Transportation Research Part E 48 (2012) 545–554

Contents lists available at SciVerse ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Latent class nested logit model for analyzing high-speed rail access mode choice

Chieh-Hua Wen [a,*], Wei-Chung Wang [a], Chiang Fu [b]

[a] Department of Transportation Technology and Management, Feng Chia University, 100 Wenhwa Road, Seatwen, Taichung 40724, Taiwan, ROC
[b] Institute of Traffic and Transportation, National Chiao Tung University, 4F, 114 Chung Hsiao W. Road, Sec. 1, Taipei 10012, Taiwan, ROC

ARTICLE INFO

Article history:
Received 7 November 2010
Received in revised form 12 June 2011
Accepted 16 August 2011

Keywords:
Rail
Access mode
Logit
Latent class
Segmentation

ABSTRACT

This paper explores access mode choice behavior, using a survey data collected in Taiwan. The latent class nested logit model is used to capture flexible substitution patterns among alternatives and preference heterogeneity across individuals while simultaneously identifying the number, sizes, and characteristics of market segments. The results indicate that a four-segment latent class nested logit model with individual characteristics in segment membership functions is the most preferred specification. Most high-speed rail travelers were cost-sensitive to access modes, and thus strategies that reduce the access costs can be more effective than reducing the access times.

1. Introduction

Studies on intercity travel behavior often involve mode choice analyses, which provide predictions of mode shares with respect to changes in service levels. Some main intercity travel modes (e.g., train and air) with stations or terminals located at a distance from the origins and destinations of travelers require access and egress modes for the completion of a journey. Accordingly, the service levels of these access and egress modes influence travelers' choices of the main travel mode. Therefore, the improvement of an access mode to/from a rail station, for instance, is likely to encourage travelers to switch from other access/egress modes and even attract travelers of other main modes to use railways. Understanding access and egress mode choice behavior would offer valuable insights that can be used to develop effective strategies to improve ground transportation to/from stations or terminals.

Previous studies have addressed access mode choice behaviors and provided conceptual and methodological insights on how to model user behavior (e.g., Sobieniak et al., 1979; Korf and Demetsky, 1980, 1981). A discrete choice model such as a multinomial logit (MNL) (McFadden, 1973) is a standard approach for determining crucial variables affecting access mode choice. The estimation results aid in deducing policy implications for service improvement on access modes. However, a conventional access mode choice model uses identical parameter values for all the decision makers and does not consider individual preference heterogeneity toward access modes. Therefore, such a model may not properly explain the choice behaviors of all users.

To account for the heterogeneous preferences of travelers, some access mode choice studies have incorporated a market segmentation scheme into their models (e.g., Harvey, 1986; Psaraki and Abacoumkin, 2002). In such a segmentation procedure, data samples are first classified into a finite number of segments based on a single variable or a set of socioeconomic and trip characteristics. Subsequently, separate choice models are developed for different segments. This segmentation

* Corresponding author. Tel.: +886 4 24517250x4679; fax: +886 4 24520678.
E-mail address: chwen@fcu.edu.tw (C.-H. Wen).

1366-5545/$ - see front matter
doi:10.1016/j.tre.2011.09.002

approach, however, has a number of shortcomings (Bhat, 1997): (1) the number of segments increases significantly when a large number of segmentation variables is used; (2) determining the cut-off values for separating segments is rather arbitrary, especially for continuous segmentation variables. In contrast, a latent class choice model, which is a finite mixture model for the joint modeling selection of discrete alternatives and market segmentation, produces segment-specific parameter estimates to capture heterogeneous preferences for choice alternatives and identify decision-maker profiles for each segment (Kamakura and Russell, 1989; Gupta and Chintagunta, 1994). The latent class model overcomes the limitations of traditional segmentation approaches by distinguishing each segment in terms of a large number of segmentation variables and not requiring arbitrary cut-off values for defining segments.

The standard latent class choice model uses a mixture of two MNL probabilities: this exhibits the property of independence from irrelevant alternatives (IIA) and may fail to account for the existence of similarities among choice alternatives. The latent class nested logit (NL) model, an extension of the latent class MNL model, defines the choice probability using a standard NL formulation (McFadden, 1978) that allows for individual preference heterogeneity and similarities among alternatives (Kamakura et al., 1996). The latent class NL model groups alternatives into nests with dissimilarity (or inclusive value) parameters that capture flexible substitution patterns.

Although the latent class MNL models have been popularly applied in many fields, the use of a latent class NL model is still very limited. This study reports the development of latent class MNL and NL models to examine access mode choice for high-speed rail (HSR). The data source used to test the proposed models is an access mode choice survey, conducted by the Taiwan HSR Corporation. Our findings can be used to establish effective operational and marketing plans for improving access mode services.

2. Previous literature

The development of a disaggregate logit choice model for access modes (e.g., automobile driver, automobile passenger, transit, and taxi) to various terminals of intercity modes such as air, rail, and bus was initially reported in Sobieniak et al. (1979). The results indicate that the attributes of an access mode (e.g., travel cost, line-haul time, and waiting time), individual socioeconomic characteristics, and trip characteristics are important explanatory factors in modeling the choice of access modes. Among the options for improving access, shared-ride taxi services, in particular, were found to be highly favored by bus and rail passengers. Korf and Demetsky (1980, 1981) also applied the MNL models to analyze the choice of access modes to transit stations, which were classified into five types. In the context of airport access, Harvey (1986) estimated distinct MNL models based on trip purpose (business versus non-business) as the nature of the trip caused significant differences in choice behavior; travel time and cost are significant explanatory variables in an access mode decision. These studies successfully implemented the MNL models to identify the dominant factors affecting access mode choice.

While earlier access mode choice studies focused on one-dimensional access mode choice, later studies used flexible choice models to integrate access mode choices with other dimensional choices such as station choice and intercity mode choice. Fan et al. (1993) used the NL framework to develop an access mode and station choice model that places access mode choice at the upper level and access station choice at the lower level. Debrezion et al. (2009) confirmed that identical NL structures were appropriate for modeling access mode choice and departure-station choice. To analyze urban or intercity travel behavior, main and access mode choice models can be developed using the NL model, which puts the main mode choice at the upper level and access mode choice at the lower level (Algers, 1993; Polydoropoulou and Ben-Akiva, 2001). Although access mode choice can be integrated with other choices, the present study only focuses on access mode choice.

To uncover the heterogeneous preferences of users, access mode choice studies have often applied a market segmentation approach to produce a small number of segments. Segmentation and choice modeling are typically implemented in a two-step process. For segmentation, samples are divided into a finite number of segments, each containing heterogeneous characteristics. Separate choice models are further estimated to produce segment-specific parameters. Two distinct segmentation techniques have been used in the literature (Pas and Huber, 1992). The standard approach comprises a priori segmentation, which defines segments based on one or more variables. For example, Tsamboulas et al. (1992) reported the development of MNL models combined with market segmentation to examine metro access modes; he concluded that trip purpose is an appropriate variable for market segmentation. Bekhor and Elgar (2007) used a lifestyle variable (i.e., investment in car mobility) to define three segments and estimated separate MNL models for these segments. In order to study commuters' access modes to rail transit, Rastogi and Rao (2009) considered four segmentation variables, namely, household income, type of accommodation, dependency factor, and occupation level of commuters; they estimated different NL models using each segmentation variable. An alternative approach comprises post hoc segmentation, which determines the number and profiles of segments through multivariate statistical methods such as cluster analysis (e.g., Psaraki and Abacoumkin, 2002; Outwater et al., 2004a, 2004b; Shiftan et al., 2008). When a large set of segmentation variables are used (e.g., socioeconomic, trip, and attitude), this approach often results in numerous segments.

The latent class choice model developed by Kamakura and Russell (1989) makes it possible to simultaneously perform choice modeling and market segmentation to identify the segment-specific preference parameters, individual profiles of each segment, and segment sizes. The latent class model captures the variations in preference parameters with a finite set of different values. Alternatively, the mixed logit model specifies a continuous probability density function (e.g., normal distribution) for preference parameters (Train, 2003). Greene and Hensher (2003) contrasted the latent class with the mixed logit in a model formulation and estimation approach; their study had inconclusive results about which one had superior model performance. Both the latent class and mixed logit models accommodate individual taste heterogeneity, but the latent class can explicitly identify the number, sizes, and characteristics of segments.

Recently, the latent class modeling approach has become popular in market segmentation analyses of individual choice behavior. For example, Bhat (1997) used the latent class model (referred to in that study as the endogenous market segmentation approach) to describe intercity mode choice behavior. The results indicated that level-of-service attributes (e.g., travel times and costs) affect mode choice and that traveler socioeconomic and trip characteristics (e.g., income and trip distance), as segmentation variables, determine the profiles of each segment. Zhang et al. (2009) incorporated different types of group decision-making mechanisms as latent classes into a household car choice model to enhance model accuracy. In the context of airline choice, Wen and Lai (2010) illustrated that the latent class model outperforms the standard MNL model and can further improve the goodness-of-fit if individual socioeconomic and trip characteristics are included as segmentation variables. Arunotayanun and Polak (2011) applied both latent class and mixed logit models in their study of shippers' mode choice; the empirical results showed that the traditional segmentation approach that uses a single segmentation variable fails to completely capture taste heterogeneity.

All the studies on mode and travel-related choice behaviors have adopted the MNL formulation in both an unconditional choice probability function that captures individual preferences for alternatives and a membership probability function that maps individual characteristics to a segment profile. The latent class MNL model still has the IIA property within the segments. To improve the latent class MNL model, Kamakura et al. (1996) proposed a latent class NL model that allows flexible substitution patterns among alternatives in a common group. Up to this point, the applications of the latent class NL model have been limited (Swait, 2003; Bodapati and Gupta, 2004).

3. Model structure

The latent class model calibrates segment-specific sets of parameters to consider preference heterogeneity across individuals (Kamakura and Russell, 1989). Given a finite and fixed number of S segments, and given that a particular traveler t belongs to segment s (s = 1, 2,..., S), the utility function of t for any access mode m can be expressed as

$$U_{tm|s} = \beta_s' X_{tm} + \varepsilon_{tm|s} \tag{1}$$

where X_{tm} is a vector of observable attributes, β_s is a vector of unknown segment-specific parameters, and $\varepsilon_{tm|s}$ expresses the random error of the utility function.

The formulation of the latent class choice model consists of two probabilities:

$$P_t(m) = \sum_s P_t(m|s) \cdot H_t(s) \tag{2}$$

Within segment s, $P_t(m|s)$ is the conditional probability that traveler t chooses alternative m. The segment membership function $H_t(s)$ represents the probability that traveler t belongs to segment s.

The latent class MNL model adopts the standard MNL formulation for these two probabilities (Gupta and Chintagunta, 1994):

$$P_t(m|s) = \frac{\exp(\beta_s' X_{tm})}{\sum_{m'} \exp(\beta_s' X_{tm'})} \tag{3}$$

$$H_t(s) = \frac{\exp(w_s' z_n)}{\sum_{s'} \exp(w_{s'}' z_n)} \tag{4}$$

where z_n is a vector of membership function variables that consists of individual characteristics and w_s is a vector of unknown parameters for segment s.

The probability formulation of the latent class NL model (a two-level NL model) can be expressed as follows (Kamakura et al., 1996):

$$P_t(m) = \sum_s [P_t(m|n,s) \cdot P_t(n|s)] \cdot H_t(s) \tag{5}$$

$$P_t(m|n,s) = \frac{\exp(\beta_s' X_{tm}/\lambda_n^s)}{\sum_{m' \in N_n^s} \exp(\beta_s' X_{tm'}/\lambda_n^s)} \tag{6}$$

$$P_t(n|s) = \frac{\exp(\lambda_n^s \Gamma_{tn}^s)}{\sum_{n'} \exp(\lambda_{n'}^s \Gamma_{tn'}^s)} \tag{7}$$

$$\Gamma_{tn}^s = \ln \sum_{m' \in N_n^s} [\exp(\beta_s' X_{tm'}/\lambda_n^s)] \tag{8}$$

$$H_t(s) = \frac{\exp(w_s' z_n)}{\sum_{s'} \exp(w_{s'}' z_n)} \tag{9}$$

Within segment s, $P_t(m|n,s)$ is the probability that traveler t chooses alternative m in nest n. Within segment s, $P_t(n|s)$ is the probability that traveler t is in nest n. Within segment s, N_n^s is the set of all alternatives included in nest n, λ_n^s is the dissimilarity parameter for nest n, and Γ_{tn}^s is the logsum variable of nest n.

The dissimilarity parameter captures the similarities between pairs of alternatives in the nest. Similar to the standard NL model, if the condition $0 < \lambda_n^s \leqslant 1$ for all s and n holds, the model is consistent with utility maximization for all possible values of the explanatory variables and will not yield counterintuitive results (Ortúzar and Willumsen, 2001; Train, 2003). The latent class MNL model is a restriction of the latent class NL model; that is, when all the dissimilarity parameters in all the segments are equal to one in the latent class NL model, this model collapses to the latent class MNL model. Moreover, the standard NL model can be regarded as a special case in which the latent class NL has only one segment.

The segment membership function may contain a set of segment-specific constants, but the coefficients of one segment should be set to zero for identification. The inclusion of individual characteristics in the segment membership functions allows profiles of typical members to be obtained for each segment. Given any individual and any segment, a segmentation analysis will be able to estimate the probability of that individual belonging to that segment. The segment size is calculated as the average of the individual membership probabilities.

The latent class MNL and NL models with different numbers of segments should be estimated and their model performance should be assessed in order to determine the best number of segments. Bayesian information criterion (BIC), Akaike Information Criterion (AIC), and the adjusted likelihood ratio index can be used to evaluate the performance of various models (Walker and Li, 2007). BIC is defined as $-2LL + K\ln(N)$, where LL is the final log-likelihood value, K is the number of parameters, and N is the sample size. The AIC formula is equal to $-2(LL - K)$. The adjusted likelihood ratio index is defined as $1 - (LL - K)/LL^*$, where LL^* is the log-likelihood value when all the parameters are zero. Lower BIC and AIC values indicate more preferred latent class models, and a model with a higher adjusted likelihood ratio index fits the data well.

Initially, estimations of the standard MNL and NL models could identify important explanatory variables as well as plausible and statistically acceptable nested structures. Analysts can estimate the latent class NL models using parameter estimates obtained by the standard NL models as the starting values. Likewise, analysts can use the latent class MNL parameter estimates as starting values to estimate the latent class NL model. The latent class MNL model requires the simultaneous calibration of the utility function and membership parameters; the latent class NL model requires additional dissimilarity parameters to be estimated. The present study calibrated the model parameters using a GAUSS statistical software package (Aptech Systems, 2008).

4. Data

The empirical case deals with an access mode choice of the Taiwan HSR. Its 345-km route passes through the western corridor of the island. This HSR has the advantage of speed; however, some of the stations are located far from metropolitan areas. Therefore, HSR travelers have to use other travel modes to get from their points of origin to the stations or from the stations to their final destinations. If the HSR access modes do not operate at high levels of service, few travelers will use the HSR. To improve the access mode service, the Taiwan High Speed Rail Corporation conducted both revealed-preference (RP) and stated-preference (SP) surveys in 2007 (Taiwan High Speed Rail Corporation, 2007). The survey data were collected at six of the eight terminals; two stations were excluded because both have an extensive bus/Metro network with convenient transit access services. Each sampled terminal yielded 200 valid responses, and the final number of respondents was 1200.

The respondents' background information (e.g., gender, age, education, occupation, and income), trip characteristics (trip purpose, travel group size, and access distance), and mode choice preferences were obtained. In the RP data set, seven existing access modes were available to HSR travelers: city bus, train, car driver, car passenger, motorcycle driver, motorcycle passenger, and taxi. Each car traveler was either a driver or a passenger. The car driver alternative involved the fuel cost of the access trip and the parking fee at the HSR station. The car passenger alternative involved a driver offering a lift to the HSR station and dropping the passenger off at the curbside, which included no parking. Similar definitions were also applied to motorcycles. The taxi mode was defined as an automobile that transported passengers for a fare determined by the travel mileage. The RP data included access travel costs and times for the chosen access mode and previously used access mode(s).

In addition to the currently available access modes, respondents were asked about their possible use of one hypothetical mode (i.e., respondents were asked if they would use express shuttle buses if such a service was made available). The SP data consisted of eight access modes, and the SP experimental design included eight access modes and four attributes for each mode, namely access travel cost, parking cost, access travel time, and waiting time. Each attribute had three levels in the experiments, and a full factorial design produced a large number of possible combinations. A fractional factorial design was used to reduce the total number of combinations to 27, using an orthogonal table of $L_{27}(3^{13})$. Blocked designs were then employed to classify the 27 scenarios into 9 subsets. Each respondent was asked to evaluate three hypothetical scenarios in the SP experiments, and the number of observations used in the model estimation was 3600.

5. Estimation results

5.1. Standard MNL and NL models

The SP data were used to estimate access mode choice models. Eight mode-specific constants were available to be specified, and the motorcycle passenger alternative was selected as the reference mode because of its low percentage of market share. Four level-of-service attributes (i.e., access cost/income, parking fee/income, access time/distance, and waiting time) were specified as generic variables based on the likelihood ratio test versus a specification with alternative-specific variables. The access cost and parking fee variables, combined with income, were found to improve the model fit, indicating that the access cost and parking fee sensitivities of travelers decrease with an increase in personal income. Similarly, the access time divided by access distance had better explanatory power, indicating that the access time sensitivity decreases with an increase in access distance of travelers.

Table 1 presents the estimation results of the standard MNL and NL models. The result of the standard MNL model indicates that four level-of-service variables associated with costs and times had negative signs, as expected, while, except for the waiting time, the other three variables were significantly different from zero at the 5% significance level (t-value > 1.96). However, the standard MNL model did not have a satisfactory goodness-of-fit because the likelihood ratio was only 0.0878. Notably, if travelers' socioeconomic and trip characteristics were included as alternative-specific variables, the MNL model fit would be slightly increased. However, in order to demonstrate the superiority of latent class models, the specifications of the utility functions were kept simple.

After analyzing various nested structures that are behaviorally interpretable, the results of three standard NL models with dissimilarity estimates within a logical range are reported in Table 1. NL Model 1 in the second column included the public transport modes (i.e., city bus, train, and express shuttle bus) in a single nest, while the other alternatives were not grouped into nests. NL Model 2 consists of one nest with the car-driver and car-passenger alternatives, while the other modes are included as single alternatives. NL Model 3 includes two nests: car modes in one nest and public transport modes in the other. The coefficients of the access cost, parking fee, access time, and waiting time have the expected signs and magnitudes comparable to those of the standard MNL model.

In all three NL models, the t-values of the dissimilarity estimates were significantly different from one at the 5% level of significance. In addition, the likelihood ratio tests indicated that the three NL models significantly outperformed the standard MNL model. NL Model 3 contained the preferred specification, with the other NL models rejected based on the likelihood ratio tests at the 5% significance level (NL Model 3 versus NL Model 1: $\chi^2 = 6.68 > 3.84$; NL Model 3 versus NL Model 2: $\chi^2 = 35.84 > 3.84$).

5.2. Latent class MNL models

Based on the utility specification of the standard MNL models, this study estimated the latent class MNL models with only segment-specific constants in the membership functions. Table 2 shows some measures that can aid in the selection of the proper number of segments for the latent class MNL models. Compared with the standard MNL model (one-segment solution), the latent class MNL models have a significantly improved goodness-of-fit. However, as the number of segments increases, the improvement in the fit diminishes. The four-segment solution is preferred because it has the lowest BIC and AIC values as well as the largest log-likelihood and likelihood ratio. The latent class MNL models with five or more segments were not estimable because the membership probabilities in some segments were relatively small, leading to convergence problems.

Table 3 shows the parameter estimates of the preferred four-segment latent class MNL model. The coefficient of membership function constant for Segment 4 is set to zero for identification. Segment 1, the largest segment with 35% of the total,

Table 1
Estimation results for standard MNL and NL models (t-values in parentheses).

	MNL Model	NL Model 1	NL Model 2	NL Model 3
City bus constant	0.549 (5.91)	0.749 (8.22)	0.547 (5.86)	0.752 (8.29)
Train constant	0.623 (5.59)	0.965 (10.68)	0.661 (5.96)	0.994 (11.35)
Car driver constant	1.118 (10.21)	1.076 (9.53)	1.614 (9.46)	1.651 (10.43)
Car passenger constant	1.599 (23.29)	1.551 (22.61)	1.788 (19.08)	1.777 (19.31)
Motorcycle driver constant	0.307 (3.67)	0.310 (3.54)	0.260 (2.92)	0.248 (2.79)
Taxi constant	1.086 (10.75)	0.934 (9.62)	1.060 (9.62)	0.902 (9.20)
Express shuttle bus constant	0.053 (0.50)	0.584 (4.98)	0.035 (0.34)	0.586 (5.02)
Access cost/income	−0.096 (−8.45)	−0.072 (−7.27)	−0.093 (−8.86)	−0.068 (−6.91)
Parking fee/income	−0.159 (−4.29)	−0.161 (−4.01)	−0.093 (−2.06)	−0.750 (−1.67)
Access time/distance	−0.061 (−3.32)	−0.035 (−2.55)	−0.048 (−2.87)	−0.026 (−2.14)
Waiting time	−0.037 (−1.75)	−0.020 (−1.29)	−0.045 (−2.22)	−0.023 (−1.56)
Dissimilarity (t-value vs. 1)				
Public modes nest		0.490 (7.28)		0.473 (7.52)
Car modes nest			0.451 (2.64)	0.345 (3.26)
Number of parameters	11	12	12	13
Log-likelihood at zero	−7165.514	−7165.514	−7165.514	−7165.514
Final log-likelihood	−6536.598	−6520.576	−6534.158	−6517.237
Likelihood ratio	0.0878	0.0900	0.0881	0.0905
Adjusted likelihood ratio	0.0862	0.0883	0.0864	0.0887
Likelihood ratio test vs. MNL		$\chi^2 = 32.0 > 3.84$	$\chi^2 = 4.9 > 3.84$	$\chi^2 = 38.7 > 5.99$

has significant values for the access cost and waiting time variables. The most favored mode in this segment is car passenger (as indicated by the mode-specific constant). Segment 2, with a 27% share, has relatively significant parking fee and waiting time variables. The travelers in Segment 2 generally prefer to reach HSR stations by driving cars, riding the express shuttle bus, or taking taxis. The car-driving travelers are obliged to pay parking fees, and were therefore sensitive to this variable. The travelers in this segment appear to care about waiting time. Travelers in Segment 3 (23%) are very sensitive to access cost; the parking fee variable had a counterintuitive sign. Travelers in Segment 3 prefer low-cost modes such as the train or motorcycle driving, and they are less likely to choose a taxi or express shuttle bus. Segment 4 (15% of travelers) has significant access cost and access time variables; it consists of cost-sensitive and time-sensitive travelers who prefer the city bus, taxi, or express shuttle bus.

5.3. Latent class NL models

Table 4 lists the results for the preferred three-segment latent class NL model, which corresponds to the preferred standard NL model (NL Model 3) and includes only segment-specific constants in the membership functions. These latent class NL models allow differential dissimilarity parameters across segments and impose restrictions on the dissimilarity estimates within reasonable range. This three-segment latent class NL model outperformed the three-segment latent class MNL model in terms of the goodness-of-fit measures and the likelihood ratio test at the 5% significance level. While some dissimilarity estimates are either insignificant (e.g., car-mode nest in segment 1) or constrained to one, nested structures still hold for all the segments. Most dissimilarity parameters are statistically significant, indicating that a consideration of the similarities among alternatives is critical. Travelers in Segment 1 (the largest segment with 48% of the total) are sensitive to the access cost; their preferred modes are car passengers and drivers. Segment 2 (28%) has relatively significant values for the access cost, waiting time, and access time variables; these time-sensitive travelers generally prefer taxis, driving cars, or the express shuttle bus. Segment 3 (24%) has many cost-sensitive travelers who prefer the train or motorcycle (driving alone); they are less likely to choose a taxi.

Table 5 presents the estimation result for the preferred four-segment latent class NL model. The four-segment latent class NL model has the best likelihood ratio and AIC values compared with the three-segment latent class NL and four-segment latent class MNL models, while the BIC suggests that the four-segment latent class MNL model is superior. Although increasing the number of segments would decrease the significance of dissimilarity parameters, the four-segment latent class NL model still captures the inter-alternative correlation in some segments. The four-segment latent class NL is preferred over the latent class MNL model because it jointly accommodates flexible substitution patterns among alternatives and variations in taste parameters.

Table 6 presents the estimation results for the preferred four-segment latent class NL model, including the individual characteristics in the segment membership function. Although many segmentation variables were tested, only three variables were statistically significant in at least one segment. The personal income and trip distance are continuous variables, while the trip purpose is a dummy variable (business = 1; non-business = 0). Because Segment 4 is chosen as a base, all of the membership coefficients in Segment 4 are normalized to zero. The estimates for other segments are interpreted relative to Segment 4. This four-segment latent class NL model with individual characteristics in segment membership functions statistically outperforms the corresponding latent class NL model without individual characteristics (in Table 5). Travelers in Segment 1 (34%) are insensitive to the access cost, compared with travelers in the other segments; they prefer to be car passengers, indicated by the alternative specific constants. The individual profiles are medium-income business travelers with long access distances. The travelers in Segment 2 (28% of the total) prefer to drive cars, take taxis, or ride an express shuttle bus; they are very sensitive to the access time and waiting time. The results provide evidence that most travelers in this segment are high-income individuals who cross medium distances to access the rail stations in order to travel for business. Segment 3 consists of cost-sensitive and time-insensitive travelers who prefer low-cost travel modes such as trains or motorcycles. In this segment, most individuals have low incomes; they move over short distances to access the rail stations, and travel for purposes other than business. Segment 4 comprises the smallest percentage of travelers (15%) who have a relatively high income. They are sensitive to the access cost and parking fee and prefer to use a city bus, express shuttle bus, or taxi.

5.4. Discussion

As expected, the most important determinants of access mode choice include the access cost, access time, parking cost, and waiting time, which is consistent with the findings of previous studies. With regard to access modes to HSR stations, the

Table 2
Goodness-of-fit measures of latent class MNL models.

Segment	Number of parameters	Final log-likelihood	Likelihood ratio	Adjusted likelihood ratio	BIC	AIC
1	11	−6536.598	0.0878	0.0862	13,163	13,095
2	23	−5643.180	0.2125	0.2093	11,475	11,332
3	35	−5145.240	0.2819	0.2771	10,577	10,360
4	47	−4716.195	0.3418	0.3353	9817	9526

Table 3
Estimation results for four-segment latent class MNL model.

	Segment 1	Segment 2	Segment 3	Segment 4
City bus constant	1.868 (4.41)	4.226 (5.28)	−1.995 (−5.92)	3.712 (7.60)
Train constant	0.640 (1.01)	2.742 (2.76)	0.558 (4.54)	−0.373 (−0.31)
Car driver constant	1.440 (2.67)	5.332 (8.13)	−3.395 (−6.40)	2.218 (3.02)
Car passenger constant	4.781 (13.48)	2.073 (3.00)	−1.677 (−5.82)	1.802 (3.58)
Motorcycle driver constant	−1.091 (−1.52)	0.600 (0.77)	0.165 (1.33)	−0.679 (−0.78)
Taxi constant	1.763 (3.59)	5.265 (8.02)	−2.716 (−2.32)	2.785 (5.45)
Express shuttle bus constant	1.268 (2.63)	5.381 (7.99)	−3.230 (−4.01)	3.417 (7.32)
Access cost/income	−0.115 (−2.45)	−0.071 (−4.30)	−0.393 (−2.96)	−0.338 (−7.56)
Parking fee/income	−0.346 (−1.64)	−0.097 (−2.15)	0.085 (0.74)	−0.607 (−1.62)
Access time/distance	−0.073 (−1.57)	−0.071 (−1.44)	−0.079 (−1.73)	−0.119 (−2.26)
Waiting time	−0.244 (−2.10)	−2.655 (−6.51)	−0.098 (−1.73)	0.024 (0.55)
Membership function constant	0.847 (8.36)	0.604 (5.81)	0.451 (4.31)	
Segment size (%)	35	27	23	15

Table 4
Estimation results for three-segment latent class NL model without individual characteristics in membership functions.

	Segment 1	Segment 2	Segment 3
City bus constant	3.137 (10.97)	4.340 (6.54)	−1.529 (−5.05)
Train constant	3.051 (10.68)	2.367 (2.48)	0.574 (4.76)
Car driver constant	3.694 (1.59)	5.203 (8.67)	−1.436 (−5.24)
Car passenger constant	4.187 (14.66)	4.329 (4.35)	−1.501 (−5.70)
Motorcycle driver constant	−1.625 (−2.07)	0.482 (0.65)	0.296 (4.44)
Taxi constant	1.203 (3.62)	5.163 (8.55)	−2.928 (−2.45)
Express shuttle bus constant	3.154 (11.04)	5.143 (8.51)	−2.128 (−3.68)
Access cost/income	−0.036 (−6.77)	−0.063 (−4.02)	−0.382 (−2.96)
Parking fee/income	−0.084 (−0.23)	−0.054 (−1.29)	−0.059 (−1.48)
Access time/distance	0.001 (0.19)	−0.080 (−1.66)	−0.061 (−1.48)
Waiting time	0.003 (1.05)	−1.767 (−7.70)	−0.088 (−1.76)
Dissimilarity (*t*-value vs. 1)			
Public modes nest	0.050 (–)	1.000 (–)	0.774 (1.88)
Car modes nest	0.180 (0.99)	0.265 (2.57)	0.050 (–)
Membership function constant	0.684 (8.95)	0.207 (2.51)	
Segment size (%)	48	28	24
Number of parameters	38		
Final log-likelihood	−5032.973		
Likelihood ratio	0.2976		
Adjusted likelihood ratio	0.2923		
BIC	10,377		
AIC	10,142		
Likelihood ratio test vs. three-segment latent class MNL	$\chi^2 = 224.53 > 7.81$		

results of the NL models identify two behaviorally interpretable nests (private cars in one nest and public modes in the other). By considering the similarity among access modes, the NL models provide additional behavioral insights and improve the goodness-of-fit.

The latent class model identified market segments in terms of access mode attributes and travelers' characteristics. Each segment has a unique set of taste parameters to capture preference heterogeneity across individuals. The use of latent class MNL models significantly improves the goodness-of-fit relative to the standard MNL and NL models, indicating that modeling access mode choice must account for individual heterogeneity.

While most studies have used the latent class model with the MNL formulation, this research avoided the shortcoming of the IIA property by using the latent class NL model. The latent class NL model simultaneously accounts for flexible substitution patterns among alternatives and preference heterogeneity across individuals. Interestingly, as the number of segments increases, the joint effects of the dissimilarity and taste parameters appear to decrease the significance of the dissimilarity parameters. This result is similar to the phenomenon of confounding between inter-alternative correlation and inter-agent taste heterogeneity when the mixed generalized extreme value models (e.g., mixed cross-NL model) are used (Hess et al., 2005). The preferred latent class NL model is the four-segment solution with the city bus, train, and express shuttle bus in one nest and private cars in another. By jointly accommodating a flexible structure for the similarity among alternatives and variations in taste parameters, the four-segment latent class NL model is preferred over the other segment results. The four-segment latent class NL model with individual characteristics in segment membership functions has the best likelihood ratio, BIC, and AIC values and can reveal the individual characteristics of each segment.

The preferred latent class NL model captures the variations in preference parameters with four sets of estimates. The mixed logit model specifies a distributional function (e.g., normal distribution) for the coefficients of observable explanatory

Table 5
Estimation results for four-segment latent class NL model without individual characteristics in membership functions.

	Segment 1	Segment 2	Segment 3	Segment 4
City bus constant	1.877 (4.75)	4.027 (5.12)	−1.998 (−5.28)	3.753 (8.35)
Train constant	1.703 (4.36)	2.536 (2.59)	0.524 (3.17)	1.090 (0.76)
Car driver constant	1.236 (2.25)	5.216 (8.66)	−3.320 (−4.74)	2.082 (3.16)
Car passenger constant	4.627 (12.50)	3.600 (1.44)	−1.734 (−6.21)	1.637 (3.27)
Motorcycle driver constant	−1.174 (−1.65)	0.462 (0.64)	0.169 (0.99)	−0.737 (−0.84)
Taxi constant	1.337 (2.91)	5.132 (8.56)	−2.907 (−2.70)	2.454 (4.21)
Express shuttle bus constant	1.861 (4.76)	5.306 (8.16)	−2.932 (−3.57)	3.639 (7.65)
Access cost/income	−0.065 (−1.93)	−0.076 (−3.37)	−0.468 (−2.86)	−0.247 (−3.00)
Parking fee/income	−0.337 (−1.82)	−0.106 (−1.31)	0.086 (0.52)	−0.596 (−1.99)
Access time/distance	−0.090 (−2.07)	−0.074 (−1.85)	−0.070 (−1.33)	−0.074 (−1.49)
Waiting time	−0.047 (−1.62)	−2.667 (−5.65)	−0.085 (−1.26)	0.017 (0.54)
Dissimilarity (*t*-value vs. 1)				
Public modes nest	0.060 (29.56)	1.000 (−)	0.929 (0.48)	0.635 (1.46)
Car modes nest	1.000 (−)	0.491 (0.59)	1.000 (−)	1.000 (−)
Membership function constant	0.856 (8.40)	0.629 (5.90)	0.448 (4.19)	
Segment size	35%	27%	23%	15%
Number of parameters	51			
Final log-likelihood	−4707.503			
Likelihood ratio	0.3430			
Adjusted likelihood ratio	0.3359			
BIC	9833			
AIC	9517			
Likelihood ratio test vs. four-segment latent class MNL	χ^2 = 17.38 > 9.49			

variables and estimates the parameters (e.g., mean and standard deviation) of the specific distribution. The latent class and mixed logit models account for taste variations in different ways, but the latent class can explicitly identify the number, sizes, and characteristics of segments.

This study identified separate market segments for access services to HSR stations. Strategic plans for improving access modes will be effective when accounting for the needs of each market segment. Most HSR travelers were cost-sensitive to access modes, and thus strategies that reduce the access costs can be more effective than reducing the access times. In particular, low-income and non-business travelers (approximately 23% of travelers) are very sensitive to access costs. Low-fare strategies associated with public access modes (e.g., free access buses) are likely to be successful if these are aimed at HSR travelers. Segment 4 (15%) was sensitive to parking fees. Raising parking fees will discourage many travelers from driving their cars to the stations. Segment 2 (28%) was sensitive to waiting time; travelers in this segment had high regard for their time, enjoyed high incomes, and traveled for business. These travelers will seek out public transportation if it offers high-frequency services.

The Taiwan HSR has not offered express shuttle bus access in the past, but express shuttle bus services are under consideration. They resemble trains and city buses in terms of common unobserved attributes such as comfort and convenience. This study suggests that express shuttle buses must differentiate their service quality from the services offered by trains and city buses. HSR travelers who have higher incomes and travel for business purpose would prefer to use these express shuttle buses. In order to attract sufficient ridership, express shuttle buses must provide high-frequency services with shorter travel times and reasonable fares.

6. Conclusions

This study explored the access mode choices of HSR travelers using the conventional MNL and NL models and unconventional latent class MNL and NL models. The data used to test these models were obtained from an access mode choice survey for Taiwan HSR. The latent class models identified HSR travelers' heterogeneous preferences toward access modes and market segments in terms of individual socioeconomic and trip characteristics. The latent class approach has become popular in modeling individual choice behavior, but most works have applied a latent class MNL model that exhibits the IIA property. The contribution of this paper is the development of latent class NL models to capture the flexible correlation structure between access modes and the preference heterogeneity across travelers. The latent class NL model overcomes the shortcoming of the IIA property and can be feasibly estimated.

The empirical results indicated that some access modes have a certain degree of correlation in observed utility and should be grouped into nests so as to capture their similarities. This provides evidence that the standard NL model is preferred over the MNL model. The latent class MNL and NL models significantly improved the goodness-of-fit over the standard MNL and NL models, indicating the existence of individual preference heterogeneity. The four-segment latent class NL model with individual socioeconomic and trip variables in membership functions was the most preferred because it had the best goodness-of-fit and the ability to accommodate a flexible structure for the similarity among alternatives and variations in taste parameters.

Table 6
Estimation results for four-segment latent class NL model with individual characteristics in membership functions.

	Segment 1	Segment 2	Segment 3	Segment 4
City bus constant	1.874 (4.45)	4.068 (4.90)	−1.935 (−4.69)	3.783 (8.77)
Train constant	1.687 (3.97)	2.536 (2.58)	0.535 (2.32)	1.118 (0.49)
Car driver constant	1.235 (2.09)	5.195 (8.79)	−3.339 (−3.36)	2.090 (3.30)
Car passenger constant	4.630 (11.64)	3.632 (0.73)	−1.745 (−6.38)	1.643 (3.04)
Motorcycle driver constant	−1.182 (−1.65)	0.465 (0.67)	0.185 (0.73)	−0.739 (−0.85)
Taxi constant	1.317 (2.82)	5.130 (8.88)	−2.900 (−3.75)	2.438 (3.66)
Express shuttle bus constant	1.858 (4.46)	5.313 (7.76)	−2.949 (−3.68)	3.649 (6.71)
Access cost/income	−0.064 (−1.78)	−0.073 (−2.21)	−0.431 (−2.04)	−0.254 (−2.24)
Parking fee/income	−0.328 (−1.86)	−0.092 (−0.65)	0.065 (0.26)	−0.586 (−2.20)
Access time/distance	−0.090 (−1.54)	−0.072 (−2.06)	−0.071 (−1.12)	−0.075 (−1.02)
Waiting time	−0.047 (−1.21)	−2.601 (−4.36)	−0.089 (−1.08)	0.017 (0.45)
Dissimilarity (*t*-value vs. 1)				
Public modes nest	0.060 (25.07)	1.000 (−)	0.919 (0.50)	0.647 (1.67)
Car modes nest	1.000 (−)	0.489 (0.29)	1.000 (−)	1.000 (−)
Membership function constant	0.371 (0.98)	0.236 (0.58)	1.652 (4.10)	
Personal income	−0.004 (−1.36)	−0.001 (−0.44)	−0.020 (−5.40)	
Trip purpose	0.178 (0.79)	0.819 (3.47)	−0.728 (−2.81)	
Trip distance	0.411 (1.96)	0.046 (0.20)	−0.054 (−0.24)	
Segment size	34%	28%	23%	15%
Number of parameters	60			
Final log-likelihood	−4621.882			
Likelihood ratio	0.3550			
Adjusted likelihood ratio	0.3466			
BIC	9735			
AIC	9364			
Likelihood ratio test vs. four-segment latent class MNL without individual characteristics	χ^2 = 171.24 > 16.92			

Four market segments for access services to HSR stations were identified. Strategies for improving access modes should consider travelers' heterogeneous preferences across segments. Most HSR travelers were cost-sensitive to access modes. Therefore, low-fare strategies associated with public access modes, for instance, can be more effective than reducing the access times in all segments. Because private access modes have high market shares, public access modes must deliver high service quality to attract users of private access modes.

A number of directions can be considered for future research. The mixed logit model is a very flexible choice model that accounts for random taste variation. This model employs a continuous distribution to represent variations in taste parameters, while the latent class model considers a finite set of distinct values as parameters. Future research could estimate and compare both latent class and mixed NL models from the perspective of access mode choice. Future research can develop and estimate a more general structure that will be able to integrate access and egress mode choices into intercity travel mode choices.

Acknowledgements

The authors would like to thank two anonymous referees and the editor-in-chief for their valuable suggestions and comments.

References

Algers, S., 1993. Integrated structure of long-distance travel behavior models in Sweden. Transportation Research Record 1413, 141–149.
Aptech Systems, 2008. GAUSS Version 9.0 User Guide. Aptech Systems, Inc..
Arunotayanun, K., Polak, J.W., 2011. Taste heterogeneity and market segmentation in freight shippers' mode choice behaviour. Transportation Research Part E 45 (2), 138–148.
Bekhor, S., Elgar, A., 2007. Investment in mobility by car as an explanatory variable for market segmentation. Journal of Public Transportation 10 (2), 17–32.
Bhat, C.R., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34–48.
Bodapati, A.V., Gupta, S., 2004. The recoverability of segmentation structure from store-level aggregate data. Journal of Marketing Research 41 (3), 351–364.
Debrezion, G., Pels, E., Rietveld, P., 2009. Modelling the joint access mode and railway station choice. Transportation Research Part E 45 (1), 270–283.
Fan, K.-S., Miller, E.J., Badoe, D., 1993. Modeling rail access mode and station choice. Transportation Research Record 1413, 49–59.
Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681–698.
Gupta, S., Chintagunta, P.K., 1994. On using demographic variables to determine segment membership in logit mixture models. Journal of Marketing Research 13, 128–136.
Harvey, G., 1986. Study of airport access mode choice. Journal of Transportation Engineering 112 (5), 525–545.
Hess, S., Bierlaire, M., Polak, J.W., 2005. Capturing correlation and taste heterogeneity with mixed GEV models. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics. Springer Publisher, Dordrecht, The Netherlands, pp. 55–76.
Kamakura, W.A., Russell, G.A., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26 (4), 379–390.

Kamakura, W.A., Kim, B.-D., Lee, J., 1996. Modeling preference and structural heterogeneity in consumer choice. Marketing Science 15 (2), 152–172.
Korf, J.L., Demetsky, M.J., 1980. Transit station classification for access mode analysis. Journal of Advanced Transportation 14 (3), 275–300.
Korf, J.L., Demetsky, M.J., 1981. Analysis of rapid transit access mode choice. Transportation Research Record 817, 29–35.
McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zaremmbka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.
McFadden, D., 1978. Modeling the choice of residential location. Transportation Research Record 672, 72–77.
Ortúzar, J.de D., Willumsen, L.G., 2001. Modelling Transport. John Wiley & Sons, Ltd..
Outwater, M.L., Castleberry, S., Shiftan, Y., Ben-Akiva, M., Zhou, Y.S., Kuppam, A., 2004a. Attitudinal market segmentation approach to mode choice and ridership forecasting. Transportation Research Record 1854, 32–42.
Outwater, M.L., Modugula, V., Castleberry, S., Bhatia, P., 2004b. Market segmentation approach to mode choice and ferry ridership forecasting. Transportation Research Record 1872, 71–79.
Pas, E.I., Huber, J.C., 1992. Market segmentation analysis of potential inter-city rail travelers. Transportation 19 (2), 177–196.
Polydoropoulou, A., Ben-Akiva, M., 2001. Combined revealed and stated preference nested logit access and mode choice model for multiple mass transit technologies. Transportation Research Record 1771, 38–45.
Psaraki, V., Abacoumkin, C., 2002. Access mode choice for relocated airports: the new Athens international airport. Journal of Air Transport Management 8 (2), 89–98.
Rastogi, R., Rao, K.V.K., 2009. Segmentation analysis of commuters accessing transit: Mumbai study. Journal of Transportation Engineering 135 (8), 506–515.
Shiftan, Y., Outwater, M.L., Zhou, Y., 2008. Transit market research using structural equation modeling and attitudinal market segmentation. Transport Policy 15 (3), 186–195.
Sobieniak, J., Westin, R., Rosapep, T., Shin, T., 1979. Choice of access mode to intercity terminals. Transportation Research Record 728, 47–53.
Swait, J., 2003. Flexible covariance structures for categorical dependent variables through finite mixtures of generalized extreme value models. Journal of Business Economic and Statistics 21 (1), 80–87.
Taiwan High Speed Rail Corporation, 2007. A Report on Service Quality of Taiwan High Speed Rail Transfer Facilities: Analysis of Station Access Mode Choice. Everest Engineering Consultants, Inc..
Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.
Tsamboulas, D., Golias, J., Vlahoyannis, M., 1992. Model development for metro station access mode choice. Transportation 19 (3), 231–244.
Walker, J.L., Li, J., 2007. Latent lifestyle preferences and household location decisions. Journal of Geographical Systems 9 (1), 77–101.
Wen, C.-H., Lai, S.-C., 2010. Latent class models of international air carrier choice. Transportation Research Part E 46 (2), 211–221.
Zhang, J., Kuwano, M., Lee, B., Fujiwara, A., 2009. Modeling household discrete choice behavior incorporating heterogeneous group decision-making mechanisms. Transportation Research Part B 43 (2), 230–250.