


Article

Passengers' Demand Characteristics Experimental Analysis of EMU Trains with Sleeping Cars in Northwest China

Yuzhao Zhang * , Jianqiang Wang and Wenjuan Cai

School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China; wjq@mail.lzjtu.cn (J.W.); 0218058@stu.lzjtu.edu.cn (W.C.)

* Correspondence: zhangyuzhao-81@mail.lzjtu.cn

Received: 20 July 2019; Accepted: 25 September 2019; Published: 27 September 2019



Abstract: Passenger demand characteristics for electrical multiple unit (EMU) trains with sleeping cars will directly affect the train operation scheme in a long transportation corridor. Descriptive statistics of individual attributes and passenger choice intentions for EMU trains with sleeping cars are calculated based on the revealed preference (RP) and stated preference (SP) survey data in Northwest China to illustrate the overall conditions of passengers' demands. Considering the higher dimensionality and multi-collinearity in the dataset of influencing factors, the factor analysis method was first adopted to reduce the number of dimensions of the raw dataset and obtain orthogonal common factors. Then, the ordinal logistic regression model was adopted to test and perform a regression analysis based on multinomial logit theory. The analysis shows that these influencing factors, such as income, profession, educational background and residence, would have a greater impact on the choice of an EMU train with sleeping cars. It is significant that passengers' choice intentions are positively correlated with income and educational background. The result can provide some reference for the decision-making regarding operating an EMU train with sleeping cars in Northwest China. In addition, the proposed method can be applied to the analysis of passengers' demand characteristics in similar situations.

Keywords: passenger transport; transportation corridor; data analysis; EMU trains with sleeping cars; multinomial logit

1. Introduction

The purpose of the operation of railway passenger trains is to satisfy the demand of passengers. Therefore, the operation mode of passenger trains depends on the characteristics of passenger flow [1]. Trains running on high-speed railways are highly competitive in the passenger transportation market because of the advantages of high speed, short travel time, high operating density, low energy consumption and low pollution [2,3]. Because high-speed railways operate in the daytime and are stationary at night, high-speed trains seldom include sleeping cars. For those passengers who spend more than five hours travelling, their level of fatigue increases sharply. Therefore, the service quality for passengers will be influenced, and the demand for night travel cannot be met. Electrical multiple unit (EMU) trains with sleeping cars operate with a sunset departure and a sunrise arrival, and this operation is an attempt by the China railways to improve the operation mode of high-speed railways based on market demand. To make full use of the transport capacity of high-speed rails in the evening and meet the market demand better, since 1 January 2015 eight pairs of sunset-departure and sunrise-arrival EMU trains with sleeping cars have been running between Beijing, Shanghai and Guangzhou, Shenzhen. From the beginning, the passenger flows of these trains have been increasing

steadily, and the operation effects are good [4]. Subsequently, according to the actual demand, some railway bureaus in central and eastern China added a certain number of EMU trains with sleeping cars.

The sleeping berths of new type EMU train with sleeping cars are a longitudinal arrangement with two layers. The interior and exterior photos of a sleeping car are shown as Figure 1. With this kind of train, a car can accommodate up to 60 berths, while a train can carry 880 passengers. So its transport capacity is higher than that of the first-class berth of an ordinary passenger train. Furthermore, the travel time and the ticket price of EMU train with sleeping cars are highly competitive too. The comparison among different kinds of passenger trains is shown in Figure 2.



Figure 1. Interior and exterior photos of a sleeping car.

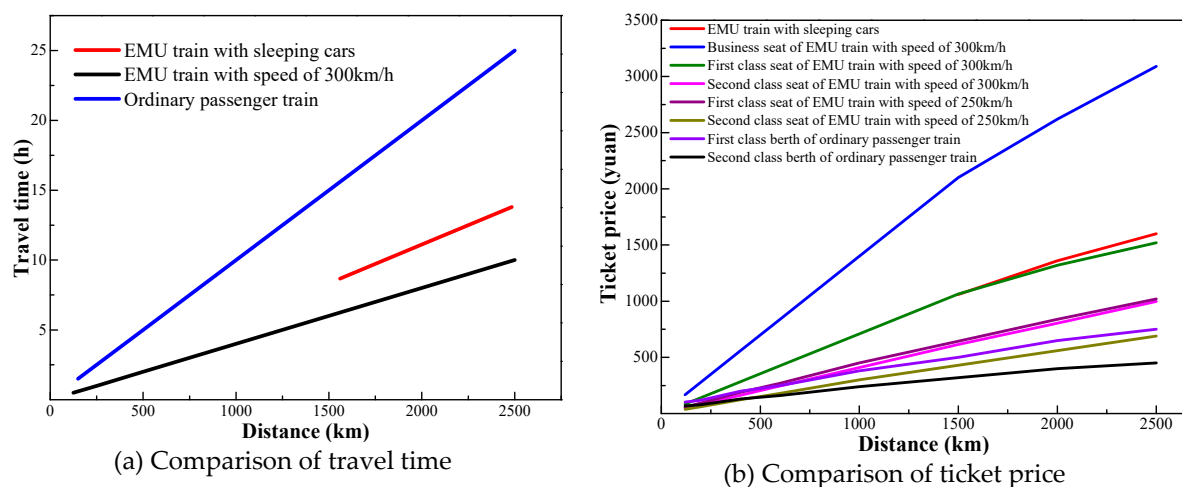


Figure 2. Comparison among different kinds of trains.

In Figure 2a, when the distance is from 1500 km to 2500 km, the travel time of EMU trains with sleeping cars is between 8 and 13 hours. Moreover, the average speed of EMU trains with sleeping cars is far greater than that of ordinary passenger trains, while slightly less than that of EMU trains with speed of 300 km/h. Figure 2b shows that the average ticket price of EMU trains with sleeping cars is approximately equal to that of first class seat of EMU trains with speed of 300 km/h, and it is evidently lower than that of business class seat of EMU trains with speed of 300 km/h, while it is obviously higher than that of EMU trains with speed of 250 km/h and ordinary passenger trains. Although the comfort degree of different passenger trains for passengers is difficult to quantify, it is generally agreed that the traveling comfort by seat is much lower than that of sleeping car when travel time exceeds 8 hours. The EMU trains with sleeping cars can run more than 2000 kilometers at night with higher long-distance service quality [5].

According to the operation status of EMU trains with sleeping cars, they can enhance the service quality and transport capacity of railway passenger transportation. In addition, they help to optimize the product structure of passenger transportation, as well as promote the development of sustainable transportation.

Compared with the developed regions in East China, high-speed rail construction and operation in Northwest China is obviously lagging behind. The Xi'an–Baoji high-speed railway, the Lanzhou–Urumqi passenger dedicated line, and the Baoji–Lanzhou passenger dedicated line opened, respectively, in December 2013, December 2014, and July 2017, so the entire high-speed railway in the Northwest region has been connected (as shown in Figure 3). The distance of the entire line from Xi'an to Urumqi is approximately 2500 kilometers, while the running time of the EMU train is approximately 13–15 hours, and it objectively has the capability to operate an EMU train with sleeping cars [6].



Figure 3. High-speed railways in Northwest China.

Given the nature of various considerations in Northwest China, such as economic factors, consumption behavior, culture, geography, and so forth, it is generally believed that the passenger flow is insufficient and the consumption capacity is on the low side, so the demand for high-level passenger transportation products would be low in these areas. Therefore, to predict passengers' demand characteristics, this paper is dedicated to data investigation, modeling and analysis of passengers' inclination to use EMU trains with sleeping cars in Northwest China.

2. Literature Review

Passenger demand is the basis of product design and organization for passenger transportation. In the past several decades, various attempts have been made to obtain the demand characteristics of

passengers. These attempts primarily include the following aspects: analysis of passenger demand characteristics, investigation and study of the travel satisfaction of passengers, and design of passenger transport products based on passenger demands. In terms of research methods, there is not only statistical analysis of actual survey data but also theoretical modeling.

In the 1970s, research on passenger demand was carried out. Ben-Akiva systematically studied the structure of passenger travel demand using a theoretical reasoning method and empirical analysis [7]. This is the representative achievement of the early studies of passenger travel demand, and it provides a good reference for the following research on travel behavior of passengers. Subsequently, passenger demand characteristics of various transport modes were studied, such as railways [1,8], highways [9], public transport [10,11], and air transport [12]. For instance, Owen and Phillips analyzed the travel demand characteristics of railway passengers based on British Rail's monthly ticket data and propose a demand function considering responses to changes in economic variables [1]. Buehler and Pucher compared the public transport demand in America with that in Germany using several indexes, and their conclusion is that public transport in Germany developed more successfully [10].

With the rapid development of high-speed railways, an increasing number of researchers have focused on the travel demand characteristics of high-speed railway passengers. Gunn et al. analyze potential passenger transport markets of high-speed railways in Australia based on investigation, and the results provide a reference for the government to decide whether to construct high-speed railways [13]. Hsiao and Yang collect the data of students in a university of Northern Taiwan and then study their willingness to travel by high-speed railway trains [14]. Moreover, there are also some achievements reporting the impact of high-speed railway services on aviation demand [15,16] and on the tourism market [17,18].

The travel satisfaction of passengers often reflects the degree of matching between passengers' demand attributes and transport supply attributes, and it is commonly measured by the comprehensive quality of passenger service. Shen et al. consider the evaluation method for passenger satisfaction for urban rail transit and establish an evaluation model based on structural equation modeling [19]. The results show that ticket price, information distribution, safety and staff service are the most crucial factors affecting passenger satisfaction. According to nearly half a million data records of travel satisfaction, Abenoza et al. analyze the satisfaction of travelers with public transport in Sweden, and they obtain the service attributes that have the greatest impact on passenger satisfaction [20]. Regarding the satisfaction of railway passengers, Aydin et al. adopt a fuzzy evaluation method to evaluate the level of passenger satisfaction based on a massive amount of survey data, and the results can provide recommendations not only for future investment but also for the improvement of rail transit operation [21,22]. Chou et al. studied relationships among customer loyalty, service quality and customer satisfaction of high-speed railways in Taiwan using a structural equation modeling method [23]. The results show that there is a positive correlation between customer loyalty and service quality, while the relationship between customer loyalty and customer satisfaction is also positive.

Passenger demand data are often obtained by applying the stated preference (SP) survey method and the revealed preference (RP) survey method [24,25]. The results of passenger demand characteristics are usually used to predict future passenger demand [26] or design better passenger transport products [27–30]. EMU trains with sleeping cars are one of the passenger transport products on high-speed railways, and this kind of train currently operates in China. Thus, related studies are mainly from Chinese scholars. Zhang analyzes the potential passenger market of EMU trains with sleeping cars and then proposes several marketing strategies for this kind of train [4]. An (2016) discusses the pricing strategy of an EMU train with sleeping cars under market-oriented conditions to improve the railway's competitiveness [5]. By analyzing the determinants that influence the operation of EMU trains with sleeping cars, Zhang et al. propose that the reasonable operation distance of EMU trains with sleeping cars should be within 2400 kilometers, while the time value of their major customers should be under 50 Yuan/h [6].

These studies provided useful references for the analysis of passenger demand characteristics and the design of passenger transport products. Most present studies are aimed at traditional passenger transport modes, such as railways, aviation, and urban rail transit. Regrettably, research on the demand characteristics for new passenger transport products is scarce, especially for high-speed railways. EMU trains with sleeping cars have operated for several years on high-speed railways in central and eastern China, and good operation effects have been achieved. Whether there are corresponding passenger demands in Northwest China is worth studying. Thus, based on a survey, this paper studies the passenger demand characteristics of EMU trains with sleeping cars in Northwest China using factor analysis and regression analysis methods.

The remainder of this paper is organized as follows. In Section 3, statistical analysis of the survey data is introduced. Section 4 states the modeling process of passenger choice intention for EMU trains with sleeping cars, and includes three parts: a factor analysis, multinomial logit modeling, and an ordinal logistic regression. Parameter calibration of the model is carried out in Section 5, and then, the analysis of results is also given in this section. Section 6 provides some conclusions and discussions.

3. Statistical Analysis of the Data

The purpose of this study is to reveal the preferences of different passengers regarding EMU trains with sleeping cars. In Northwest China, taking the travelling group on a data survey day as the research object, there are 5 kinds of transportation modes that can meet the travel distance of about 2000 km, which are air travel, high-speed railway, ordinary railway, long-distance bus, and self-driving travel. We have investigated the first three modes, and abandoned the other modes. The reason is that in the northwestern part of China, high-speed railway has not yet been opened in areas accessible by long-distance buses, and the passenger volume of long-distance buses between cities of 2000 km is very small. The self-driving passengers are not potential passenger group of EMU train with sleeping cars, and the volume is fewer. According to the actual situation, the sampling survey of passenger flows using a one-to-one intentional survey method was conducted by research assistants at major railway stations and airports in the study area including Xi'an Xianyang International Airport, Xi'an Railway Station, Lanzhou Zhongchuan International Airport, Lanzhou West Railway Station and Lanzhou Railway Station in September 2017. The research assistants randomly selected respondents from passengers who were either waiting for a train or a flight in waiting room or departure lounge to guarantee the representativeness and randomness of the samples. In the process of investigation, several railway stations and airports were selected to reflect different places. Sometimes, surveys were conducted on different dates to reflect different times, and the respondents were from several trains and flights to reflect different trains (flights).

In the process of the data survey, we clearly informed passengers that travel by EMU trains with sleeping cars was safer and more comfortable and could achieve sunset departure and sunrise arrival, but the ticket price is higher. The proportion of questionnaires distributed to passengers of different transportation modes was roughly the same as the proportion of passengers carried by various modes on the survey day. When choosing respondents within each transportation mode, we selected them randomly according to the factors that can be taken into account at different times, different places and different trains (flights) to ensure the representativeness of the sample data as far as possible.

In the survey (see Appendix A), 3500 questionnaires were distributed, and 3005 questionnaires were returned, 2966 of which were valid questionnaires. Among them, there are 1738 questionnaires from ordinary railway passengers, 58.6%; 540 questionnaires from high-speed railway passengers, 18.2%; and 688 questionnaires from aviation passengers, 23.2%.

The questionnaire contains both RP and SP survey questions. The RP questions mainly involve some personal attributes of the respondents, including gender, age, educational background, profession, income, and place of residence. SP questions are willingness surveys, and they are mainly used to obtain data that cannot be directly observed. The potential impact of travel cost, travel distance, speed and convenience on the travel mode choice of respondents was further studied through the subjective

preference selection of multiple scenarios under hypothetical conditions. According to the statistical analysis of the data, the general characteristics of the sample are shown in Table 1.

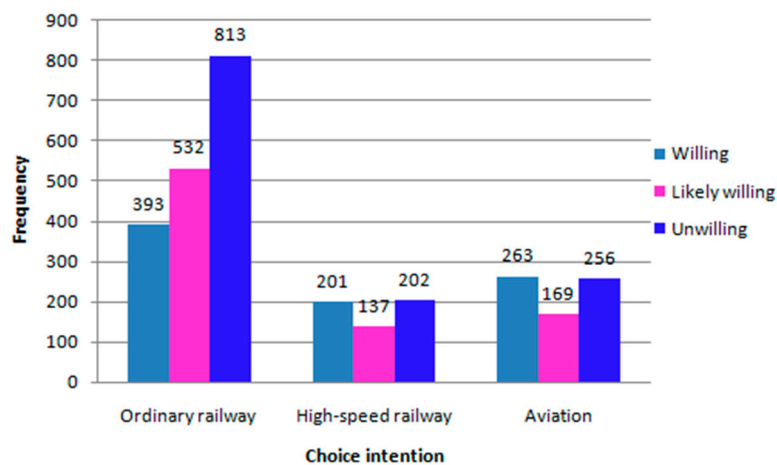
Table 1. Characteristics of survey respondents.

Variable	Classification	Ordinary Railway		High-Speed Railway		Aviation	
		Frequency (#)	Proportion(%)	Frequency (#)	Proportion (%)	Frequency (#)	Proportion (%)
Gender (x_1)	1. Male	1216	70.0	372	68.9	470	68.3
	2. Female	522	30.0	168	31.1	218	31.7
Age (x_2)	1. Less than 26 years old	380	21.9	72	13.3	108	15.7
	2. 26–35 years old	572	32.9	226	41.9	244	35.5
	3. 36–45 years old	410	23.6	132	24.4	166	24.1
	4. 46–60 years old	320	18.4	98	18.1	158	23.0
	5. Over 60 years old	56	3.2	12	2.2	12	1.7
Educational background (x_3)	1. Junior middle school	510	29.3	52	9.6	104	15.1
	2. Senior middle school	500	28.8	120	22.2	110	16.0
	3. Junior college	330	19.0	158	29.3	152	22.1
	4. Undergraduate course	358	20.6	186	34.4	268	39.0
	5. Postgraduate	40	2.3	24	4.4	54	7.8
Profession (x_4)	1. Student	156	9.0	20	3.7	30	4.4
	2. Migrant worker	592	34.1	60	11.1	60	8.7
	3. Staff of private enterprise	218	12.5	128	23.7	124	18.0
	4. Public functionary	208	12.0	128	23.7	122	17.7
	5. Staff of state owned enterprise	218	12.5	88	16.3	122	17.7
	6. Self-employed person	214	12.3	86	15.9	158	23.0
	7. Other	132	7.6	30	5.6	72	10.5
Income (x_5)	1. Less than 3000 yuan	468	26.9	84	15.6	64	9.3
	2. 3000 to 4000 yuan	626	36.0	162	30.0	166	24.1
	3. 4001 to 6000 yuan	380	21.9	146	27.0	200	29.1
	4. 6001 to 10000 yuan	198	11.4	104	19.3	150	21.8
	5. More than 10000 yuan	66	3.8	44	8.1	108	15.7
Place of residence (x_6)	1. Shanxi	440	25.3	26	4.8	120	17.4
	2. Xinjiang	208	12.0	28	5.2	176	25.6
	3. Gansu	708	40.7	342	63.3	166	24.1
	4. Qinghai	56	3.2	36	6.7	56	8.1
	5. Beijing, Tianjin, Hebei	6	0.3	16	3.0	22	3.2
	6. Jiangsu, Zhejiang, Shanghai	54	3.1	14	2.6	38	5.5
	7. Pearl River Delta	4	0.2	12	2.2	8	1.2
	8. Other	262	15.1	66	12.2	102	14.8
Trip purpose (x_7)	1. Be away on official business	328	18.9	230	42.6	258	37.5
	2. Tourism	180	10.4	66	12.2	110	16.0
	3. Visit relatives or friends	318	18.3	116	21.5	140	20.3
	4. Commute to work	554	31.9	58	10.7	46	6.7
	5. Conduct business	88	5.1	30	5.6	56	8.1
	6. Other	270	15.5	40	7.4	78	11.3

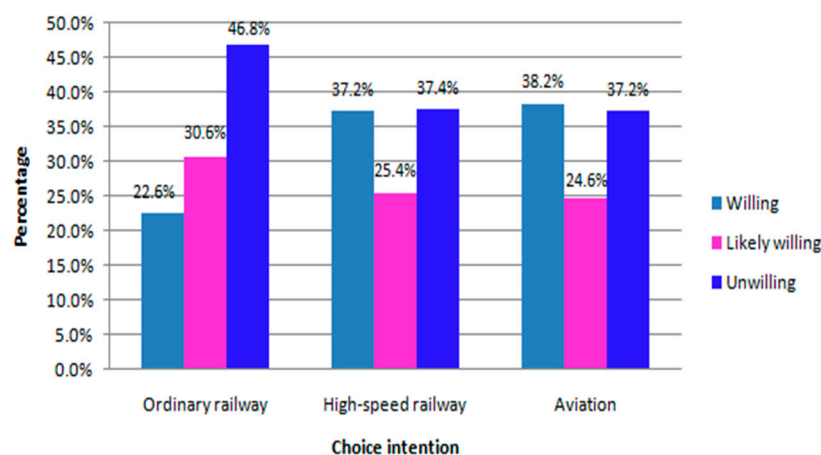
From Table 1, one notable characteristic is that the proportion of male to female respondents is extremely unbalanced as male passengers are more than twice of female passengers. In fact, the survey data is approximately consistent with real statistical data. Another obvious characteristic is that respondents living in Northwest China are far more than those living in other areas, this is also similar to real data.

The choice intentions for EMU trains with sleeping cars are investigated among passengers of ordinary railways, high-speed railways, and aviation. The passenger choice intention levels for the EMU trains with sleeping cars are expressed by three grades: willing, perhaps willing, and unwilling. The corresponding frequency and percentage analyses are shown in Figure 4. According to the analysis, the ratio of those who are willing to travel by EMU trains with sleeping cars and those who are not willing to travel by EMU trains with sleeping cars is 0.482 among ordinary railway passengers, while the ratios of high-speed railway passengers and aviation passengers are 0.995 and 1.027, respectively.

The result shows that the proportion of ordinary railway passengers who are willing to travel by EMU trains with sleeping cars is the lowest, while the proportions of high-speed railway passengers and air passengers are similar and far higher than that of ordinary railway passengers.



(a) Frequency analysis



(b) Percentage analysis

Figure 4. Passenger choice intention for electrical multiple unit (EMU) trains with sleeping cars.

4. Mathematical Models

4.1. Problem Analysis

To analyze the quantitative relationship between passenger willingness to travel by EMU trains with sleeping cars and their personal attributes, regression models are considered for quantitative analysis and prediction. The methods of the binary logit model (BL), multiple logit model (MLN), nested logit model (NL) and mixed logit model (Mixed Logit) are often adopted to research the choice of transportation mode and travel satisfaction [31,32]. If the object of study is transformed from the probability of selecting an event to the ratio of the probability of selecting an event and the corresponding probability of not selecting an event, which is also known as the odds ratio, then the corresponding logistic regression model can be obtained. Furthermore, the numerical relationship between dependent variables and independent variables can be quantitatively analyzed and predicted.

However, if a logistic model is used for regression analysis, there can be no multiple collinearity among the corresponding independent variables. Otherwise, the variance and covariance of the parameter estimates will increase, and the accuracy of the regression analysis will be affected in more severe cases. As for the survey data of the passenger choice intention for EMU trains with sleeping

cars, we may discover that there is significant multicollinearity among the passenger's multi-attribute data by testing, so processing of the data is required.

Therefore, the factor analysis can be considered to reveal the inherent common factors and special factors among the diverse attributes of many passengers. In other words, the construction of the factor model decomposes much of the original observed variable into a linear combination of a few factors, and these new factors are orthogonal. That is, there is no multicollinearity among the factors, and the total number of new factors is less than the original variable data dimension. More importantly, the extracted factors reflect the essential characteristic attributes of passenger decision-making more simply and more directly.

Meanwhile, because the dependent variables are ordered as multiclass variables, an ordinal regression method is selected to analyze the choice willingness because it is more accurate. Compared with ordinary logistic regression, ordered regression considers the continuous changes of the internal logic among dependent variables, which avoids the irrationality caused by the discretization choice, so it is more suitable to the actual situation of the selection of an EMU train with sleeping cars.

Consequently, focusing on the problem of passenger choice intention for EMU trains with sleeping cars, this paper adopts a factor analysis method to reduce the data dimensionality of possible factors and obtain orthogonal common factors. Then, an ordered regression method is used to perform parameter estimation and quantitative analysis. The technical roadmap of the methodology is shown as Figure 5.

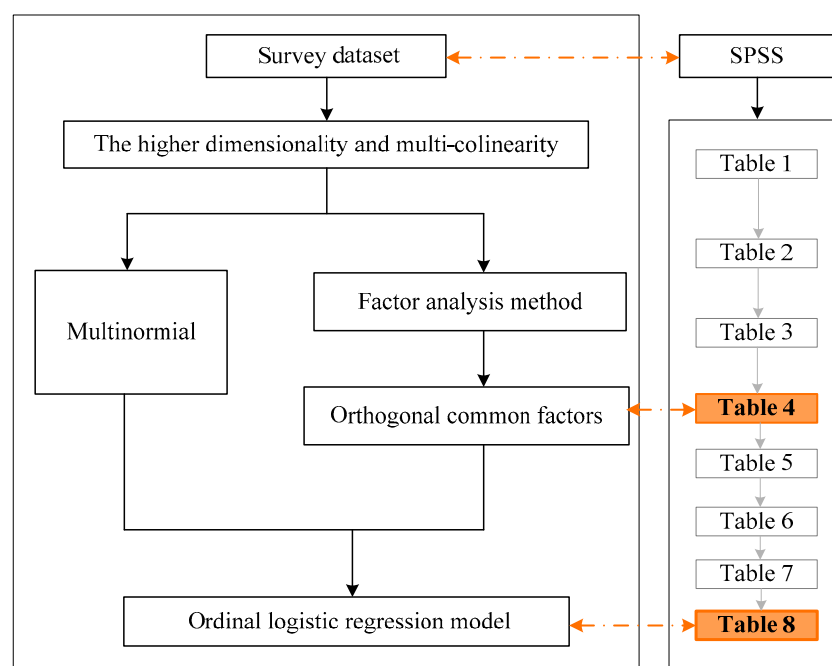


Figure 5. Technical roadmap of the methodology.

4.2. Factor Analysis

If the number of factors that may influence the passenger choice intention for EMU trains with sleeping cars is g , and the data set of all samples is X , then $X = (x_1, x_2, \dots, x_g)$, where x_i ($i = 1, 2, \dots, g$) indicates the i -th column data in the dataset, representing all sample data of the i -th factor in the dataset, and it can be further expressed as $x_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T$. If the number of common factors is m , ($m < g$) and the number of special factors is g , then the vector of common and special factors can be expressed as $f = (f_1, f_2, \dots, f_m)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_g)$, respectively. We can obtain the factor model as follows:

$$\begin{aligned}
x_1 &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 \\
x_2 &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2 \\
&\vdots \\
x_g &= a_{g1}f_1 + a_{g2}f_2 + \cdots + a_{gm}f_m + \varepsilon_g
\end{aligned}$$

In the model, f and ε consist of independent variables, and each f_i is also orthogonal. a_{ij} is called the factor load and represents the load of the i -th factor on the j -th ($j = 1, 2, \dots, m$) common factor, and it reflects the corresponding weight. The specific steps of factor analysis are as follows:

Step 1: All data in the sample set are normalized according to the factor column, and the expression is shown in Equation (1):

$$x'_i = \frac{x_i - \bar{x}_i}{\sigma_i} \quad (1)$$

Among them, \bar{x}_i and σ_i represent the mean value and standard deviation of all sample data for the i -th factor, respectively.

Step 2: Calculate the correlation matrix R of the sample, and the expression is as Equation (2).

$$R = \frac{1}{n-1} X \cdot X' \quad (2)$$

Step 3: Calculate the characteristic root and characteristic vector of the correlation matrix R . According to the equation $|R - \lambda I| = 0$, the characteristic root λ_i can be obtained, and we suppose that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let L express the characteristic vector of the correlation matrix R , and l_{ij} express the elements of L . Then, the equation $L \cdot L' = I$ can be obtained.

Step 4: The number of principal factors expressed by m is determined by the cumulative contribution rate. Generally, the value of m is calculated according to the ratio of the sum of the information amount of the selected principal factor, and the total information amount is not less than 85%, that is, $\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i \geq 85\%$.

Step 5: Compute the factor load matrix A . If a_{ij} is an element in matrix A , then $a_{ij} = l_{ij} \sqrt{\lambda_j}$.

Step 6: Determine the factor model, whose matrix expression is $X = Af + \varepsilon$.

Step 7: Estimate the score function of factors, and the matrix expression is shown as follows.

$$f = (A^T A)^{-1} A^T X \quad (3)$$

4.3. Multinomial Logit Model

According to random utility theory, it is assumed that the choice intention of travelers to select EMU trains with sleeping cars is Y , ($Y = k, k = 1, 2, 3$), which represents willing, perhaps willing, and not willing, while the utility value of the different choice intentions is U_k , then $U_k = V_k + \varsigma_k$, where V_k is a fixed item in the utility function and ς_k is random. When ς_k obeys the Gumbel distribution, it is converted to the ordinary logit model. V_k is shown in Equation (4).

$$V_k = M_k + \sum_{s=1}^S \theta_s^k z_s^k \quad (4)$$

In Equation (4), M_k represents the constant term corresponding to the k -th choice in the utility function, z_s^k represents the s -th explanatory variable corresponding to the k -th selection, and θ_s^k represents the parameter value of the s -th explanatory variable corresponding to the k -th travel mode, where $s = 1, 2, \dots, S$.

Based on the multiple logit model, the probability that the traveler chooses intention k is expressed by Equation (5).

$$p(Y = k) = p_k = \frac{\exp(V_k)}{\sum_{k'=1}^3 \exp(V_{k'})} \quad (5)$$

4.4. Ordinal Logistic Regression

As the development of a multiple logistic model, ordered logistic regression is a more accurate regression analysis method aiming to produce ordinal variables, and it requires that the regression dependent variables must be ordered multiclass variables. The problem of passenger choice intention for EMU trains with sleeping cars studied in this paper can be solved by the ordered regression method. Then, a function is defined as follows:

$$H_k(z) = \ln\left(\frac{p(Y \leq k)}{1 - p(Y \leq k)}\right) = -\alpha^k + \sum_{l=1}^S \beta_l^k z_l^k \quad (6)$$

The function $H_k(z)$ is actually a logarithmic transformation of the ratio of the cumulative probability of $(Y \leq k)$ and the cumulative probability of $(Y > k)$, and it forms a linear equation consisting of a set of parameters $(\alpha^k, \beta_1^k, \beta_2^k, \dots, \beta_s^k)$, where, α^k is a constant term and β_l^k is the parameter of the l -th explanatory variable in the linear equation corresponding to the k -th choice intention. Regardless of where the break point of the dependent variable is in the model, the coefficient of each independent variable β_l^k remains unchanged, while the constant term α^k changes.

After determining $H_k(z)$, the probability p_k that the dependent variable Y has a value of k can be obtained, which is shown in Equation (7).

$$p_k = \frac{\exp(H_k)}{1 + \exp(H_k)} - \sum_{l=1}^{k-1} p_l \quad (7)$$

5. Parameter Calibration of the Model

First, the explanatory variables of the passengers' willingness to choose the EMU train with sleeping cars can be divided into two types, which are unordered multiple classified variables and ordered multiple classified variables. Unordered multiple classified variables include place of residence, gender, trip purpose, and so on, while ordered multiple classified variables refer to the results of the factor extraction. To define the model, a dummy variable must be introduced as an unordered multiple classified variable.

In this paper, SPSS20 is used to perform factor analysis and ordered logistic regression analysis on the survey data. There are many extreme values in potential explanatory variables when performing ordered regression, so the Cauchit link function is adopted. A parallel line test in SPSS is needed to determine whether each independent variable has the same effect on the dependent variable in each regression equation, and the test is passed if all effects are the same.

5.1. Results of the Factor Analysis

Four variables, including age, educational background, profession and income, are selected for factor analysis, and these four variables reflect the travel choice characteristics of passengers prominently.

The results of the Kaiser–Meyer–Olkin (KMO) test and Bartlett spherical test are shown in Table 2. The KMO test is used to study the partial correlation among variables; generally, the value should be greater than 0.5, and it is 0.563 for our survey data set. The Sig. value of the Bartlett spherical test is 0.000, and thus is less than 0.01. Thus, the null hypothesis that the correlation matrix is a unit matrix is rejected, that is to say, there is significant correlation between the variables.

Table 2. Results of the Kaiser–Meyer–Olkin (KMO) test and Bartlett sphere test.

Test Method	Result	
Kaiser-Meyer-Olkin test	0.563	
Bartlett sphere test	Approximate chi-square	1167.376
	df	6
	Sig.	0.000

Table 3 shows the variance explained by each common factor and its cumulative sum. It can be seen from the cumulative percentage of the initial eigenvalue column that the cumulative variance explained by the first three common factors is more than 85%, so they can explain the information contained in the original variables well.

Table 3. Interpretation output of variance.

Component	Initial Eigenvalue			Extracting Sum of Squares			Rotation Sum of Squares		
	Total	Percentage of Variance	Cumulative Percentage	Total	Percentage of Variance	Cumulative Percentage	Total	Percentage of Variance	Cumulative Percentage
1	1.659	41.479	41.479	1.659	41.479	41.479	1.347	33.678	33.678
2	1.154	28.842	70.322	1.154	28.842	70.322	1.058	26.457	60.135
3	0.599	14.971	85.293	0.599	14.971	85.293	1.006	25.158	85.293
4	0.588	14.707	100.000						

Note: Extraction method: principal component analysis; Rotation method: orthogonal rotation method with Kaiser standardization.

The coefficient matrix of the factor score is shown in Table 4, so the final factor expressions are given by Equations (8), (9), and (10).

$$f_1 = -0.099x_2 - 0.216x_3 + 0.520x_4 + 0.763x_5 \quad (8)$$

$$f_2 = 0.929x_2 + 0.095x_3 + 0.243x_4 - 0.276x_5 \quad (9)$$

$$f_3 = 0.080x_2 + 1.076x_3 - 0.019x_4 - 0.240x_5 \quad (10)$$

Among them, the common factor f_1 is more representative of profession and income factors, f_2 mainly represents the factor of age, and f_3 primarily represents the factor of educational background. The common factors f_1 , f_2 and f_3 obtained by factor analysis are orthogonal, and can be used as independent variables of an ordinal regression to perform regression analysis.

Table 4. Coefficient matrix of the factor score.

Variable	Factor 1	Factor 2	Factor 3
Age	−0.099	0.929	0.080
Educational background	−0.216	0.095	1.076
Profession	0.520	0.243	−0.019
Income	0.763	−0.276	−0.240

Note: Extraction method: principal component analysis; Rotation method: orthogonal rotation method with Kaiser standardization.

5.2. Results of Ordinal Logistic Regression

An ordinal regression test is carried out taking the passenger's intention to travel by the EMU train with sleeping cars as dependent variables while taking f_1 , f_3 and place of residence x_6 as regression independent variables.

The fitting information of the model and the result of the fitting degree for the data model are shown in Tables 5 and 6, respectively. Table 5 shows that the significance value (Sig.) of the chi-square

test for the ordered regression model is 0.000 and is far less than 0.01, which means that the final model is well established. Table 6 shows that the significance values (Sig.) of the Pearson statistics and deviation statistics are both 0.000 for the fitting degree test, which indicates that the fitting degree of the model is good.

Table 5. Fitting information of the model.

Model	−2 Log Likelihood	Chi-square	df	Significance
Intercept only	4650.626			
Final	3256.907	1393.719	9	0.000

Table 6. Result of fitting degree for data model.

Statistic	Chi-square	Df	Significance
Pearson	2390.179	1595	0.000
Deviation	2403.311	1595	0.000

Link function: Cauchit.

From Table 7, we can see that the significance value (Sig.) is 0.072, which is greater than 0.05, and the parallel line test is passed. This indicates that the regression equations are parallel to each other, in other words, each independent variable has the same effect on the dependent variable in each regression equation.

Table 7. Result of parallel test.

Model	−2 Log Likelihood	Chi-square	df	Significance
Null hypothesis	3256.907			
Generalized	3241.123	15.784	9	0.072

Note: Null hypothesis specifies that the location parameters are the same in all response categories. Link function: Cauchit.

The final results of regression analysis are shown in Table 8; where, position represents the corresponding relationship between the location and name of regression parameters, FAC1_2 and FAC3_2 represent the clustered factor 1 and factor 2 respectively, and A9 represents the classification variable x_6 .

Table 8. Value of parameter estimation.

		Estimation	Standard Error	Wald	df	Significance	95% Confidence Interval	
							Lower Limit	Upper Limit
Threshold	[A11 = 1]	−1.691	0.120	197.041	1	0.000	−1.927	−1.454
	[A11 = 2]	0.282	0.105	7.211	1	0.007	0.076	0.488
Position	FAC1_2	−1.446	0.067	466.403	1	0.000	−1.577	−1.315
	FAC3_2	−0.641	0.045	201.067	1	0.000	−0.730	−0.553
	[A9 = 1]	−0.202	0.137	2.191	1	0.139	−0.470	0.065
	[A9 = 2]	−0.323	0.144	5.032	1	0.025	−0.605	−0.041
	[A9 = 3]	−0.353	0.118	8.908	1	0.003	−0.584	−0.121
	[A9 = 4]	−0.369	0.192	3.697	1	0.055	−0.745	0.007
	[A9 = 5]	−0.313	0.327	0.918	1	0.338	−0.954	0.328
	[A9 = 6]	−0.040	0.225	0.031	1	0.859	−0.481	0.401
	[A9 = 7]	−0.216	0.421	0.263	1	0.608	−1.040	0.609
	[A9 = 8]	0 ^a	.	.	0	.	.	.

Link function: Cauchit. a. The parameter is redundant, so the value is set to 0.

According to Table 8, we obtain the choice intention model of passengers for EMU trains with sleeping cars, which is shown as follows.

$$H_1 = -1.691 - 1.446f_1 - 0.641f_3 - 0.202x_{61} - 0.323x_{62} - 0.353x_{63} - 0.369x_{64} - 0.313x_{65} - 0.040x_{66} - 0.216x_{67} \quad (11)$$

$$H_2 = 0.282 - 1.446f_1 - 0.641f_3 - 0.202x_{61} - 0.323x_{62} - 0.353x_{63} - 0.369x_{64} - 0.313x_{65} - 0.040x_{66} - 0.216x_{67} \quad (12)$$

In formula (11) and (12), H_1 and H_2 respectively represent the upper and lower bounds of regression results. They are regression results expressed by mathematical expressions, indicating the degree of influence of different factors on the choice willingness of EMU train with sleeping cars.

To analyze the relationship between passengers' choice intention for EMU trains with sleeping cars and the characteristics of passengers' personal attributes more clearly, Equations (13) and (14) are obtained according to Equations (8), (9), (10), (11), and (12).

$$H_1 = -1.691 + 0.092x_2 - 0.377x_3 - 0.740x_4 - 0.949x_5 - 0.202x_{61} - 0.323x_{62} - 0.353x_{63} - 0.369x_{64} - 0.313x_{65} - 0.040x_{66} - 0.216x_{67} \quad (13)$$

$$H_2 = 0.282 + 0.092x_2 - 0.377x_3 - 0.740x_4 - 0.949x_5 - 0.202x_{61} - 0.323x_{62} - 0.353x_{63} - 0.369x_{64} - 0.313x_{65} - 0.040x_{66} - 0.216x_{67} \quad (14)$$

5.3. Result Analysis

From the level of significance, we can see that f_1 (mainly representing profession and income) and f_3 (mainly representing educational background) are significant factors influencing the choice of passengers for EMU trains with sleeping cars. Meanwhile, passengers from different places of residence also have different choice intentions for EMU trains with sleeping cars. However, the factor f_2 which mainly represents age is not a significant factor in choosing EMU trains with sleeping cars.

Each influencing factor affects the dependent variable to a different degree. The coefficients corresponding to each influencing factor reflect the degree of their influence on the acceptance of EMU trains with sleeping cars by passengers, while the sign of coefficient values represents the changing trend of the probability of passengers accepting EMU trains with sleeping cars with the influencing factors. When the sign of the corresponding coefficient value of an influencing factor is positive, it shows that with the increase of the value of this variable, a passenger's attraction to EMU trains with sleeping cars will gradually decrease, that is to say, passengers are more reluctant to choose this kind of product. Conversely, a decrease indicates that passengers are more willing to choose this kind of product.

The following results can be concluded according to formulas (13) and (14).

- (1) The major factors influencing the choice intention of EMU trains with sleeping cars are successively income, profession, educational background, place of residence and age.
- (2) Passengers who have higher income tend to choose the EMU trains with sleeping cars.
- (3) The choice tendency for EMU trains with sleeping cars gradually, successively increase for passengers whose professions are student, migrant worker, staff of private enterprise, public functionary, staff of state-owned enterprise, and self-employed person.
- (4) Passengers from different places of residence have different understandings of and inclinations to choose EMU trains with sleeping cars. Passengers from Xinjiang, Gansu, Qinghai, as well as from Beijing, Tianjin and Hebei look forward to EMU trains with sleeping cars. Meanwhile, passengers from Shanxi and the Pearl River Delta region also show a certain willingness to use the product. Nevertheless, passengers from Jiangsu, Zhejiang and Shanghai have a lower propensity for this product, and they show greater interest in aviation.
- (5) Passengers with a higher education level are more inclined to choose EMU trains with sleeping cars. However, the degree of inclination is lower than that of the income factor.

6. Discussion and Conclusions

An EMU train with sleeping cars is a new type of passenger transportation product in the long-distance passenger transport market which adopts electric traction. Compared with air transport, it has the advantages of large transport capacity, low unit energy consumption, low environmental pollution, and so on. When conditions permit, the operation of an EMU train with sleeping cars can optimize passenger transport structure and promote sustainable transport development. Through experimental data analysis, it can be seen that passengers traveling in Northwest China have a certain demand for EMU trains with sleeping cars; meanwhile, passengers with different incomes or different educational backgrounds or different places of residence have a different level of willingness to use this kind of train. In this paper, factor analysis and ordered logistic regression are adopted to quantitatively analyze the relationship between passenger choice intention for EMU trains with sleeping cars and the individual attributes of the passengers, and the analysis results can be used to predict the tendency of passengers' choice inclination for EMU trains with sleeping cars based on individual attributes.

To address the problem that the survey data cannot be directly subjected to regression analysis, we propose a new solution approach. First, a factor analysis is used to reduce the dimension of multidimensional data, and a few orthogonalized common factors are obtained. Then, the ordinal regression method is used to analyze and predict the passenger's choice intention for the EMU trains with sleeping cars. From the regression results of the data set in this paper, it can be seen that the fitting degree and regression accuracy of this method are higher, and it can better address this type of problem.

According to our research, the following conclusions can be obtained:

- (1) The characteristics of potential passenger groups are obvious, and the proportion of passengers being willing and likely willing to select EMU trains with sleeping cars is generally large.
- (2) The high-income group, highly educated group, government and institutional staff, and individual and private owners are potential customers of EMU trains with sleeping cars in the future in Northwest China.
- (3) The expectation of local passengers for EMU trains with sleeping cars is significantly higher than that of other passengers.
- (4) The ticket price of an EMU train with sleeping cars is the primary factor that passengers are concerned about. Furthermore, travel purpose, source of travel expenses, departure and arrival time, travel destinations, and convenience in reaching the high-speed rail station are all factors that will influence a passenger's choice of EMU trains with sleeping cars.

Moreover, according to the result of demand analysis and forecast, compared with the actual data of passenger flow between Xi'an and Urumqi, it is estimated that the direct passenger flow from Xi'an, Lanzhou to Urumqi will be at least 1000 persons every day. So we conclude that the passenger conditions can meet the requirement of operating EMU trains with sleeping cars between Xi'an and Urumqi.

The result can provide a reference for decision making regarding operating EMU train with sleeping cars in Northwest China. Meanwhile, the proposed method can be applied to the analysis of passenger-flow characteristics in similar areas. How the three factors, namely, the ticket price of the EMU train with sleeping cars, the degree of convenience of travel and the quality of travel service affect whether passengers choose the EMU train with sleeping cars, is further research work that we need to carry out in the future.

Author Contributions: All the authors have contributed a lot to this work. Y.Z. and J.W. contributed to the project administration, conceptualization, formal analysis, methodology, funding acquisition, original draft preparation, review, and editing. W.C. contributed to software, validation, investigation, and visualization. All authors approve the final manuscript.

Funding: This research was funded by the National Natural Science Foundation Project of China, Grant No. 71761025, and was also supported by the State Social Science Fund Project of China under Grant No. 14XGL011.

Acknowledgments: Thanks to the whole team of passenger-flow investigation for their hard word. Thanks to Philip Kofi Alimo and Huanhuan Lv, School of Traffic and Transportation, Lanzhou Jiaotong University, for their help in proofreading. Special thanks to Shiwei Li, School of Traffic and Transportation, Lanzhou Jiaotong University, for his help in investigation and data processing. The authors would also like to thank the reviewers and editors for their nice work on this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Questions of stated preference (SP) survey for passengers of high-speed railway:

1. What is your gender? ()
 A. Male B. Female
2. What is your age group? ()
 A. Less than 26 years old B. 26–35 years old C. 36–45 years old
 D. 46–60 years old E. Over 60 years old
3. What's your educational background? ()
 A. Junior middle school B. Senior middle school C. Junior college
 D. Undergraduate course E. Postgraduate
4. What's your occupation? ()
 A. Student B. Public functionary C. Staff of state owned enterprise
 D. Staff of private enterprise E. Self-employed person
 F. Migrant worker G. Other
5. What's your monthly income (RMB)? ()
 A. Less than 3000 Yuan B. 3000 to 4000 Yuan C. 4001 to 6000 Yuan
 D. 6001 to 10000 Yuan E. More than 10000 Yuan
6. What's your current place of residence? ()
 A. Shanxi B. Xinjiang C. Gansu D. Qinghai E. Beijing, Tianjin, Hebei
 F. Jiangsu, Zhejiang, Shanghai G. Pearl River Delta H. Other
7. What's your purpose for this trip? ()
 A. Be away on official business B. Tourism C. Visit relatives or friends
 D. Commute to work E. Conduct business F. Other
8. Which factor will make you switch to air travel in the future? ()
 A. Lower air ticket price B. Longer travel distance
 C. More convenient and quicker to airport D. Better air service quality
9. "An EMU train with sleeping cars" is running on high-speed railway, and operates with a sunset departure and a sunrise arrival. It is safer and more comfortable than an ordinary train, but with a higher ticket price. Would you like to take the EMU train with sleeping cars when you travel long distances?
 A. Willing B. Likely willing C. Unwilling
10. What's your trip destination this time? (Mark "√" on the corresponding option)

Serial number	1	2	3	4	5	6
City	Xi'an	Baoji	Tianshui	Lanzhou	Wuwei	Xi'ning
Destination						
Serial number	7	8	9	10	11	12
City	Zhangye	Jiuquan (Jiayuguan)	Hami	Turpan	Urumqi	Other regions in Xinjiang
Destination						

References

- Owen, A.D.; Phillips, G.D.A. The characteristics of railway passenger demand: An econometric investigation. *J. Transp. Econ. Policy* **1987**, *21*, 231–253.
- Zhao, J.; Zhao, Y.; Li, Y. The variation in the value of travel-time savings and the dilemma of high-speed rail in China. *Transp. Res. Part A Policy Pract.* **2015**, *82*, 130–140. [\[CrossRef\]](#)
- Albalade, D.; Fageda, X. High speed rail and tourism: Empirical evidence from Spain. *Transp. Res. Part A Policy Pract.* **2016**, *85*, 174–185. [\[CrossRef\]](#)
- Zhang, T.W.; Yang, X.X.; Dong, Z.S. Demand analysis of sunset-departure and sunrise-arrival train on high-speed railway. *Railw. Transp. Econ.* **2017**, *39*, 61–65.
- An, Z.W. Study on strategy of market-oriented pricing system of high-speed railway emu with sleeping cars. *Railw. Transp. Econ.* **2016**, *38*, 5–9.
- Zhang, T.W.; Nie, L.; He, Z.H. Optimization model of organization modes choice of sunset-departure and sunrise-arrival train based on simultaneous integrated maintain time on high-speed railway. *J. China Railw. Soc.* **2015**, *37*, 1–9.
- Ben-Akiva, M.E. Structure of Passenger Travel Demand Models. Ph.D Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1973.
- Wang, Y.; Han, B.M.; Wang, J.K. A passenger flow routing model for high-speed railway network in different transportation organization modes. *Promet Traffic Transp.* **2018**, *30*, 671–682. [\[CrossRef\]](#)
- Wang, J.; Zhou, W.; Li, S.; Shan, D. Impact of personalised route recommendation in the cooperation vehicle-infrastructure systems on the network traffic flow evolution. *J. Simul.* **2018**, 1–15. [\[CrossRef\]](#)
- Buehler, R.; Pucher, J. Demand for public transport in Germany and the USA: An analysis of rider characteristics. *Transport. Rev.* **2012**, *32*, 541–567. [\[CrossRef\]](#)
- Liang, Q.; Weng, J.; Zhou, W.; Santamaria, S.B.; Ma, J.; Rong, M. Individual travel behavior modeling of public transport passenger based on graph construction. *J. Adv. Transp.* **2018**. [\[CrossRef\]](#)
- Dantas, T.M.; Oliveira, F.L.C.; Repolho, H.M.V. Air transportation demand forecast through bagging holt winters methods. *J. Air Transp. Manag.* **2017**, *59*, 116–123. [\[CrossRef\]](#)
- Gunn, H.F.; Bradley, M.A.; Hensher, D.A. High speed rail market projection: Survey design and analysis. *Transportation* **1992**, *19*, 117–139. [\[CrossRef\]](#)
- Hsiao, C.H.; Yang, C. Predicting the travel intention to take high speed rail among college students. *Transp. Res. Part F Traffic Psychol. Behav.* **2010**, *13*, 277–287. [\[CrossRef\]](#)
- Fu, X.; Zhang, A.; Lei, Z. Will China's airline industry survive the entry of high-speed rail. *Res. Transp. Econ.* **2012**, *35*, 13–25. [\[CrossRef\]](#)
- Li, H.; Strauss, J.; Lu, L. The impact of high-speed rail on civil aviation in China. *Transp. Policy* **2019**, *74*, 187–200. [\[CrossRef\]](#)
- Pagliara, F.; La Pietra, A.; Gomez, J.; Vassallo, J.M. High speed rail and the tourism market: Evidence from the Madrid case study. *Transp. Policy* **2015**, *37*, 187–194. [\[CrossRef\]](#)
- Chen, Z.; Haynes, K.E. Impact of high-speed rail on international tourism demand in China. *Appl. Econ. Lett.* **2015**, *22*, 57–60. [\[CrossRef\]](#)
- Shen, W.; Xiao, W.; Wang, X. Passenger satisfaction evaluation model for urban rail transit: A structural equation modeling based on partial least squares. *Transp. Policy* **2016**, *46*, 20–31. [\[CrossRef\]](#)
- Abenoza, R.F.; Cats, O.; Susilo, Y.O. Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transp. Res. Part A Policy Pract.* **2017**, *95*, 64–84. [\[CrossRef\]](#)

21. Aydin, N.; Celik, E.; Gumus, A.T. A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. *Transp. Res. Part. A Policy Pract.* **2015**, *77*, 61–81. [[CrossRef](#)]
22. Aydin, N. A fuzzy-based multi-dimensional and multi-period service quality evaluation outline for rail transit systems. *Transport. Policy* **2017**, *55*, 87–98. [[CrossRef](#)]
23. Chou, P.F.; Lu, C.S.; Chang, Y.H. Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan. *Transportmetr. A Transp. Sci.* **2014**, *10*, 917–945. [[CrossRef](#)]
24. De Dios Ortuzar, J.; Iacobelli, A. Mixed modelling of interurban trips by coach and train. *Transp. Res. Part A Policy Pract.* **1998**, *32*, 345–357. [[CrossRef](#)]
25. Ji, Y.; Mishalani, R.G.; McCord, M.R. Transit passenger origin–destination flow estimation: Efficiently combining onboard survey and large automatic passenger count datasets. *Transp. Res. Part. C Emerg. Technol.* **2015**, *58*, 178–192. [[CrossRef](#)]
26. Ke, J.; Zheng, H.; Yang, H.; Chen, X. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transp. Res. Part. C Emerg. Technol.* **2017**, *85*, 591–608. [[CrossRef](#)]
27. Niu, H.; Zhou, X.; Gao, R. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transp. Res. Part. B Methodol.* **2015**, *76*, 117–135. [[CrossRef](#)]
28. Yin, J.; Yang, L.; Tang, T.; Gao, Z.; Ran, B. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. *Transp. Res. Part. B Methodol.* **2017**, *97*, 182–213. [[CrossRef](#)]
29. Robenek, T.; Azadeh, S.S.; Maknoon, Y.; et al. Train timetable design under elastic passenger demand. *Transp. Res. Part. B Methodol.* **2018**, *111*, 19–38. [[CrossRef](#)]
30. Lv, H.; Zhang, Y.; Huang, K.; Yu, X.; Wu, J. An energy-efficient timetable optimization approach in a bi-direction urban rail transit line: A mixed-integer linear programming model. *Energies* **2019**, *12*, 2686. [[CrossRef](#)]
31. Wang, J.; Niu, H. A distributed dynamic route guidance approach based on short-term forecasts in cooperative infrastructure-vehicle systems. *Transp. Res. Part. D Transport. Environ.* **2019**, *66*, 23–34. [[CrossRef](#)]
32. Seyedabrishami, S.; Shafahi, Y. A joint model of destination and mode choice for urban trips: A disaggregate approach. *Transp. Plann. Technol.* **2013**, *36*, 703–721. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).