

Research Article

Cross-Border Travel Behavior Analysis of Hong Kong-Zhuhai-Macao Bridge Using MXL-BMA Model

Yajie Zou ¹, Wanbing Han ¹, Bo Lin ², Bing Wu ¹, Linbo Li ¹, Shubo Wu ¹,
and Malik Muneeb Abid³

¹Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China

²Shanghai Municipal Engineering Design Institute (Group) Co., Ltd., Shanghai 200092, China

³Department of Civil Engineering, Wah Engineering College, University of Wah, Wah 47040, Punjab, Pakistan

Correspondence should be addressed to Shubo Wu; shubowu@tongji.edu.cn

Received 14 February 2023; Revised 12 April 2023; Accepted 17 May 2023; Published 25 May 2023

Academic Editor: Muqing Du

Copyright © 2023 Yajie Zou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Hong Kong-Zhuhai-Macao Bridge (HZMB) is an important transportation facility connecting Hong Kong, Zhuhai, and Macao. Thus, analyzing the characteristics of cross-border behavior becomes crucial for enhancing the smart travel experience of the HZMB. Discrete choice models (e.g., logit models) are commonly used to describe travel mode choice behavior. Multinomial logit (MNL) is subjected to the independence of irrelevant alternatives (IIA) assumption. Nested logit (NL) model does not consider the heterogeneity of travel individuals. Mixed logit (MXL) model can overcome the above limitations, but it may neglect model uncertainty. Therefore, a Bayesian model averaging (BMA) approach is applied to model travel mode choice behavior considering using revealed preference/stated preference (RP/SP) fusion data collected by questionnaires online. A structural equation model (SEM) is adopted to explore the potential relationship between latent variables, and two travel modes (i.e., cross-border bus and cross-border private car) are selected to analyze the cross-border travel mode choice of the HZMB. The results reveal that the MXL-BMA approach can better explain the cross-border travel mode choice behavior. And the transportation modes arriving and departing the HZMB have a significant impact on the travel mode choice of the HZMB. The findings of this study can provide suggestions for designing personalized travel services for travelers across the HZMB.

1. Introduction

The Hong Kong-Zhuhai-Macao Bridge (HZMB) serves as a crucial transportation facility that connects Hong Kong, Zhuhai, and Macao. Analyzing the travel mode choice behavior while crossing the HZMB bridge is of utmost importance for regional traffic planning, management, and control [1–3]. Travel mode choice behavior analysis can explore the cross-border travel preferences of HZMB users and identify the significant factors in travel decisions. Moreover, it explains the internal mechanism of travel demand for cross-border travelers and then guides the travelers to choose a more reliable, efficient, safe, and environmentally friendly cross-border travel mode. These can support the rational utilization of road transportation infrastructure resources [4–7].

The HZMB is a vital connection between Guangdong province, Hong Kong, and Macao. Several studies have analyzed the travel behavior between Hong Kong, Zhuhai, and Macao. Lu et al. [8] applied a nested logit (NL) model to explore the travel behavior characteristics of passengers between Hong Kong and Zhuhai via HZMB. Hu [9] adopted the K-means method to analyze the spatiotemporal characteristics of travelers entering Hong Kong using cell phone data, revealing a significant increase in entering passenger flow during holidays, with tourists predominantly originating from Shenzhen or neighboring cities. Liu and Shi [10] utilized the smart card dataset of Shenzhen Metro to describe the cross-border travel behavior between Hong Kong and Shenzhen and further explore the travel spatiotemporal characteristics.

For travel behavior analysis, the discrete choice models are the common methods, such as logit family models, including multinomial logit (MNL), NL, and mixed logit (MXL). These models aim at maximizing the random utility to represent individual choice behavior [11–13]. Broach et al. [14] proposed a MNL model to explore the travel modes behavior of walking, bike, car, and bus. However, MNL model assumes that the random terms of alternative travel modes follow the Gumbel distribution and neglect the relationship between alternative travel modes. Moreover, it is subjected to the independence of irrelevant alternatives (IIA) assumption. To overcome this issue, NL model explains the relationship between different travel modes via sharing a part of random errors, which is helpful to accurately individual travel behavior [15, 16]. Fan [17] utilized RP/SP fusion data to establish an NL model for travel time and travel mode (i.e., car and bus) choice, aiming to identify the crucial factors influencing travel mode choice behavior. However, NL model does not consider the correlation between datasets and heterogeneity among individuals. MXL model introduces a random coefficient to represent the heterogeneity of travel individual behavior choice [18]. Wu et al. [19] identified the crucial factors on transportation mode choice behavior on urban subway entrance and exit and further revealed the importance of transfer facilities and environment on travel mode choice behavior. Ye et al. [20] adopted MXL model to explore the impact of shared bikes on travel mode and quantified the influence of various factors, such as travel attributes, building environment, and travel characteristics on the willingness to use shared bikes under different scenarios.

Machine learning-based models (e.g., random forest and neural network (NN)) take travel mode choice modeling as a classification problem and have been demonstrated the feasibility of describing travel mode choice behavior [21–23]. Lindner et al. [24] applied artificial neural networks (ANN) and classification trees (CT) to model travel mode choice using multicollinear data with multiple dimensions. Lhéritier et al. [25] presented a nonparametric machine learning model to describe airline itinerary choice behavior considering the nonlinear relationship between attributes of travel mode and individual characteristics. Golshani et al. [26] explored the relative importance of exploratory variables and used ANN to model travel mode and timing. Machine learning-based models output logit models in fitting performance and predictive capability in travel mode choice behavior, but the interpretability is poor [27, 28].

To sum up, the previous studies on cross-border travel behavior analysis of Hong Kong-Zhuhai-Macao mainly focused on analyzing the characteristics of travel behavior (e.g., travel purpose and travel spatiotemporal distribution). Few studies focus on exploring the transportation mode and travel behavior characteristics. Moreover, MXL model is more capable of capturing the heterogeneity among individuals in travel behavior with RP/SP fusion data. However, there is model uncertainty, which may result in poor fitting performance. Bayesian model averaging (BMA) approach was proposed to overcome the model uncertainty for improving model performance in many fields (e.g.,

transportation safety and energy) [29–31]. BMA approach can weigh candidate models and combine the outputs into a new model to handle the model uncertainty, which takes the fitting performance and advantages of candidate models into account. Zhou et al. [32] proposed an ensemble approach, which integrated copula-based Bayesian model averaging (CBMA) and multiple deterministic artificial neural networks (ANN) to accomplish accurate probabilistic PM2.5 prediction tasks.

Therefore, with the RP/SP fusion data, this study proposed an MXL-BMA approach to interpreting the cross-border travel mode choice behavior of HZMB. This approach can solve the model uncertainty of the discrete selection model. The contributions of this study are twofold: (1) the SEM is utilized to explore the causal relationships and mechanisms between latent variables and the corresponding observation variables, as well as among latent variables, which can further support accurately described cross-border travel behavior of HZMB and (2) the MXL-BMA is proposed to analyze cross-border travel mode choice behavior of HZMB, which can handle the model uncertainty and better interpret the cross-border behavior to inform transportation planning and policy making for improving smart services.

The rest of the article is organized as follows. Section 2 describes the data source and processing procedure in detail. The model description is described in Section 3, and Section 4 analyzes the results of travel mode choice behavior. Section 5 briefly concludes the study.

2. Data Collection and Preprocessing

2.1. Sample Size. Due to the influence of COVID-19 which affects travelers between Guangdong province, Hong Kong, and Macao, the cross-border travel behavior dataset from HZMB users traveling from Zhuhai to Hong Kong was collected via an online questionnaire survey. As of October 23, 2019, the average daily passenger trips to three ports via HZMB were approximately 66900. According to the total number of survey respondents, the sample size can be calculated by the following equation:

$$n \approx \frac{Z_{\alpha/2}^2 S^2}{E^2 + (Z_{\alpha/2}^2 S^2)}, \quad (1)$$

where $Z_{\alpha/2}^2$ is the coefficient of reliability, which is set to 1.96 (the confidence interval is set as 0.95, i.e., $\alpha = 0.05$). S^2 is the variance with a value of 0.5, E is the margin of error with a value of 0.05, and N denotes the overall survey respondents with a value of 66900. Therefore, this study acquired 811 valid sample data, which exceeds the minimum sample size of 390.

2.2. Questionnaire Design. Previous studies suggested that the travel mode choice behavior is associated with various socioeconomic attributes of travelers (e.g., gender, age, and education), travel-related factors (e.g., travel purpose and travel time), and psychological perception factors (e.g., safety, reliability, and comfort). Therefore, the questionnaire

of this study includes traveler socioeconomic attributes, revealed preference (RP), and stated preference (SP).

2.2.1. Traveler Socioeconomic Attributes. Table 1 shows the socioeconomic attributes of travelers, including age, gender, education, and so on.

2.2.2. Revealed Preference. The RP survey encompasses various factors, including travel-related factors and psychological perception factors. Travel-related factors are explicit variables because they are easily observable (see Table 2). The psychological perception factors are latent variables that are harder to observe (see Table 3). The corresponding observation variables are utilized to represent latent variables to explore the service perception of travel mode for cross-border travel. These latent variables include reliability, safety, convenience, and comfort, and the corresponding observation variables are quantitatively measured with the five-component scale method.

2.2.3. Stated Preference. To investigate the factors that influence cross-border travel mode choice behavior, various attribute combinations were designed for travelers to choose from. Note that various travel modes and travel times result in varying travel patterns among public transport users. Hence, different observation variables are utilized for different travel modes and six levels are established (see Table 4). According to the uniform design experiment theory, six survey scenarios are generated considering various combinations of variables, as shown in Table 5.

2.3. Data Quality Control. To ensure the data quality collected from the questionnaire survey, the reliability and validity test are conducted. The reliability test utilized Cronbach's alpha coefficient, while the validity test used Kaiser–Meyer–Olkin (KMO) and Bartlett spherical tests. The results show that Cronbach's alpha coefficient is 0.845, which is larger than the threshold value of 0.8, indicating that the data is reliable. Moreover, the KMO value is 0.891, which is greater than 0.8, and the significance level (i.e., p value) of the Bartlett spherical test is less than 0.05, indicating each variable exhibits high independence. To sum up, the collected travel data is suitable for analyzing the travel characteristics of HZMB.

2.4. RP/SP Data Fusion. RP and SP data reflect travel behavior characteristics from different perspectives. RP data capture the actual travel behavior patterns, while SP data provide the information on travel preferences and choices of travelers. However, RP/SP fusion data provide a more comprehensive information for understanding travel behavior. In addition, previous studies demonstrated that more accurate results can be obtained in travel mode choice behavior using RP/SP fusion data rather than using RP data or SP data [33, 34]. Therefore, this study integrates RP and SP data with the help of the panel data method to analyze

cross-border travel behavior of HZMB. In the random utility theory-based discrete choice model, the relationship between random errors in the utility function of RP and SP data can be represented as follows:

$$\text{Var}(\varepsilon_i^{RP}) = \mu^2 \cdot \text{Var}(\varepsilon_i^{SP}), \quad (2)$$

where ε_i^{RP} and ε_i^{SP} denote the random errors of RP data and SP data. μ denotes the measured coefficient, which can be calculated using the random errors obtained from models based on single RP data and SP data.

3. Model Description

This study utilized the SEM model to explore the potential correlation among the latent variables, and then an MXL-BMA model is proposed to interpret the cross-border travel behavior under different travel modes.

3.1. Structural Equation Model. SEM is a statistical method, which combines exploratory factor analysis and path analysis to explain the relationships between latent variables and observation variables, as well as among latent variables, and to explore the causal relationships and mechanisms between multiple variables. It not only can handle the problems (e.g., multiple mediation and multicollinearity) to provide an accurate interpretation of the interrelationships between variables but also can allow measurement error in observation variables [35, 36]. In previous studies, SEM is utilized to investigate the relationship among latent variables that affected travel mode choice behavior, and between latent variables and corresponding observation variables. It should be noted that the observation variables solely serve to measure the latent variables, and do not impact the travel mode choice behavior.

$$\begin{aligned} \eta &= \Gamma\xi + B\eta + \zeta, \\ X &= \Lambda_X\xi + \delta, \\ Y &= \Lambda_Y\eta + \varepsilon, \end{aligned} \quad (3)$$

where ξ and η represent the exogenous latent variables and endogenous latent variables, respectively. B denotes the regression coefficient matrix of relations among endogenous latent variables, Γ denotes the regression coefficient matrix of relations between exogenous latent variables and endogenous latent variables. X and Y are the observation variables of exogenous latent variables and endogenous latent variables and Λ_X and Λ_Y are regression coefficient matrices between exogenous observation variables and exogenous latent variables and among endogenous observation variables and endogenous latent variables. ζ , δ , and ε are the error vectors.

3.2. MXL-BMA Model

3.2.1. Mixed Logit Model. The MXL model [37, 38] assumes that travelers can repeatedly choose travel mode and there is a potential relationship with the utility of unobservable items. In addition, the observation variables are not

TABLE 1: Survey on the socioeconomic attribute of the traveler.

Observation variable	Variable definition
Gender	Male and female
Age	<18; 18–31; 30–40; 40–50; >50
Education level	High school or below (technical secondary school, technical school, and vocational high school); junior college; undergraduate degree; and postgraduate or above
Occupation	Student; employees of enterprises and public institutions; cadres and managers of enterprises and institutions; government workers; individual industrial and commercial enterprises; production personnel in agriculture, forestry, animal husbandry, fishing, and water resources; a freelancer; and others
Annual household income	<100000; 100000–150000; 150000–200000; 200000–250000; 250000–300000; >300000

TABLE 2: Survey on travel-related factors.

Observation variable	Variable definition
Travel purpose	Work; school; visiting friend and relatives; tourism, shopping, and other recreational activities; hospital; business; and others
Transportation mode (origin-Zhuhai)	Bus; taxi; private car; online car-hailing; and others
Travel time (origin-Zhuhai)	<15 min; 15–30 min; 30–45 min; 45–60 min; >60 min
Transportation mode (section of HZMB)	Bus; taxi; private car; and others
Travel time (section of HZMB)	<40 min; 40–50 min; 50–60 min; >60 min
Transportation mode (Hong Kong-destination)	Bus; taxi; private car; online car-hailing; subway; and others
Travel time (Hong Kong-destination)	<15 min; 15–30 min; 30–45 min; 45–60 min; >60 min
The residence time of cross-border	1; 2; 3–5; >5
Number of entourages	0; 1; 2–3; ≥ 4
Whether Hong Kong and Macao travel together	Yes or no
Whether the return travel also crosses the HZMB	Yes or no
Number of transfers	0; 1–2; >3
Annual frequency of cross-border travel	1; 2–10; 11–30; >30

restricted to conforming to normal distribution or being independent. In the MXL model, the coefficients of explanatory variables are the random parameters of a distribution, thus effectively addressing the IIA defect that exists in the traditional MNL model. Let β_q denotes the coefficient of a traveling individual q , which follows a distribution with parameter θ , that is $\beta_q \sim f(\beta_q|\theta)$. Then, the constructed MXL model can approximate any random utility model. The utility function can reflect the travel preference characteristics of individuals and provide a more comprehensive explanation of cross-border travel mode choice behavior. Stochastic utility theory uses the utility function to represent utility value and measure the travel mode choice. For MXL model, the utility value can be calculated as follows:

$$U_{qit} = \beta X_{qit} + \theta'_q \left[(1 - \delta_{qt,RP}) \times \left(\sum_{s=1}^{T_q} \delta_{qt,RP^s} \cdot Y_{qis} \right) \right] + \varepsilon_{qit}, \quad (4)$$

where β is the coefficient of the explanatory variable, θ'_q denotes the estimated parameter of factors affecting state-dependent heterogeneity. $\delta_{qt,RP}$ and Y_{qis} are data binary indicator variables. If the travel choice of travel individual matches the actual choice of the RP survey, $\delta_{qt,RP} = 1$, otherwise, $\delta_{qt,RP} = 0$. If the travel individual q chooses travel mode i under scenario t , $Y_{qis} = 1$, otherwise, $Y_{qis} = 0$.

The random parameter is utilized to capture individual differences in preference, and it is typically modeled using a probability distribution, such as normal distribution, log-normal distribution, uniform distribution, t distribution, and so on. This study selected four latent variables as random parameters and assumed that the parameters follow t distribution. Hence, the utility function can be rewritten as the sum of random utility and error as follows:

$$U_{qit} = V_{qit} + \varepsilon_{qit}, \quad (5)$$

where V_{qit} is the random utility of the utility function and ε_{qit} is the error. The probability of traveler q choosing travel mode i under scenario t can be calculated as follows:

$$p(y_{qit} = i) = \frac{e^{\mu V_{qit}}}{\sum_{j=1}^J e^{\mu V_{qit}}}, \quad (6)$$

where μ denotes the measured coefficient of RP/SP data fusion.

3.2.2. Bayesian Model Averaging. MXL model can avoid the bias from IIA attribute, but it neglects the model uncertainty, that is, it cannot determine whether an explanatory variable should be taken into account when analyzing travel behavior. These result in low accuracy of interpretability. BMA

TABLE 3: Survey on psychological perception factors.

Latent variable	Observation variable	Variable definition
Reliability	Reliability_1	I will consider punctuality when I choose the travel mode
	Reliability_2	If the travel mode is usually delayed, I will choose another travel mode, even if it costs more money (more than 5 CNY)
	Reliability_3	If the travel mode is usually delayed, I will choose another travel mode, even if it costs more time (more than 10 minutes)
Safety	Safety_1	Personal and property safety is guaranteed during the ride
	Safety_2	Low probability of vehicle failure
	Safety_3	The vehicle safety warning message is obvious
Convenience	Convenience_1	It is convenient from the origin to the location of the travel mode
	Convenience_2	The ride time is free
	Convenience_3	There are various forms of ticket purchase during cross-border travel
Comfort	Comfort_1	The vehicle stable driving
	Comfort_2	The vehicle seats are comfortable to ride
	Comfort_3	If the travel mode is not comfortable, I will choose a more comfortable travel mode, even if it costs more money (more than 5 CNY)
	Comfort_4	If the travel mode is not comfortable, I will choose a more comfortable travel mode, even if it costs more time (more than 10 minutes)

TABLE 4: The levels of factors for different travel modes.

	Cross-border private car		Cross-border bus			
	Travel time in HZMB (min ^a)	Travel cost (CNY ^b)	Parking cost (CNY/0.5 h ^c)	Travel time in HZMB (min)	Travel cost (CNY)	Waiting time (min)
42	120		5	50	55	5
45	135		7.5	54	60	10
48	150		10	58	65	15
51	165		12.5	62	70	20
54	180		15	66	75	25
57	195		17.5	70	80	30

Note. ^amin denotes minute; ^bCNY denotes Chinese Yuan; ^ch denotes hour.

TABLE 5: The details of travel time and cost for different travel modes under different scenarios.

Scenario	Cross-border private car			Cross-border bus		
	Travel time in HZMB (min)	Travel cost (CNY)	Parking cost (CNY/0.5 h)	Travel time in HZMB (min)	Travel cost (CNY)	Waiting time (min)
1	42	135	10	62	75	30
2	45	165	17.5	50	65	25
3	48	195	7.5	66	55	20
4	51	120	15	54	80	15
5	54	150	5	70	70	10
6	57	180	12.5	58	60	5

approach chooses a set of candidate models as model space and weights the candidate models with posterior probability to take their fitting performance into account. BMA approach integrates the outputs of candidate models into a new model to depict the travel behavior. Therefore, BMA approach is applied to overcome the model uncertainty of constructed MXL models and improve the interpretability of travel behavior analysis [39, 40].

BMA weights the models in model space M with the help of posterior probability and combines the results into a deterministic model to depict the travel choice behavior. The model space contains K models that can be denoted by $M = \{M_1, M_2, \dots, M_K\}$, which is a set of candidate models. For the given data set D , the posterior probability of quantity of interest y can be represented as follows:

$$p(y|D) = \sum_{k=1}^K p(y|M_k, D)p(M_k|D), \quad (7)$$

where $p(y|M_k, D)$ denotes the posterior probability of y under the candidate model M_k , $p(M_k|D)$ is the posterior

model probability and $\sum_{k=1}^K p(M_k|D) = 1$. According to the Bayes rule, $p(M_k|D)$ can be calculated by the following equation:

$$p(M_k|D) = \frac{p(M_k)p(D|M_k)}{\sum_{i=1}^K p(M_i)p(D|M_i)}, \quad (8)$$

where $p(M_k)$ denotes the prior probability of candidate model M_k when it is regarded as a “true” model and $p(D|M_k)$ is the marginal model likelihood of candidate model M_k , that is,

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k, \quad (9)$$

where θ_k is the parameter vector of model M_k , $p(\theta_k|M_k)$ denotes the prior probability distribution of parameter θ_k under model M_k , and $p(D|\theta_k, M_k)$ is the likelihood under the model M_k with parameter θ_k . Therefore, the posterior mean and variance can be calculated as follows:

$$E[y|D] = \sum_{k=1}^K E[y|D, M_k]p(M_k|D),$$

$$\text{Var}[y|D] = \sum_{k=1}^K (\text{Var}[y|D, M_k] + E[y|D, M_k]^2)p(M_k|D) - E[y|D]^2. \quad (10)$$

One difficulty in implementing BMA approach is how to determine a proper model space. A proper model space not only can save efforts on low-performing models but also improve interpretability. A popular method is Occam’s window, also known as the principle of parsimony [41]. It advocates for selecting the simplest models with the least number of assumptions, which can handle the problems of overfitting, low universality, and low interpretability in complex models. Therefore, Occam’s window is applied to determine a proper model space. It mainly follows these two principles:

First, if the posterior probability of a model is much smaller than the given optimal model, this model should be removed from the model space, as shown in (11), $\max_i \{p(M_i|D)\}$ is the maximum posterior probability of model. The C value is determined according to the actual

situation. After many attempts, the C is set to 20 in this study.

$$A' = \left\{ M_k: \frac{\max_i \{p(M_i|D)\}}{p(M_k|D)} \leq C \right\}. \quad (11)$$

Second, if the posterior probability of complex model is smaller than the sample model, the complex model should not be considered, as shown in equation (12). This principle is used to exclude some complex models and is generally reduced to one or two models.

$$B = \left\{ M_k: \exists M_l \in M, M_l \subset M_k, \frac{p(M_l|D)}{p(M_k|D)} \geq 1 \right\}. \quad (12)$$

To sum up, equation (7) can be rewritten as follows:

$$p(y|D) = \sum_{M_k \in A} p(y|M_k, D)p(M_k|D), \quad (13)$$

where $A = A' \setminus B \in M$.

4. Results

4.1. Latent Variables Analysis. The SEM is utilized to explore the internal relationship between latent variables, as well as latent variables and the corresponding observation variables. The calibration results of SEM are listed in Table 6. It can be observed that the coefficients of latent variables associated with the corresponding observation variables are all greater than 0.85 (the minimum value is 0.858), which indicates that the observation variables have superior capacity in representing the latent variables. Moreover, the absolute values of correlation coefficient between different latent variables exceed 0.5, indicating that there is robust association. Note that the positive values of coefficient denote the positive relationship between latent variables, and the negative values represent the negative relationship. For example, the coefficient between reliability and convenience is -1.397 , indicating that the convenience will decrease when the reliability increases. In addition, all p values are below 0.05, revealing that the calibration results are justifiable.

To further verify the calibration results from SEM, the chi-square (χ^2), the goodness of fit indices (GFI), root mean square of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and incremental fit index (IFI) are selected as evaluation metrics, and the results are listed in Table 7. It can be seen that all the values of evaluation metrics meet the respective threshold, illustrating that the constructed SEM is well-calibrated and can effectively capture the potential relationship between latent variables and observation variables. Note that the threshold of χ^2 varies with the sample size.

4.2. Factors Analysis of Travel Mode Choice. The MXL-BMA model is applied to explain the characteristics of cross-border travel behavior, and Occam's window method is utilized to remove the complex models with poor performance. The results are shown in Table 8. Four high-performance MXL models (denote as model #1, model #2, model #3, and model #4) with posterior probability larger than 0.05 are selected, and the total posterior probability is 96.7%. Notably, model #1 has the highest posterior probability of 56.8%, suggesting the presence of model uncertainty in analyzing travel behavior.

Taking the travel mode of the cross-border taxi as an example, the parameters of travel modes for cross-border bus and cross-border private car are estimated, and the fitting results of the four high-performance models are listed in Tables 9 and 10. It should be mentioned that the variables with a slight impact on travel modes are not listed in the tables. The measured coefficient μ for RP/SP data fusion is calculated by equation (2) under the models based on RP data and SP data. The value is 0.245, which is in the range [0,

1], indicating that the proposed choice models for cross-border travel match the utility maximization principle. Note that the parameter estimates of SP variables are revised values, which are multiplied by the measured equation (12) of RP/SP data fusion.

For socioeconomic attributes of travelers, gender has a significant effect on the cross-border bus. The coefficient is negative, indicating that female travelers are more inclined to choose buses for cross-border travel. Travel purpose shows a noticeable impact on cross-border private car, tourists tend to take cross-border bus (the coefficient is negative). The transportation mode from origin to Zhuhai plays a salient role in cross-border travel, with the choice behavior of bus and private car being particularly pronounced. Specifically, the coefficient for the transportation mode of bus in cross-border bus travel mode is 4.272, while that of private car in cross-border private car travel mode is 5.075. For the transportation mode from Hong Kong to destination, the travelers who prefer to take the bus or subway are more inclined to choose cross-border bus travel mode, which is consistent with the transportation mode from origin to Zhuhai. As for the travel mode of cross-border private car, the choice of transportation mode is consistent with the last distance (i.e., from origin to Zhuhai), which tends to take a private car for cross-border travel. In addition, the travel time from Hong Kong to destination shows a more significant influence on travel mode, particularly for cross-border private car travel mode, with a coefficient of 0.238, suggesting that time-conscious travelers intend to consider using a private car for cross-border travel. Moreover, whether traveling together between Hong Kong and Macao and the number of transfers significantly affect the choice of cross-border bus travel mode.

For the variables of the travel scenario, the travel time, travel cost, and waiting time show a significant impact on travel mode choice behavior, while having a negligible effect on parking time (or parking cost). The coefficient of travel time is -0.149 , indicating a more pronounced impact on travel mode choice behavior than travel time and waiting time. This suggests that travelers are more sensitive to travel time when deciding on travel mode. Additionally, the travel scenario preference reflects the dependence effect with a coefficient of $-0.894 < 0$. This illustrates a noticeable difference in travel behavior between the SP and RP scenarios, implying that travelers tend to seek alternative cross-border travel modes when the services and policies of the current travel mode vary.

For latent variables, safety, comfort, and convenience have a noteworthy impact on cross-border travel choice behavior. Furthermore, the coefficient of safety for travel mode of cross-border bus is positive, indicating that travelers pay more attention to safety for cross-border travel and tend to take buses. In addition, the coefficients of comfort are positive for both travel modes of cross-border bus and cross-border private car, with a higher coefficient value for cross-border private car. This reflects that travelers who prioritize comfort tend to choose private cars for cross-border travel. The variable of convenience has a significant impact on all travel modes, as supported by both negative coefficients. This

TABLE 6: Calibration results of SEM.

Latent variable		Coefficient	S.D. ^a	z value	p value
Reliability	Reliability_1	1.000	—	—	—
	Reliability_2	0.909	0.047	19.417	<0.001
	Reliability_3	0.869	0.051	16.941	<0.001
Safety	Safety_1	1.000	—	—	—
	Safety_2	1.029	0.043	24.173	<0.001
	Safety_3	0.966	0.041	23.454	<0.001
Convenience	Convenience_1	1.000	—	—	—
	Convenience_2	0.989	0.057	17.221	<0.001
	Convenience_3	0.940	0.055	17.018	<0.001
Comfort	Comfort_1	1.000	—	—	—
	Comfort_2	0.862	0.037	23.309	<0.001
	Comfort_3	0.858	0.039	21.813	<0.001
	Comfort_4	0.875	0.040	21.779	<0.001
Reliability	Safety	0.525	0.128	4.109	<0.001
	Convenience	-1.397	0.589	-2.370	0.018
	Comfort	1.462	0.473	3.092	0.002
Safety convenience	Convenience	1.013	0.061	16.660	<0.001
	Comfort	0.775	0.046	16.954	<0.001

Note. ^aS.D. denotes standard deviation.

TABLE 7: Results of evaluation metrics for SEM.

Metrics	χ^2	GFI	RMSEA	CFI	TLI	IFI
Result	527.687	0.978	0.068	0.925	0.948	0.956
Threshold	—	>0.9	<0.08	>0.9	>0.9	>0.9

TABLE 8: Results of the posterior probability.

Model	Model #1 (%)	Model #2 (%)	Model #3 (%)	Model #4 (%)
Posterior probability	56.8	29.5	5.3	5.1

TABLE 9: Parameter estimation results for travel mode of cross-border bus.

Variable	P.M.	Model #1	Model #2	Model #3	Model #4
Gender	-0.879	-0.904**	-0.844**	-0.866**	-0.823**
Whether Hong Kong and Macao travel together	-1.145	-1.146***	-1.147***	-1.083***	-1.162***
Transportation mode (origin-Zhuhai)					
Bus	4.272	0.095	4.281***	4.257***	4.314***
Private car	2.945	0.106	2.998***	2.852***	2.938***
Transportation mode (Hong Kong-destination)					
Bus	1.165	1.124*	1.220*	1.161*	1.292*
Taxi	-1.560	-1.538***	-1.622***	-1.587***	-1.425***
Subway	2.137	2.216**	1.978*	2.208**	2.157**
SP variables					
Travel time	-0.149	-0.148***	-0.150***	-0.152***	-0.141***
Travel cost	-0.067	-0.067***	-0.066***	-0.061***	-0.071***
Waiting time	-0.083	-0.082***	-0.084***	-0.096***	-0.071***
Travel preference	-0.894	-0.878***	-0.926***	-0.904***	-0.876***
Latent variables					
Safety	0.583	0.045	0.555***	0.638***	0.596***
Comfort	1.389	0.068	1.344***	1.474***	1.327***
Reliability	-0.121	0.174	—	-0.367**	—
Convenience	-1.419	0.033	-1.430***	-1.391***	-1.460***

Note. The significant interval of p value [0, 0.001] is denoted by ***, (0.001, 0.01] is denoted by **, (0.01, 0.05] is denoted by *, and (0.05, 0.1] is not denoted in the coefficients of model #1, #2, #3, and #4. P.M. is the posterior mean, which represents the coefficients of the MXL-BMA model.

TABLE 10: Parameter estimation results for travel mode of cross-border private car.

Variable	P.M.	Model #1	Model #2	Model #3	Model #4
Travel purpose/tourism	-1.048	-1.008**	-1.139**	-0.912**	-1.044**
Travel time (Hong Kong-destination)	0.238	0.233*	0.291*	—	0.214
Transportation mode (origin-Zhuhai)					
Bus	2.531	2.554***	2.465***	2.691***	2.498***
Private car	5.075	5.081***	5.012***	5.344***	5.150***
Transportation mode (Hong Kong-destination)					
Bus	-2.134	-2.208***	-1.986***	-2.080***	-2.223***
Private car	4.165	4.178***	4.122***	4.336***	4.151***
Taxi	-3.137	-3.129***	-3.124***	-3.142***	-3.278***
Subway	-0.356	-0.288	-0.425	-0.358	-0.629
SP variables					
Travel time	-0.149	-0.148***	-0.150***	-0.152***	-0.141***
Travel cost	-0.067	-0.067***	-0.066***	-0.061***	-0.071***
Waiting time	-0.083	-0.082***	-0.084***	-0.096***	-0.071***
Travel preference	-0.894	-0.878***	-0.926***	-0.904***	-0.876***
Latent variables					
Safety	0.411	0.389**	0.462**	0.411**	0.321*
Comfort	1.580	1.536***	1.659***	1.528***	1.598***
Reliability	-0.128	—	-0.388*	—	—
Convenience	-1.471	-1.486***	-1.439***	-1.519***	-1.468***

Note. The significant interval of p value, $[0, 0.001]$ is denoted by ***, $(0.001, 0.01]$ is denoted by **, $(0.01, 0.05]$ is denoted by *, and $(0.05, 0.1]$ is not denoted in the coefficients of model #1, #2, #3, and #4. P.M. is the posterior mean, which represents the coefficients of MXL-BMA model.

TABLE 11: Goodness of fit metrics.

Metric	MXL-BMA	MXL
BIC	6093.514	6130.229
AIC	5891.433	5895.310

indicates travelers who choose the travel mode of cross-border bus place a higher emphasis on convenience than the travelers who take cross-border private cars. To sum up, convenience shows a more remarkable impact on cross-border travel mode choice behavior compared to other latent variables.

4.3. Discussion. The BMA approach integrates the outputs of candidate models into a certain model to improve the interpretability of travel mode choice behavior. For quantitative evaluation of the fitting performance of MXL-BMA model and the MXL models in the model space, the Bayesian information criterion (BIC) and Akaike information criterion (AIC) are considered as the metrics, the results are listed in Table 11. The goodness of fit results demonstrate that the MXL-BMA model provides a superior description of cross-border travel behavior compared to MXL model. Notably, the MXL model selected in this section is the one with the best-fitting performance among the candidate models.

To comprehensively compare the explanatory performance of MXL-BMA model and MXL model in analyzing cross-border travel mode choice behavior, the variables with a high contribution to utility function (i.e., has a great impact on travel mode choice behavior) and show a larger difference between the two models are taken as examples to analyze, that is the transportation modes (i.e., bus and private car) from origin to Zhuhai and from Hong Kong to destination

TABLE 12: Comparative results of explanatory variables for two models.

Variable	Cross-border bus		Cross-border private car	
	MXL	MXL-BMA	MXL	MXL-BMA
Transportation mode (origin-Zhuhai)				
Bus	4.257	4.272 (+0.5%)	2.465	2.531 (+2.7%)
Private car	2.852	2.945 (+3%)	5.012	5.075 (+1%)
Transportation mode (Hong Kong- destination)				
Bus	1.220	1.165 (-4.5%)	-1.986	-2.134 (+7.5%)
Private car	—	—	4.158	4.165 (+0.2%)
Subway	1.978	2.137 (+8%)	—	—

(i.e., bus, private car, and subway), as shown in Table 12. It can be observed that the parameter coefficients of MXL-BMA model are more effective than those of MXL model in capturing the influence of connecting transportation modes on cross-border travel behavior. For the MXL-BMA model, the transportation mode of private car from the origin to Zhuhai shows the difference from other transportation modes, indicating that travelers who choose the cross-border bus travel mode do so due to the absence of a private license plate to pass through the HZMB. As for the transportation mode of the bus from Hong Kong to destination, MXL-BMA model exhibits a 7.5% increase in coefficient compared to MXL model for cross-border private car travel mode, while showing a 4.5% decrease for cross-border bus travel mode. Moreover, regarding the transportation mode of subway, the coefficient of MXL-BMA model increases by 8%, indicating that MXL-BMA model provides a more accurate representation of the influence of the developed public transport system of Hong Kong on cross-border travel. In addition, the coefficient difference of other variables between the two travel modes is small, which

affects that they have a comparatively lower impact on the overall utility function.

5. Conclusions

This study proposed an MXL-BMA model to describe the cross-border travel behavior of HZMB travelers using RP/SP survey data. An SEM model is constructed to explore the potential correlations between latent variables, and the BMA approach is applied to model the MXL model for explaining the cross-border travel mode choice characteristics. The main conclusions are summarized as follows:

- (1) The MXL-BMA model effectively captures the cross-border travel mode choice behavior on HZMB. Specifically, the number of transfers and the connecting transportation modes during cross-border travel on HZMB exhibit a significant influence on cross-border bus travel mode, whereas time-conscious travelers tend to select the cross-border private car. To sum up, the transportation mode arriving and departing the HZMB has a noticeable impact on cross-border travel modes.
- (2) The MXL-BMA model demonstrates a superior ability to explain cross-border travel mode choice behavior compared to the MXL model. Furthermore, for the section from origin to Zhuhai, the MXL-BMA model reveals the travelers who choose the travel mode of cross-border bus instead of the private car due to restricted access to HZMB. Conversely, for the section from Hong Kong to destination, the MXL-BMA model can more accurately explore the characteristics of the travelers who choose the private car.

Note that the dataset used in this study is collected from a questionnaire survey online, which may not entirely represent the actual travel behavior. For future work, the data collection range can be expanded to improve data quality, and machine learning-based methods (e.g., Bayesian-related approach and decision tree-based approach) can be applied to explore cross-border travel mode choice behavior.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Data curation was performed by Bo Lin; Formal analysis was performed by Yajie Zou, Wanbing Han, and Shubo Wu; Funding acquisition was performed by Bing Wu and Linbo Li; Investigation was performed by Bo Lin and Malik Muneeb Abid; Methodology was performed by Yajie Zou,

Wanbing Han, and Bo Lin; Supervision was performed by Yajie Zou, Bing Wu, and Linbo Li; Validation was performed by Wanbing Han, Bo Lin, and Malik Muneeb Abid; Yajie Zou, Bo Lin, and Shubo Wu wrote the original draft; Yajie Zou, Wanbing Han, and Malik Muneeb Abid reviewed and edited the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

This research was funded by the National Key Research and Development Program (2019YFB1600703) and the Shanghai Science and Technology Committee (19210745700).

References

- [1] X. Yang, Y. Zou, and L. Chen, "Operation analysis of freeway mixed traffic flow based on catch-up coordination platoon," *Accident Analysis & Prevention*, vol. 175, Article ID 106780, 2022.
- [2] X. Li, J. Tang, X. Hu, and W. Wang, "Assessing intercity multimodal choice behavior in a touristy city: a factor analysis," *Journal of Transport Geography*, vol. 86, Article ID 102776, 2020.
- [3] S. Wu, X. Chen, C. Shi, J. Fu, Y. Yan, and S. Wang, "Ship detention prediction via feature selection scheme and support vector machine (SVM)," *Maritime Policy & Management*, vol. 49, no. 1, pp. 140–153, 2022.
- [4] J. Zhao, Z. Yan, X. Chen, B. Han, S. Wu, and R. Ke, "K-GCN-LSTM: A k-hop graph convolutional network and long-short-term memory for ship speed prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 606, Article ID 128107, 2022.
- [5] C. Ding, Y. Wang, T. Tang, S. Mishra, and C. Liu, "Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice," *Transportation Research Part D: Transport and Environment*, vol. 60, pp. 28–40, 2018.
- [6] Y. Zou, T. Zhu, Y. Xie, Y. Zhang, and Y. Zhang, "Multivariate analysis of car-following behavior data using a coupled hidden Markov model," *Transportation Research Part C: Emerging Technologies*, vol. 144, Article ID 103914, 2022.
- [7] W. Li, L. Bao, Y. Li, H. Si, and Y. Li, "Assessing the transition to low-carbon urban transport: a global comparison," *Resources, Conservation and Recycling*, vol. 180, Article ID 106179, 2022.
- [8] J. Lu, C. Lv, T. Feng, and Z. Wang, "Evaluating the effects of Hong Kong-Zhuhai-Macao bridge on international air travel: demand analysis of the air-bridge-air path," *European Journal of International Management*, vol. 14, no. 4, pp. 595–616, 2020.
- [9] Z. Hu, "Analysis of passenger flow behavior based on big data related to Shenzhen-Hongkong ports," Wuhan University of Technology, Wuhan, China, ph.d, 2018.
- [10] J. Liu and W. Shi, "A cross-boundary travel tale: unraveling Hong Kong residents' mobility pattern in Shenzhen by using metro smart card data," *Applied Geography*, vol. 130, Article ID 102416, 2021.
- [11] J. Shin and N. Tilahun, "The role of residential choice on the travel behavior of young adults," *Transportation Research Part A: Policy and Practice*, vol. 158, pp. 62–74, 2022.
- [12] Z. Tao, Q. Nie, and W. Zhang, "Research on travel behavior with car sharing under smart city conditions," *Journal of*

- Advanced Transportation*, vol. 2021, Article ID 8879908, 13 pages, 2021.
- [13] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT press, Cambridge, MA, USA, 1985.
 - [14] J. Broach, J. Dill, and N. W. McNeil, "Travel mode imputation using GPS and accelerometer data from a multi-day travel survey," *Journal of Transport Geography*, vol. 78, pp. 194–204, 2019.
 - [15] T. Bai, X. Li, and Z. Sun, "Effects of cost adjustment on travel mode choice: analysis and comparison of different logit models," *Transportation Research Procedia*, vol. 25, pp. 2649–2659, 2017.
 - [16] Y. Han, W. Li, S. Wei, and T. Zhang, "Research on Passenger's travel mode choice behavior waiting at bus station based on SEM-logit integration Model," *Sustainability*, vol. 10, no. 6, pp. 1996–2018, 2018.
 - [17] B. Fan, "Research on transfer model of car-commuter trip based on Nested Logit," Beijing Jiaotong University, Beijing, China, ph.d, 2014.
 - [18] H. Aziz, N. N. Nagle, A. M. Morton, M. R. Hilliard, D. A. White, and R. N. Stewart, "Exploring the impact of walk-bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data," *Transportation*, vol. 45, no. 5, pp. 1207–1229, 2018.
 - [19] J. Wu, M. Yang, S. Sun, and J. Zhao, "Modeling travel mode choices in connection to metro stations by mixed logit models: a case study in Nanjing, China," *Promet - Traffic & Transportation*, vol. 30, no. 5, pp. 549–561, 2018.
 - [20] M. Ye, Y. Chen, G. Yang, B. Wang, and Q. Hu, "Mixed logit models for travelers' mode shifting considering bike-sharing," *Sustainability*, vol. 12, no. 5, pp. 2081–2097, 2020.
 - [21] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and logit models," *Travel behaviour and society*, vol. 20, pp. 22–35, 2020.
 - [22] L. Cheng, X. Chen, J. De Vos, X. Lai, and F. Witlox, "Applying a random forest method approach to model travel mode choice behavior," *Travel behaviour and society*, vol. 14, pp. 1–10, 2019.
 - [23] Y. Zou, L. Ding, H. Zhang, T. Zhu, and L. Wu, "Vehicle acceleration prediction based on machine learning models and driving behavior analysis," *Applied Sciences*, vol. 12, no. 10, p. 5259, 2022.
 - [24] A. Lindner, C. S. Pitombo, and A. L. Cunha, "Estimating motorized travel mode choice using classifiers: an application for high-dimensional multicollinear data," *Travel Behaviour and Society*, vol. 6, pp. 100–109, 2017.
 - [25] A. Lhéritier, M. Bocamazo, T. Delahaye, and R. Acuna-Agost, "Airline itinerary choice modeling using machine learning," *Journal of choice modelling*, vol. 31, pp. 198–209, 2019.
 - [26] N. Golshani, R. Shabanpour, S. M. Mahmoudifard, S. Derrible, and A. Mohammadian, "Modeling travel mode and timing decisions: comparison of artificial neural networks and copula-based joint model," *Travel Behaviour and Society*, vol. 10, pp. 21–32, 2018.
 - [27] F. Wang and C. L. Ross, "Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model," *Transportation Research Record*, vol. 2672, no. 47, pp. 35–45, 2018.
 - [28] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Systems with Applications*, vol. 78, pp. 273–282, 2017.
 - [29] G. Wang, R. Jia, J. Liu, and H. Zhang, "A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning," *Renewable Energy*, vol. 145, pp. 2426–2434, 2020.
 - [30] Y. Zhang, W. Li, H. Deng, and Y. Li, "Evaluation of public transport-based accessibility to health facilities considering spatial heterogeneity," *Journal of Advanced Transportation*, vol. 2020, Article ID 7645153, 2020.
 - [31] S. Wu, Y. Zou, L. Wu, and Y. Zhang, "Application of Bayesian model averaging for modeling time headway distribution," *Physica A: Statistical Mechanics and Its Applications*, vol. 620, Article ID 128747, 2023.
 - [32] Y. Zhou, F.-J. Chang, H. Chen, and H. Li, "Exploring copula-based bayesian model averaging with multiple ANNs for PM2.5 ensemble forecasts," *Journal of Cleaner Production*, vol. 263, Article ID 121528, 2020.
 - [33] L. V. Coote, J. Swait, and W. Adamowicz, "Separating generalizable from source-specific preference heterogeneity in the fusion of revealed and stated preferences," *Journal of choice modelling*, vol. 40, Article ID 100302, 2021.
 - [34] J. Le and J. Teng, "Understanding influencing factors of travel mode choice in urban-suburban travel: a case study in Shanghai," *Urban Rail Transit*, pp. 1–20, 2023.
 - [35] P. Tarka, "An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences," *Quality and Quantity*, vol. 52, no. 1, pp. 313–354, 2018.
 - [36] J. F. Hair, C. M. Ringle, S. P. Gudergan, A. Fischer, C. Nitzl, and C. Menictas, "Partial least squares structural equation modeling-based discrete choice modeling: an illustration in modeling retailer choice," *Business Research*, vol. 12, no. 1, pp. 115–142, 2019.
 - [37] Y. Jiang, J. Zhang, Y. Wang, and W. Wang, "Capturing ownership behavior of autonomous vehicles in Japan based on a stated preference survey and a mixed logit model with repeated choices," *International Journal of Sustainable Transportation*, vol. 13, no. 10, pp. 788–801, 2019.
 - [38] H. Li, X. Li, X. Xu, J. Liu, and B. Ran, "Modeling departure time choice of metro passengers with a smart corrected mixed logit model—a case study in Beijing," *Transport Policy*, vol. 69, pp. 106–121, 2018.
 - [39] Y. Zou, B. Lin, X. Yang, L. Wu, M. Muneeb Abid, and J. Tang, "Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management," *Journal of Advanced Transportation*, vol. 2021, Article ID 6671983, 2021.
 - [40] T. M. Fragoso, W. Bertoli, and F. Louzada, "Bayesian model averaging: a systematic review and conceptual classification," *International Statistical Review*, vol. 86, no. 1, pp. 1–28, 2018.
 - [41] A. B. Astuti, N. Iriawan, and H. Kuswanto, "Model components selection in bayesian model averaging using occam's window for microarray data," *Journal of Natural A*, vol. 1, no. 2, pp. 67–74, 2014.