Hindawi
Complexity
Volume 2023, Article ID 4551091, 15 pages
https://doi.org/10.1155/2023/4551091

*Research Article*

# Pricing for Refund Service Fee of High-Speed Railway: An Optimization Approach with Uncertain Demand

**Yu Wang**[1] **Jiafa Zhu**[2] **Hongye Wang,**[1] **and Xiaoyan Lv**[1]

[1]*Institute of Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China*
[2]*Institute of Transportation and Economics, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China*

Correspondence should be addressed to Jiafa Zhu; luoyu167@163.com

Received 20 June 2023; Revised 24 September 2023; Accepted 28 September 2023; Published 18 October 2023

Academic Editor: Zhen Zhang



Compared to single ticket purchase behavior, the impact of ticket cancellation behavior on revenue is full of complexity. Due to the cancelled tickets will be resold with uncertain demand, ticket cancellation behaviors and ticket purchase behaviors are intertwined and influenced, composed of a dynamic and complex system in the presale period. How to charge for ticket cancellation behavior in the name of refund service fee to reduce losses as much as possible is an urgent problem for high-speed railway enterprises. However, there has been little research on this issue. Therefore, a pricing optimization approach for refund service fee based on negative binomial distribution was proposed in this article. Firstly, we proved that the probability of passengers arriving based on the accumulated ticket sales obeyed the negative binomial distribution, which was used to fit the uncertainty demand of passengers. Then, we categorized the passengers to build an optimization model with the objective of maximizing compensation for losses caused by ticket cancellation. A case study was implemented to show that the proportion of refund service fee to ticket price is generally higher than 50%. The refund service fee remains monotonously nondecreasing as the departure date approaches. It also indicated that the current charging standard for refund service fees was too low to offset the losses. In addition, passenger preferences and passenger flow have significant impacts on the dynamic pricing strategy for refund service fees.

## 1. Introduction

In recent years, China's high-speed railway has gained worldwide recognition for its rapid development. As of 2022, China railway enterprises have operated over 40000 kilometers of high-speed railway network and accumulated rich management experience. High-speed railway enterprises set market-oriented goals, followed the requirements of revenue management, took price discrimination as one of the bases, and adopted dynamic pricing or differentiated pricing to maximize revenue. Ticket cancellation behavior has a significant impact on the revenue of high-speed railway (Table 1). The refund amount accounted for 16.3% of the total income from July 2023 to August 2023. However, for the high-speed railway managers or scholars in the pricing field, the primary focus is on the ticket pricing, and the charging for ticket cancellation is less involved.

Furthermore, the influence of ticket cancellation behavior on the revenue of high-speed railway is obviously more complex than ticket purchase behavior. As the cancelled tickets can be resold with uncertain demand, it is necessary to optimize ticket cancellation behaviors and ticket purchase behaviors as a whole. On the other hand, ticket cancellation behavior will certainly cause losses to high-speed railway enterprises and significantly weaken revenue. In order to reduce the occurrence of ticket cancellation behavior and compensate for their own profits, high-speed railway enterprises charge a certain proportion of the corresponding ticket price as refund service fee. To have a more intuitive understanding and ease of explanation of the refund service fee, the refund service fee is generally represented by the proportion of the refund service fee to the ticket price (refund service fee/ticket price) in the following content such as tables and figures.

Table 1: The influence of refund on revenue.

| Period | The proportion of the number of refunds to passenger flow (%) | The proportion of refund amount to total income (%) |
|---|---|---|
| July 2023–August 2023 | 15.1 | 16.3 |

High-speed railway enterprises actually adopted a tiered charging standard during the presale period, as shown in Figure 1. 20% of the corresponding ticket price will be charged as the refund service fee within 24 hours before departure. From 24 hours to 48 hours before departure, 10% of the corresponding ticket price will be charged, and 5% of the corresponding ticket price will be charged from 48 hours to 8 days before departure. The refund service fee will not be charged if exceeds 8 days before departure.

Unfortunately, the above charging standards mainly rely on the personal experience of high-speed railway managers and lack of support from a theoretical model with actual data; hence, the rationality cannot be verified. This is the first significance of this study. The other is as follows. If railway enterprises pursue maximum revenue, they can directly increase the refund service fee and, even in extreme cases, make the refund service fee to be equal to the ticket price (the amount refunded to the passenger is 0), while the above extreme measures cannot be implemented because of passenger complaints and public opinion. Therefore, the key is what extent should the refund service fee being increased to. In reality, railway enterprises aim to make up for losses as much as possible as the objective that can be accepted by passengers as well as reducing ticket cancellation behaviors of passengers more or less. This goal is lower than simply pursuing maximum revenue and is easier to achieve. Therefore, this manuscript proposed a new optimization approach for pricing of refund service fee to solve the aforementioned challenges.

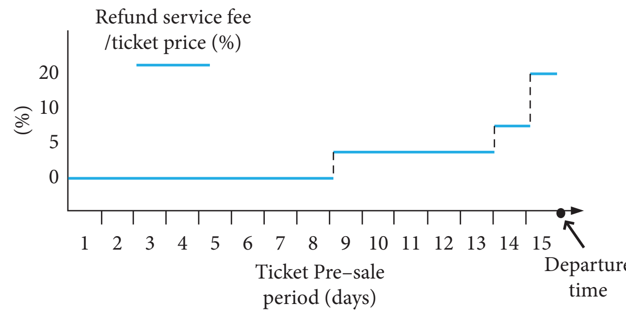


Figure 1: Current charging standard of refund service fee.

## 2. Literature Review

The pricing for refund service fee can be regarded as part of pricing in revenue management. The core of revenue management in high-speed railway is pricing and it mainly refers to dynamic pricing or differential pricing based on price discrimination in economics. Among the three levels of price discrimination, the first level is hard to realize because the price of a commodity equals the buyer's maximum willingness to pay. The second level and third level are easier to apply in practice. The typical application scenario of the second-level price discrimination is the mobile communication market; the price depends on the number of units to be purchased [1]. The third-level price discrimination reflects pricing policy, according to the price elasticity of demand, which is changed to adapt to segment market. Price discrimination is introduced into the field of transportation pricing research such as aviation [2, 3] and is shown to have a positive effect on earnings. Puller and Taylor [4] analyzed the price discrimination adopted by airlines based on the time of ticket purchase. Luo and Peng [5] developed a continuous-time dynamic pricing model for two competitive flights. Zhang and Cooper [6] proposed a Markov decision process formulation of a dynamic pricing problem for multiple substitutable flights between the same origin and destination, taking customers' choice behaviors among different flights into account. Santos and Gillis [7] proposed a data-driven modeling framework to estimate the flight pass price, which is a new concept in airlines.

Railway transport enterprises have also introduced pricing strategies such as differential pricing and dynamic pricing to maximize the revenue. Voss [8] proposed a pricing model for different types of fares, such as student tickets. Wang et al. [9] constructed an optimization model for differential pricing of high-speed railway based on discrete price and solving the model by particle swarm optimization. Zhang et al. [10] suggested a revenue-maximization model that integrated both operation planning and pricing dimensions, in terms of dynamic ticket pricing, elasticity in passenger demand, and flexible dispatching. Qin et al. [11] divided the passenger market according to the different factors affecting passenger choice behaviors, maximized ticketing revenue with expected travel cost as the reference point, and used prospect theory to construct a differentiated pricing model under elastic demand. Lin and Sibdari [12] made use of the game theory model to study the dynamic pricing problem between related substitutes. Zhao [13] built a comprehensive optimization model of ticket amount and ticket price for multiple trains and multiple stops and designed a hybrid heuristic algorithm to search for the optimal solution according to the characteristics of the model with the aim of getting different pricing for different trains in different sections. Li and Cao [14] took enterprise profits and passenger welfare as the pricing objectives of high-speed railway, decomposed the multiobjective problem by using the idea of hierarchical sequence, and finally determined the optimal multiobjective pricing strategy. Cai and Ou [15] categorized high-speed railway passengers and obtained different service attributes of passengers for parallel trains and then introduced the concept of revenue management and built a dynamic pricing model for parallel trains with the goal of maximizing the overall revenue of multiple trains.

Compared to the larger amount of literature studies on high-speed railway ticket pricing, few scholars have paid attention to the issue of how to charge for ticket cancellation behavior (refund service fee). Cirillo et al. [16] proposed an intertemporal choice model of ticket cancellation and exchange for railway passengers where customers were assumed to be forward-looking agents. A dynamic discrete choice model (DDCM) was applied to predict the timing in which ticket exchange or cancellation occurred in response to fare and trip schedule uncertainty. Iiescu et al. [17] explored the use of discrete choice methods for airline passenger cancellation behavior and estimated a discrete time proportional odds model with a prospective time scale based on the occurrence of cancellations in a sample of tickets provided by the Airline Reporting Corporation. Zhao et al. [18] designed a dynamic pricing model for refund fees that meet the expectations of passengers and airlines and chose the equilibrium point between them to solve refund service fees. Zhong et al. [19, 20] used the expected marginal value of air tickets at the time of ticket sales and refunds to determine refund service fee based on the dynamic pricing theory of revenue management and finally established a refund service fee model.

On the other hand, passenger arrival probability (ticket sale intensity) is an important foundation and is often applied to high-speed railway pricing. Talluri and Ryzin [21] integrated the multiple logit model into the dynamic model to estimate passenger arrival intensity. Vulcano et al. [22] used the method of expectation-maximization (EM) simultaneously to estimate passenger arrival intensity parameters and passenger selection preferences. Vulcano et al. [23] adopted and expanded this method on the basis of the expectation-maximization method. Many authors mainly used normal distribution or Poisson distribution to describe the probability of passenger arrival. Lee and Hersh [24] were the first to describe the arrival behavior of passengers by introducing the assumption that their arrival process was a Poisson's flow. The assumption was relaxed for establishing a dynamic pricing programming model to solve the expected revenue. Talluri et al. [25] pointed out that, in the multistage pricing problem, the Poisson distribution could be used as a reasonable assumption. Gallego and van Ryzin [26] and Bitran and Caldentey [27] assumed that the arrival of customers followed the Poisson distribution; each customer had an independent and identically distributed reservation price (willingness to pay), and then a new demand function was established. Zhao and Zheng [28] thought that the dynamic pricing problem of the single product under the condition of the demand following a nonhomogeneous Poisson process was studied over time. Feng and Gallego [29] assumed that the demand for product with each price level was a Poisson process in a limited sales period; the optimal time of charging price for a single product on an admissible time-varying path was discussed. For normal distribution, Song et al. [30] simulated the passenger demand function with normal distribution to get the mathematical relation between the demand of passengers and ticket prices and then gradually constructed and solved the dynamic ticket price optimization model of high-speed railway trains. However, the facts indicate that neither normal distribution nor Poisson distribution is accurate enough to fit the passenger arrival probability in most situations.

Based on the review of the above literature studies, it can be summarized as follows:

(1) There are many studies on differential pricing or dynamic pricing for tickets based on price discrimination in the academic circle. However, research of applying dynamic pricing to refund service fee is limited. Furthermore, the pricing for refund service fee which will also greatly affect the goals of pursuing maximum revenue is as important as ticket pricing in revenue management. Unfortunately, most studies always ignore the refund service fee. It is meaningful to study how to charge for ticket cancellation.

(2) As the cancelled tickets will be resold, the impact of ticket cancellation behavior on revenue is full of complexity, that is, ticket cancellation behaviors and ticket purchase behaviors are intertwined and influenced, composed of a dynamic and complex system in the presale period. This increases the research difficulties.

(3) The probability of passenger arrival is one of the foundations for building a pricing optimization model, but neither Poisson distribution nor normal distribution is accurate enough to describe the passenger arrival probability in most situations in which some tickets are already sold.

In response to the three points mentioned above, we proposed a dynamic pricing model for refund service fee with the following advantages:

(1) This is one of the limited numbers of articles that study pricing for refund service fee so as to figure out whether the current refund service fee based on the personal experience of the high-speed railway manager is reasonable.

(2) We proved that negative binomial distribution is more suitable for fitting the arrival probability of passengers under the situation of having tickets sold instead of Poisson distribution or normal distribution. The arrival probability of passengers is the key to solve the complex system that couples ticket cancellation behaviors and ticket purchase behaviors.

(3) We introduced the concept of psychological expectations to innovate the utility functions representing decision-making processes of passengers.

## 3. Model

*3.1. Model Assumption.* To simplify the problem and facilitate the establishment of models, some reasonable assumptions need to be made as follows:

(1) Only take second-class seat as the research object

(2) A passenger cannot purchase tickets and refund in the same time period

(3) The number of tickets sold is equivalent to the number of passengers arriving in the presale period

(4) Only one ticket can be refunded at a time

*3.2. Variable Definition.* Table 2 defines all symbols representing variables or parameters in this manuscript.

*3.3. Probability of Passengers Arriving Based on Negative Binomial Distribution.* It is the randomness of passenger arrival that leads to the uncertain demand. Thus, the precondition for studying uncertain demand is to get the number of passenger arrivals. Once passengers have the willingness to travel, they will search for the information of travel through the Internet. On each day of the presale period, there will be the occurrence of passenger arrival events, which is generally described by Poisson distribution. Referring to literature studies [16, 27–31], the probability of passenger arrival can be fitted with the Poisson distribution. Set $x$ as the number of passenger arrivals, and it obeys the independent Poisson distribution with parameter $\lambda$. Then, the arrival probability density function can be given as follows:

$$P(x = X) = \frac{(\lambda)^X e^{-\lambda}}{X!}. \tag{1}$$

$\lambda$ is unknown and is generally estimated through the mean of samples. Due to the estimated value heavily dependent on the scope of samples, sometimes there may be obvious errors. In order to describe the distribution pattern of passengers arriving more accurately, we need to dynamically update $\lambda$. If more tickets have been sold in the presale period, it is believed that the probability of passenger arrival is increasing. On the contrary, fewer ticket sale is likely to indicate a decrease in passenger arrival probability. Then, the Bayesian formula is used to dynamically update the value of $\lambda$ based on the accumulated ticket sales so that a more accurate posterior distribution of passenger arrival probability will be obtained. Assume that $\lambda$ obeys Gamma distribution with parameters $(b, a)$ [32], and the probability function of $\lambda$ can be expressed as follows:

$$f_\lambda(x) = \frac{\mathrm{ae}^{-\mathrm{ax}} (\mathrm{ax})^{b-1}}{\Gamma(b)}. \tag{2}$$

Parameters $(b, a)$ can be estimated through mean $\mu$ and variance $\sigma^2$ of samples:

$$\begin{aligned} \mu(\lambda) &= \frac{b}{a}, \\ \sigma^2(\lambda) &= \frac{b}{a^2}. \end{aligned} \tag{3}$$

Taking $A(t)/A(t-0)$ as an event representing the number of passengers arriving from the beginning to the $t^{\text{th}}$ day in the presale period, the probability of $A(t) = n$ can be derived as follows:

$$\begin{aligned} P(A(t) = n) &= \int_0^{+\infty} \frac{(xt)^n}{n!} e^{-xt} f_\lambda(x) \mathrm{d}x \\ &= \int_0^{+\infty} \frac{(xt)^n}{n!} e^{-xt} \frac{ae^{-ax} (ax)^{b-1}}{\Gamma(b)} \mathrm{d}x \\ &= \frac{t^n a^b}{n!\Gamma(b)} \int_0^{+\infty} x^{n+b-1} e^{-(a+t)x} \mathrm{d}x. \end{aligned} \tag{4}$$

Set $q = (a+t)x$ and bring it into formula (4). Then, we continue to derive the following formula:

$$\begin{aligned} &\frac{t^n a^b}{n!\Gamma(b)} \int_0^{+\infty} x^{n+b-1} e^{-(a+t)x} \mathrm{d}x \\ &= \frac{t^n a^b}{n!\Gamma(b)} \left(\frac{1}{a+t}\right)^{n+b} \int_0^{+\infty} x^{n+b-1} e^{-q} \mathrm{d}q \\ &= \frac{\Gamma(b+n)}{n!\Gamma(b)} \left(\frac{a}{a+t}\right)^b \left(\frac{t}{a+t}\right)^n \\ &= \frac{(n+b-1)(n+b-2)\cdots b}{n!} \left(\frac{a}{a+t}\right)^b \left(\frac{t}{a+t}\right)^n \\ &= C_{n+b-1}^n \left(\frac{a}{a+t}\right)^b \left(\frac{t}{a+t}\right)^n. \end{aligned} \tag{5}$$

It can be seen from formula (5) that the number of passengers arriving from the beginning to the $t^{\text{th}}$ day in the presale period obeys the negative binomial distribution with parameters $[b, a/(a+t)]$. If $n$ tickets have been sold when a passenger refunds on the $t^{\text{th}}$ day in the presale period, according to the Bayesian formula, the posterior probability function of $\lambda$ is as follows:

$$\begin{aligned} f_\lambda[x | A(t) = n] &= \frac{P(A(t) = n | x) f_\lambda(x)}{P(A(t) = n)} \\ &= \frac{((xt)^n/n!) e^{-xt} \left(ae^{-ax} (ax)^{b-1}/\Gamma(b)\right)}{(\Gamma(b+n)/n!\Gamma(b)) (a/a+t)^b (t/a+t)^n} \\ &= \frac{ae^{-ax} (ax)^{b-1} (xt)^n e^{-xt}}{\Gamma(b+n)} \left(\frac{a+t}{a}\right)^b \left(\frac{a+t}{t}\right)^n \\ &= \frac{(a+t) e^{-(a+t)x} ((a+t)x)^{n+b-1}}{\Gamma(n+b)}. \end{aligned} \tag{6}$$

Formula (6) shows that the posterior probability of $\lambda$ obeys the Gamma distribution with parameters $(b+n, a+t)$. Then, the probability of $z$ passenger arriving from refund time to departure time during the presale period is given as follows:

Table 2: Parameters and variables.

| Variables or parameters | Definition |
|---|---|
| $\text{od}(r, s)$ | Origin-destination (OD) pair in the transportation service which is composed of the train stops, from station $r$ to station $s$, equivalent to segment |
| $\text{od}(\hat{r}, \hat{s})$ | OD pair where ticket cancellation (refund) behavior occurs |
| $T$ | Number of days (periods) included in the presale period |
| $t$ | $t^{\text{th}}$ day (periods) in the presale period |
| $A(t)/A(t-0)$ | Event, passenger arrival from the beginning of the presale period to the $t^{\text{th}}$ day. Equal to time period $(t-0)$ |
| $n/z$ | Number of passengers arriving |
| $\Gamma(\text{b})$ | Gamma function |
| a, b | Parameters of negative binomial distribution |
| $k$ | High-speed railway train, $k = 1, 2 \ldots K$ |
| $\mu_{rs}^{k}$ | 0-1 judgment variable, when train $k$ provides service between $\text{od}(r, s)$, then $\mu_{rs}^{k} = 1$; otherwise, $\mu_{rs}^{k} = 0$ |
| $i$ | Passenger group, $i = 1, 2$ |
| $m_{rs}^{i}$ | The proportion of passenger group $i$ to all passengers between $\text{od}(r, s)$ |
| $w_{rs}^{k}$ | Travel time of train $k$ between $\text{od}(r, s)$ |
| $\overline{p}_{rs}$ | Average price between $\text{od}(r, s)$ |
| $\overline{w}_{rs}$ | Average travel time between $\text{od}(r, s)$ |
| $w_k$ | Travel time of train $k$ |
| $U_{rs}^{ik}$ | Utility obtained from train $k$ for passenger group $i$ between $\text{od}(r, s)$ |
| $\text{Pb}_{rs}^{ik}$ | Selection probability of passenger group $i$ choosing train $k$ between $\text{od}(r, s)$ |
| $\varphi_1, \varphi_2$ | Parameters of utility function |
| $c_{rs}^{k}, c_{\hat{r}\hat{s}}^{k}$ | Number of remaining tickets between $\text{od}(r, s)$ or $\text{od}(\hat{r}, \hat{s})$ |
| $c_k$ | Capacity of train $k$ |
| $\beta_t$ | Upper limit of the proportion of the refund service fee to the ticket price on the $t^{\text{th}}$ day |
| **Decision variable** | **Definition** |
| $R_{\hat{r}\hat{s}}^{tk}$ | Refund service fee of train $k$ on the $t^{\text{th}}$ day |

$$
\begin{aligned}
P(A(T-t) = z \mid A(t) = n) &= \int_{0}^{+\infty} \frac{[x(T-t)]^{z} e^{-x(T-t)}}{z!} \bullet \frac{(a+t)e^{-(a+t)x}((a+t)x)^{b+n-1}}{\Gamma(b+n)} \mathrm{d}x \\
&= \frac{\Gamma(b+n+z)}{z!\Gamma(b+n)} \left(\frac{a+t}{a+T}\right)^{b+n} \left(\frac{T-t}{a+T}\right)^{z}.
\end{aligned} \tag{7}
$$

Formula (7) means that the number of passengers arriving obeys the negative binomial distribution with parameters $[b+n, (a+t)/(a+T)]$ during the time period $T-t$ under the condition $A(t) = n$. Through the above proof, the negative binomial distribution is more accurate than the traditional Poisson distribution in fitting the probability of passenger arrival with the condition of some tickets already being sold.

*3.4. Passenger Choice Behavior.* Passengers first decide whether to purchase tickets after arrival and then make decisions about which train to take. Passengers' main focus is on travel time and ticket price. But different passenger groups have diverse preferences. In the prospect theory, passengers psychologically set a benchmark called psychological expectation in advance. They make choices that meet their expectations by comparing them with the benchmark. This is the mechanism by which psychological expectations affect passenger choices. Thus, we divided the passengers into two categories: price-sensitive passengers $(i = 1)$ and time-sensitive passengers $(i = 2)$. For time-sensitive passengers, they are more concerned about the time cost that they spend on the journey. When the travel time exceeds psychological expectations, the utility brought by travel time is 0. Correspondingly, the ticket price will not bring any utility to price-sensitive passengers when it is higher than their psychological expectations [32–34]. Passengers' psychological expectations can be concretized as average travel time $\overline{w}_{rs}$ and average ticket prices $\overline{p}_{rs}$ [35]. Thus, the utility function for price-sensitive passengers and time-sensitive passengers, respectively, can be expressed as follows:

$$U_{rs}^{ik} = \varphi_1 \bullet \min\left(p_{rs}^k - \overline{p}_{rs}, 0\right) + \varphi_2 \bullet \left(w_{rs}^k - \overline{w}_{rs}\right), \quad i = 1, \tag{8}$$

$$U_{rs}^{ik} = \varphi_1 \bullet \left(p_{rs}^k - \overline{p}_{rs}\right) + \varphi_2 \bullet \min\left(w_{rs}^k - \overline{w}_{rs}, 0\right), \quad i = 2, \tag{9}$$

$$\overline{p}_{rs} = \frac{1}{K}\sum_{k=1}^{K} p_{rs}^k \bullet \mu_{rs}^k, \tag{10}$$

$$\overline{w}_{rs} = \frac{1}{K}\sum_{k=1}^{K} w_{rs}^k \bullet \mu_{rs}^k. \tag{11}$$

On the basis of formulas (8)–(11), the probabilities of passenger group $i$ selecting train $k$ can be obtained as follows:

$$Pb_{rs}^{ik} = \frac{\mu_{rs}^k \bullet \exp\left(U_{rs}^{ik}\right)}{1 + \sum_{k=1}^{K} \mu_{rs}^k \bullet \exp\left(U_{rs}^{ik}\right)}. \tag{12}$$

Correspondingly, the probability that the passenger group $i$ give up purchasing tickets can be described by the following formula:

$$\widehat{P}b_{rs}^{ik} = \frac{1}{1 + \sum_{k=1}^{K} \mu_{rs}^k \bullet \exp\left(U_{rs}^{ik}\right)}. \tag{13}$$

*3.5. Dynamic Optimization Model.* On the $t^{\text{th}}$ day in the presale period, a ticket cancellation occurs between od$(\hat{r}, \hat{s})$, leading to the number of remaining tickets being added by 1. The ticket amount between other od$(r, s), \forall r, s$ remains unchanged. Set $D_z^n(t) = P(A(T-t) = z \mid A(t) = n)$. The expected ticket sales of train $k$ after a ticket cancellation can be expressed as the following formula:

$$\text{Sale}(\hat{r}, \hat{s}) = \sum_{z=1}^{+\infty}\left[D_z^{n+1}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right], \tag{14}$$

where $z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}$ refers to the passenger demand for train $k$ between $(\hat{r}, \hat{s})$ when the number of passengers arriving is $z$. $\min(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik})$ means passenger demand will not exceed the remaining tickets. The expected ticket sales between any nonrefund od$(r, s) \forall r, s$ are shown as follows:

$$\text{Sale}(r, s) = \sum_{z=1}^{+\infty}\left[D_z^{n}(t) \bullet \min\left(c_{rs}^k, z \bullet \sum_{i=1}^{2} m_{rs}^i \bullet Pb_{rs}^{ik}\right)\right]. \tag{15}$$

The expected revenue of train $k$ from the occurrence of a ticket cancellation to the end of the presale period is $B_{\text{refund}}$:

$$\begin{aligned} B_{\text{refund}} &= R_{\hat{r}\hat{s}}^{tk} + \sum_{(r,s)}\left[\text{Sale}(r, s) \bullet p_{rs}^k\right] + \text{Sale}(\hat{r}, \hat{s}) \bullet p_{\hat{r}\hat{s}}^k \\ &= R_{\hat{r}\hat{s}}^{tk} + \sum_{(r,s)}\left\{\sum_{z=1}^{+\infty}\left[D_z^n(t) \bullet \min\left(c_{rs}^k, z \bullet \sum_{i=1}^{2} m_{rs}^i \bullet Pb_{rs}^{ik}\right)\right] \bullet p_{rs}^k\right\} \\ &\quad + \sum_{z=1}^{+\infty}\left[D_z^{n+1}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] \bullet p_{\hat{r}\hat{s}}^k. \end{aligned} \tag{16}$$

Consistent with formula (16), if there is no ticket cancellation, the expected revenue of train $k$ is $\widehat{B}_{\text{refund}}$:

$$\begin{aligned} \widehat{B}_{\text{refund}} &= \sum_{(r,s)}\left\{\sum_{z=1}^{+\infty}\left[D_z^n(t) \bullet \min\left(c_{rs}^k, z \bullet \sum_{i=1}^{2} m_{rs}^i \bullet Pb_{rs}^{ik}\right)\right] \bullet p_{rs}^k\right\} \\ &\quad + \sum_{z=1}^{+\infty}\left[D_z^{n}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] \bullet p_{\hat{r}\hat{s}}^k + p_{\hat{r}\hat{s}}^k. \end{aligned} \tag{17}$$

The existing literature studies on high-speed railway pricing always take revenue maximization as the objective function. It should have been constructed to maximize revenue for the objective function of refund service fees. However, based on experience, ticket cancellation behavior must definitely lead to losses to the high-speed railway enterprises, that is, the expected revenue after ticket cancellation is very likely lower than it without ticket refund. In general revenue management, pursuing maximum revenue is the primary goal. But in order to avoid passengers' complaints or public pressure, it should be chosen to make up for losses by charging refund service fees to maximize the profits rather than obtaining higher revenue. In other words, making up losses as much as possible is cost-effective. Therefore, refund service fee should be equal to the difference of expected revenue caused by refunds or as much as possible. Thus, the objective has been transformed from maximizing revenue to maximizing compensation for losses generated by ticket cancellation behavior.

$$\begin{aligned} &\max B_{\text{refund}}\left(R_{\hat{r}\hat{s}}^{tk}\right) \leq \widehat{B}_{\text{refund}} \\ \Longrightarrow &\max B_{\text{refund}}\left(R_{\hat{r}\hat{s}}^{tk}\right) = \widehat{B}_{\text{refund}} \\ \Longrightarrow &B_{\text{refund}}\left(R_{\hat{r}\hat{s}}^{tk}\right) = \widehat{B}_{\text{refund}}. \end{aligned} \tag{18}$$

Equation (18) is equivalent to the following equation:

$$\begin{aligned} &R_{\hat{r}\hat{s}}^{tk} + \sum_{z=1}^{+\infty}\left[D_z^{n+1}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] \bullet p_{\hat{r}\hat{s}}^k \\ &= \sum_{z=1}^{+\infty}\left[D_z^{n}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] \bullet p_{\hat{r}\hat{s}}^k + p_{\hat{r}\hat{s}}^k. \end{aligned} \tag{19}$$

Formula (19) is the objective function and has the following constraints. Constraint (20) is the limit of train capacity. Constraint (21) sets an upper limit on the proportion of refund service fee to ticket price. As our goal is to reduce losses as much as possible, so $R_{\hat{r}\hat{s}}^{tk} \leq p_{\hat{r}\hat{s}}^{tk}$. Besides, restricted by practical application scenarios, especially potential passenger complaints and public pressure, we had better made $\beta_t \leq 1$ as the upper limit on the proportion of refund service fee to ticket price. Constraint (22) means that refund service fee is monotonically not reducing in the presale period.

$$0 \leq c_{\hat{r}\hat{s}}^k\left(c_{\hat{r}\hat{s}}^k\right) \leq c_k, \forall (r, s), \tag{20}$$

$$0 \leq \frac{R_{\hat{r}\hat{s}}^{tk}}{p_{\hat{r}\hat{s}}^k} \leq \beta_t, \forall t, \tag{21}$$

$$R_{\hat{r}\hat{s}}^{\overrightarrow{t}k} \leq R_{\hat{r}\hat{s}}^{\overleftarrow{t}k}, \overrightarrow{t} \leq \overleftarrow{t}, 1 \leq \forall \overrightarrow{t}, \overleftarrow{t} \leq T. \tag{22}$$

## 4. Solution

To solve decision variable $R_{\hat{r}\hat{s}}^{tk}$, objective function (19) can be transformed into the following formula:

$$\begin{aligned} R_{\hat{r}\hat{s}}^{tk} = &\left\{\sum_{z=1}^{+\infty}\left[D_z^{n}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] + 1\right\} \bullet p_{\hat{r}\hat{s}}^k \\ &- \sum_{z=1}^{+\infty}\left[D_z^{n+1}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}\right)\right] \bullet p_{\hat{r}\hat{s}}^k. \end{aligned} \tag{23}$$

Record $O_{\hat{r}\hat{s}}^k$ as the border amount of passengers arrival exceeding the remaining tickets. Thus, $O_{\hat{r}\hat{s}}^k = (c_{\hat{r}\hat{s}}^k + 1)/(\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik})$. Set $d_z^n(t)$ as the cumulative probability that less $z$ passenger arrival from refund time to departure time during the presale period on the basis of $n$ tickets being sold. That is, $d_z^n(t) = P(A(T-t) \leq z | A(t) = n)$.

Then, put $O_{\hat{r}\hat{s}}^k = (c_{\hat{r}\hat{s}}^k + 1)/(\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik})$ and $d_z^n(t) = P(A(T-t) \leq z | A(t) = n)$ into the first half of formula (23).

$$\begin{aligned} &\sum_{z=1}^{+\infty}\left[D_z^{n+1}(t) \bullet \min\left(c_{\hat{r}\hat{s}}^k + 1, z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{i}\right)\right] \\ &= \sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^{n+1}(t) \bullet \left(z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{i}\right)\right] + \sum_{z=O_{\hat{r}\hat{s}}^k+1}^{+\infty}\left[D_z^{n+1}(t) \bullet \left(c_{\hat{r}\hat{s}}^k + 1\right)\right] \\ &= \sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^{n+1}(t) \bullet \left(z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{i}\right)\right] + \sum_{z=O_{\hat{r}\hat{s}}^k+1}^{+\infty}\left[D_z^{n+1}(t)\right] \bullet \left(c_{\hat{r}\hat{s}}^k + 1\right) \\ &= \sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^{n+1}(t) \bullet \left(z \bullet \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{i}\right)\right] + \left[1 - d_{O_{\hat{r}\hat{s}}^k}^n(t)\right] \bullet \left(c_{\hat{r}\hat{s}}^k + 1\right). \end{aligned} \tag{24}$$

Similarly, for these OD pairs $(r, s)$ with no ticket cancellation, set $\widehat{O}_{rs}^k = c_{\hat{r}\hat{s}}^k / \sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i \bullet Pb_{\hat{r}\hat{s}}^{ik}$. Input it into the second half of formula (23).

$$\begin{aligned}
&\sum_{z=1}^{+\infty}\left[D_z^n(t)\bullet \min\left(c_{\hat{r}\hat{s}}^k, z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]+1\\
&=\sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^n(t)\bullet\left(z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]+\left[\sum_{z=O_{\hat{r}\hat{s}}^k+1}^{+\infty}\left(D_z^n(t)\bullet c_{\hat{r}\hat{s}}^k\right)\right]+1\\
&=\sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^n(t)\bullet\left(z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]+\sum_{z=O_{\hat{r}\hat{s}}^k+1}^{+\infty}\left[D_z^n(t)\right]\bullet c_{\hat{r}\hat{s}}^k+1\\
&=\sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^n(t)\bullet\left(z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]+\left[1-d_{O_{\hat{r}\hat{s}}^k}^n(t)\right]\bullet c_{\hat{r}\hat{s}}^k+1.
\end{aligned}\tag{25}$$

Finally, by bringing formulas (24) and (25) into formula (23), we can get $R_{\hat{r}\hat{s}}^{tk}$ by equation (26). It can be solved directly by elementary mathematics.

$$\mathbf{R}_{\hat{r}\hat{s}}^{\mathbf{tk}}=\begin{cases}\sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^n(t)\bullet\left(z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]-\sum_{z=1}^{O_{\hat{r}\hat{s}}^k}\left[D_z^{n+1}(t)\bullet\left(z\bullet\sum_{i=1}^{2} m_{\hat{r}\hat{s}}^i\bullet Pb_{\hat{r}\hat{s}}^i\right)\right]+\left[1-d_{O_{\hat{r}\hat{s}}^k}^n(t)\right]\bullet c_{\hat{r}\hat{s}}^k-\left[1-d_{O_{\hat{r}\hat{s}}^k}^n(t)\right]\bullet\left(c_{\hat{r}\hat{s}}^k+1\right)+1, & \dfrac{R_{\hat{r}\hat{s}}^{tk}}{p_{\hat{r}\hat{s}}^{tk}}<\beta_t,\\ p_{\hat{r}\hat{s}}^{tk}\bullet\beta_t, & \dfrac{R_{\hat{r}\hat{s}}^{tk}}{p_{\hat{r}\hat{s}}^{tk}}>\beta_t.\end{cases}\tag{26}$$

# 5. Case Analysis

*5.1. Basic Data.* Beijing-Shanghai high-speed railway line in China was just selected as a numerical example. The Beijing-Shanghai high-speed railway line mainly involves four stations, namely, Beijing South (BS), Jinan West (JS), Nanning South (NS), and Shanghai Hongqiao (SH). In order to simulate and verify the applicability and accuracy of our model, we randomly choose two trains (G19 and G21) with different stop plans on the line to construct a situation of multiple trains and multiple stops. These two trains both depart from Beijing South Station and finally arrive at Shanghai Hongqiao station. G21 only stops at Nanning South Station. Differently, G19 stops at Jinan West station additionally and provides service between all OD pairs. All the above information can be seen clearly in Figure 2. We take the data of two trains departing on April 1st as the case for empirical analysis.

Table 3 details the ticket prices and travel time of two trains on each segment. It is obvious that the ticket prices of G21 and G19 are the same between any OD pair. Because of fewer stops and less travel time between the same OD pair, G21 has a stronger competitiveness and has been welcomed by passengers, taking second-class seat as the research object. Table 4 describes the capabilities of two trains. The maximum number of tickets allocated to any OD pair is 1113 for both trains.

According to the capacity of the two trains, the initial results of tickets allocated to all OD pairs are set as shown in Table 5.

As the presale period lasts 15 days, set $T = 15$. Through analysis of the ticket sale data on April 1st, we have obtained the accumulated ticket sale intensity curves between different OD pairs. Because the accumulated ticket sale intensity curve of each OD pair has a high similarity, we select statistical results of several representative OD pairs to display to understand the trend and characteristics of the accumulated ticket sale intensity curves, as shown in Figure 3.

It can be concluded that the accumulated ticket sales of each segment in the presale period have certain similarities; that is, few passengers purchase tickets from 15 days to 5 days before departure (the accumulated proportion of the ticket sale does not exceed 20% in Figure 3). However, the accumulated ticket sales begin to increase rapidly in the last 5 days before departure. Table 6 details that the ticket cancellations were mainly concentrated within 2 days before departure or 5 days before departure based on the data from July 2023 to August 2023. On the other hand, it is reasonable for passengers to refund tickets for free in the early presale period and the reality is exactly like this (Figure 1). Therefore, we focus on the ticket cancellation behavior during the peak of the presale period, that is, set $11 \le t \le 15$. So, the ticket sale on each day in the peak of the presale period is obtained in Table 7.

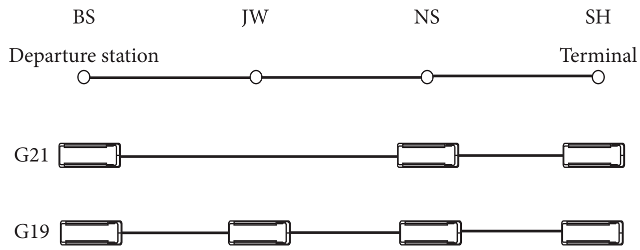


Figure 2: The train stop schemes.

Table 3: Travel time and ticket price.

| Train | From station | To station | Travel time (unit: min) | Ticket price (unit: RMB) |
|---|---|---|---|---|
| G19 | BS | NS | 204 | 504 |
| | JW | NS | 120 | 315 |
| | BS | SH | 268 | 626 |
| | JW | SH | 184 | 453 |
| | NS | SH | 62 | 153 |
| | BS | JW | 82 | 211 |
| G21 | BS | NS | 194 | 504 |
| | BS | SH | 258 | 626 |
| | NS | SH | 62 | 153 |

Table 4: Capacity of each train (second-class seat).

| Train | Capacity (unit: seat) |
|---|---|
| G19 | 1113 |
| G21 | 1113 |

Table 5: Number of tickets allocated to each OD pair.

| Train | From station | To station | Number of tickets allocated |
|---|---|---|---|
| G19 | BS | NS | 100 |
| | JW | NS | 113 |
| | BS | SH | 800 |
| | JW | SH | 100 |
| | NS | SH | 213 |
| | BS | JW | 213 |
| G21 | BS | NS | 313 |
| | BS | SH | 800 |
| | NS | SH | 313 |

Based on empirical estimation, price-sensitive passengers account for 80% of the total passengers, and the remaining 20% is the travel time-sensitive passenger. Thus, $m^1_{\hat{r}\hat{s}} = 0.8, m^2_{\hat{r}\hat{s}} = 0.2$. The values of utility function parameters, for example, are as follows: $\phi_1 = -0.9$ and $\phi_2 = -0.1$. Regardless of the current charging standard for refund service fee, set $\beta_t = 1, \forall t$.

*5.2. Computational Results.* We used Python software to solve the model. In order to calculate the probability of arrival of passengers based on the negative binomial distribution. Method of moment estimation is used to get the values of two variables $\mu$ and $\sigma$ through inputting sample data. Table 8 presents the values of key parameters $a$ and $b$ which determine the specificity of negative binomial distribution obtained by formulas (3) and (4).

Continue to calculate the passenger selection probabilities. According to the definition of psychological expectations proposed in this article, the average values of travel time and ticket price between each OD pair are shown in Table 9.

Based on the data in Table 3, the probabilities of different passenger groups choosing from train set between each OD pair are calculated by formulas (9)–(14), which can be seen in Table 10. Price-sensitive passengers or travel time-sensitive passengers prefer G21 between OD pairs such as (BS, NS), (BS, SH), and (NS, SH), where two trains provide service. Because of less travel time and the same ticket price, G21 is more competitive than G19 on the same segment.

Set $t = 11$ to $t = 15$, respectively, and input them into formula (25) to solve the decision variable $R^{tk}_{\hat{r}\hat{s}}$. The results can be seen as follows.

Table 11 shows that the dynamic optimization results of refund service fee in this article account for over 50% of the ticket price for all OD pairs. It also indicates that ticket cancellation behavior of passengers will cause significant losses to high-speed railway enterprises and seriously damage their goals of maximizing revenue. Figures 4 and 5 provide comparisons of the trend of the refund service fee fluctuating between different OD pairs for the same train. Whether G19 or G21, the refund service fees monotonically increase and have certain linear characteristics in the peak of the presale period approaching the departure date due to the probability of ticket sales decreasing. The high-speed railway enterprises have to raise the refund service fee to offset the loss of expected revenue.

*5.3. The Impact of Passenger Choice Behavior.* Figure 6 suggests that passenger selection preferences play an important role in influencing refund service fee. We take the OD pair (Beijing South, Nanning South) as an example. Because of less travel time and the same price (Table 4), G21 possesses stronger competitiveness than G19 between Beijing South and Nanning South segments. Passengers will give priority to purchase tickets of G21 (50.6% vs. 18.6% for price-sensitive passengers, 45.2% vs. 27.4% for travel time-sensitive passengers in Table 10), which results in higher probability of selling tickets for G21. Therefore, G21 only charging for a lower refund service fee can minimize the loss of expected revenue. However, the refund service fee curve of G21 quickly converges to that of G19, as the presale period approaches the end. Considering the actual charging standard for refund service fee (black curve in Figure 6), the actual refund service fee curve remains far below the curves of G21 and G19, which demonstrates the current standard of

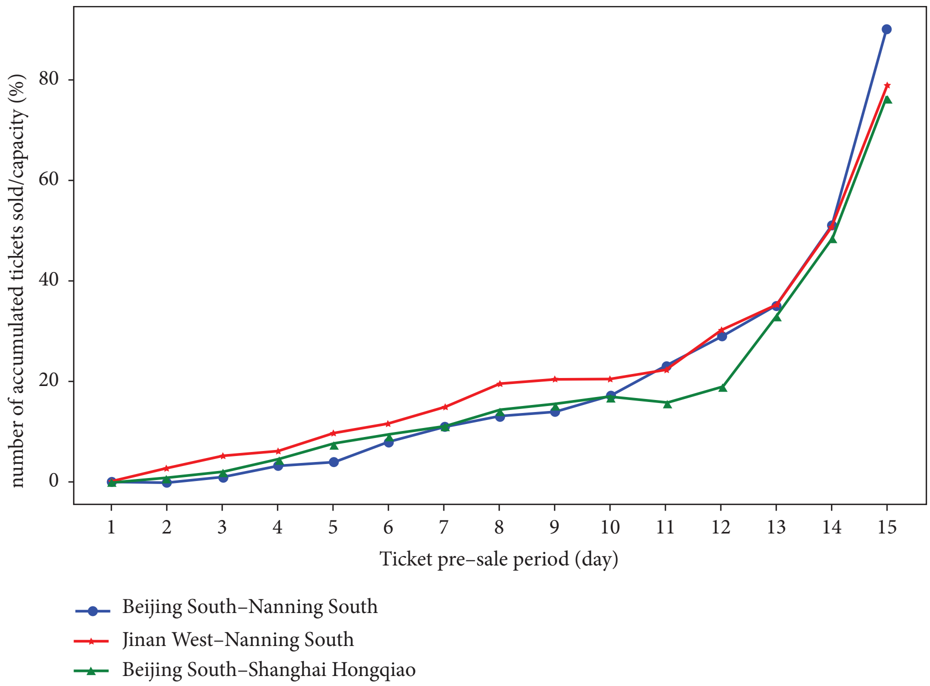


Figure 3: Accumulated ticket sale intensity curves of main segments.

Table 6: Statistics of the number of refunds per day in the presale period.

| The proportion of daily refunds to the total number in the presale period (%) | | | | | | | |
|---|---|---|---|---|---|---|---|
| $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ |
| 1.04% | 0.91% | 0.84% | 0.89% | 1.03% | 1.20% | 1.44% | 1.97% |
| $t = 9$ | $t = 10$ | $t = 11$ | $t = 12$ | $t = 13$ | $t = 14$ | $t = 15$ | |
| 2.57% | 3.14% | 4.30% | 6.15% | 8.43% | 19.14% | 46.95% | |

Table 7: Accumulated ticket sales on each day.

| Train | From station | To station | Accumulated ticket sales ($n$) ($t^{th}$ day in the presale period) | | | | |
|---|---|---|---|---|---|---|---|
| | | | $t = 11$ | $t = 12$ | $t = 13$ | $t = 14$ | $t = 15$ |
| G19 | BS | NS | 23 | 29 | 35 | 51 | 90 |
| | JW | NS | 25 | 34 | 40 | 57 | 89 |
| | BS | SH | 125 | 151 | 264 | 387 | 611 |
| | JW | SH | 17 | 28 | 32 | 54 | 78 |
| | NS | SH | 39 | 55 | 70 | 123 | 173 |
| | BS | JW | 45 | 67 | 90 | 129 | 182 |
| G21 | BS | NS | 85 | 101 | 133 | 167 | 265 |
| | BS | SH | 198 | 255 | 302 | 377 | 759 |
| | NS | SH | 99 | 121 | 147 | 166 | 257 |

Table 8: Values of parameters $a$ and $b$.

| From station | To station | $a$ | $b$ |
|---|---|---|---|
| BS | SH | 0.0064 | 0.4259 |
| BS | JW | 0.0068 | 0.6404 |
| BS | NS | 0.0066 | 0.5500 |
| JW | SH | 0.0521 | 2.9948 |
| JW | NS | 0.0655 | 1.8006 |
| NS | SH | 0.0100 | 0.5328 |

Table 9: Average travel time and average ticket price.

| From station | To station | Travel time (unit: min) | Ticket price (unit: RMB) |
|---|---|---|---|
| BS | SH | 263 | 626 |
| BS | JW | 82 | 211 |
| BS | NS | 199 | 504 |
| JW | SH | 184 | 453 |
| JW | NS | 120 | 315 |
| NS | SH | 62 | 153 |

Table 10: The probabilities of different passenger groups choosing from train set.

| Train | From station | To station | Price-sensitive passengers (%) | Travel time-sensitive passengers (%) |
|---|---|---|---|---|
| G19 | BS | NS | 18.6 | 27.4 |
| | JW | NS | 50.0 | 50.0 |
| | BS | SH | 18.6 | 27.4 |
| | JW | SH | 50.0 | 50.0 |
| | NS | SH | 33.3 | 33.3 |
| | BS | JW | 50.0 | 50.0 |
| G21 | BS | NS | 50.6 | 45.2 |
| | BS | SH | 50.6 | 45.2 |
| | NS | SH | 33.3 | 33.3 |

Table 11: Refund service fee and proportion to the ticket price.

| Train | From station | To station | Refund service fee (yuan)/proportion of refund service fee to ticket price (%) ($t^{th}$ day in the presale period) | | | | |
|---|---|---|---|---|---|---|---|
| | | | $t = 11$ | $t = 12$ | $t = 13$ | $t = 14$ | $t = 15$ |
| G19 | BS | NS | 362/71.9% | 394/78.1% | 423/84% | 451/89.5% | 478/94.9% |
| | JW | NS | 171/54.5% | 201/64% | 231/73.4% | 260/82.5% | 315/100% |
| | BS | SH | 473/75.6% | 507/81.1% | 540/86.3% | 570/91.1% | 599/95.7% |
| | JW | SH | 231/51.1% | 273/60.4% | 320/70.7% | 363/80.3% | 408/90.3% |
| | NS | SH | 102/67.2% | 114/74.5% | 124/81.4% | 134/87.9% | 143/94.1% |
| | BS | JW | 122/57.9% | 141/67.1% | 160/75.9% | 177/84.2% | 194/92.2% |
| G21 | BS | NS | 303/60.2% | 348/69.1% | 390/77.5% | 430/85.4% | 468/92.9% |
| | BS | SH | 396/63.3% | 449/71.9% | 498/79.7% | 544/86.9% | 544/86.9% |
| | NS | SH | 102/67.2% | 114/74.5% | 124/81.4% | 134/87.9% | 143/94.1% |

charging for ticket refund is too low and it is not sustainable for the financial balance-oriented development of high-speed railway enterprises.

*5.4. The Impact of Passenger Flow.* Passenger flow is also a crucial factor affecting the refund service fee and fluctuates periodically with the departure date. Even on the two adjacent departure dates, the optimization result of refund service fee varies significantly under the influence of passenger flow. To illustrate this, April 2nd was selected as the departure date to compare with April 1st above from Beijing South to Nanjing South for G21. Table 12 shows that the ticket sale on April 2nd is faster than that on April 1st between main OD pairs (Beijing South, Nanning South) and (Beijing South, Shanghai Hongqiao) during the presale period with the same seat capacity. This indicates that April 2nd has a stronger passenger flow. The final refund service fee solved on the two adjacent departure dates obtained though the optimization model proposed in this article is shown in Figure 7. The refund service fee curve on April 2nd is always more than that on April 1st on the same day during the presale period, which means that the refund service fee varies inversely with passenger flow. The higher the passenger flow intensity, the greater the probability that high-speed railway manager can resell the cancelled tickets, resulting in a lower refund service fee charged. In summary, we can reduce the proportion of the refund service fee to the ticket price when demand exceeds supply.

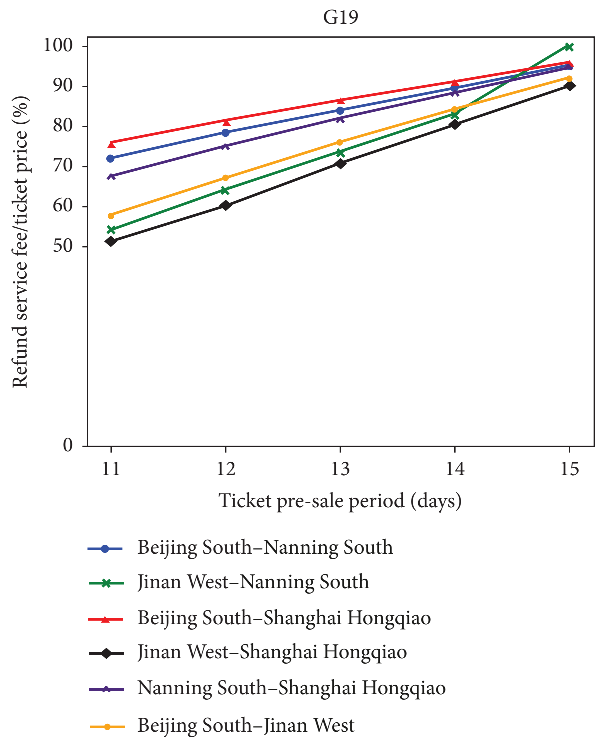


Figure 4: Proportion of refund service fee to ticket price between each segment for G19.

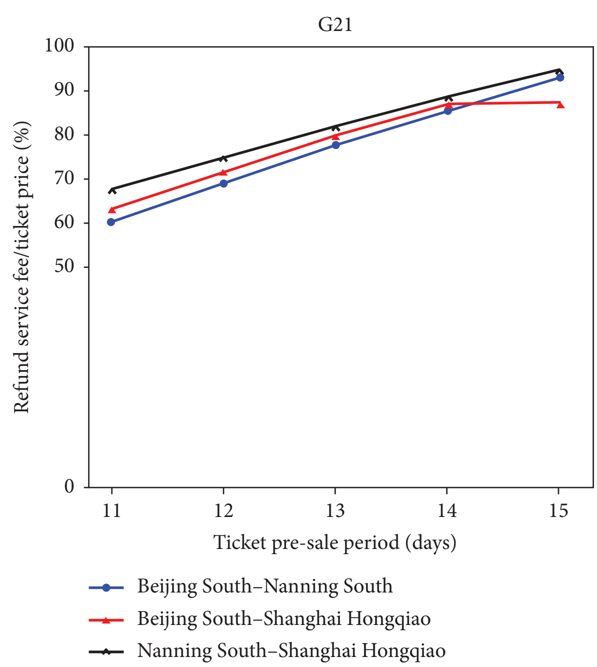


Figure 5: Proportion of refund service fee to ticket price between each segment for G21.

*5.5. Discussion.* In the light of above, we can make some conclusions as follows:

(1) The refund service fee solved by the optimization model monotonically increases as the presale period gradually ends, following a linear pattern.

(2) It is evident that refund service fee, which aims to minimize losses incurred due to ticket cancellation, is significantly higher than the actual value. This suggests that high-speed railway enterprises have set a low charging standard for refund service fees, and measures such as refund restrictions and ticket

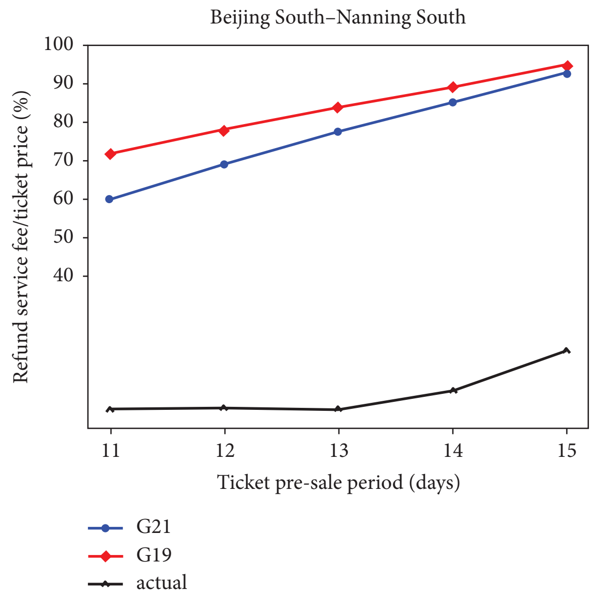


Figure 6: The impact of passenger choice on refund service fee.

Table 12: Accumulated ticket sales for G21 on different departure dates.

| Departure date | From station | To station | Accumulated ticket sale ($n$) ($t^{th}$ day in the presale period) | | | | |
|---|---|---|---|---|---|---|---|
| | | | $t = 11$ | $t = 12$ | $t = 13$ | $t = 14$ | $t = 15$ |
| April 1st | BS | NS | 85 | 101 | 133 | 167 | 265 |
| April 1st | BS | SH | 334 | 389 | 502 | 689 | 734 |
| April 2nd | BS | SH | 398 | 422 | 515 | 760 | 775 |
| April 2nd | BS | NS | 155 | 186 | 211 | 257 | 299 |

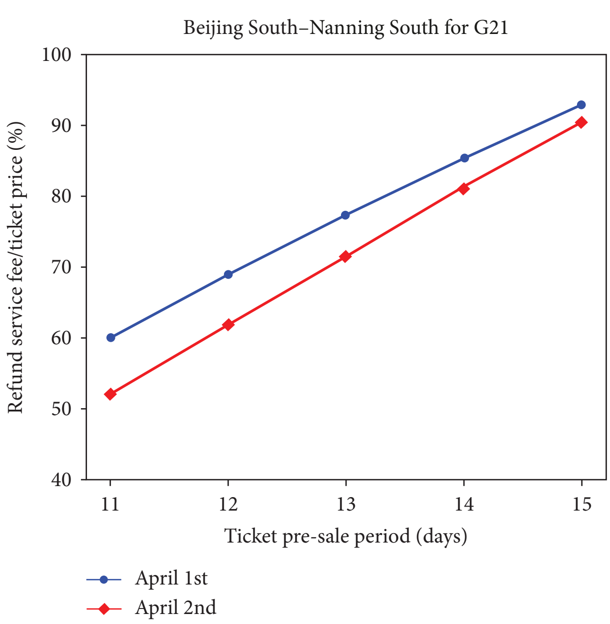


Figure 7: The impact of passenger flow on refund service fee.

cancellation penalties should be implemented to reduce passenger ticket refunds.

(3) Passenger preferences have a significant impact on refund service fee. The more competitive the train is, the lower the proportion of refund service fee to ticket price should be.

(4) Another factor affecting refund service fee is passenger flow intensity, and it is negatively correlated with the refund service fee. High-speed railway enterprises should charge for refund service fee reasonably according to passenger flow regularity.

## 6. Conclusions

We proposed a pricing optimization approach for refund service fees based on the negative binomial distribution under the situation of complexity in ticket sale as well as address the issues of overreliance on personal experience and lack of data support in the current pricing strategies of high-speed railway. Our model was designed to compensate for revenue loss as much as possible. It was proved that the arrival conditional probability of passengers conforms to the negative binomial distribution based on tickets sold in the presale period. We also weighed passenger preferences in terms of ticket prices and travel time to derive the selection probability of different passenger groups for high-speed trains. Finally, we established and solved the dynamic pricing model to obtain refund service fee for any train and any segment on $t^{th}$ day of the presale period.

In the case analysis, the current charging standard for refund service fee is insufficient to make up for the revenue loss. In fact, the refund service fee should not exceed 50% of the ticket price to avoid legal issues caused by passenger complaints. Therefore, the results of this article are based on the goal of compensating for the revenue losses and are higher than the actual situation. Considering the reality, railway enterprises will not raise the current charging standard of the refund service fee to the level in the case analysis. As the presale period nears its end, the refund service fees gradually increase in a linear trend. At the same time, passenger selection preferences and passenger flow intensity have dominant impacts on the results.

Our study has only chosen a less focused route to construct an objective function and explore its feasibility. Of course, it has limitations, especially without taking the essential idea of making the fees charged for refunds at different periods and subsequently optimizing it. Furthermore, there is always a gap between theory and reality. We modeled the issue of pricing for refund service fee theoretically and still need to overcome application challenges. Considering the complexity of model construction and solution, how to organize a large-scale information system to solve the model in parallel to satisfy timeliness will be the first challenge. In addition, the big data required for parameter estimation is also a key obstacle. Our further studies will consider more ideas and practical limitations to obtain more applicable results. In addition, the application scenario can be extended from a single line with multiple trains and multiple stops in this article to a complex high-speed railway network composed of multiple lines and trains.

## Data Availability

The data can be obtained from China Railway Marketing System.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Authors' Contributions

Y.W. conceptualized the study; Y.W. and J.Z. contributed to methodology; H. W. and X. Y. were responsible for software; Y.W. and J.Z. analyzed the data and performed result analysis. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This research was funded by the research plan of China Academy of Railway Sciences Group Co., Ltd, grant no. 2023YJ148.

## References

[1] D. Flath, "Second-degree price discrimination by Japanese newspapers," *Japan and the World Economy*, vol. 44, pp. 14–25, 2017.

[2] S. Borenstein, "Hubs and high fares. Dominance and market power in the U.S. Airline industry," *The RAND Journal of Economics*, vol. 20, no. 3, pp. 344–365, 1989.

[3] S. Giaume and S. Guillou, "Price discrimination and concentration in European airline markets," *Journal of Air Transport Management*, vol. 10, no. 5, pp. 305–310, 2004.

[4] S. L. Puller and L. M. Taylor, "Price discrimination by day-of-week of purchase: evidence from the U.S. Airline industry," *Journal of Economic Behavior and Organization*, vol. 84, no. 3, pp. 801–812, 2012.

[5] L. Luo and J. H. Peng, "Dynamic pricing model for airline revenue management under competition," *Systems Engineering- Theory and Practice*, vol. 27, no. 11, pp. 15–25, 2007.

[6] D. Zhang and W. L. Cooper, "Pricing substitutable flights in airline revenue management," *European Journal of Operational Research*, vol. 197, no. 3, pp. 848–861, 2009.

[7] B. F. Santos and M. M. D. Gillis, "Optimizing the prices for airline flight passes," *Transportation Research Procedia*, vol. 37, pp. 266–273, 2019.

[8] A. Voss, "Collective public-transport tickets and anticipated majority choice: a model of student tickets," *Transportation Research Part A: Policy and Practice*, vol. 80, pp. 263–276, 2015.

[9] Y. Wang, Y. Liu, X. Li et al., "Preparation of nano-$Mg(OH)_2$ and its flame retardant and antibacterial modification on polyethylene terephthalate fabrics," *Polymers*, vol. 15, no. 1, pp. 7–14, 2022.

[10] X. Zhang, L. Li, S. Le Vine, and X. Liu, "An integrated pricing planning strategy to optimize passenger rail service with uncertain demand," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 1, pp. 435–448, 2019.

[11] J. Qin, W. Qu, X. Wu, and Y. Zeng, "Differential pricing strategies of high speed railway based on prospect theory: an empirical study from China," *Sustainability*, vol. 11, no. 14, p. 3804, 2019.

[12] K. Y. Lin and S. Y. Sibdari, "Dynamic price competition with discrete customer choices," *European Journal of Operational Research*, vol. 197, no. 3, pp. 969–980, 2009.

[13] X. Zhao, *Research on Joint Optimization of Pricing and Seat Allocation for High-Speed Railway*, Beijing Jiaotong University, Beijing, China, 2019.

[14] X. M. Li and H. Z. Cao, "Multi-objective pricing of high-speed railway passenger tickets based on epsilon-constraint method," *Journal of Transportation Systems Engineering and Information Technology*, vol. 1, pp. 6–11, 2020.

[15] J. M. Cai and Y. S. Ou, "Dynamic differential pricing of high-speed railway parallel trains considering renvenue management," *Journal of Transportation Systems Engineering and Information Technology*, vol. 5, pp. 2–8, 2020.

[16] C. Cirillo, F. Bastin, and P. Hetrakul, "Dynamic discrete choice model for railway ticket cancellation and exchange decisions," *Transportation Research Part E: Logistics and Transportation Review*, vol. 110, pp. 137–146, 2018.

[17] D. C. Iiescu, L. A. Garrow, and R. A. Parker, "A hazard model of US airline passengers' refundand exchange behavior," *Transportation Reaearch Part B*, vol. 42, pp. 229–242, 2007.

[18] G. H. Zhao, D. Wang, and J. Y. Deng, "The study on dynamic pricing method of airline refund service fee," *Price: Theory and Practice*, vol. 1, pp. 139–142, 2020.

[19] N. Zhong and P. Tian, "Refunds design based on the theory of dynamic pricing," *Journal of Shanghai Jiaotong University*, vol. 12, no. 41, pp. 1993–2000, 2007.

[20] N. Zhong, *Research on Dynamic Pricing Strategies for Revenue Management*, Shanghai Jiaotong University, Shanghai, China, 2008.
[21] K. T. Talluri and G. J. V. Ryzin, *The Theory and Practice of Revenue Management*, Springer Netherlands, Dordrecht, Netherlands, 2004.
[22] G. Vulcano, G. van Ryzin, and W. Chaar, “OM practice—choice-based revenue management: an empirical study of estimation and optimization,” *Manufacturing and Service Operations Management*, vol. 12, no. 3, pp. 371–392, 2010.
[23] 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.
[24] T. C. Lee and M. A. Hersh, “A model for dynamic airline seat inventory control with multiple seat bookings,” *Transportation Science*, vol. 27, no. 3, pp. 252–265, 1993.
[25] K. T. Talluri, G. J. V. Ryzin, and I. Z. Karaesmen, “Revenue management: models and methods,” in *Proceedings of the Winter Simulation Conference*, pp. 148–161, Hilton Austin Hotel, Austin, TX, USA, December, 2009.
[26] G. Gallego and G. van Ryzin, “A multi-product dynamic pricing problem and its applications to network yield management,” *Operations Research*, vol. 45, no. 1, pp. 24–41, 1997.
[27] G. Bitran and R. Caldentey, “An overview of pricing models for revenue management,” *IEEE Engineering Management Review*, vol. 44, no. 4, pp. 134–229, 2016.
[28] W. Zhao and Y. S. Zheng, “Optimal dynamic pricing for perishable assets with nonhomogeneous demand,” *Management Science*, vol. 46, no. 3, pp. 375–388, 2000.
[29] Y. Feng and G. Gallego, “Perishable asset revenue management with markovian time dependent demand intensities,” *Management Science*, vol. 46, no. 7, pp. 941–956, 2000.
[30] W. B. Song, P. Zhao, and B. Li, “Research on conprehensive optimization of dynamic pricing and seat allocation for high-speed single train,” *Journal of the China Raliway Society*, vol. 7, pp. 10–16, 2018.
[31] L. X. Cui and Q. Hou, “Dynamic pricing model of high-speed railway based on opportiunity,” *Journal of the China Raliway Society*, vol. 3, no. 43, pp. 9–17, 2021.
[32] E. Avineri, “A cumulative prospect theory approach to passengers’ behavior modeling: waiting time paradox revisited,” *J. Intell. Transp. Syst.*, vol. 8, pp. 195–204, 2004.
[33] G. M. Ramos, W. Daamen, and S. Hoogendoorn, “A state-of-the-art review: developments in utility theory, prospect theory and regret theory to investigate travelers’ behaviour in situations Involving travel timeuncertainty,” *Transport Reviews*, vol. 34, pp. 46–67, 2014.
[34] S. Gao, E. Frejinger, and M. Ben-Akiva, “Adaptive route choices in risky traffic networks: a prospect theory approach,” *Transportation Research Part C: Emerging Technologies*, vol. 18, pp. 727–740, 2010.
[35] R. C. Jou, R. Kitamura, M. C. Weng, and C. C. Chen, “Dynamic commuter departure time choice under uncertainty,” *Transportation Research Part A: Policy and Practice*, vol. 42, no. 5, pp. 774–783, 2008.