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