{rs}(m-1) \setminus \{p_{rs}^h\}$;
Return 1
End

Algorithm 2: The Last Booking Phase Partition algorithm.

sold-out time point of preferable itinerary may be at the end of different booking phases, we proposed 2 typical situations of all preferable itineraries' tickets sold out order to estimate the time-varying demand respectively.

Typical Situation 1. For any itinerary $p_{rs}^k \in P_{rs}$, its tickets would be sold out in its first booking phase $\check{m}_{rs}^k$.

Typical Situation 2. For any itinerary $p_{rs}^k \in P_{rs}$, the sale of its tickets would last from the first booking phase $\check{m}_{rs}^k$ to the last booking phase $\widehat{m}_{rs}^k$.

For each typical situation, the HSR passenger time-varying demand estimation can be described as follows: Firstly, partition the ticket-booking process into several booking phases and figure out the set of preferable itineraries in each booking phase. Secondly, in each booking phase, divide the total expected departure time period $[T_{rs}^0, T_{rs}^1]$ into expected departure intervals based on the set of preferable itineraries. Thirdly, ticket booking volume of each preferable itinerary is reversely assigned to its corresponding expected departure interval to obtain the time-varying demand distribution in each booking phase. At last, sum the time-varying demand distributions of all booking phases to obtain the time-varying demand.

The following content are based on *Typical Situation 1* and *2* to design two corresponding time-varying demand estimation algorithms.

3. Single Booking Phase Reverse Assignment Algorithm

3.1. Framework of Single Booking Phase Reverse Assignment Algorithm. According to *Typical Situation 1*, all itineraries in P_{rs} would only be preferable in their first booking phase. Denote $\overline{P}_{rs}(m)$ as the set of preferable itineraries in the booking phase $m = 1, 2, \cdots, M$. Based on Assumption (*A3*), (*A5*), (*A6*) and *Typical Situation 1*, the booking phases can be obtained by the first booking phase partition algorithm. Hence, the booking phase scheme is $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$, and $\overline{P}_{rs}(m) = \check{P}_{rs}(m)$, $m = 1, 2, \cdots, M$.

In the booking phase $m = 1, 2, \cdots, M$, for any preferable itinerary $\overline{p}_{rs}^i \in \overline{P}_{rs}(m)$, denote its corresponding expected departure interval as $t_m(\overline{p}_{rs}^i)$. According to *Typical Situation 1*, the ticket booking volume of preferable itinerary $\overline{p}_{rs}^i$ in booking phase m is equal to its flow $\overline{q}_{rs}^i$. We reversely assign the ticketing volume $\overline{q}_{rs}^i$ of $\overline{p}_{rs}^i$ to its expected departure interval $t_m(\overline{p}_{rs}^i)$. $\overline{q}_{rs}^i$ is evenly assigned to expected departure interval $t_m(\overline{p}_{rs}^i)$, which leads to a distribution in $t_m(\overline{p}_{rs}^i)$, i.e.,

$$Q_{rs}^i(x) = \frac{\overline{q}_{rs}^i}{\left|t_m\left(\overline{p}_{rs}^i\right)\right|}$$
$$x \in t_m\left(\overline{p}_{rs}^i\right), \ \overline{p}_{rs}^i \in \overline{P}_{rs}(m), \ m = 1, 2, \cdots, M \tag{8}$$

where $|t_m(\overline{p}_{rs}^i)| = t_{r,s,i} - t_{r,s,i-1}$.

For the example in Figure 1, the single booking phase reverse assignment can be described as follows. Based on Assumption (*A3*), (*A5*), (*A6*) and *Typical Situation 1*, the booking phase scheme is expressed as follows.

$$\begin{aligned} \overline{P}_{rs}(1) &= \check{P}_{rs}(1) = \left\{p_{rs}^1, p_{rs}^4, p_{rs}^6\right\} \\ \overline{P}_{rs}(2) &= \check{P}_{rs}(2) = \left\{p_{rs}^2, p_{rs}^5\right\} \\ \overline{P}_{rs}(3) &= \check{P}_{rs}(3) = \left\{p_{rs}^3\right\} \end{aligned} \tag{9}$$

In booking phase I, using Eq. (8), evenly assign the flows q_{rs}^1, q_{rs}^4 and q_{rs}^6 to their corresponding expected departure intervals $t_1(\overline{p}_{rs}^1) = t_1(p_{rs}^1), t_1(\overline{p}_{rs}^2) = t_1(p_{rs}^4)$ and $t_1(\overline{p}_{rs}^3) = t_1(p_{rs}^6)$ to obtain the time-varying distribution of booking phase I. In booking phase II, evenly assign the flows q_{rs}^2 and q_{rs}^5 to their corresponding expected departure interval $t_2(\overline{p}_{rs}^1) = t_2(p_{rs}^2)$ and $t_2(\overline{p}_{rs}^2) = t_2(p_{rs}^5)$ to obtain the time-varying distribution of booking phase II. Then in the booking phase III, evenly assign the flow q_{rs}^3 of preferable itinerary p_{rs}^3 to its corresponding expected departure interval $t_3(\overline{p}_{rs}^1) = t_3(p_{rs}^3)$ and obtain the time-varying distribution of booking phase III. At last, sum the time-varying distributions of all booking phases to obtain the time-varying demand.

Based on Assumptions (*A3*), (*A5*), (*A6*) and *Typical Situation 1*, we can use the first booking phase partition algorithm to get the booking phase scheme $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$. In the booking phase $m = 1, 2, \cdots, M$, for any preferable itinerary $\overline{p}_{rs}^i \in \overline{P}_{rs}(m)$, its corresponding expected departure interval $t_m(\overline{p}_{rs}^i)$ can be obtained by Eq. (2) and (4). The flow $\overline{q}_{rs}^i$ of $\overline{p}_{rs}^i$ is reversely assigned to $t_m(\overline{p}_{rs}^i)$ by Eq. (8). In general, we design the Single Booking Phase Reverse Assignment (SBPRA) algorithm, which is shown in Appendix D, to estimate HSR time-varying demand. The flow diagram of the SBPRA algorithm is shown in Figure 3.

3.2. Case Analysis. We apply the data (shown in Appendix E) of OD pair Beijing-Shanghai on December 1st 2015 from the RTS into the SBPRA algorithm. There are 34 itineraries for OD pair Beijing-Shanghai, and the departure time, cost and flow of each itinerary are given in Table 1. The total effective operation period of this OD pair $[T_{rs}^0, T_{rs}^1] = [6{:}00, 20{:}00]$. The average monthly residential incomes of Beijing and Shanghai are 7086 RMB and 6504 RMB in 2015 respectively [39, 40]. Based on 22 working days in a month and 8 working hours in a day, the average income can be expressed as 0.67 RMB per minute and 0.62 RMB per minute respectively. We use average residential income to express the unit time fee of adjusted expected departure time, i.e., $\theta = (0.67 + 0.62)/2 = 0.65$ RMB per minute.

We calculate the passenger time-varying demand $Q_{rs}(x), x \in [6{:}00, 20{:}00]$ for OD pair Beijing-Shanghai with SBPRA algorithm. Firstly, use the Algorithm 1 to calculate the first booking phase scheme $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$. Secondly, do $\overline{P}_{rs}(m) \longleftarrow \check{P}_{rs}(m), m = 1, 2, \cdots, M$ to obtain the booking phase scheme $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$. Thirdly, for $m = 1, 2, \cdots, M$, we calculate the expected departure interval $t_m(\overline{p}_{rs}^i)$ for all $\overline{p}_{rs}^i \in \overline{P}_{rs}(m)$ shown in Figure 3 by Eq. (2) and (4); then the ticket booking volume of each preferable itinerary in booking phase $m = 1, 2, \cdots, M$ is reversely assigned to its corresponding expected departure interval, shown in Figure 4; and the distribution of the reverse assignment in booking phase $m = 1, 2, \cdots, M$ is illustrated in Figure 5; at last, the accumulated time-varying demand distribution is shown in Figure 6.

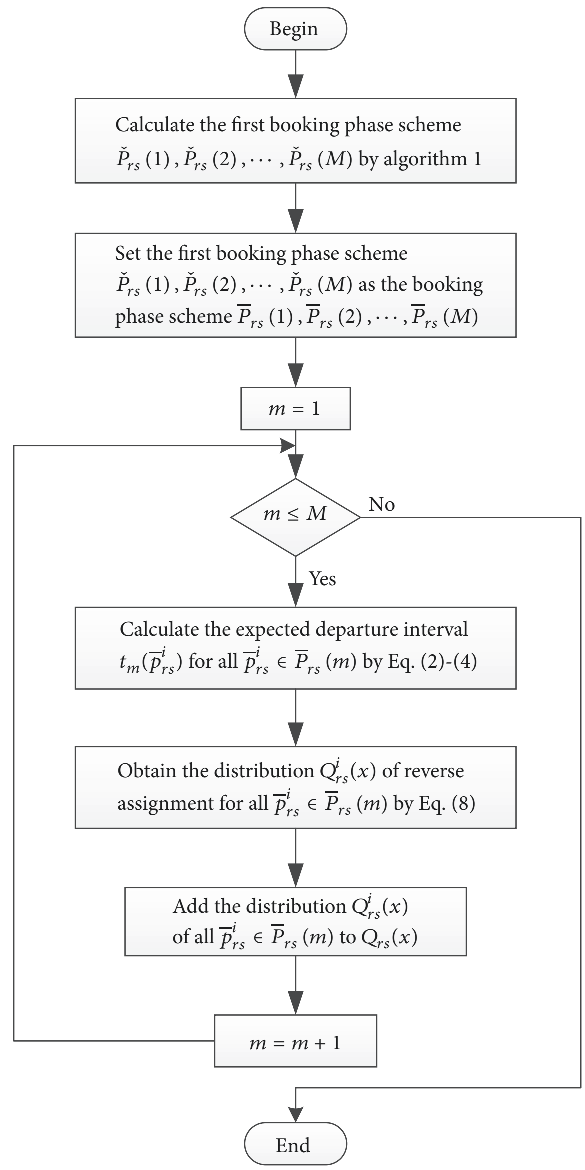

Figure 3: The flow diagram of the SBPRA algorithm.

In Algorithm 1, for any $p_{rs}^k, p_{rs}^h \in P_{rs}$, we use Eq. (6) to calculate the precedence relationship between p_{rs}^k and p_{rs}^h shown in Table 2. According to the precedence relationship in the Table 2, firstly, we set $M = 0, \check{P}_{rs}(1) = \emptyset$, and calculate the d^k of $p_{rs}^k \in P_{rs}$ by $d^k = |\{p_{rs}^h \in P_{rs} \mid p_{rs}^h \prec p_{rs}^k\}|$. For any itinerary $p_{rs}^k \in P_{rs}$, if $d^k = 0$, then $\check{m}_{rs}^k = 1$ and put it in $\check{P}_{rs}(1)$. The d^k and $\check{m}_{rs}^k$ of p_{rs}^k are shown in column $M = 0$ of Table 3. For example, G11≺G105, then d^{G105} is 1; and G1≺G115 and G13≺G115, then d^{G115} is 2. The itinerary set $\check{P}_{rs}(1)$ is shown in Table 4.

Secondly, set $M = 1, \check{P}_{rs}(2) = \emptyset$. For any itinerary $p_{rs}^h \prec p_{rs}^k, p_{rs}^k \in P_{rs} \setminus \check{P}_{rs}(1), p_{rs}^h \in \check{P}_{rs}(1)$, if $d^k = 1$, then $\check{m}_{rs}^k = M + 1 = 2$ and put this itinerary in $\check{P}_{rs}(2)$; otherwise, set $d^k = d^k - 1$. The d^k and $\check{m}_{rs}^k$ of p_{rs}^k are shown in column $M = 1$ of Table 3. For example, G11≺G105, G105 $\in P_{rs} \setminus \check{P}_{rs}(1)$, G11 $\in$

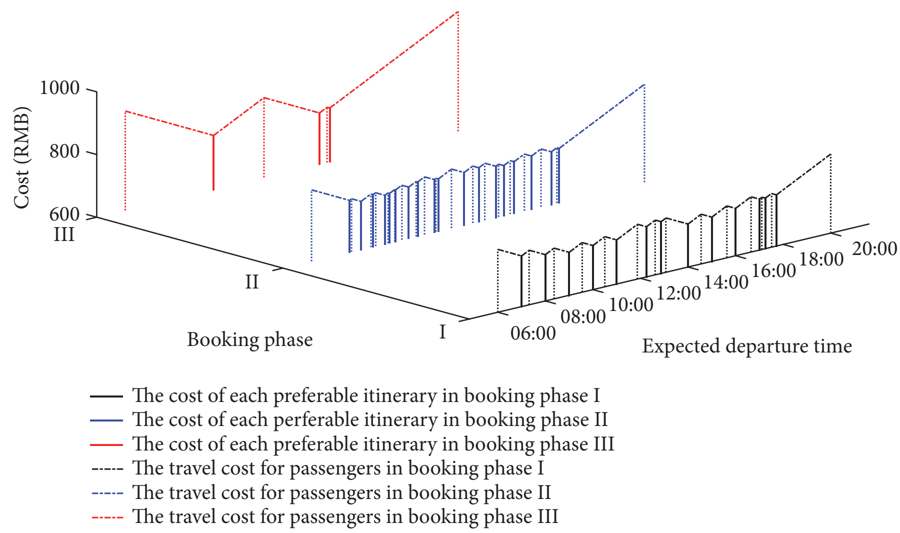

FIGURE 4: The cost and the expected departure interval of each preferable itinerary in each booking phase from the SBPRA algorithm.

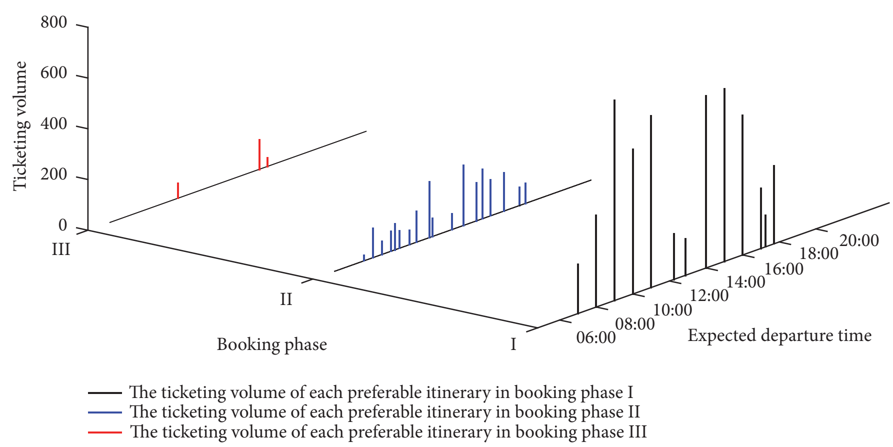

FIGURE 5: The ticketing volume of each itinerary in each booking phase from the SBPRA algorithm.

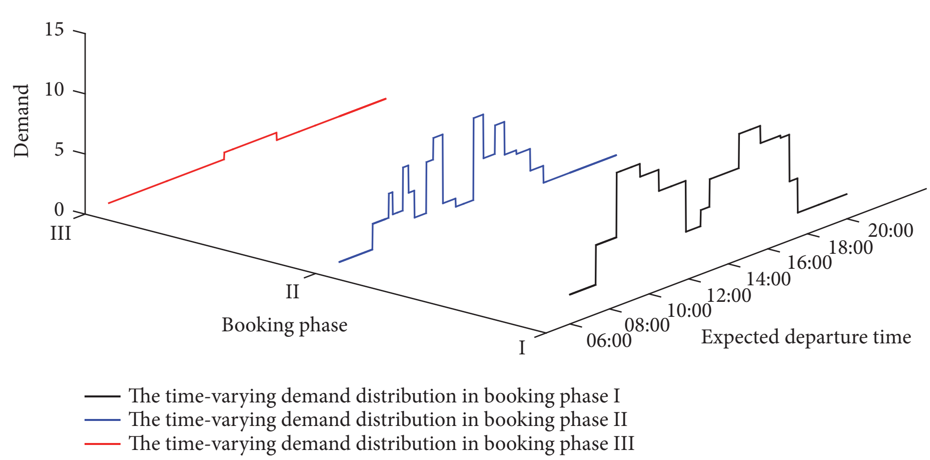

FIGURE 6: The time-varying demand distribution in each booking phase of the SBPRA algorithm.

Table 1: The data of each itinerary of OD pair Beijing-Shanghai.

p_{rs}^k	t_{rs}^k	c_{rs}^k	q_{rs}^k
G101	7:00	761.6	193
G105	7:36	765.6	21
G11	8:00	747.6	361
G107	8:05	753.6	114
G111	8:35	766.1	52
G1	9:00	737.1	788
G113	9:05	755.1	77
G41	9:17	757.6	103
G115	9:32	764.1	68
G117	9:43	772.1	58
G13	10:00	740.6	571
G119	10:05	763.1	55
G121	10:28	772.6	121
G15	11:00	740.6	676
G125	11:10	767.6	219
G411	11:20	766.6	71
G129	12:15	760.1	180
G131	12:25	767.6	61
G133	12:52	765.1	144
G135	13:02	771.6	238
G137	13:45	763.6	150
G3	14:00	737.1	678
G43	14:05	759.6	193
G139	14:10	763.6	116
G141	14:31	764.1	141
G143	14:36	771.6	35
G17	15:00	741.1	680
G145	15:15	769.1	151
G19	16:00	751.1	550
G147	16:05	767.6	71
G149	16:25	772.6	79
G21	17:00	762.6	238
G153	17:15	761.1	123
G157	17:43	763.1	307

$\check{P}_{rs}(1)$, and $d^{\mathrm{G105}} = 1$, then $\check{m}_{rs}^{\mathrm{G105}} = 2$. G1≺G117, G117 $\in P_{rs} \setminus \check{P}_{rs}(1)$, G1 $\in \check{P}_{rs}(1)$, and $d^{\mathrm{G117}} = 3$, then $d^{\mathrm{G117}} = 3 - 1 = 2$; G13≺G117 and G13 $\in \check{P}_{rs}(1)$, then $d^{\mathrm{G117}} = 2 - 1 = 1$. The itinerary set $\check{P}_{rs}(2)$ is shown in Table 4.

Thirdly, set $M = 2, \check{P}_{rs}(3) = \emptyset$. For any itinerary $p_{rs}^h \prec p_{rs}^k, p_{rs}^k \in P_{rs} \setminus \bigcup_{m=1}^{2} \check{P}_{rs}(m), p_{rs}^h \in \check{P}_{rs}(2)$, if $d^k = 1$, then $\check{m}_{rs}^k = M + 1 = 3$ and put this itinerary in $\check{P}_{rs}(3)$; otherwise, set $d^k = d^k - 1$. The d^k and $\check{m}_{rs}^k$ of p_{rs}^k are shown in column $M = 2$ of Table 3. For example, G115≺G117, G117 $\in P_{rs} \setminus \bigcup_{m=1}^{2} \check{P}_{rs}(m)$, G115 $\in \check{P}_{rs}(2)$, and $d^{\mathrm{G117}} = 1$, then $\check{m}_{rs}^{\mathrm{G117}} = 3$. The itinerary set $\check{P}_{rs}(3)$ is shown in Table 4.

Figure 4 illustrates the expected departure interval and travel cost of each preferable itinerary in each booking phase. From booking phase I to III, the number of preferable itinerary is decreasing rapidly, the time range of each preferable itinerary's expected departure interval is wider and the height of vertical solid line which represents the cost of each preferable itinerary is rising.

Figure 5 shows the change of ticketing volume of each preferable itinerary. From booking phase I to III, the height of vertical solid lines, regarded as ticketing volume of each preferable itinerary, is declining drastically. Based on this information, the feature of tickets-booking process simulated by the SBPRA algorithm can be described as follows:

(1) For OD pair (r, s), passengers choose their preferable itineraries and book their corresponding tickets with the minimum travel cost, so the sold-out order of all preferable itineraries' tickets is from low cost to high cost.

(2) For OD pair (r, s), all itineraries' tickets had sold out in the first booking phase, it causes that most passengers book tickets at early booking phase, the travel cost of those passengers are relatively low and the adjusted expected departure time ranges are relatively narrow. In contrast, a small percentage of passengers who book tickets at late booking phase have to choose those itineraries with higher cost and need to adjust their expected departure time in a wider range.

We reversely assign each preferable itinerary's ticket booking volume in Figure 5 to its corresponding expected departure interval in Figure 4 to obtain the time-varying demand distribution in each booking phase in Figure 6. Sum the distributions in Figure 6 from booking phase I to III to get the accumulated time-varying demand distribution in Figure 7. The red solid line in Figure 7 is the time-vary demand of OD pair Beijing-Shanghai calculated by SBPRA algorithm, and the Table 5 shows the numerical results of time-varying demand in details.

The method of polynomial fitting is adopted for fitting the above distribution of time-varying demand, and the result is shown in Figure 8. It can be seen that the travel demands before around 7:30 and after 17:30 are relatively low. From 9:00 to 11:00 and 14:00 to 16:00, there are two demand peaks. Around 12:00, the drop of travel demand is probably due to the approaching lunch time.

For the sensitive analysis of parameter θ in the SBPRA algorithm, we also calculate the numbers of booking phases and time-varying demand distributions for different values θ, shown in Table 6 and Figure 9 respectively. From Table 6, it can be seen that with larger value of θ, passenger is more concerned about the cost of adjusting the expected departure time, which results in the decreasing of numbers of booking phases. From Figure 9, with the increasing of parameter θ, the fluctuation of the time-varying demand distribution is increasing.

4. Multiple Booking Phases Reverse Assignment Algorithm

Since the above SBPRA algorithm is the estimation method based on *Typical Situation 1*, each itinerary flow is reversely assigned to corresponding expected departure intervals in its first booking phase. In this section, based on *Typical Situation*

Table 2: The precedence relationship.

G11<G105	G1<G115	G13<G119	G129<G131	G43<G139	G141<G143
G11<G107	G1<G117	G13<G121	G133<G135	G3<G141	G17<G145
G1<G111	G115<G117	G15<G121	G3<G137	G17<G141	G19<G147
G1<G113	G13<G115	G15<G125	G3<G43	G17<G143	G19<G149
G1<G41	G13<G117	G15<G411	G3<G139	G3<G143	

Table 3: The precedence relationship.

p_{rs}^k	$M = 0$		$M = 1$		$M = 2$		p_{rs}^k	$M = 0$		$M = 1$		$M = 2$	
	d^k	$\check{m}_{rs}^k$	d^k	$\check{m}_{rs}^k$	d^k	$\check{m}_{rs}^k$		d^k	$\check{m}_{rs}^k$	d^k	$\check{m}_{rs}^k$	d^k	$\check{m}_{rs}^k$
G101	0	1					G131	1			2		
G105	1			2			G133	0	1				
G11	0	1					G135	1			2		
G107	1			2			G137	1			2		
G111	1			2			G3	0	1				
G1	0	1					G43	1			2		
G113	1			2			G139	2		1			3
G41	1			2			G141	2			2		
G115	2			2			G143	3		1			3
G117	3		1			3	G17	0	1				
G13	0	1					G145	1			2		
G119	1			2			G19	0	1				
G121	2			2			G147	1			2		
G15	0	1					G149	1			2		
G125	1			2			G21	0	1				
G411	1			2			G153	0	1				
G129	0	1					G157	0	1				

Table 4: The first booking phase scheme.

Booking phase	Preferable itineraries
$\check{P}_{rs}(1)$	G101, G11, G1, G13, G15, G129, G133, G3, G17, G19, G21, G153, G157
$\check{P}_{rs}(2)$	G105, G107, G111, G113, G41, G115, G119, G121, G125, G411, G131, G135, G137, G43, G141, G145, G147, G149
$\check{P}_{rs}(3)$	G117, G139, G143

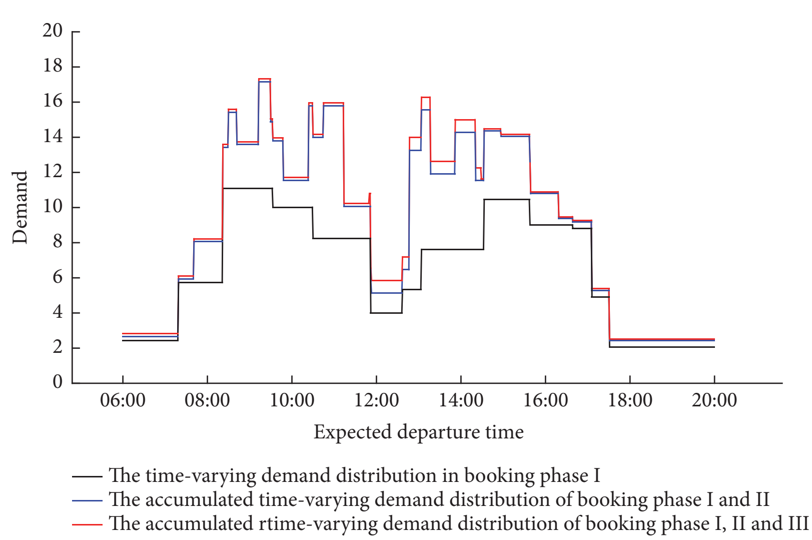

Figure 7: The accumulated time-varying demand distribution of the SBPRA algorithm.

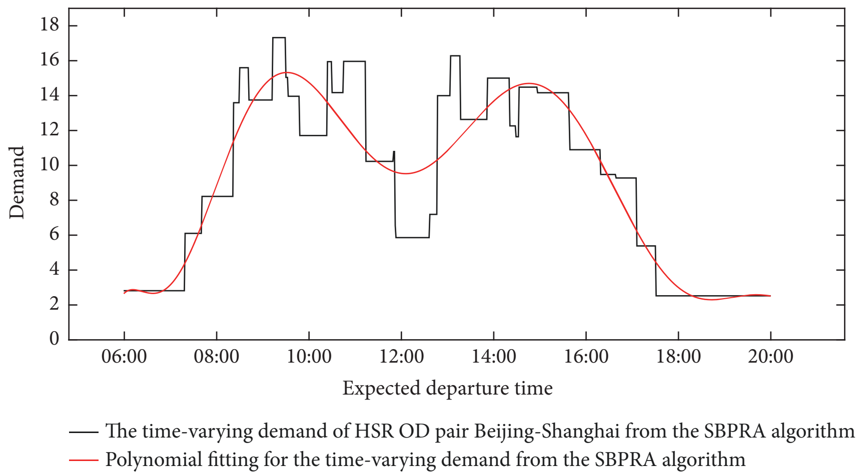

FIGURE 8: Polynomial fitting for the time-varying demand from the SBPRA algorithm.

TABLE 5: The time-varying demand from the SBPRA algorithm.

Time range	Demand volume (person/minute)	Time range	Demand volume (person/minute)
[6:00, 7:19)	2.8	[11:53, 12:37)	5.9
[7:19, 7:41)	6.1	[12:37, 12:47)	7.2
[7:41, 8:22)	8.2	[12:47, 13:04)	14
[8:22, 8:30)	13.6	[13:04, 13:17)	16.3
[8:30, 8:42)	15.6	[13:17, 13:52)	12.6
[8:42, 9:13)	13.7	[13:52, 14:21)	15
[9:13, 9:30)	17.3	[14:21, 14:29)	12.3
[9:30, 9:33)	15	[14:29, 14:33)	11.6
[9:33, 9:48)	14	[14:33, 14:57)	14.5
[9:48, 10:24)	11.7	[14:57, 15:38)	14.2
[10:24, 10:30)	15.9	[15:38, 15:39)	12.7
[10:30, 10:45)	14.2	[15:39, 16:19)	10.9
[10:45, 11:14)	16	[16:19, 16:39)	9.5
[11:14, 11:50)	10.2	[16:39, 17:06)	9.3
[11:50, 11:52)	10.8	[17:06, 17:31)	5.4
[11:52, 11:53)	6.6	[17:31, 20:00]	2.5

TABLE 6: Parameter θ and its corresponding number of partition booking phases.

θ	0.05	0.35	0.65	0.95	1.25
Number of booking phases	8	4	3	2	2

2, each itinerary flow is reversely assigned to corresponding expected departure intervals from its first booking phase to last booking phase. The MBPRA algorithm will be introduced next.

4.1. Framework of the Multiple Booking Phases Reverse Assignment Algorithm. According to *Typical Situation 2*, itinerary $p_{rs}^k \in P_{rs}$ would be preferable for passengers in the booking phase $\check{m}_{rs}^k, \check{m}_{rs}^k + 1, \cdots, \widehat{m}_{rs}^k$. As a result, the set of preferable itineraries $\overline{P}_{rs}(m)$ in the booking phase m can be calculated as follows.

$$\overline{P}_{rs}(m) = \bigcup_{j=1}^{m} \check{P}_{rs}(j) \cap \bigcup_{j=m}^{M} \widehat{P}_{rs}(j), \quad m = 1, 2, \cdots, M, \ (r, s) \in RS \tag{10}$$

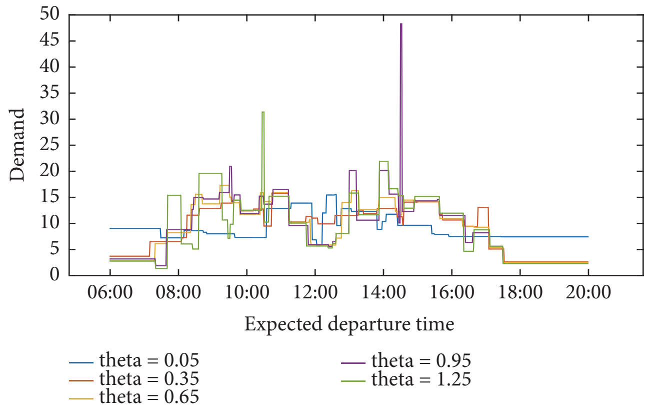

FIGURE 9: The time-varying demand distribution from SBPRA algorithm with different parameter θ.

Hence, the booking phase scheme can be expressed as $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$, and $\overline{P}_{rs}(m) = \{p^k_{rs} \in P_{rs} \mid \check{m}^k_{rs} \leq m \leq \widehat{m}^k_{rs}\}, m = 1, 2, \cdots, M$.

For any itinerary $p^k_{rs} \in P_{rs}$, its flow q^k_{rs} is evenly assigned to its all corresponding expected departure intervals in booking phase $\check{m}^k_{rs}, \check{m}^k_{rs} + 1, \cdots, \widehat{m}^k_{rs}$, i.e., the ticket booking volume of preferable itinerary $\overline{p}^i_{rs} \in \overline{P}_{rs}(m)$ in booking phase m is allocated by a proportion of this itinerary's flow $\overline{q}^i_{rs}$. The proportion is equal to the ratio of its expected departure interval's time range $|t_m(\overline{p}^i_{rs})|$ in booking phase m to its all corresponding expected departure intervals' time range $\sum_{j=\check{m}^i_{rs}}^{\widehat{m}^i_{rs}} |t_j(\overline{p}^i_{rs})|$ in all booking phase $\check{m}^k_{rs}, \check{m}^k_{rs} + 1, \cdots, \widehat{m}^k_{rs}$. Hence, the ticket booking volume $\overline{q}^i_{rs}|t_m(\overline{p}^i_{rs})| / \sum_{j=\check{m}^i_{rs}}^{\widehat{m}^i_{rs}} |t_j(\overline{p}^i_{rs})|$ is evenly assigned to its expected departure interval $t_m(\overline{p}^i_{rs})$ in booking phase $m(\check{m}^i_{rs} \leq m \leq \widehat{m}^i_{rs})$, which leads to a distribution in $t_m(\overline{p}^i_{rs})$. This distribution can be expressed as:

$$Q^i_{rs}(m, x) = \frac{q^i_{rs}}{\sum_{j=\check{m}^i_{rs}}^{\widehat{m}^i_{rs}} \left|t_j\left(\overline{p}^i_{rs}\right)\right|} \quad x \in t_m\left(\overline{p}^i_{rs}\right), \ \overline{p}^i_{rs} \in \overline{P}_{rs}(m), \ m = 1, 2, \cdots, M, \ \check{m}^i_{rs} \leq m \leq \widehat{m}^i_{rs} \ (r, s) \in RS \tag{11}$$

Besides, $|t_j(\overline{p}^i_{rs})| = t_{r,s,i} - t_{r,s,i-1}$ is the time range of $t_j(\overline{p}^i_{rs})$.

The example of the multiple booking phases reverse assignment in Figure 1 can be described as follows. Based on Assumption (*A3*), (*A5*), (*A6*) and *Typical Situation 2*, the booking phase scheme which is calculated by Eq. (10) is expressed as follows.

$$\begin{aligned} \overline{P}_{rs}(1) &= \bigcup_{j=1}^{1} \check{P}_{rs}(j) \cap \bigcup_{j=1}^{3} \widehat{P}_{rs}(j) = \left\{p^1_{rs}, p^4_{rs}, p^6_{rs}\right\} \\ \overline{P}_{rs}(2) &= \bigcup_{j=1}^{2} \check{P}_{rs}(j) \cap \bigcup_{j=2}^{3} \widehat{P}_{rs}(j) = \left\{p^2_{rs}, p^5_{rs}, p^6_{rs}\right\} \\ \overline{P}_{rs}(3) &= \bigcup_{j=1}^{3} \check{P}_{rs}(j) \cap \bigcup_{j=3}^{3} \widehat{P}_{rs}(j) = \left\{p^3_{rs}, p^5_{rs}, p^6_{rs}\right\} \end{aligned} \tag{12}$$

In booking phase I, the preferable itinerary set $\overline{P}_{rs}(1) = \{p^1_{rs}, p^4_{rs}, p^6_{rs}\}$. Using Eq. (11) can obtain the time-varying distribution of booking phase I, i.e., evenly assign the ticket booking volumes q^1_{rs}, q^4_{rs} and $q^6_{rs}|t_1(p^6_{rs})| / \sum_{j=1}^{3} |t_j(p^6_{rs})|$ of p^1_{rs}, p^4_{rs} and p^6_{rs} to their corresponding expected departure intervals $t_1(p^1_{rs}), t_1(p^4_{rs})$ and $t_1(p^6_{rs})$ respectively. In booking phase II, the preferable itinerary set $\overline{P}_{rs}(2) = \{p^2_{rs}, p^5_{rs}, p^6_{rs}\}$. Using Eq. (11) can obtain the time-varying distribution of booking phase II, i.e., evenly assign the ticket booking volumes $q^2_{rs}, q^5_{rs}|t_2(p^5_{rs})| / \sum_{j=2}^{3} |t_j(p^5_{rs})|$ and $q^6_{rs}|t_2(p^6_{rs})| / \sum_{j=1}^{3} |t_j(p^6_{rs})|$ to their corresponding expected departure intervals $t_2(p^2_{rs}), t_2(p^5_{rs})$ and $t_2(p^6_{rs})$ respectively. Then in the following booking phase III, the preferable itinerary set $\overline{P}_{rs}(3) = \{p^3_{rs}, p^5_{rs}, p^6_{rs}\}$. Evenly assign the ticket booking volumes $q^3_{rs}, q^5_{rs}|t_3(p^5_{rs})| / \sum_{j=2}^{3} |t_j(p^5_{rs})|$ and $q^6_{rs}|t_3(p^6_{rs})| / \sum_{j=1}^{3} |t_j(p^6_{rs})|$ to their corresponding expected departure intervals $t_3(p^3_{rs}), t_3(p^5_{rs})$ and $t_3(p^6_{rs})$ respectively, this can obtain the time-varying distribution of booking phase III. At last, sum the time-varying distributions of all booking phases to obtain the time-varying demand.

Based on Assumption (*A3*), (*A5*), (*A6*) and *Typical Situation 2*, we can obtain the first booking phase scheme $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$ and the last booking phase scheme $\widehat{P}_{rs}(1), \widehat{P}_{rs}(2), \cdots, \widehat{P}_{rs}(M)$ by the first and the last booking phase partition algorithm respectively. After that, we can obtain the booking phase scheme $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$ by Eq. (10). In the booking phase $m = 1, 2, \cdots, M$, for any preferable itinerary $\overline{p}^i_{rs} \in \overline{P}_{rs}(m)$, its corresponding expected departure interval $t_m(\overline{p}^i_{rs})$ can be calculated by Eq. (2), (4).

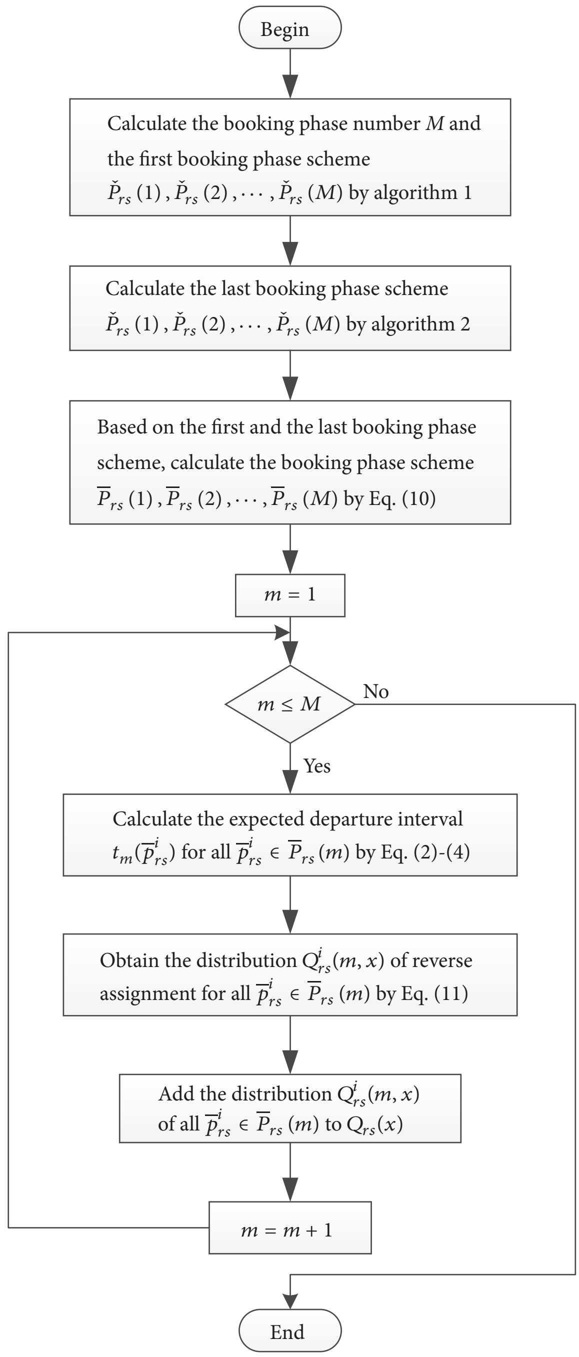

Figure 10: The flow diagram of the MBPRA algorithm.

Then, the tickets booking volume $\overline{q}^i_{rs}|t_m(\overline{p}^i_{rs})|/\sum_{j=\check{m}^i_{rs}}^{\widehat{m}^i_{rs}}|t_j(\overline{p}^i_{rs})|$ of $\overline{p}^i_{rs}$ is reversely assigned to $t_m(\overline{p}^i_{rs})$ by Eq. (11). In conclusion, we design the Multiple Booking Phases Reverse Assignment (MBPRA) algorithm, which is shown in Appendix F, to estimate HSR time-varying demand, and the flow diagram of MBPRA algorithm is shown in the Figure 10.

4.2. Case Analysis. We apply the data (shown in Appendix E) of Beijing-Shanghai HSR on December 1st 2015 from the RTS to the MBPRA algorithm, and the parameters setting are the same as in Section 3.2. The first and the last booking phase scheme calculated by Algorithms 1 and 2 respectively are shown in Tables 4 and 7. Based on the first and last booking phase scheme, the booking phase scheme is obtained by Eq. (10), shown in Table 8. We can see that the continuous ticket-booking process is partitioned into 3 booking phases.

Figure 11 illustrates the expected departure interval of each preferable itinerary in each booking phase. From booking phase I to III, the cost of each preferable itinerary is rising, and there are no other obvious trend of changes. Figure 12 shows the change of ticketing volume of each preferable itinerary. From booking phase I to III, the ticketing volume of each preferable itinerary is declining gradually, and the declining speed is slower than SBPRA algorithm.

For OD pair (r, s), passengers choose their preferable itineraries with the minimum travel cost and book their corresponding tickets, and each itinerary's tickets remain on-sale from its first to the last booking phase. Those conditions cause most passengers to book tickets at the early booking phase with a relatively lower travel cost. As the booking process goes on, few passengers purchase tickets at the late booking phase with a relatively higher travel cost. The decline speed of ticketing volume and the increase speed of travel cost simulated by the MBPRA algorithm are gentler than the SBPRA algorithm. In conclusion, comparing the two solutions by the MBPRA algorithm and the SBPRA algorithm, passengers are less sensitive to the changes of travel cost in the former algorithm.

Figure 13 shows the time-varying demand distribution in each booking phase, and the accumulated time-varying demand distribution of the MBPRA is illustrated in Figure 14. Table 9 shows the numerical results of time-varying demand by the MBPRA algorithm.

Polynomial fitting is adopted for the above distribution of time-varying demand results, and we get demand distribution curve of Beijing-Shanghai, shown in Figure 15. We can see that the fluctuation trends of time-varying demand distribution from the MBPRA algorithm and the SBPRA algorithm are similar. It means that the solution space between the MBPRA algorithm and the SBPRA algorithm is relative narrow.

For the sensitive analysis of Parameter θ in the MBPRA algorithm, the time-varying demand distributions with different parameters θ are shown in Figure 16. From the SBPRA and MBPRA algorithms, it is obvious that they have the same number of booking phases. From Figure 16, it can be seen that the change trend of the time-varying demand distribution have the same characteristic comparing with Figure 9.

5. Validity Analysis

In order to analyze the validity of the above two time-varying demand estimation algorithms, we compare the results of those two algorithms with the real time-varying demands. However, due to the difficulty of obtaining real time-varying demand, we adopts some special data from RTS, which are close to the real time-varying demand, to analyze the validity. We choose some special OD pairs of HSR which are served by high train frequencies (headway is less than

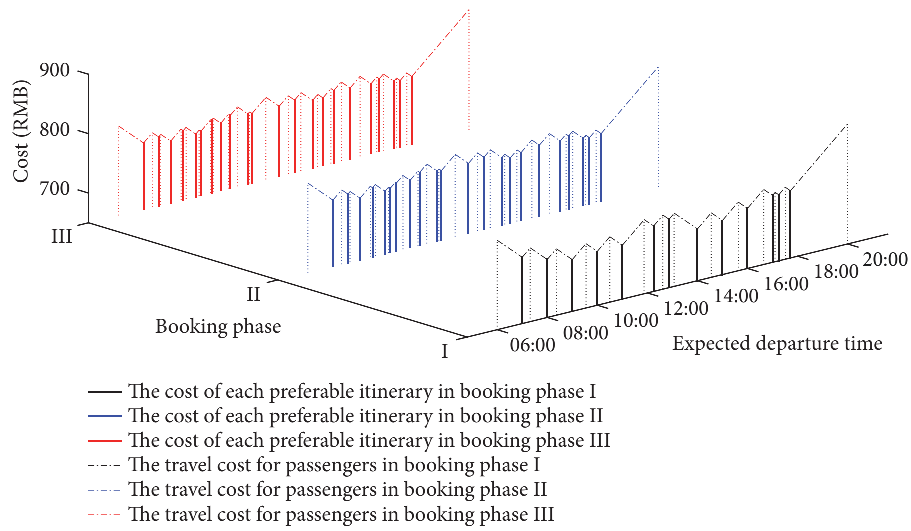

Figure 11: The cost and the expected departure interval of each preferable itinerary in each booking phase from the MBPRA algorithm.

Table 7: The last booking phase scheme.

Booking phase	Preferable itineraries
$\widehat{P}_{rs}(1)$	G11, G1, G13, G15, G129, G133, G3, G17, G19
$\widehat{P}_{rs}(2)$	G115, G43, G141
$\widehat{P}_{rs}(3)$	G101, G105, G107, G111, G113, G41, G117, G119, G121, G125, G411, G131, G135, G137, G139, G143, G145, G147, G149, G21, G153, G157

Table 8: The booking phase scheme.

Booking phase	Preferable itineraries
$\overline{P}_{rs}(1)$	G101, G11, G1, G13, G15, G129, G133, G3, G17, G19, G21, G153, G157
$\overline{P}_{rs}(2)$	G101, G105, G107, G111, G113, G41, G115, G119, G121, G125, G411, G131, G135, G137, G43, G141, G145, G147, G149, G21, G153, G157
$\overline{P}_{rs}(3)$	G101, G105, G107, G111, G113, G41, G117, G119, G121, G125, G411, G131, G135, G137, G139, G143, G145, G147, G149, G21, G153, G157

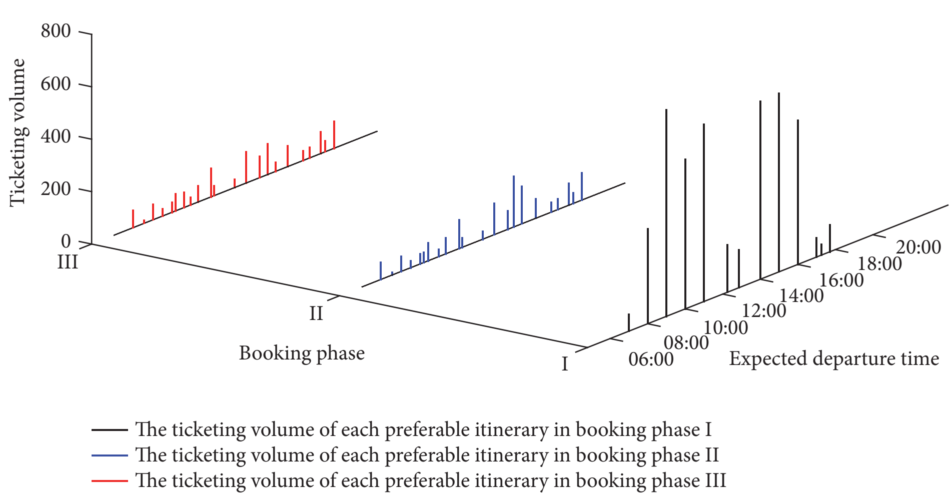

Figure 12: The ticketing volume of each itinerary in each booking phase from the MBPRA algorithm.

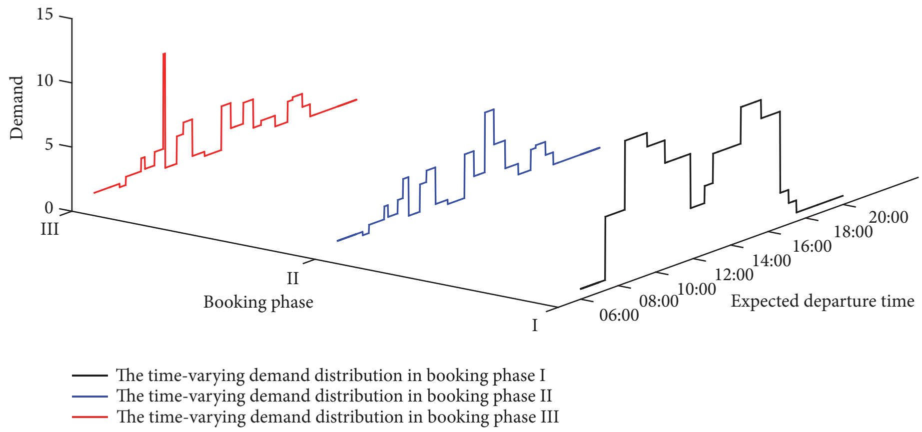

FIGURE 13: The time-varying demand distribution in each booking phase of the MBPRA algorithm.

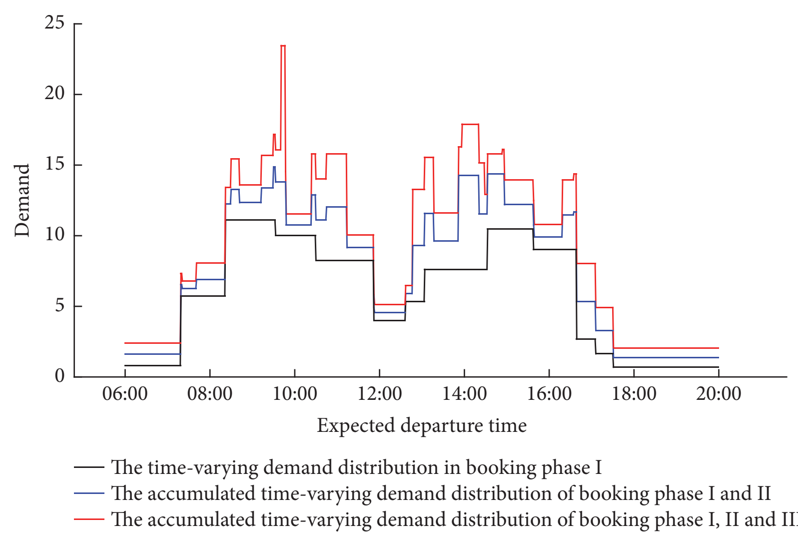

FIGURE 14: The accumulated time-varying demand distribution of the MBPRA algorithm.

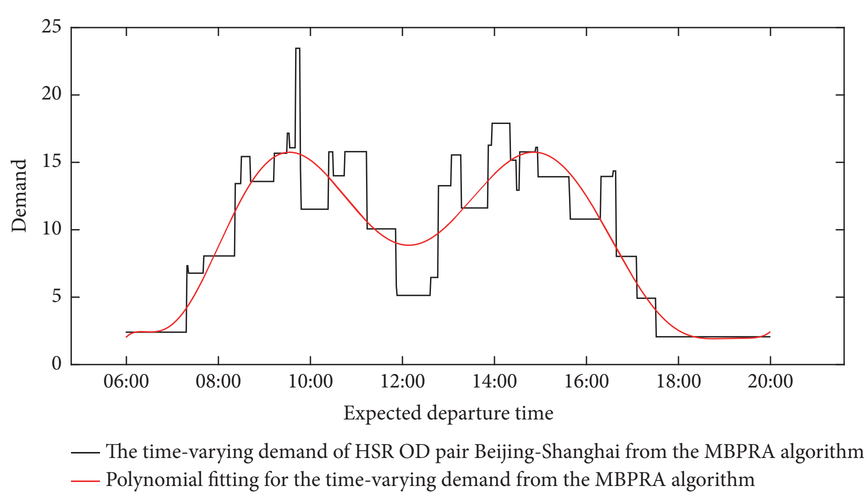

FIGURE 15: Polynomial fitting for the result of the MBPRA algorithm of OD pair Beijing-Shanghai.

Table 9: The time-varying demand from the MBPRA algorithm.

Time range	Demand volume (person/minute)	Time range	Demand volume (person/minute)
[6:00, 7:19)	2.4	[12:37, 12:47)	6.5
[7:19, 7:21)	7.3	[12:47, 13:04)	13.3
[7:21, 7:41)	6.8	[13:04, 13:17)	15.6
[7:41, 8:22)	8.1	[13:17, 13:52)	11.6
[8:22, 8:30)	13.4	[13:52, 13:57)	16.3
[8:30, 8:42)	15.4	[13:57, 14:21)	17.9
[8:42, 9:13)	13.6	[14:21, 14:29)	15.2
[9:13, 9:30)	15.7	[14:29, 14:33)	12.9
[9:30, 9:33)	17.2	[14:33, 14:54)	15.8
[9:33, 9:41)	16.1	[14:54, 14:57)	16.1
[9:41, 9:47)	23.5	[14:57, 15:38)	13.9
[9:47, 9:48)	14.5	[15:38, 15:39)	12.5
[9:48, 10:24)	11.5	[15:39, 16:19)	10.8
[10:24, 10:30)	15.8	[16:19, 16:35)	14
[10:30, 10:45)	14	[16:35, 16:39)	14.4
[10:45, 11:14)	15.8	[16:39, 17:06)	8
[11:14, 11:52)	10.1	[17:06, 17:31)	4.9
[11:52, 11:53)	5.8	[17:31, 20:00]	2.1
[11:53, 12:37)	5.1		

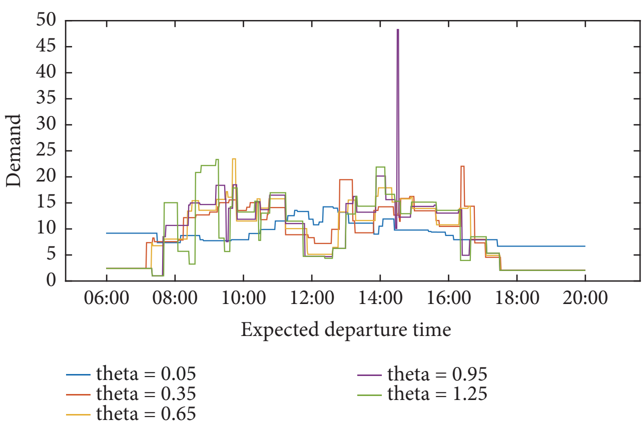

Figure 16: The time-varying demand distributions from MBPRA algorithm with different θ.

1 hour) with sufficient train capacities, so the transport volumes in each hour can be seen as its real hourly demand of this OD pair. Comparing them with the time-varying demands calculated by the SBPRA algorithm and the MBPRA algorithm respectively, we can test the accuracy of the above two algorithms approximatively.

We apply the data (shown in Appendix G) of OD pair Beijing-Tianjin in December 1^{st} 2015 to analyze the validity of the SBPRA algorithm and the MBPRA algorithm. There are 129 itineraries of OD pair Beijing-Tianjin on that day. The effective operation period of this OD pair is $[T_{rs}^0, T_{rs}^1] = [6{:}00, 23{:}00]$. The average monthly residential income of Beijing and Tianjin are 7086 RMB and 4944 RMB in 2015 respectively [40, 41], and the average income can be expressed as 0.67 RMB per minute and 0.47 RMB per minute respectively. We set $\theta = (0.67+0.47)/2 = 0.57$ RMB per minute. The comparison between hourly transport volumes from RTS and the results from the SBPRA algorithm and MBPRA algorithm are shown in Table 10.

From the Table 10, we can see that the error rates of the SBPRA algorithm and the MBPRA algorithm are 8.64% and 6.37% respectively, which are relatively low and verifies those two algorithms. Besides, the MBPRA algorithm has a lower error rate than the SBPRA algorithm, which implies that ticket-booking process of this OD pair on December 1^{st} 2015 is closer to *Typical Situation 2*.

6. Conclusion and Further Studies

This paper focuses on the problem of HSR time-varying demand estimation. By simulating ticket-booking process, we reversely assign the ticketing volume of each preferable itinerary to its corresponding expected departure interval in each ticket-booking phase, and then sum the demand distributions in all booking phases to obtain the time-varying demand. Owing to the variety of the sold-out orders for all preferable itineraries' tickets and only the data of the itinerary flow, the precedence relationship is introduced to constrain the ticket sold-out order of all itineraries for each OD pair. Based on the precedence relationship of itineraries, two typical situations are proposed, and the SBPRA algorithm and the MBPRA algorithm are designed. The case analysis shows that the results of those two algorithms can better reflect the time-varying characteristics of HSR passenger demand, and the fluctuation of those two distributions are similar, but the SBPRA algorithm results are more relevant

Table 10: Error analysis for the SBPRA algorithm and the MBPRA algorithm.

Time range	Hourly transport volumes	SBPRA algorithm		MBPRA algorithm	
		Hourly demand volumes	Error	Hourly demand volumes	Error
[6:00, 7:00)	509	610.3	101.3	563.2	54.2
[7:00, 8:00)	1022	967.4	54.6	1072.1	50.1
[8:00, 9:00)	1472	1717.5	245.5	1587.1	115.1
[9:00, 10:00)	2282	2147.6	134.4	2201.9	80.1
[10:00, 11:00)	2152	1962.1	189.9	1948.9	203.1
[11:00, 12:00)	1619	1826.8	207.8	1836.4	217.4
[12:00, 13:00)	1756	1706.8	49.2	1753.0	3.0
[13:00, 14:00)	2065	1842.6	222.4	1846.8	218.2
[14:00, 15:00)	1833	2034.1	201.1	1983.4	150.4
[15:00, 16:00)	1937	1774.0	163.0	1854.3	82.7
[16:00, 17:00)	2317	2381.2	64.2	2350.6	33.6
[17:00, 18:00)	2351	2265.5	85.5	2363.1	12.1
[18:00, 19:00)	1635	1762.5	127.5	1666.2	31.2
[19:00, 20:00)	1057	1156.6	99.6	1149.4	92.4
[20:00, 21:00)	1073	860.4	212.6	853.3	219.7
[21:00, 22:00)	638	721.7	83.7	714.5	76.5
[22:00, 23:00]	442	423.0	19.0	415.9	26.1
Sum	26160	26160	2261.2	26160	1665.9
Error rates	——	——	8.64%	——	6.37%

to the itinerary cost differences. Numerical analysis have shown that the error rates of the SBPRA algorithm and the MBPRA algorithm are 8.64% and 6.37% respectively. They have rather good estimation accuracy, which validate those two algorithms.

The current research, as a first step to estimate time-varying demand in HSR, can be extended along several avenues as follows: (1) This paper only considers travel cost for passengers with the same unit time fee, but the unit time value may vary for different passengers. Further studies can classify passengers into several categories with different socio-economic characteristics (e.g. income level). Besides, different class seats could be considered. (2) This paper uses simulative method to estimate time-varying demand by partitioning continuous ticket-booking process into discrete booking phases according to two typical situations. If more detailed information is accessible, such as ticket-booking time of each passenger, then the time-varying demand can be estimated by the actual sold-out order of preferable itineraries. (3) We will study the estimation problem of the day-to-day dynamic demand for HSR system in the further research.

Appendix

A. Notation

Sets

RS Set of OD pair

P_{rs} Set of itineraries of OD pair (r, s)

P'_{rs} Current set of available itineraries of OD pair (r, s)

$\overline{P}_{rs}$ Set of preferable itineraries for passengers of OD pair (r, s)

$\check{P}_{rs}(m)$ Set of preferable itineraries whose first booking phase is m

$\hat{P}_{rs}(m)$ Set of preferable itineraries whose last booking phase is m

$\overline{P}_{rs}(m)$ Set of preferable itineraries in the booking phase m

P_{rs}^{L} The itinerary set of the longest precedence relationship chain L_{rs}

Indexes

(r, s) Index of OD pair, $(r, s) \in RS$

m Index of booking phase, $m = 1, 2, \cdots, M$

i Index of preferable itinerary, $i = 1, 2, \cdots, I$

k, h, g Index of itinerary, $p_{rs}^{k}, p_{rs}^{h}, p_{rs}^{g} \in P_{rs}$

Parameters

$[T_{rs}^{0}, T_{rs}^{1}]$ Effective operation period of OD pair (r, s)

p_{rs}^{k} Itinerary of OD pair (r, s), $p_{rs}^{k} \in P_{rs}$

$\overline{p}_{rs}^{i}$ Preferable itinerary of OD pair (r, s), $\overline{p}_{rs}^{i} \in \overline{P}_{rs}$

c_{rs}^{k} Cost of itinerary p_{rs}^{k}

t_{rs}^{k} Depart time of itinerary p_{rs}^{k} from station r

q_{rs}^{k} Flow of itinerary p_{rs}^{k}

x Expected departure time for passenger, $x \in [T_{rs}^{0}, T_{rs}^{1}]$

θ Unit time fee for passengers who adjust expected departure time

t_{rsi} Time division point between $t(\overline{p}_{rs}^{i})$ and $t(\overline{p}_{rs}^{i+1})$

$\check{m}_{rs}^{k}$ The first booking phase of p_{rs}^{k}

Table 11: The ticket booking data of each itinerary of OD pair Beijing-Shanghai on December 1st 2015.

Train Number	Origin Station	Destination Station	Depart time	Length/km	time/min	Flow/person
G101	BeijingNan	Shanghaihongqiao	7:00	1318	337	193
G105	BeijingNan	Shanghaihongqiao	7:36	1318	345	21
G11	BeijingNan	Shanghaihongqiao	8:00	1318	309	361
G107	BeijingNan	Shanghaihongqiao	8:05	1318	321	114
G111	BeijingNan	Shanghaihongqiao	8:35	1318	346	52
G1	BeijingNan	Shanghaihongqiao	9:00	1318	288	788
G113	BeijingNan	Shanghaihongqiao	9:05	1318	324	77
G41	BeijingNan	Shanghaihongqiao	9:17	1318	329	103
G115	BeijingNan	Shanghaihongqiao	9:32	1318	342	68
G117	BeijingNan	Shanghaihongqiao	9:43	1318	358	58
G13	BeijingNan	Shanghaihongqiao	10:00	1318	295	571
G119	BeijingNan	Shanghaihongqiao	10:05	1318	340	55
G121	BeijingNan	Shanghaihongqiao	10:28	1318	359	121
G15	BeijingNan	Shanghaihongqiao	11:00	1318	295	676
G125	BeijingNan	Shanghaihongqiao	11:10	1318	349	219
G411	BeijingNan	Shanghaihongqiao	11:20	1318	347	71
G129	BeijingNan	Shanghaihongqiao	12:15	1318	334	180
G131	BeijingNan	Shanghaihongqiao	12:25	1318	349	61
G133	BeijingNan	Shanghaihongqiao	12:52	1318	344	144
G135	BeijingNan	Shanghaihongqiao	13:02	1318	357	238
G137	BeijingNan	Shanghaihongqiao	13:45	1318	341	150
G3	BeijingNan	Shanghaihongqiao	14:00	1318	288	678
G43	BeijingNan	Shanghaihongqiao	14:05	1318	333	193
G139	BeijingNan	Shanghaihongqiao	14:10	1318	341	116
G141	BeijingNan	Shanghaihongqiao	14:31	1318	342	141
G143	BeijingNan	Shanghaihongqiao	14:36	1318	357	35
G17	BeijingNan	Shanghaihongqiao	15:00	1318	296	680
G145	BeijingNan	Shanghaihongqiao	15:15	1318	352	151
G19	BeijingNan	Shanghaihongqiao	16:00	1318	316	550
G147	BeijingNan	Shanghaihongqiao	16:05	1318	349	71
G149	BeijingNan	Shanghaihongqiao	16:25	1318	359	79
G21	BeijingNan	Shanghaihongqiao	17:00	1318	339	238
G153	BeijingNan	Shanghaihongqiao	17:15	1318	336	123
G157	BeijingNan	Shanghaihongqiao	17:43	1318	340	307

$\widehat{m}_{rs}^{k}$ The last booking phase of p_{rs}^{k}

$t(\overline{p}_{rs}^{i})$ Expected departure interval of $\overline{p}_{rs}^{i}$

$t_m(\overline{p}_{rs}^{i})$ Expected departure interval of $\overline{p}_{rs}^{i}$ in booking phase m

Variables

M The itinerary number of the longest precedence relationship chain L_{rs}

$Q_{rs}(x)$ Time-varying demand of OD pair (r, s), $x \in [T_{rs}^{0}, T_{rs}^{1}]$

B. The First Booking Phase Partition Algorithm

See Algorithm 1.

C. The Last Booking Phase Partition Algorithm

See Algorithm 2.

D. Single Booking Phase Reverse Assignment Algorithm

See Algorithm 3.

E. The Ticket Booking Data of Each Itinerary of OD Pair Beijing-Shanghai on December 1st 2015

See Table 11.

Input The effective operation period $[T_0, T_1]$ and the itinerary set P_{rs} of OD pair (r, s), the cost c_{rs}^k, the flow q_{rs}^k, and the departure time t_{rs}^k of itinerary $p_{rs}^k \in P_{rs}$, the unit time fee θ of adjusting expected departure time for passengers;
Output Passenger time-varying demand $Q_{rs}(x), x \in [T_{rs}^0, T_{rs}^1]$ for OD pair (r, s).
Begin
  Calculate M and $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$ for P_{rs} by the first booking phase partition algorithm;
  do $\overline{P}_{rs}(m) \longleftarrow \check{P}_{rs}(m), m = 1, 2, \cdots, M$, obtain $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$;
  $Q_{rs}(x) \longleftarrow 0, x \in [T_{rs}^0, T_{rs}^1]$;
  For $m = 1, 2, \cdots, M$, do
  Begin 1
    Obtain the expected departure interval $t_m(\overline{p}_{rs}^i)$ for all $\overline{p}_{rs}^i \in \overline{P}_{rs}(m)$ by Eq. (2) and (4);
    Obtain the distribution $Q_{rs}^i(x)$ of reverse assignment for all $\overline{p}_{rs}^i \in \overline{P}_{rs}(m)$ by Eq. (8);
    do $Q_{rs}(x) \longleftarrow Q_{rs}(x) + Q_{rs}^i(x), x \in t_m(\overline{p}_{rs}^i), \overline{p}_{rs}^i \in \overline{P}_{rs}(m)$;
  Return 1
End

Algorithm 3: Single Booking Phase Reverse Assignment algorithm.

Input The effective operation period $[T_{rs}^0, T_{rs}^1]$ and the itinerary set P_{rs} of OD pair (r, s), the cost c_{rs}^k, the flow q_{rs}^k, and the departure time t_{rs}^k of itinerary $p_{rs}^k \in P_{rs}$, the unit time fee θ of adjusting expected departure time for passengers;
Output Passenger time-varying demand $Q_{rs}(x), x \in [T_{rs}^0, T_{rs}^1]$ for OD pair (r, s).
Begin
  Calculate M and $\check{P}_{rs}(1), \check{P}_{rs}(2), \cdots, \check{P}_{rs}(M)$ for P_{rs} by the first booking phase partition algorithm;
  Calculate $\widehat{P}_{rs}(1), \widehat{P}_{rs}(2), \cdots, \widehat{P}_{rs}(M)$ for P_{rs} by the last booking phase partition algorithm;
  Based on the first and the last booking phase scheme to obtain the booking phase scheme
  $\overline{P}_{rs}(1), \overline{P}_{rs}(2), \cdots, \overline{P}_{rs}(M)$ by Eq. (10);
  $Q_{rs}(x) \longleftarrow 0, x \in [T_{rs}^0, T_{rs}^1]$;
  For $m = 1, 2, \cdots, M$, do
  Begin 1
    Obtain the expected departure interval $t_m(\overline{p}_{rs}^i), \overline{p}_{rs}^i \in \overline{P}_{rs}(m)$ by Eq. (2) and (4);
    Obtain the distribution $Q_{rs}^i(m, x), x \in t_m(\overline{p}_{rs}^i), \overline{p}_{rs}^i \in \overline{P}_{rs}(m)$ by Eq. (11);
    do $Q_{rs}(x) \longleftarrow Q_{rs}(x) + Q_{rs}^i(m, x), x \in t_m(\overline{p}_{rs}^i), \overline{p}_{rs}^i \in \overline{P}_{rs}(m)$;
  Return 1
End

Algorithm 4: Multiple Booking Phases Reverse Assignment algorithm.

F. Multiple Booking Phases Reverse Assignment Algorithm

See Algorithm 4.

G. The Ticket Booking Data of Each Itinerary of OD Pair Beijing-Tianjin on December 1st 2015

See Table 12.

Data Availability

The relevant data used to support the findings of this study are in the main body and Appendix of the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Funding

This study was supported by the National Natural Science Foundation of China [grant number 71701216, U1334207] and

Table 12: The ticket booking data of each itinerary of OD pair Beijing-Tianjin on December 1st 2015.

Train Number	Origin Station	Destination Station	Depart time	Length/km	time/min	Flow/person
C2001	BeijingNan	Tianjin	6:13	120	33	264
C2003	BeijingNan	Tianjin	6:21	120	33	43
C2201	BeijingNan	Tianjin	6:32	120	38	80
C2005	BeijingNan	Tianjin	6:42	120	33	69
C2007	BeijingNan	Tianjin	6:50	120	33	53
C2117	BeijingNan	Tianjin	7:00	120	33	112
C2011	BeijingNan	Tianjin	7:08	120	33	88
G471	BeijingNan	Tianjinnan	7:10	122	34	37
C2283	BeijingNan	Tianjin	7:13	120	36	72
G261	BeijingNan	Tianjinnan	7:15	122	34	7
G57	BeijingNan	Tianjinnan	7:20	122	34	15
C2009	BeijingNan	Tianjin	7:24	120	33	151
G177	BeijingNan	Tianjinnan	7:25	122	34	18
G383	BeijingNan	Tianjin	7:30	127	46	18
G383	BeijingNan	Tianjinxi	7:30	122	36	8
C2203	BeijingNan	Tianjin	7:32	120	38	112
G265	BeijingNan	Tianjinnan	7:48	122	34	54
C2013	BeijingNan	Tianjin	7:49	120	33	296
G381	BeijingNan	Tianjin	7:53	127	47	23
G381	BeijingNan	Tianjinxi	7:53	122	36	11
G219	BeijingNan	Tianjin	8:00	120	35	148
C2241	BeijingNan	Tianjin	8:08	120	38	242
C2017	BeijingNan	Tianjin	8:21	120	33	280
G387	BeijingNan	Tianjin	8:25	127	40	24
C2019	BeijingNan	Tianjin	8:31	120	33	392
G321	BeijingNan	Tianjinnan	8:40	122	34	106
C2205	BeijingNan	Tianjin	8:49	120	38	280
C2021	BeijingNan	Tianjin	9:01	120	33	475
C2023	BeijingNan	Tianjin	9:06	120	33	271
G395	BeijingNan	Tianjin	9:10	127	51	23
C2025	BeijingNan	Tianjin	9:11	120	33	141
C2027	BeijingNan	Tianjin	9:19	120	33	219
C2029	BeijingNan	Tianjin	9:29	120	33	416
C2207	BeijingNan	Tianjin	9:40	120	38	239
G117	BeijingNan	Tianjinnan	9:43	122	34	87
G27	BeijingNan	Tianjinnan	9:48	122	34	35
C2031	BeijingNan	Tianjin	9:52	120	33	376
C2033	BeijingNan	Tianjin	10:00	120	33	348
C2209	BeijingNan	Tianjin	10:08	120	38	184
C2285	BeijingNan	Tianjin	10:18	120	36	120
G121	BeijingNan	Tianjinnan	10:28	122	41	71
C2035	BeijingNan	Tianjin	10:29	120	33	490
G301	BeijingNan	Tianjinnan	10:40	122	34	25
C2037	BeijingNan	Tianjin	10:42	120	33	459
G181	BeijingNan	Tianjinnan	10:45	122	34	33
C2039	BeijingNan	Tianjin	10:52	120	33	276
C2211	BeijingNan	Tianjin	10:57	120	38	146
G125	BeijingNan	Tianjinnan	11:10	122	34	30
C2041	BeijingNan	Tianjin	11:10	120	33	416

Table 12: Continued.

Train Number	Origin Station	Destination Station	Depart time	Length/km	time/min	Flow/person
G33	BeijingNan	Tianjinnan	11:15	122	34	35
C2213	BeijingNan	Tianjin	11:19	120	38	298
C2043	BeijingNan	Tianjin	11:29	120	33	367
C2045	BeijingNan	Tianjin	11:37	120	33	218
C2215	BeijingNan	Tianjin	11:47	120	38	255
G325	BeijingNan	Tianjinnan	12:05	122	34	62
C2047	BeijingNan	Tianjin	12:05	120	33	451
C2217	BeijingNan	Tianjin	12:18	120	38	322
G131	BeijingNan	Tianjinnan	12:25	122	34	47
C2049	BeijingNan	Tianjin	12:33	120	33	386
C2051	BeijingNan	Tianjin	12:41	120	33	277
G133	BeijingNan	Tianjinnan	12:52	122	34	36
C2287	BeijingNan	Tianjin	12:52	120	36	148
G167	BeijingNan	Tianjinnan	12:57	122	34	27
C2053	BeijingNan	Tianjin	13:00	120	33	301
C2055	BeijingNan	Tianjin	13:07	120	33	201
C2219	BeijingNan	Tianjin	13:15	120	38	208
G217	BeijingNan	Tianjin	13:30	120	35	283
G29	BeijingNan	Tianjinnan	13:35	122	34	52
C2057	BeijingNan	Tianjin	13:39	120	33	451
G59	BeijingNan	Tianjinnan	13:40	122	34	13
C2221	BeijingNan	Tianjin	13:44	120	38	154
G137	BeijingNan	Tianjinnan	13:45	122	34	44
G397	BeijingNan	Tianjin	13:55	127	51	45
G397	BeijingNan	Tianjinxi	13:55	122	35	16
C2059	BeijingNan	Tianjin	13:56	120	33	297
C2223	BeijingNan	Tianjin	14:07	120	39	311
C2061	BeijingNan	Tianjin	14:23	120	33	454
C2225	BeijingNan	Tianjin	14:31	120	38	310
C2063	BeijingNan	Tianjin	14:44	120	33	412
G193	BeijingNan	Tianjinnan	14:46	122	41	66
C2227	BeijingNan	Tianjin	14:54	120	38	280
G393	BeijingNan	Tianjin	15:05	127	46	30
G393	BeijingNan	Tianjinxi	15:05	122	35	11
C2065	BeijingNan	Tianjin	15:05	120	33	353
C2067	BeijingNan	Tianjin	15:15	120	33	307
G399	BeijingNan	Tianjin	15:20	127	49	49
G399	BeijingNan	Tianjinxi	15:20	122	35	6
C2069	BeijingNan	Tianjin	15:25	120	33	285
C2071	BeijingNan	Tianjin	15:35	120	33	407
G475	BeijingNan	Tianjinnan	15:36	122	34	75
G45	BeijingNan	Tianjinnan	15:41	122	34	31
C2289	BeijingNan	Tianjin	15:45	120	36	117
G195	BeijingNan	Tianjinnan	15:46	122	41	18
C2229	BeijingNan	Tianjin	15:51	120	38	248
G37	BeijingNan	Tianjinnan	16:10	122	34	74
C2231	BeijingNan	Tianjin	16:15	120	38	437
C2073	BeijingNan	Tianjin	16:27	120	33	499
C2075	BeijingNan	Tianjin	16:35	120	33	500
G197	BeijingNan	Tianjinnan	16:42	122	34	92
C2077	BeijingNan	Tianjin	16:45	120	33	499

Table 12: Continued.

Train Number	Origin Station	Destination Station	Depart time	Length/km	time/min	Flow/person
C2233	BeijingNan	Tianjin	16:54	120	38	216
C2079	BeijingNan	Tianjin	17:09	120	33	504
C2081	BeijingNan	Tianjin	17:18	120	33	481
C2235	BeijingNan	Tianjin	17:31	120	38	271
C2083	BeijingNan	Tianjin	17:41	120	33	491
C2085	BeijingNan	Tianjin	17:49	120	33	420
G201	BeijingNan	Tianjinnan	17:55	122	41	101
C2613	BeijingNan	Tianjinxi	17:59	117	35	83
C2087	BeijingNan	Tianjin	18:07	120	33	440
C2089	BeijingNan	Tianjin	18:15	120	33	235
C2237	BeijingNan	Tianjin	18:25	120	38	173
G481	BeijingNan	Tianjinnan	18:28	122	34	67
C2091	BeijingNan	Tianjin	18:35	120	33	314
G271	BeijingNan	Tianjinnan	18:39	122	41	4
C2291	BeijingNan	Tianjin	18:43	120	36	75
C2093	BeijingNan	Tianjin	18:51	120	33	327
C2095	BeijingNan	Tianjin	19:02	120	33	282
C2239	BeijingNan	Tianjin	19:10	120	38	174
C2015	BeijingNan	Tianjin	19:25	120	33	298
C2097	BeijingNan	Tianjin	19:39	120	33	303
C2099	BeijingNan	Tianjin	20:05	120	33	381
C2101	BeijingNan	Tianjin	20:15	120	33	158
C2243	BeijingNan	Tianjin	20:27	120	38	186
C2105	BeijingNan	Tianjin	20:46	120	33	264
D401	BeijingNan	Tianjinnan	20:58	97.6	48	84
C2245	BeijingNan	Tianjin	21:19	120	38	356
C2109	BeijingNan	Tianjin	21:50	120	33	282
C2111	BeijingNan	Tianjin	22:00	120	33	162
C2115	BeijingNan	Tianjin	22:52	120	33	280

the Humanities and Social Sciences Foundation of Jiangxi China [grant number GL1516].

References

[1] Y.-H. Chang, C.-H. Yeh, and C.-C. Shen, "A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line," *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 91–106, 2000.

[2] J.-W. Goossens, S. van Hoesel, and L. Kroon, "A branch-and-cut approach for solving railway line-planning problems," *Transportation Science*, vol. 38, no. 3, pp. 379–393, 2004.

[3] J. H. M. Goossens, C. P. M. van Hoesel, and L. G. Kroon, "On solving multi-type railway line planning problems," *European Journal of Operational Research*, vol. 168, no. 2, pp. 403–424, 2006.

[4] H. L. Fu, L. Nie, L. Y. Meng, B. R. Sperry, and Z. H. He, "A hierarchical line planning approach for a large-scale high speed rail network: the China case," *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 61–83, 2015.

[5] M. Kaspi and T. Raviv, "Service-oriented line planning and timetabling for passenger trains," *Transportation Science*, vol. 47, no. 3, pp. 295–311, 2013.

[6] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.

[7] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.

[8] E. Barrena, D. Canca, L. C. Coelho, and G. Laporte, "Exact formulations and algorithm for the train timetabling problem with dynamic demand," *Computers & Operations Research*, vol. 44, no. 3, pp. 66–74, 2014.

[9] L. R. Weatherford and S. Pölt, "Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues," *Journal of Revenue and Pricing Management*, vol. 1, no. 3, pp. 234–254, 2002.

[10] B. C. Smith, J. F. Leimkuhler, and R. M. Darrow, "Yield management at American airlines," *Interfaces*, vol. 22, no. 1, pp. 8–31, 1992.

[11] G. Vulcano, G. van Ryzin, and R. Ratliff, "Estimating primary demand for substitutable products from sales transaction data," *Operations Research*, vol. 60, no. 2, pp. 313–334, 2012.

[12] Y. Sun, Z. Jiang, J. Gu, M. Zhou, Y. Li, and L. Zhang, "Analyzing high speed rail passengers' train choices based on new online booking data in China," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 96–113, 2018.

[13] J. I. McGill, "Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints," *Annals of Operations Research*, vol. 60, no. 1, pp. 209–240, 1995.

[14] S. Mukhopadhyay, S. Samaddar, and G. Colville, "Improving revenue management decision making for airlines by evaluating analyst-adjusted passenger demand forecasts," *Decision Sciences*, vol. 38, no. 2, pp. 309–327, 2007.

[15] R. M. Ratliff, B. Venkateshwara Rao, C. P. Narayan, and K. Yellepeddi, "A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings," *Journal of Revenue and Pricing Management*, vol. 7, no. 2, pp. 153–171, 2008.

[16] C.-H. Wen and P.-H. Chen, "Passenger booking timing for low-cost airlines: a continuous logit approach," *Journal of Air Transport Management*, vol. 64, pp. 91–99, 2017.

[17] Y.-C. Chiou and C.-H. Liu, "Advance purchase behaviors of air tickets," *Journal of Air Transport Management*, vol. 57, pp. 62–69, 2016.

[18] Y.-C. Chiou and C.-H. Liu, "Advance purchase behaviors of air passengers: a continuous logit model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 93, pp. 474–484, 2016.

[19] D. Escobari, "Estimating dynamic demand for airlines," *Economics Letters*, vol. 124, no. 1, pp. 26–29, 2014.

[20] K. Ashok and M. E. Ben-Akiva, "Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows," *Transportation Science*, vol. 34, no. 1, pp. 21–36, 2000.

[21] M. Bierlaire and F. Crittin, "An efficient algorithm for real-time estimation and prediction of dynamic OD tables," *Operations Research*, vol. 52, no. 1, pp. 116–127, 2004.

[22] X. S. Zhou and H. S. Mahmassani, "A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework," *Transportation Research Part B: Methodological*, vol. 41, no. 8, pp. 823–840, 2007.

[23] V. Marzano, A. Papola, and F. Simonelli, "Limits and perspectives of effective O–D matrix correction using traffic counts," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 120–132, 2009.

[24] E. Cascetta, A. Papola, V. Marzano, F. Simonelli, and I. Vitiello, "Quasi-dynamic estimation of o–d flows from traffic counts: formulation, statistical validation and performance analysis on real data," *Transportation Research Part B: Methodological*, vol. 55, pp. 171–187, 2013.

[25] T. Djukic, G. Flötteröd, H. Van Lint, and S. Hoogendoorn, "Efficient real time OD matrix estimation based on Principal Component Analysis," in *Proceedings of the International IEEE Conference on Intelligent Transportation Systems*, pp. 115–121, Anchorage, Alaska, USA, September 2012.

[26] M. L. Hazelton, "Statistical inference for time varying origin-destination matrices," *Transportation Research Part B: Methodological*, vol. 42, no. 6, pp. 542–552, 2008.

[27] H. Shao, W. H. K. Lam, A. Sumalee, A. Chen, and M. L. Hazelton, "Estimation of mean and covariance of peak hour origin-destination demands from day-to-day traffic counts," *Transportation Research Part B: Methodological*, vol. 68, pp. 52–75, 2014.

[28] W. Wang, J. P. Attanucci, and N. H. M. Wilson, "Bus passenger origin-destination estimation and related analyses using automated data collection systems," *Journal of Public Transportation*, vol. 14, no. 4, pp. 131–150, 2011.

[29] J. Chan, *Rail OD Estimation and Journey Time Reliability Metrics Using Automated Fare Data*, Fare Data. Massachusetts Institute of Technology, Cambridge, Mass, USA, 2007.

[30] S. C. Wong and C. O. Tong, "Estimation of time-dependent origin-destination matrices for transit networks," *Transportation Research Part B: Methodological*, vol. 32, no. 1, pp. 35–48, 1998.

[31] Y. W. Li and M. J. Cassidy, "A generalized and efficient algorithm for estimating transit route ODs from passenger counts," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 114–125, 2007.

[32] A. Cui, *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2006.

[33] J. H. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.

[34] B. B. Li, "Markov models for Bayesian analysis about transit route origin-destination matrices," *Transportation Research Part B: Methodological*, vol. 43, no. 3, pp. 301–310, 2009.

[35] X. M. Yao, P. Zhao, and D. D. Yu, "Real-time origin-destination matrices estimation for urban rail transit network based on structural state-space model," *Journal of Central South University*, vol. 22, no. 11, pp. 4498–4506, 2015.

[36] Y. X. Ji, R. G. Mishalani, and M. R. McCord, "Transit passenger origin-destination flow estimation: efficiently combining onboard survey and large automatic passenger count datasets," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 178–192, 2015.

[37] N. J. Douglas, L. Henn, and K. Sloan, "Modelling the ability of fare to spread AM peak passenger loads using rooftops," in *Proceedings of the Australasian Transport Research Forum*, Adelaide, Australia, September 2011.

[38] H. Y. Su, F. Shi, G. M. Xu, J. Qin, and X. H. Shan, "Schedule-based passenger assignment for high-speed rail networks considering the ticket-booking process," *Mathematical Problems in Engineering*, vol. 2016, Article ID 1650839, 15 pages, 2016.

[39] Shanghai Municipal Statistics Bureau & Survey Office of the National Bureau of Statistics in Shanghai, *Shanghai Statistical Yearbook*, China Statistics Press, 2016.

[40] Beijing Municipal Bureau of Statistics & Survey Office of the National Bureau of Statistics in Beijing, *Beijing Statistical Yearbook*, China Statistics Press, 2016.

[41] Tianjin Municipal Bureau of Statistics & Survey Office of the National Bureau of Statistics in Tianjin, *Tianjin Statistical Yearbook*, China Statistics Press, 2016.