

Research Article

Mode Choice Model for Public Transport with Categorized Latent Variables

Jian Chen and Shoujie Li

School of Traffic & Transportation, Chongqing Jiaotong University, No. 66 Xuefu Rd., Nan'an District, Chongqing City 40074, China

Correspondence should be addressed to Shoujie Li; shoujieli@gmail.com

Received 2 May 2017; Accepted 10 July 2017; Published 14 August 2017

Academic Editor: Luca D'Acerno

Copyright © 2017 Jian Chen and Shoujie Li. 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.

Mode choice model for public transport, which integrates structural equation model (SEM) and discrete choice model (DCM) with categorized latent variables, was presented in this paper. Apart from identifying those important latent variables that affect mode choice for public transport, the objective of this study was also to develop an improved disaggregative model that better explains travel behavior of those decision-makers in choosing public transport. After extensive observations, selective latent variable sets which consist of latent variable components were chosen together with explicit variables in formulating the utility functions. Data collected in Chengdu city, China, were used to calibrate and validate the model. Results showed that the impact of fare on mode choice of public transport escalated in the SEM-DCM integrated model compared with the traditional logit model. The goodness of fit for the integrated model with latent variable sets is 0.201 higher than that of the traditional logit model, which proves that latent variables have an obvious impact on mode choice behavior, and the SEM-DCM integrated model has higher accuracy and stronger explanatory ability. The results are especially helpful for public transport operators to achieve higher mode share split by improving the service quality of public transport in terms of providing more convenience and better service environment for public transport users.

1. Introduction

Mode choice behavior for public transport is a key element in public transport planning, as it has a direct impact on the design of urban transport system structure, and is also the basis for urban public transport planning and management policymaking. Discrete choice models (DCMs) are a typical method of research on consumer choice behavior originally applied in economics. Paying for using public transport, public transport users are actually consumers. Therefore, their mode choice behavior is actually a consumer choice behavior. Discrete choice models (DCMs) explain individual choice behavior as the consequence of preferences that an individual makes with the assumption that the consumer chooses the most preferred option. Under certain assumptions, consumer preferences can be represented by a utility function in which the choice process somehow seems to be a sort of “black box.” Its inputs are the attributes of available options and the individual's characteristics, and its output is the utility

value of each alternative. Koppelman and Pas [1] carried out a preliminary study of DCMs to analyze travel characteristics of mode choice and travel purposes. Ben-Akiva et al. [2] borrowed the concept of utility function in economics and applied it to transport mode choice behavior from developing DCMs. Later, Koppelman and Wen [3] selected variables of frequency, travel cost, in-vehicle time, and out-of-vehicle time in mode choices among transport modes of air, train, car, and bus. Hensher and Reyes [4] and Schwanen [5] considered spatial variation in building DCMs on mode choice behavior. Mitra and Buliung [6] applied Multinomial Logit Models (MNL) to analyze the differences of mode choice to school between children aged 11 and 14-15 years in Toronto, Canada, and the result shows that distance is the biggest problem for younger children to walk to school, while boys are more willing to walk than girls. Liu et al. [7] investigated the Swedish mode choice behavior under different climate conditions. Palma and Rochat [8] analyzed mode choice for commuters in Geneva using Nested Logit Model and

found out that car ownership is the key factor that influences mode choice behavior. Apparently, it can be concluded that those DCMs considered only objective/measurable attributes from the alternatives and socioeconomic characteristics of the individuals as explanatory variables.

Based on consumer choice behavior theory, consumers' preference is influenced by consumers' satisfaction to available transport modes and their importance to the consumers. It is obvious that individual perception and aptitude to different transport modes differ due to their different socioeconomic status and different intrinsic characteristics of transport modes, which cannot be directly observed or measured. In other words, transport mode choice behavior is influenced not only by measurable factors (travel time, fare, gender, age, occupation, income, etc.), but also by unmeasurable factors (service quality, convenience, safety, etc.). Therefore, latent variables should be added in the model to include the subjective factors' influence on transport mode choice. The initial research on latent variables can be traced back to Spearman [9] that conducted factor analysis for human intelligence test. Latent variables were widely applied in many fields, such as social sciences [10, 11], psychology [12, 13], and market and economics research [14, 15].

Unmeasured variables, factors, unobserved variables, constructs, or true scores were several terms that researchers used to refer to latent variables in psychology and the social sciences [13]. In transport modeling, latent variables include service reliability, environment perception, and potential preferences for transport modes, which bring about new approaches in modeling transport mode behavior. Those research studies on extended framework incorporated with latent variables for DCMs were developed by Train et al. [16] and Ben-Akiva and Boccara [17]. Traditional mode choice models have been enriched by introducing latent variables in modeling transport mode choice behavior [14, 18–20]. Morikawa et al. [21] and Morikawa et al. [20] included modal comfort and convenience in their analyses of transport mode choice. In their studies, latent variables were measured and modeled through attitudes (attitudinal indicator variables) towards the chosen transport mode and its alternatives. Golob [22] used a series of models to explain how mode choice and attitudes regarding HOV (high-occupancy vehicle toll) lanes free of charge in San Diego differed across the population. However, attitudes towards fairness of the HOT (HOV-Toll) lanes to carpoolers exhibit the greatest number of significant explanatory variables. Mitra et al. [23] and Lee et al. [24] analyzed the relationship between potential safety measures and severity of road accidents in transport planning. Gopinath [25], Morikawa et al. [21], Walker and Ben-Akiva [26], Johansson et al. [27], Yáñez et al. [28], and Politis et al. [29] added the influence of traveler's aptitude and perception into their models and integrated it into demand forecasting models. Shiftan et al. [30] explored the causal relationship between latent variables (time sensibility, determined travel plan, aspiration to use transport vehicle, etc.) and their corresponding measured variables and categorized the transport market based on these latent variables. Deutsch et al. [31] analyzed the relationship between travel behavior of residents and their perception on surrounding environment,

using structural equation model based on 719 sample data points from a survey in Santa Barbara and California, USA. Nurlaela and Curtis [32] established a mode choice model in downtown, which has not been developed and analyzed empirically before.

To capture the impact of latent variables on the decision process, during the last few decades, a hybrid choice model (HCM) has been developed [33] using psychometric data to explicitly model attitudes and perceptions and their influences on mode choices [28, 34, 35]. Kim et al. [36] expanded the HCM by including social influence as well as attitudes on choice probabilities to investigate the effects of these factors on the intention to purchase electric cars. Among the numerous HCMs that explicitly model latent psychological factors such as attitudes and perceptions (latent variables), the Integrated Choice and Latent Variable (ICLV) model is easy to be understood and exhibits better predictive power [37–42]. There are serious problems when arbitrary values are used for normalization and when data variability is low, especially regarding the generation of the latent variables. Hence, Raveau et al. [43] suggested normalizing variances associated with the hybrid discrete choice model's structural equations instead of the parameters of its measurement equations.

With high economic development rate and fast urbanization in China in the last few decades, it was observed that more passengers were willing to pay for taking transport mode with higher service quality, more convenience, and higher personal security. This paper aims to extend Bhat and Dubey's [42] model formulation to develop an ICLV model for public transport that incorporates five latent variable sets associated with convenience, personal safety, modal comfort, service environment, and waiting feelings which were more sensitive to public transit users according to data collected in Chengdu, China. Five fundamental latent variable sets were proposed that represent a combination of underlying affective value norms/beliefs, lifestyle orientations, and personality traits rather than a more generic single "willingness to walk/cycle/drive" attitude as the latent construct [44]. In other words, five latent variable sets were employed to explain the public transport mode choice behavior to achieve a more precise interpretation of the individual factors affecting public transport mode choice. In summary, this paper has improved model formulations based on models proposed by Bhat and Dubey [42] and Yáñez et al. [28] and made comparisons among three models that incorporated different numbers of explicit variables with latent variable sets. A new approach of estimating goodness of fit when comparing the ICLV model with a model without latent variable sets was applied as well, which has not received the attention it deserves. The latent variables were identified based on data collected in Chengdu, China, which is a typical urban city; therefore, the results would be beneficial to public transit policymaking and operation especially for those urban cities in China.

The rest of the paper is organized as follows. Section 2 presents the model formulation and estimation. Section 3 presents the data and sample characteristics. The estimation results of the models are presented and discussed in Section 4.

TABLE 1: Model variables description.

Latent variables	Index	Measured variables
Convenience	y_1	Travel time from departure place to bus stop is short.
	y_2	Convenient in transferring to other buses or transport modes.
	y_3	Information of bus stop on route is clear.
Personal safety	y_4	If there is a safety hammer, fire extinguisher, and so forth in the vehicle that make passengers feel safe.
	y_5	If propaganda on a solution to an emergency situation has been well carried out.
	y_6	Overall satisfaction to feeling safe in a bus.
Modal comfort	y_7	Feeling comfortable on bus seats.
	y_8	If entertainment facilities, like TV and radio broadcast, are pleasant.
	y_9	Clear and correct bus stop information in broadcast in the bus.
Service environment	y_{10}	Clean inside the bus.
	y_{11}	Fresh air in the bus.
	y_{12}	If passengers are in good order when they are boarding or alighting.
	y_{13}	If the bus driver and conductor are polite and willing to serve passengers.
	y_{14}	If interference among passengers is little.
Waiting feelings	y_{15}	Overall satisfaction to in-vehicle environment.
	y_{16}	Estimated arrival time of incoming bus is accurate and in time shown on a variable message sign (VMS) board at bus stops.
	y_{17}	Actual waiting time is very close to the estimation on the VMS board.
	y_{18}	Newspapers or other information provided at the bus stop is interesting.

Section 5 concludes the paper by summarizing the key findings and providing directions for further research.

2. Model Formulation and Estimation

2.1. Assumptions

- (i) Travelers are rational in making mode choices as they will choose the travel scheme with the highest utility value.
- (ii) Options of mode choice are categorized into public transport and nonpublic transport modes.
- (iii) Error distribution of each utility function follows Gumbel distribution with mean of 0 independently, while error distribution of the rest of the stochastic factor functions follows a normal distribution.

2.2. Model Variables. Apart from social status and travel characteristics, which can be directly observed and measured, other factors included in the model are latent variables, which cannot be directly observed or measured. Therefore, it requires corresponding measured variables to describe latent variables. In this paper, 7 Likert scales were adopted to measure latent variables. Five latent variable sets are used to model the public transport mode choice behavior, which are convenience, personal safety, modal comfort, service environment, and waiting feelings, as shown in Table 1.

2.3. Model Formulation. There are two stages in formulating this model shown in Figure 1: (1) the latent variable structural

equation model and the latent variable measurement equation model and (2) the mode choice model.

Index i refers to alternatives ($i = 1, 2, \dots, I$), index l to latent variables ($l = 1, 2, \dots, L$), index k to objective attributes ($k = 1, 2, \dots, K$), and index q to individuals ($q = 1, 2, \dots, Q$).

(1) Latent Variable Structural Equation Model. It is assumed that the latent variable η_l^* is a linear function of covariates as follows:

$$\eta_l^* = \lambda_l' \mathbf{x} + \zeta_l, \quad (1)$$

where \mathbf{x} denotes a $(\bar{D} \times 1)$ vector of observed covariates without including a constant, λ_l is a corresponding $(\bar{D} \times 1)$ vector of coefficients, and ζ_l is a random error term assumed to be normally distributed. In our notation, the same exogenous vector \mathbf{x} is used for all latent variables; however, this is in no way restrictive, since one may place the value of zero in the equation discussed by Stapleton (1978) and used by Bolduc et al. [37] by assuming that ζ_l is standard normally distributed. Next, define $L \times \bar{D}$ matrix $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_L)'$, $(L \times 1)$ vectors $\mathbf{z}^* = (z_1^*, z_2^*, \dots, z_L^*)'$, and $\boldsymbol{\zeta}^* = (\zeta_1, \zeta_2, \dots, \zeta_L)'$. To allow correlation among the latent variables, $\boldsymbol{\zeta}$ is assumed to be standard multivariate normally distributed: $\boldsymbol{\zeta} \sim (0_L, \Sigma_\zeta)$, where Σ_ζ is a correlation matrix. Equation (1) can be written in matrix form as

$$\boldsymbol{\eta}^* = \boldsymbol{\lambda} \mathbf{x} + \boldsymbol{\zeta}. \quad (2)$$

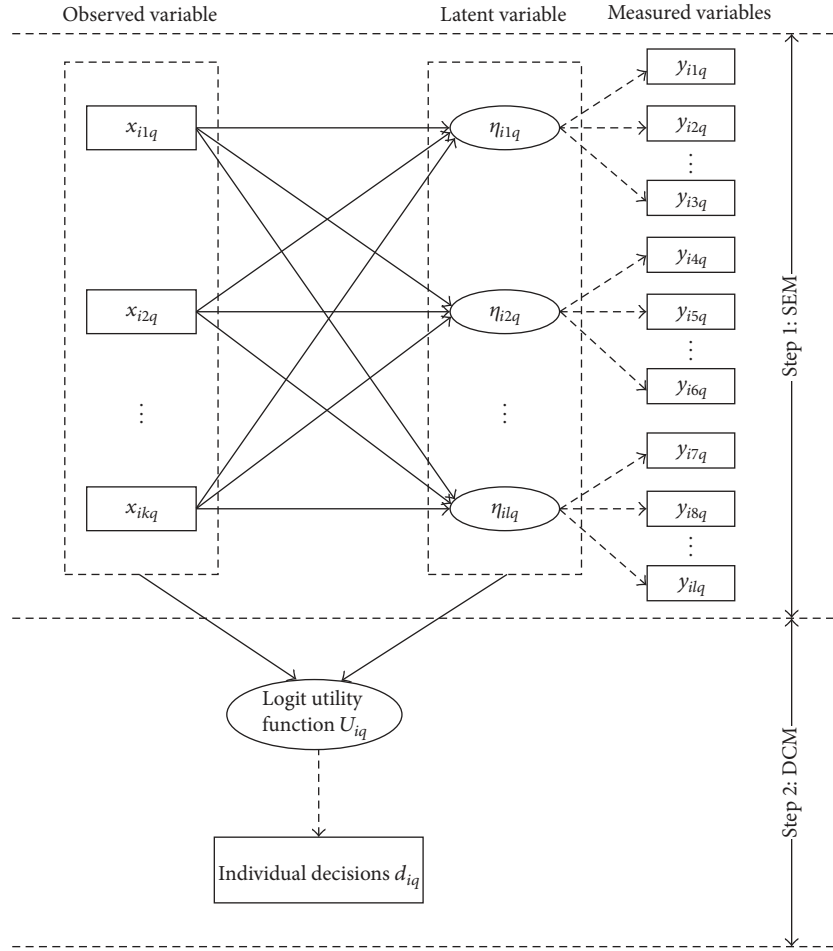


FIGURE 1: Mathematical model.

(2) *Latent Variable Measurement Equation Model.* Here, we define

$$y_h = \delta_h + \gamma'_h \boldsymbol{\eta}^* + \xi_h, \quad (3)$$

where H refers to continuous variables (y_1, y_2, \dots, y_H) with an associated index h ($h = 1, 2, \dots, H$), δ_h to a scalar constant, γ_h to an $(L \times 1)$ vector of latent variable loadings on the h th continuous indicator variable, and ξ_h to a normally distributed measurement error term.

Stack the H continuous variables into an $(H \times 1)$ vector \mathbf{y} , the H constants δ_h into an $(H \times 1)$ vector $\boldsymbol{\delta}$, and the H error terms into another $(H \times 1)$ vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_H)$. Similarly, let Σ_y be the covariance matrix of $\boldsymbol{\xi}$ and define the $(H \times L)$ matrix of latent variable loadings $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_H)'$. Equation (3) can be rewritten in matrix form as follows:

$$\mathbf{y}^* = \boldsymbol{\delta} + \boldsymbol{\gamma} \boldsymbol{\eta}^* + \boldsymbol{\xi}. \quad (4)$$

(3) *Choice Model.* According to random utility theory, it is assumed that individuals make decisions with maximum utility U_{iq} . It is also assumed that the analyst, who is an observer without perfect information, is only able to define/observe a

representative utility V_{iq} . Therefore, it is necessary to associate an error term ε_{iq} with each alternative [45] shown as follows:

$$U_{iq} = V_{iq} + \varepsilon_{iq}. \quad (5)$$

Latent variables can be included in a utility function shown in (6), where θ_{ik} and β_{il} are parameters to be estimated, and they are correlated, respectively, to the tangible attributes and the latent variables:

$$V_{iq} = \sum_k \theta_{ik} \cdot x_{ikq} + \sum_l \beta_{il} \cdot \eta_{ilq}. \quad (6)$$

A binary variable d_{iq} is used to define the individual decision-making as shown in

$$d_{iq} = \begin{cases} 1 & \text{if } U_{iq} \geq U_{jq}, \forall j \in A(q) \\ 0 & \text{in othe case.} \end{cases} \quad (7)$$

As the latent variable η_{ilq} is actually unknown, the discrete choice model should be estimated jointly with the structural equation model (2) and measurement equation model (4).

2.4. Estimation. As Yáñez et al. [28] mentioned, currently, there are two approaches to estimate hybrid choice models; they differ mainly in how the available information is used. Although there is experimental software for simultaneous estimation [46], it can only estimate multinomial logit (MNL) models without accommodating heterogeneity or correlation among individuals and/or observations through random parameters or error components. Therefore, we followed Yáñez adopting sequential estimation approach.

According to Yáñez et al. [28], this sequential estimation approach consists of two separate stages in sequential estimation approach: the latent variable modeling and the discrete choice modeling. In the first stage, a structural equation model is solved to obtain estimators (and their standard deviation errors) for the parameters in the equations containing the latent variables with the explanatory variables and the perception indicators. Using these parameters, it is possible to calculate expected values for the latent variable of each alternative for individuals, through (2), and eventually include them directly in the discrete choice model in the second stage. In this stage, a proper procedure requires integrating with the variation of the latent variable [34]. With the interactions between the structural equation model and choice model, the latent variable can be added in any explanatory variable. Therefore, we estimate these parameters together with those of the traditional variables in the second stage, in order to guarantee unbiased estimators for the parameters involved [13] as the expected values of the latent variables have measurement error. Estimating a Mixed Logit (ML) model with random parameters on the latent variables can solve this issue.

In this paper, we adopted the sequential estimation approach that Yáñez et al. [28] proposed to solve the problem. Based on survey data of individual preferences in southwest China, latent variable sets were carefully selected. The aim is to explore the unique characteristics of latent variables in the fast economic development era in China and help the public transit policymakers and operators to improve share of public transport in daily trips.

3. Data Collection and Analysis

Data on predesigned questionnaires were collected from those respondents of commuters in Chengdu, China. This survey collected 570 questionnaires, and the valid amount was 497, which is 87.06%, excluding those samples: (1) samples in which respondents did not seriously answer the question; (2) samples with 3 or more questions void of answers; (3) samples with 5 continuous extreme values. All these valid questionnaires are summarized in Table 2.

According to structural equation model, data should follow multivariate normal distribution. Maximum likelihood method is adopted to estimate parameters in the structural equation model. Multivariate normality test covers two main indices, skewness and kurtosis, and makes judgements based on their absolute values. (1) If the absolute value of $S > 3.0$, then it is in the range of extreme skewness; (2) if the absolute value of $K > 10.0$, then it means there is a problem with the

value of kurtosis; and (3) if the absolute value of $K > 20.0$, then it is in the range of extreme kurtosis.

Table 3 shows that values of skewness of 19 measured variables are from -0.762 to 0.576 , which are all less than 3, while values of kurtosis vary from -0.993 to 0.702 , which means that all measured variables follow normal distribution.

4. Estimation Results

Use the data from the questionnaire to conduct the following analysis: (1) exploratory factor analysis. Verify whether the 5 measurement indices of latent variable sets of public transport service quality are qualified to be effective factors that influence passengers' mode choice behavior. (2) Compare calibration results from mode choice model considering both measured variables and latent variables with those from mode choice model considering measured variables only and analyze the differences.

4.1. Exploratory Factor Analysis. There is no general standard for measurement index of public transport service quality; therefore, exploratory factor analysis is used to verify whether the latent variables and their corresponding measurement indices are suitable or not. And the factors with eigenvalue more than 1 are selected based on principal component analysis; there are 5 selected latent variable sets in total: convenience, personal safety, modal comfort, service environment, and waiting feelings. All the Cronbach α values which are a coefficient for 5 latent variable sets are above 0.69, which is very reliable.

4.2. Analysis of Calibration Results of Two Models. First integrate variables of travelers' personal characteristics and mode choice options into mode choice behavior model, the Binary Logit (BL) in Stage II, and get the basic transport mode choice model, which does not consider LVs. Use NLOGIT (NLOGIT is a suite of software for the estimation of discrete choice models by Econometric Software Inc.) software to estimate parameters of this model and verify the results, remove those variables meeting $|t| < 1.96$, and get parameter estimation and check results after various tests shown in Table 5. Variables of travelers' personal characteristics include gender, occupation, transport vehicles ownership, and fare. If t values of gender, occupation, and traffic tools are positive, then the larger their variables are, the more possible it is to choose public transport mode; if t values are negative, then the higher the price is, the less possible it is to choose public transport mode. And the goodness of fit of this model is 0.245, which proves that the model has certain explanatory ability with high accuracy.

Put 5 latent variable sets into choice model; we get public transport mode choice model considering latent variable sets and use NLOGIT to calibrate and verify the parameter of the integrated model. As latent variables cannot be directly measured, they are described by measured variables. Different integrated models could be worked out based on different description methods of latent variable sets and measured variables: (1) SEM-BLI: the relationships among

TABLE 2: Summary of collected data.

	Item	Proportion	Note
Gender	Male	60.1%	
	Female	39.9%	
Age	Under 18	0.7%	
	19 to 35	87.8%	
	36 to 55	10.1%	
	Over 56	1.4%	
Occupation	Student	44.6%	The sum of these three proportions of commenters is 93.9%, which is in line with the fact that commuters for work or school account for 75%.
	Staff in government or public institution	25.0%	
	Staff in enterprise	24.3%	
	Other	6.1%	
Education background	Completion of specialized or general sec. ed. & under	2.0%	
	Technical secondary school, junior college, and university	52.7%	
	Master degree or above	45.3%	
Income/month	Under ¥1,000	32.4%	Nearly half of the respondents are students, so the income is quite low.
	¥1,000 to ¥2,000	23.0%	
	¥2,000 to ¥3,000	19.6%	
	¥3,000 to ¥4,000	15.5%	
	Over ¥4,000	9.5%	
Vehicle ownership	Car	21.6%	
	Bicycle	27.0%	
	Electric bicycle	8.2%	
	Other vehicle	43.2%	
Trips/day	0 to 1 time	28.4%	
	2 to 4 times	65.5%	
	Over 5 times	6.1%	
Trip purpose	Work or school	75.0%	
	Shopping	15.5%	
	Recreation	4.7%	
	Other	4.7%	

latent variable sets and measured variables and between latent variable sets and measured variables are described through a structural equation based on the calculation of (2) and (4); (2) SEM-BL2 (the relationship between latent variable sets and measured variables through factor analysis): put all measured variables into integrated model based on results of Table 4; (3) SEM-BL3: only put measured variables of high-load factors into the integrated model after factor analysis. All calculation results of parameters in each model are illustrated in Table 6.

SEM of public transport service quality calculated by AMOS (AMOS stands for Analysis of Moment Structures, which is software that solves structural equations model developed by SPSS, Inc.) includes two parts: (1) Causal relationship between latent variable sets and measured variables of passengers' personal characteristics shown in the structural pattern in Table 7 and (2) relationship between latent variables and their measured variables shown in the measurement pattern in Table 7. It can be shown in Table 7 that the older the passenger's age is, the more satisfied

he/she is with the public transport service; influence of occupation on service environment is also positive, and it is more obvious than that of age, as government officers and employees in commerce and service industry demand higher service quality than people of other occupations. Educational background and modal comfort are positively correlated. Well-educated people demand high-level comfort; monthly income is also in positive correlation with convenience and personal safety, which means that people with higher income level pay more attention to convenience and personal safety. Women care more about waiting feelings, reflecting women's lack of patience compared with men.

It can be seen from Table 6 that results of estimated parameters are different among three different estimation methods in the ICLV model. In SEM-BL1, the variable set that has the greatest impact on mode choice is convenience, while the least impact is from modal comfort, but in SEM-BL2 and SEM-BL3, the variable set that has the greatest impact on mode choice is waiting feelings, while the least

TABLE 3: Measured variables normality test results.

Number	Measured variable	Average	Standard deviation	Skewness		Kurtosis	
				Value	CR	Value	CR
1	y_1	4.0676	1.66461	-0.109	0.199	-0.773	0.396
2	y_2	4.0473	1.55329	-0.002	0.199	-0.780	0.396
3	y_3	3.4459	1.57482	0.122	0.199	-0.687	0.396
4	y_4	4.9459	1.53766	-0.739	0.199	0.078	0.396
5	y_5	3.9662	1.51373	-0.097	0.199	-0.563	0.396
6	y_6	4.5000	1.24267	-0.485	0.199	-0.054	0.396
7	y_7	3.6284	1.41556	0.158	0.199	-0.484	0.396
8	y_8	3.3243	1.44851	0.233	0.199	-0.715	0.396
9	y_9	3.4189	1.50741	0.152	0.199	-0.793	0.396
10	y_{10}	4.0270	1.42354	-0.292	0.199	-0.560	0.396
11	y_{11}	3.1351	1.37840	0.149	0.199	-0.674	0.396
12	y_{12}	3.1351	1.64399	0.404	0.199	-0.699	0.396
13	y_{13}	3.5946	1.43256	0.011	0.199	-0.464	0.396
14	y_{14}	3.0541	1.47422	0.370	0.199	-0.739	0.396
15	y_{15}	3.8446	1.36383	-0.253	0.199	-0.704	0.396
16	y_{16}	2.7635	1.44919	0.421	0.199	-0.818	0.396
17	y_{17}	3.0338	1.32693	0.203	0.199	-0.658	0.396
18	y_{18}	2.7838	1.30692	0.576	0.199	0.028	0.396

TABLE 4: Rotated component matrix.

Variable sets	Measured variable	Convenience	Personal safety	Modal comfort	Service environment	Waiting feelings
Convenience	y_2	0.857	0.164	0.206	0.007	-0.003
	y_1	0.743	0.091	-0.015	0.156	0.139
	y_3	0.653	0.075	-0.011	0.067	0.353
Personal safety	y_5	0.082	0.852	0.058	-0.056	0.179
	y_4	0.047	0.850	0.077	0.110	-0.006
	y_6	0.230	0.695	0.242	0.099	0.033
Modal comfort	y_7	0.107	0.127	0.800	0.246	0.000
	y_8	0.057	0.070	0.727	0.129	0.326
	y_9	-0.003	0.196	0.695	0.324	0.191
Service environment	y_{12}	0.007	-0.012	0.007	0.833	0.079
	y_{15}	0.109	0.163	0.363	0.771	0.076
	y_{11}	0.010	0.073	0.303	0.735	0.124
	y_{13}	0.075	0.045	0.122	0.733	0.052
	y_{14}	0.084	-0.126	0.028	0.724	0.265
Waiting feelings	y_{10}	0.164	0.260	0.357	0.632	0.051
	y_{17}	0.204	0.019	0.026	0.159	0.826
	y_{16}	0.031	0.127	0.172	0.090	0.783
	y_{18}	0.191	0.053	0.228	0.178	0.666

TABLE 5: Reliability test results.

Item	Latent variable sets	Cronbach α
1	Convenience	0.696
2	Personal safety	0.768
3	Modal comfort	0.759
4	Service environment	0.869
5	Waiting feelings	0.742

is from modal comfort and personal safety, as the values of the goodness of fit for latent variables obtained by factor analysis in the latter two models do not include other explicit measured variables (age, education background, income, occupation, gender, etc.). Therefore, estimated parameters are different, and explanatory ability and accuracy of the latter two models are relatively low. When comparing with the latter two models, the goodness of fit for SEM-BL2 that

TABLE 6: Calibration results of SEM-DCM model of mode choice for bus.

Influence factor	BL model		SEM-BL1 model		SEM-BL2 model		SEM-BL3 model	
	Parameter	<i>t</i>	Parameter	<i>t</i>	Parameter	<i>t</i>	Parameter	<i>t</i>
Constant	-2.939	-5.599	-8.314	-3.281	-8.891	-4.391	-8.724	-4.445
Gender	0.592	3.386	0.828	1.929	1.049	1.980	0.947	1.992
Occupation	0.483	2.303	0.758	2.436	0.649	2.558	0.684	2.657
Vehicle ownership	0.410	4.313	0.354	2.482	0.394	2.807	0.410	2.911
Price	-0.162	-2.378	-0.242	-2.304	-0.206	-2.369	-0.170	-2.156
Convenience	—	—	0.673	2.645	0.340	2.046	0.374	2.416
Personal safety	—	—	0.532	2.637	0.584	3.142	0.357	1.849
Modal comfort	—	—	0.420	2.125	0.330	1.981	0.389	1.998
Service environment	—	—	0.639	2.241	0.552	2.815	0.454	2.111
Waiting feelings	—	—	0.423	3.332	0.777	4.049	0.599	3.264
Sample size	497		497		497		497	
$L(0)$	-102.586		-102.586		-102.586		-102.586	
$L(\hat{\theta})$	-77.425		-56.818		-57.122		-57.875	
ρ^2	0.245		0.446		0.443		0.436	

Note. All values meet the verification requirements, except *t* value of personal Safety in SEM-BL3 which is less than 1.96.

includes all measured variables is 0.007 which is higher than that for SEM-BL3 which only includes measured variables of large load matrices. Therefore, it can be concluded that those models that only include measured variables of large load matrices may miss certain information for estimation.

Of all the parameters in the calibration results shown in Table 7, y_2 has the highest impact on convenience, while y_3 has the lowest impact on it; y_4 has the highest impact on personal safety; y_7 has the highest impact on modal comfort, while y_9 has the lowest impact on it; y_{10} and y_{13} have the lowest impact on service environment; y_{17} has the highest impact on waiting feelings while y_{18} has the lowest impact on it. The result of this research provides basis and inspiration for public transport companies to improve the service quality on the perspective of passengers.

Based on the final estimation and verification of model parameters, the ICLV model with 5 latent variable sets is more accurate than those BL models without including latent variable sets, and the goodness of fit has improved from 0.245 to 0.446, with $L(\theta)$ increasing by 20.607, which means that explanatory ability of the ICLV has greatly improved, and latent variables play an important role in decision-making of mode choice. In the ICLV model, all parameters estimated meet the verification requirements. After adding latent variables, the constant value in *t* obviously declines, which means that BL model is less accurate if latent variable sets are not added to the model.

Of all the latent variable sets, convenience and service environment have a major impact on mode choice for public transport, while modal comfort has a minor impact on it. In the integrated model that includes latent variables, fare obviously influences the travelers' choice, which means significance of fare will be underestimated if latent variables are not added to the model.

5. Conclusions

Borrowing the concepts of latent variables and measured variables in the theory of consumer behavior, this research introduced them into an ICLV model to analyze those variables that affect mode choice for public transport. Measured and unmeasured factors influencing mode choice for public transport were analyzed to build the structural equation model. Quantitative causal relationships between various factors and degree of influence on mode choices were estimated.

Different estimation methods were applied to compare the final results. It is found that the ICLV model is stronger in explanatory ability and accuracy than the model without including latent variables. Therefore, it is more appropriate to apply structural equation model to build public transport mode choice model and calculate the value of the goodness of fit for latent variables, which can be described by measured variables by public transport users. The results also revealed that latent variable sets (e.g., service environment and waiting feelings) have a major impact on mode choice behavior. Therefore, it is obvious that, to achieve higher share of mode split, public transport operators should improve service quality by increasing those latent factors (e.g., convenience, personal safety, and service environment).

Further research in the future should focus on universality of types and quantity of latent variables and their degree of influence on mode choice behavior. Meanwhile, more efforts should be made for the collection of questionnaire survey data through social networking services to reduce the bias on the data.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

TABLE 7: Results of structural equation model of public transport services quality.

Pattern	Structural equation model		Convenience		Personal safety		Modal comfort		Service environment		Waiting feelings		Standard deviation of error term	
	Parameter	t	Parameter	t	Parameter	t	Parameter	t	Parameter	t	Parameter	t	Parameter	t
Structural pattern	Age	—	—	—	—	—	—	—	0.07	1.99	—	—	0.10	8.59
	Occupation	—	—	—	—	—	—	—	0.22	3.00	—	—	1.83	8.57
	Education background	—	—	—	—	—	—	0.06	1.98	—	—	—	0.29	8.57
	Monthly income	0.13	2.38	0.11	2.25	—	—	—	—	—	—	—	1.69	8.84
	Gender	—	—	—	—	—	—	—	—	—	—	0.08	2.06	8.57
	γ_1	0.58	—	—	—	—	—	—	—	—	—	—	1.78	6.58
	γ_2	0.85	4.57	—	—	—	—	—	—	—	—	—	0.64	1.98
	γ_3	0.47	4.71	—	—	—	—	—	—	—	—	—	1.85	7.72
	γ_4	—	—	0.71	—	—	—	—	—	—	—	—	1.17	6.08
	γ_5	—	—	0.83	7.41	—	—	—	—	—	—	—	0.72	3.86
	γ_6	—	—	0.57	6.49	—	—	—	—	—	—	—	0.93	7.54
	γ_7	—	—	—	—	—	0.81	—	—	—	—	—	0.68	4.23
	γ_8	—	—	—	—	—	0.72	7.33	—	—	—	—	0.97	5.98
	γ_9	—	—	—	—	—	0.31	3.55	—	—	—	—	2.39	8.45
	γ_{10}	—	—	—	—	—	—	—	0.64	—	—	—	1.13	7.86
	γ_{11}	—	—	—	—	—	—	—	0.73	9.41	—	—	0.86	7.40
	γ_{12}	—	—	—	—	—	—	—	0.70	7.35	—	—	1.41	7.57
	γ_{13}	—	—	—	—	—	—	—	0.64	6.79	—	—	1.16	7.97
γ_{14}	—	—	—	—	—	—	—	0.65	6.93	—	—	1.21	7.88	
γ_{15}	—	—	—	—	—	—	—	0.87	8.41	—	—	0.42	5.17	
γ_{16}	—	—	—	—	—	—	—	—	—	0.64	—	1.16	6.78	
γ_{17}	—	—	—	—	—	—	—	—	—	0.82	6.60	0.57	3.88	
γ_{18}	—	—	—	—	—	—	—	—	—	0.61	6.01	1.07	7.24	

All values of t are larger than 1.96, which means that $P < 0.05$ meets the verification requirement.

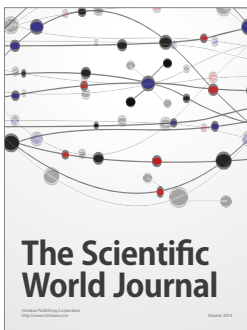
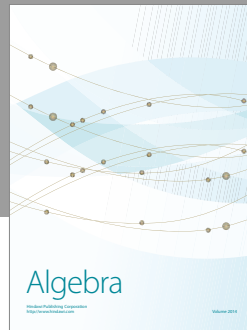
Acknowledgments

The authors wish to express their appreciation to the National Natural Science Foundation of China (51308569), Education Committee of Chongqing, China (14SKG03), Social Science Foundation of Chongqing, China (2015PY36), and Initial Research Fund of Chongqing Jiaotong University, China (2021500029), for sponsoring this research.

References

- [1] F. S. Koppelman and E. I. Pas, "Travel-choice behavior: models of perceptions, preference, and choice," *Transportation Research Record*, vol. 765, no. 1, pp. 26–33, 1980.
- [2] M. Ben-Akiva, A. de Palma, and P. Kanaroglou, "Dynamic model of peak period traffic congestion with elastic arrival rates," *Transportation Science*, vol. 20, no. 3, pp. 164–181, 1986.
- [3] F. S. Koppelman and C. H. Wen, "Alternative nested logit models: structure, properties and estimation," *Transportation Research Part B: Methodological*, vol. 32B, no. 5, pp. 289–298, 1998.
- [4] D. A. Hensher and A. J. Reyes, "Trip chaining as a barrier to the propensity to use public transport," *Transportation*, vol. 27, no. 4, pp. 341–361, 2000.
- [5] T. Schwanen, "The determinants of shopping duration on workdays in The Netherlands," *Journal of Transport Geography*, vol. 12, no. 1, pp. 35–48, 2004.
- [6] R. Mitra and R. N. Buliung, "Exploring differences in school travel mode choice behaviour between children and youth," *Transport Policy*, vol. 42, pp. 4–11, 2015.
- [7] C. Liu, Y. O. Susilo, and A. Karlström, "The influence of weather characteristics variability on individual's travel mode choice in different seasons and regions in Sweden," *Transport Policy*, vol. 41, pp. 147–158, 2015.
- [8] A. D. Palma and D. Rochat, "Mode choices for trips to work in Geneva: an empirical analysis," *Journal of Transport Geography*, vol. 8, no. 1, pp. 43–51, 2000.
- [9] C. Spearman, "General intelligence objectively determined and measured," *The American Journal of Psychology*, vol. 15, no. 2, pp. 201–293, 1904.
- [10] T. P. Gerber, "Market, state, or don't know? education, economic ideology, and voting in contemporary Russia," *Social Forces*, vol. 79, no. 2, pp. 477–521, 2000.
- [11] M. Mezzetti and F. C. Billari, "Bayesian correlated factor analysis of socio-demographic indicators," *Statistical Methods & Applications. Journal of the Italian Statistical Society*, vol. 14, no. 2, pp. 223–241, 2005.
- [12] L. R. Fabrigar, R. C. MacCallum, D. T. Wegener, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," *Psychological Methods*, vol. 4, no. 3, pp. 272–299, 1999.
- [13] K. A. Bollen, "Latent variables in psychology and the social sciences," *Annual Review of Psychology*, vol. 53, pp. 605–634, 2002.
- [14] K. Ashok, W. R. Dillon, and S. Yuan, "Extending discrete choice models to incorporate attitudinal and other latent variables," *Journal of Marketing Research*, vol. 39, no. 1, pp. 31–46, 2002.
- [15] P. Ebbes, M. Wedel, U. Böckenholt, and T. Steerneman, "Solving and testing for regressor-error (in)dependence when no instrumental variables are available: With new evidence for the effect of education on income," *Quantitative Marketing and Economics*, vol. 3, no. 4, pp. 365–392, 2005.
- [16] K. Train, D. McFadden, and A. Goett, "The incorporation of attitudes in econometric models of consumer choice," Working Paper, Cambridge Systematics, 1986.
- [17] M. Ben-Akiva and B. Boccara, "Integrated framework for travel behavior analysis," in *Proceedings of the The International Association of Travel Behavior Research (IATBR) conference*, Aix-en-Provence, France, 1987.
- [18] M. Ben-Akiva, D. McFadden, T. Gärling et al., "Extended framework for modeling choice behavior," *Marketing Letters*, vol. 10, no. 3, pp. 187–203, 1999.
- [19] L. H. Pendleton and J. S. Shonkwiler, "Valuing bundled attributes: a latent characteristics approach," *Land Economics*, vol. 77, no. 1, pp. 118–129, 2001.
- [20] T. Morikawa, M. Ben-Akiva, and D. McFadden, "Discrete choice models incorporating revealed preferences and psychometric data," in *Econometric Models in Marketing Advances in Econometrics: A Research Annual*, vol. 16, Elsevier Science, 2002.
- [21] T. Morikawa, M. Ben-Akiva, and D. McFadden, "Incorporating psychometric data in econometric choice models," Working Paper, Massachusetts Institute of Technology, 1996.
- [22] T. F. Golob, "Joint models of attitudes and behavior in evaluation of the San Diego I-15 congestion pricing project," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 6, pp. 495–514, 2001.
- [23] S. Mitra, S. Washington, E. Dumbaugh, and M. D. Meyer, "Governors highway safety associations and transportation planning: exploratory factor analysis and structural equation modeling," *Journal of Transportation and Statistics*, vol. 8, no. 1, pp. 57–75, 2005.
- [24] J.-Y. Lee, J.-H. Chung, and B. Son, "Analysis of traffic accident size for Korean highway using structural equation models," *Accident Analysis and Prevention*, vol. 40, no. 6, pp. 1955–1963, 2008.
- [25] D. A. Gopinath, *Modeling heterogeneity in discrete choice processes: application to travel demand. [Ph. D. Dissertation]*, Massachusetts Institute of Technology, 1995.
- [26] J. Walker and M. Ben-Akiva, "Generalized random utility model," *Mathematical Social Sciences*, vol. 43, no. 3, pp. 303–343, 2002.
- [27] M. V. Johansson, T. Heldt, and P. Johansson, "The effects of attitudes and personality traits on mode choice," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp. 507–525, 2006.
- [28] M. F. Yáñez, S. Raveau, and J. D. D. Ortúzar, "Inclusion of latent variables in mixed logit models: modelling and forecasting," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 9, pp. 744–753, 2010.
- [29] I. Politis, P. Papaioannou, and S. Basbas, "Integrated choice and latent variable models for evaluating flexible transport mode choice," *Research in Transportation Business & Management*, vol. 3, pp. 24–38, 2012.
- [30] Y. Shiftan, M. L. Outwater, and Y. Zhou, "Transit market research using structural equation modeling and attitudinal market segmentation," *Transport Policy*, vol. 15, no. 3, pp. 186–195, 2008.
- [31] K. Deutsch, S. Y. Yoon, and K. Goulias, "Modeling travel behavior and sense of place using a structural equation model," *Journal of Transport Geography*, vol. 28, pp. 155–163, 2013.

- [32] S. Nurlaela and C. Curtis, "Modeling household residential location choice and travel behavior and its relationship with public transport accessibility," *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 56–64, 2012.
- [33] D. McFadden, "The choice theory approach to market research," *Marketing Science*, vol. 5, no. 4, pp. 275–297, 1986.
- [34] M. Ben-Akiva, J. L. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou, "Integration of choice and latent variable models," in *Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*, H. S. Mahmassani, Ed., Pergamon, Amsterdam, Netherlands, 2002.
- [35] T. F. Golob, "Structural equation modeling for travel behavior research," *Transportation Research B: Methodological*, vol. 37, no. 1, pp. 1–25, 2003.
- [36] J. Kim, S. Rasouli, and H. Timmermans, "Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes. Application to intended purchase of electric cars," *Transportation Research Part A: Policy and Practice*, vol. 69, pp. 71–85, 2014.
- [37] D. Bolduc, M. Ben-Akiva, J. Walker, and A. Michaud, "Hybrid choice models with logit kernel: applicability to large scale models," in *Integrated Land-Use and Transportation Models: Behavioral Foundations*, M. Lee-Gosselin and S. Doherty, Eds., pp. 275–302, Elsevier, Oxford, UK, 2005.
- [38] M. Abou-Zeid, M. Ben-Akiva, M. Bierlaire, C. F. Choudhury, and S. Hess, "Attitudes and value of time heterogeneity," in *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2011.
- [39] A. Daly, S. Hess, B. Patruni, D. Potoglou, and C. Rohr, "Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behaviour," *Transportation*, vol. 39, no. 2, pp. 267–297, 2012.
- [40] M. Kamargianni and A. Polydoropoulou, "Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior," *Transportation Research Record*, no. 2382, pp. 151–161, 2013.
- [41] R. A. Daziano and D. Bolduc, "Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model," *Transportmetrica A: Transport Science*, vol. 9, no. 1, pp. 74–106, 2013.
- [42] C. R. Bhat and S. K. Dubey, "A new estimation approach to integrate latent psychological constructs in choice modeling," *Transportation Research Part B: Methodological*, vol. 67, pp. 68–85, 2014.
- [43] S. Raveau, M. F. Yáñez, and J. D. D. Ortúzar, "Practical and empirical identifiability of hybrid discrete choice models," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1374–1383, 2012.
- [44] D. Temme, M. Paulssen, and T. Dannewald, "Incorporating latent variables into discrete choice models — a simultaneous estimation approach using SEM software," *Business Research*, vol. 1, no. 2, pp. 220–237, 2008.
- [45] J. D. Ortúzar and L. G. Willumsen, *Modelling Transport*, John Wiley & Sons, Chichester, UK, 3rd edition, 2001.
- [46] D. Bolduc and A. Giroux, *The Integrated Choice and Latent Variable (ICLV) Model: Handout to Accompany the Estimation Software*, Département d'économique, Université Laval, 2005.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

